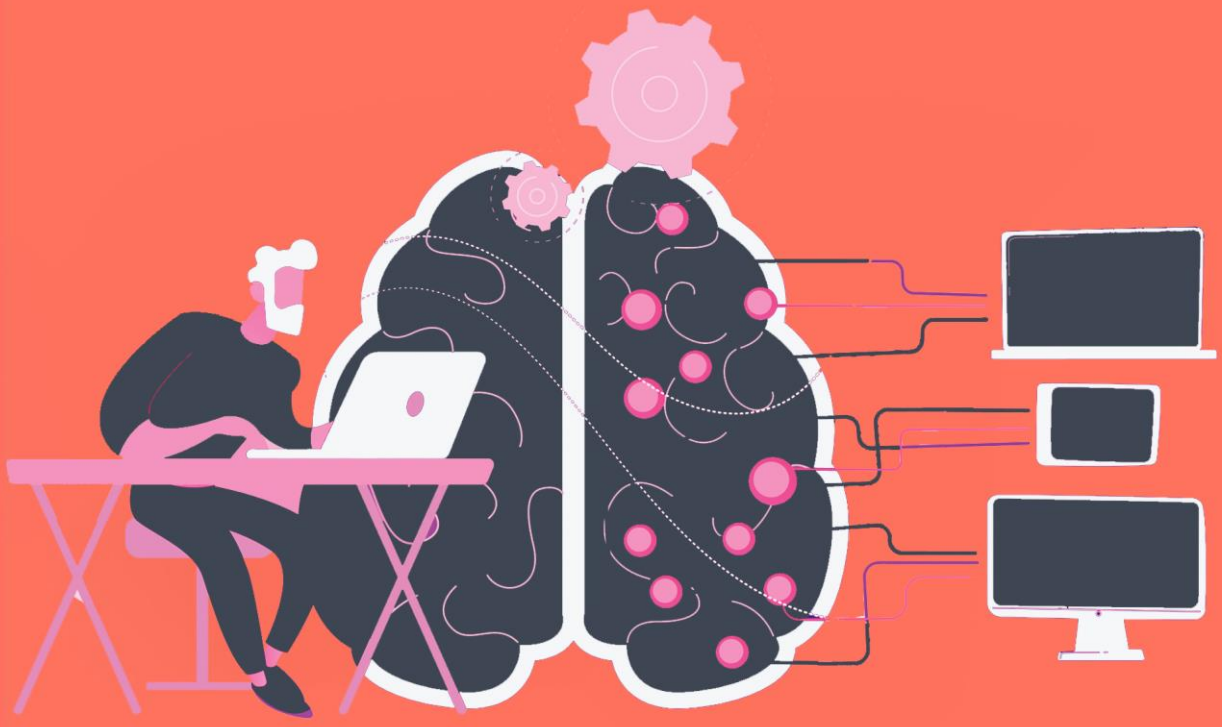


512
سؤال وجواب
في
التعلم الآلي والتعلم العميق وعلم البيانات

اعداد:
د. علاء طعيمة

بـــمه تعالى

512

ـــؤال وجواب في التعلم الآلي والتعلم العميق وعلم البيانات

اعداد:
د. علاء طعيمة

مقدمة المؤلف

يستهلك الباحث الكثير من الوقتفي البحث عن إجابات لأسئلة حول التعلم الآلي والتعلم العميق وعلم البيانات تندرج من السهولة إلى الأكثر صعوبة وقد تكون شديدة التنوع مما يزيد البحث عنها تعقيداً. لذا أصدرنا هذه الكتاب التي يضع بين يدي الباحث الإجابة عن الأسئلة الأكثر شيوعًا حول الجوانب المختلفة المتعلقة بالتعلم الآلي والتعلم العميق وعلم البيانات تسهيلاً للتعلم وتوفيراً للجهد.

لقد حاولت قدر المستطاع ان اجمع الأسئلة الأكثر طرحاً مع الأجوبة المناسبة والكافية، ومع هذا يبقى عملاً بشرياً يحتمل النقص، فاذا كان لديك أي ملاحظات حول هذا الكتاب، فلا تتردد بمراسلتنا عبر بريدنا الالكتروني alaa.taima@qu.edu.iq .

نأمل ان يساعد هذا الكتاب كل من يريد ان يدخلفي مجالات التعلم الآلي والتعلم العميق وعلم البيانات ومساعدة القارئ العربي على تعلم هذا المجالات. اسأل الله التوفيقفي هذا العمل لأثراء المحتوى العربي الذي يفتقر أشد الافتقار إلى محتوى جيد ورصينفي مجال التعلم الآلي والتعلم العميق وعلم البيانات. ونرجو لك الاستمتاع مع الكتاب ولا تنسونا من صالح الدعاء.

د. علاء طعيمة

كلية علوم الحاسوب وتكنولوجيا المعلومات

جامعة القادسية

العراق

سؤال

اشرح مفهوم التعلم الآلي
وافترض أنك تشرح ذلك لطفل
يبلغ من العمر 5 سنوات؟

المحتويات

س1/اشرح مفهوم التعلم الآلي Machine Learning (ML) وافترض أنك تشرح ذلك لطفل يبلغ من العمر 5 سنوات؟ 27

س2/كيف أبدأ مسيرتي المهنية في التعلم الآلي؟ 27

س3/ما مدى صعوبة التعلم الآلي؟ 27

س4/كيف يمكنك الدخول الى مجال التعلم الآلي؟ 27

س5/ما هي أفضل طريقة لتعلم التعلم الآلي؟ 28

س6/ما هو التعلم الآلي للمبتدئين؟ 28

س7/هل يتطلب التعلم الآلي البرمجة؟ 28

س8/ما الدرجة والشهادة العلمية التي تحتاجها للتعلم الآلي؟ 28

س9/كيف يكون التعلم العميق أفضل من التعلم الآلي؟ 28

س10/ما هي الأنواع المختلفة للشبكات العصبية العميقة؟ 28

س11/ ما الفرق بين التعلم الآلي Machine Learning (ML) والتعلم العميق Deep Learning (DL)؟ 29

س12/ما هي بعض التطبيقات الأكثر استخدامًا في التعلم العميق؟ 29

س13/ما الفرق بين القيمة المتوسطة mean value والقيمة المتوقعة expected value؟ 30

س14/ما هي بعض قيود التعلم العميق؟ 30

س15/ما هي وظائف التعلم الخاضع للإشراف Supervised Learning؟ 30

س16/ما هي وظائف التعلم غير الخاضع للإشراف Unsupervised Learning؟ 31

س17/ما هي بعض الأمثلة على خوارزميات التعلم الخاضعة للاشراف supervised learning algorithms في التعلم العميق؟ 31

س18/ ما هي بعض الأمثلة على خوارزميات التعلم غير الخاضع للاشراف unsupervised learning algorithms في التعلم العميق؟ 31

س19/كيف تختار الصيغة المناسبة appropriate formula لحل مشاكل التصنيف؟ 31

س20/ ما الفرق بين مجموعة بيانات التدريب Training ومجموعة بيانات الاختبار Testing ومجموعة بيانات التحقق Validation؟ ما هي النسبة العامة لهذه المجموعات؟ 32

س21/ لماذا الشبكات العميقة Deep Networks أفضل من الشبكات الضحلة Shallow Networks؟ 32

س22/ ما هي الاختلافات بين المتوسط Mean والوسيط Median والمنوال Mode؟ ما مدى فائدة هذه في التعامل مع القيم المفقودة في مجموعة البيانات المحددة؟ 33

س23/ ما هي Backend المختلفة المتوفرة في Keras؟..... 33

س24/ ما هي بعض أطر عمل framework التعلم العميق أو الأدوات التي تُستخدم في التعلم العميق؟ 34

س25/ ما هي الموترات tensors؟..... 34

س26/ ما هي العناصر القابلة للبرمجة في TensorFlow؟..... 34

س27/ ما هي بعض مزايا استخدام TensorFlow؟..... 34

س28/ ماذا تقصد بـ Tensorboard؟ 35

س29/ هل تعتقد أن التعلم العميق يؤدي بشكل أفضل من التعلم الآلي؟ إذا كان الأمر كذلك لماذا؟..... 35

س30/ ما هي الرياضيات المستخدمة في التعلم الآلي والعميق؟..... 35

س31/ هل يمكنني دراسة التعلم العميق قبل تعلم الآلة؟ 35

س32/ اذكر مزايا وعيوب استخدام الشبكات العصبية Neural Networks؟ 35

س33/ لقد قمت ببناء نموذج DL وأثناء التدريب لاحظت أنه بعد عدد معين من الفترات، تتناقص الدقة. ما هي المشكلة وكيف يتم حلها؟ 36

س34/ ما هي أهم الانتقادات والقيود المفروضة على التعلم العميق؟..... 36

س35/ ما هي المتغيرات الفئوية categorical variables؟ 36

س36/ ما هي المقاييس التي يمكن استخدامها لقياس ارتباط correlation البيانات الفئوية categorical data؟ 37

س37/ ما الخوارزميات التي يمكن استخدامها لاختيار المتغير المهم important variable؟ 37

س38/ كيف تختار الميزات المهمة important features في مجموعة البيانات المحددة given data set؟..... 37

س39/ افترض أنك تعمل على مجموعة بيانات، واشرح كيف ستختار المتغيرات المهمة؟37

س40/ أثناء العمل على مجموعة بيانات، كيف يمكنك اختيار المتغيرات المهمة important variables؟ 37

س41/ لماذا يحظى التعلم العميق بشعبية كبيرة ومطلوب هذه الأيام؟..... 38

س42/ ما هي أفضل الكتب عن التعلم العميق؟ 38

س43/ كيف يمكنك تغيير الشبكة العصبية من التوقع regression إلى التصنيف classification؟..... 38

س44/ هناك العديد من خوارزميات التعلم الآلي حتى الآن. إذا أعطيت مجموعة بيانات dataset، فكيف يمكن للمرء تحديد الخوارزمية التي سيتم استخدامها لذلك؟..... 39

س45/ قم بتسمية بعض مكتبات التعلم الآلي لأغراض مختلفة؟..... 39

س46/ كيف يمكنك تطوير نموذج لتحديد الانتحال والسرقة الفكرية plagiarism؟ 39
س47/ ما هي مراحل بناء نموذج في التعلم الآلي؟ 40
س48/ هل يمكنك إخبارنا بكيفية تصميم فلتر للبريد الإلكتروني العشوائي Email Spam Filter؟ 40
س49/ ما هو التعلم الآلي الاستقرائي inductive machine learning؟ 40
س50/ ما هي تقنيات الخوارزمية المختلفة في التعلم الآلي؟ 40
س51/ ما هو الفرق بين نمذجة البيانات Data Modeling وتصميم قواعد البيانات Database Design؟ 41
س52/ ما هو قانون الأعداد الكبيرة Law of Large Numbers؟ 41
س53/ هل يمكنك التمييز بين البيانات ذات التنسيق الطويل Long-Format Data والبيانات ذات التنسيق الواسع Wide-Format Data؟ 41
س54/ ما هي دوال الخسارة Loss Function ودوال التكلفة Cost Function؟ اشرح الفرق الرئيسي بينهما؟ 41
س55/ ما هي الخوارزميات الخمس الشائعة لتعلم الآلة؟ 42
س56/ ماذا تقصد بالميزات Features والتسميات Labels في مجموعة البيانات؟ 42
س57/ ماذا تقصد بالمتغيرات المستقلة independent variable والتابعة dependent variable؟ 43
س58/ ماذا تقصد بالضوضاء Noise في مجموعة بيانات Dataset معينة وكيف يمكنك إزالة الضوضاء في مجموعة البيانات؟ 43
س59/ ما هي عملية تنفيذ الانحدار الخطي linear regression؟ 44
س60/ كيف تشرح الانحدار الخطي linear regression لشخص غير تقني؟ 44
س61/ استنادًا إلى مجموعة البيانات، كيف ستعرف الخوارزمية التي يجب تطبيقها؟ 44
س62/ أين تستخدم R & Python؟ 45
س63/ ما هو اختيار النموذج Model Selection في التعلم الآلي؟ 45
س64/ ما هي المراحل الثلاث لبناء الفرضيات أو النموذج في التعلم الآلي؟ 45
س65/ كيف يتم استخدام الخوارزمية الجينية في التعلم الآلي؟ 45
س66/ كيف يمكنك تحسين بُنية مصنف التعلم العميق باستخدام الخوارزميات الجينية؟ 46
س67/ ما هي الفروق بين الارتباط correlation والتغاير covariance؟ 46
س68/ عدد الأنواع المختلفة من المخططات Plots التي نستخدمها بشكل عام في التعلم الآلي والتعلم العميق؟ 46
س69/ ما هي الافتراضات المطلوبة للانحدار الخطي Linear Regression؟ 46

س70/ ماهي Specificity وSensitivity؟ 47

س71/ ما هي مقاييس الأداء التي يمكن استخدامها لتقدير كفاءة نموذج الانحدار الخطي Linear regression؟..... 47

س72/ ما هو مقياس الأداء الأفضل R2 أو R2 المعدل؟ 47

س73/ ما هو R2؟ ما هي بعض المقاييس الأخرى التي يمكن أن تكون أفضل من R2 ولماذا؟ 47

س74/ كيف تقرر ما إذا كان نموذج الانحدار الخطي الخاص بك يناسب البيانات؟..... 47

س75/ متى تستخدم الانحدار اللوجستي ومتى تستخدم الانحدار الخطي؟ 48

س76/ هل يمكن استخدام الانحدار اللوجستي logistic regression للفئات التي تزيد عن 2؟ 48

س77/ ما هو نموذج التصويت voting model؟..... 48

س78/ هل طرق الانحدار التدريجي تتقارب converge دائمًا مع نقاط متشابهة؟ 48

س79/ كيف تتعامل مع عينات بيانات قليلة جدًا؟ هل من الممكن بناء نموذج منها؟.... 48

س80/ ماهي المقاييس التي تُستخدم لتحليل الاتجاه المركزي للبيانات central tendency of data؟..... 49

س81/ كيف يمكنك بناء data pipeline؟..... 49

س82/ كيف تعتقد أن الحوسبة الكمومية quantum computing ستؤثر على التعلم الآلي؟ 49

س83/ ما هي الخطوات الضرورية والمتضمنة في مشروع التعلم الآلي؟ 49

س84/ كيف يرتبط علم البيانات والتعلم الآلي ببعضهما البعض؟ 50

س85/ ما هي السلاسل الزمنية Time series؟ 50

س86/ ما هي المهارات المهمة التي يجب ان تمتلكها في بايثون فيما يتعلق بتحليل البيانات data analysis؟..... 50

س87/ أيهما أفضل لتحليلات النص - Python أم R؟..... 51

س88/ اشرح تحليل السلاسل الزمنية Time Series Analysis؟..... 51

س89/ يتم إعطاؤك مجموعة بيانات عن اكتشاف السرطان. لقد قمت ببناء نموذج تصنيف وحققت دقة تصل إلى 96%. لماذا لا تكون سعيدًا بأداء نموذجك؟ ماذا يمكنك أن تفعل بهذا الشأن؟..... 51

س90/ في الشبكة العصبية، ماذا لو تمت تهيئة جميع الأوزان بنفس القيمة؟..... 52

س91/ لماذا يجب عليك تحديث خوارزمية تعلم الآلة بانتظام؟ كم مرة يجب تحديثها؟ . 52

س92/ إذا كنت نموذج التدريب يعطي دقة 90٪ ونموذج الاختبار يعطي دقة 60٪؟ ثم ما المشكلة التي تواجهها؟..... 52

س93/ ما هو الفرق بين الارتباط **correlation** والسببية **causality**؟..... 53
س94/ كيف يتم نشر التعلم الآلي في سيناريوهات العالم الحقيقي؟ 53
س95/ ما هي بعض الانتقادات الموجهة للشبكات العصبية؟..... 53
س96/ هل يمكنك استخدام التعلم الآلي لتحليل السلاسل الزمنية؟..... 54
س97/ ما هي تطبيقات التعلم الآلي الخاضع للإشراف؟ 54
س99/ ما هي خوارزمية التعلم الآلي البارامترية **Parametric** وغير البارامترية **Non-parametric**؟ وأهميتها؟ 54
س100/ اعط أمثلة على خوارزمية التعلم الآلي البارامترية وخوارزمية التعلم الآلي غير البارامترية؟ 55
س101/ هل تعلم الآلة علم أم فن؟ 55
س102/ ما هي النماذج التمييزية **discriminative** والتوليدية **generative**؟ 55
س103/ ما هي تطبيقات التعلم الآلي؟..... 55
س104/ ما الفرق بين التعلم الآلي الخاضع للإشراف **supervised** والتعلم غير الخاضع للإشراف **unsupervised**؟ 56
س105/ قارن بين التصنيف **classification** والانحدار **regression** في التعلم الآلي؟..... 56
س106/ كيف يتم تحديد مدى ملاءمة خوارزمية التعلم الآلي لمشكلة معينة؟..... 57
س107/ اشرح مصطلح التعلم الآلي شبه الخاضع للإشراف **Semi-Supervised Machine Learning**؟ 57
س108/ ما الفرق بين التنقيب في البيانات **Data Mining** والتعلم الآلي **Machine Learning**؟ 57
س109/ ما الفرق بين **TensorFlow** و**Keras**؟..... 58
س110/ ماذا يفهم المرء بمصطلح علم البيانات **Data Science (DS)**؟ 59
س111/ كيف يمكنك جمع وتحليل البيانات لاستخدام وسائل التواصل الاجتماعي للتنبؤ بالطقس؟ 59
س112/ هل المزيد من البيانات دائمًا أفضل؟..... 59
س113/ كيف يساعد التعلم العميق علماء البيانات **data scientists**؟ 60
س114/ ما هي مزايا رسم بياناتك قبل إجراء التحليل؟..... 60
س115/ كيف يمكنك اختيار المصنف **classifier** بناءً على حجم مجموعة التدريب؟..... 60
س116/ متى يجب استخدام التصنيف **classification** على حساب الانحدار **regression**؟ .. 61
س117/ كيف ستفرق بين مشكلة التصنيف متعدد الفئات **multi-class classification** ومتعدد التسميات **multi-label classification**؟ 61
س118/ ما الخوارزميات المستخدمة في التصنيف الثنائي **Binary classification**؟..... 61

س119/ ما هي الخوارزميات المستخدمة لتصنيف متعدد الحدود Multinomial classification؟ 61

س120/ ماذا تقصد بـ Tensor في TensorFlow؟ 62

س121/ ما الذي يجعل TensorFlow المكتبة الأكثر تفضيلاً في التعلم العميق؟ 62

س122/ ما هي خوارزمية imputation المختلفة المتاحة للتعامل مع البيانات المفقودة؟ 62

س123/ ما هي الخوارزمية التي يمكن استخدامها في احتساب القيمة value imputation في كل من فئات البيانات الفئوية والمستمرة؟ 62

س124/ لماذا يتم استخدام Python لتنظيف البيانات في علم البيانات؟ 62

س125/ لماذا يعد تنظيف البيانات data cleaning أمرًا بالغ الأهمية؟ كيف تنظف البيانات؟ 63

س126/ ما هو التحويل الثنائي البيانات Binarization؟ كيفية التحويل الثنائي؟ 63

س127/ ما هي المكتبات الأكثر شعبية المستخدمة في علم البيانات؟ 63

س128/ في علم البيانات، لماذا يتم استخدام Python لتنظيف البيانات؟ 64

س129/ ما هو تصوير البيانات Data Visualization؟ 64

س130/ ما هي المكتبات في Python المستخدمة في تحليل البيانات Data Analysis والحسابات العلمية Scientific Computations؟ 64

س131/ ماذا تقصد بـ Matplotlib؟ 64

س132/ ما هي الاختلافات بين البيانات المصنفة Labelled وغير المصنفة Unlabeled؟ . 65

س133/ كيف يتم استخدام Hadoop في علم البيانات؟ 65

س134/ قارن بين تحليلات البيانات Data Analytics وعلم البيانات Data Science؟ 65

س135/ كيف تقيم نموذج الانحدار اللوجستي logistic regression model؟ 66

س136/ لماذا يتم استخدام تحويل فورييه Fourier transform في التعلم العميق؟ 66

س137/ ما هي بعض الأمثلة على خوارزميات التعلم الخاضعة للأشراف supervised learning algorithms في التعلم العميق؟ 66

س138/ لماذا يتحسن أداء التعلم العميق حيث يتم تغذية المزيد من البيانات إليه؟ 67

س139/ بالنظر إلى وجود العديد من خوارزميات التعلم العميق، كيف ستحدد نموذج التعلم العميق التي يجب استخدامها لمجموعة البيانات؟ 67

س140/ لماذا تعتبر وحدات معالجة الرسوميات GPU مهمة لتنفيذ نماذج التعلم العميق؟ 67

س141/ هل هناك فرق بين الشبكات العصبية والتعلم العميق؟ 67

س142/ دقة النموذج Model accuracy أم أداء النموذج Model performance؟ أيهما تفضل ولماذا؟ 68

س143/ كيف يمكنك تحسين معلمات النموذج أثناء بناء النموذج؟ 69

س144/ ما هي هياكل البيانات الشائعة المستخدمة في التعلم العميق؟ 69

س145/ ما هي لغة البرمجة أو التقنيات التي يجب أن يمتلكها شخص ما أو يتعلمها ليكون مهندس التعلم العميق DL؟ 69

س146/ ما هي بعض المشاكل الشائعة التي يواجهها مهندسو التعلم العميق؟..... 69

س147/ ما هي المكتبة المفضلة في التعلم العميق ولماذا؟..... 70

س148/ ما هو الاستخدام النهائي للتعلم العميق في وقتنا الحالي وكيف يساعد علماء البيانات؟ 70

س149/ قارن بين التعلم الآلي Machine Learning والبيانات الضخمة Big Data؟ 70

س150/ يتم تقديم مجموعة بيانات لك حول كشف الاحتيال fraud detection في المرافق utilities. لقد قمت ببناء نموذج مصنف وحقق أداء بنسبة 98.5٪. هل هذا نموذج جيد؟ إذا كانت الإجابة بنعم، فبرر ذلك. إذا لم يكن كذلك، فماذا يمكنك أن تفعل حيال ذلك؟..... 71

س151/ كيف تشرح الانحدار اللوجستي لاقتصادي وطبيب وعالم أحياء؟ 71

س152/ ما هو الفرق بين محلل البيانات Data Analyst وعالم البيانات Data Scientist؟. 71

س153/ ما هي الخطوات المختلفة التي ينطوي عليها مشروع تحليلات البيانات analytics project؟..... 72

س154/ ماذا تعني البرمجة اللغوية العصبية NLP؟ 73

س155/ كيف تتعامل مع مجموعة بيانات تفتقد إلى أكثر من 30 في المائة من قيمها؟ ... 73

س156/ هل يمكنك بناء نماذج تعلم عميقة يعتمد فقط على الانحدار الخطي؟..... 73

س157/ في التعلم الآلي، ما عدد الكلاسات التي يمكن ان تستخدم الانحدار اللوجستي Logistic Regression؟ 74

س158/ أخبرني عن حالة قد تكون فيها التقنيات الجماعية ensemble techniques مفيدة؟ 74

س159/ ما هو المقصود بالتعلم الجماعي Ensemble Learning؟..... 74

س160/ أخبرني عن حالة قد تكون فيها التقنيات الجماعية ensemble techniques مفيدة؟ 74

س161/ ما معنى bagging وboosting في التعلم العميق؟..... 75

س162/ ما هو التعلم الشامل end-to-end learning؟ أعط بعض مزاياها..... 75

س163/ لماذا يُشار أحيانًا إلى خوارزميات التعلم القائم على المثيل instance-based learning باسم خوارزميات التعلم الكسول Lazy learning؟..... 76

س164/ متى يفضل التعلم متعدد المهام multi-task learning؟ 76

س165/ ما هو التعلم المعزز reinforcement learning؟ 76

س166/ اشرح كيف يعمل نظام التوصية recommender system؟ 76

س167/ ما هي أنظمة التوصية recommender systems؟ 77

س168/ اشرح التصفية على أساس المحتوى content-based filtering في أنظمة التوصية recommender systems؟ 77

س169/ ما هو التصفية التعاونية collaborative filtering؟ 77

س170/ ما نوع نظام التوصية recommendation system الذي تستخدمه أمازون للتوصية بعناصر مماثلة؟ 77

س171/ في Google، إذا قمت بكتابة "How are" فإنه يمنحك التوصية على أنها " How are you" / "How do you do"، فهذا يعتمد على ماذا؟ 78

س172/ ما هو الفرق بين box plot وhistogram؟ 78

س173/ لماذا يتم استخدام لغة R في تصوير البيانات Data Visualization؟ 79

س174/ ما هو XGBoost؟ 79

س175/ ما هو التجميع Clustering في التعلم الآلي؟ 79

س176/ لماذا يعتبر الانحدار اللوجستي نوعًا من تقنيات التصنيف وليس الانحدار؟ 80

س177/ اشرح الانحدار اللوجستي Logistic Regression؟ 80

س178/ ما هو الفرق بين مجموعة التحقق من الصحة validation set ومجموعة الاختبار Test set؟ 80

س179/ ما هي تقنية cross-validation (CV) الذي ستستخدمه في مجموعة بيانات السلاسل الزمنية time series dataset؟ 80

س180/ هل من الممكن اختبار احتمال تحسين دقة النموذج بدون تقنيات التحقق المتبادل cross-validation؟ إذا كانت الإجابة بنعم، رجاء التوضيح؟ 81

س181/ ما هو k-fold cross-validation؟ 81

س182/ عدد تقنيات التحقق المتبادل cross validation؟ 81

س183/ ما هو Cross-validation في التعلم الآلي؟ 82

س184/ ما هي الحدود الدنيا المحلية local minima والحد الأدنى العالمية Global Minima في الانحدار؟ 82

س185/ ما هي الانتروبية Entropy في التعلم الآلي؟ 83

س186/ ما هو PCA في التعلم الآلي؟ 83

س187/ هل rotation ضروري في PCA؟ 83

س188/ ماذا يحدث إذا لم يتم تدوير المكونات في PCA؟ 83

س189/ ما هو تحليل التمايز الخطي Linear discriminant analysis (LDA)؟ 83

س190/ ما هي العلاقة بين PCA وLDA؟ 83

س191/ يتم إعطاؤك مجموعة بيانات تدريب بها الكثير من الأعمدة والصفوف. كيف تقلل من أبعاد هذه البيانات؟ 84

س192/ كيف يختلف PCA عن LDA؟ 84

س193/ ما هي عيوب النموذج الخطي linear model؟ 84

س194/ يتم منحك مجموعة بيانات عن كشف الاحتيال fraud detection. حقق نموذج التصنيف دقة 95٪ هل هو جيد؟ 84

س195/ ما هو الانحدار الخطي Linear Regression في التعلم الآلي؟ 85

س196/ ماذا تقصد بالتعلم العميق Deep learning ولماذا أصبح شائعًا الآن؟ 85

س197/ كيف يمكننا استخدام مجموعة بيانات بدون المتغير المستهدف في خوارزميات التعلم الخاضعة للإشراف؟ 86

س198/ هل يمكنك تسمية بعض هياكل البيانات data structures التي يشيع استخدامها في التعلم العميق؟ 86

س199/ كيف يتم استخدام تحويل فورييه Fourier Transform لصالح التعلم العميق؟ 86

س200/ ما هي سلاسل ماركوف Markov chains؟ 86

س201/ اشرح فوائد استخدام الإحصائيات statistics بواسطة علماء البيانات Data Scientists؟ 87

س202/ ما هي آلة بولتزمان Boltzmann machine؟ 87

س203/ ما هو البيرسيبترون perceptron؟ 87

س204/ صف هيكل الشبكات العصبية الاصطناعية ANN؟ 87

س205/ ما هو البيرسيبترون Perceptron؟ وكيف يعمل؟ 88

س206/ ما هي الخطوات المتبعة في تدريب perceptron في التعلم العميق؟ 88

س207/ ما هي مزايا استخدام Perceptron متعدد الطبقات على Perceptron أحادي الطبقة؟ 89

س208/ ما هي محددات استخدام البيرسبترون Perceptron؟ 89

س209/ قارن بين البيرسيبترون أحادي الطبقة single layer perceptron والبيرسيبترون متعدد الطبقات multi-layer Perceptron؟ 89

س210/ ما هو الفرق بين Perceptron والانحدار اللوجستي Logistic Regression؟ 89

س211/ ما هو البيرسيبترون متعدد الطبقات Multilayer perceptron [MLP]؟ 90

س212/ ماهي CNN؟ 90

س213/ ما هي الطبقات المختلفة الموجودة في CNN؟ 90

س214/ ما معنى valid padding وsame padding في CNN؟ 91

س215/ قم بتسمية بعض بُنى CNNs التي تعرفها؟ 91

س216/ لماذا يُفضل CNN على ANN لمهام تصنيف الصور Image Classificationعلى الرغم من أنه من الممكن حل تصنيف الصور باستخدام ANN؟..... 91

س217/كيفية إصلاح Constant validation accuracy في تدريب نموذج CNN؟ 92

س218/ لماذا نستخدم التلافيف convolutions للصور بدلاً من استخدام طبقات متصلة بالكامل fully connected layers؟ 92

س219/ ما فائدة استخدام التجميع الأقصى max-pooling في تصنيف الشبكات العصبية التلافيفية CNN؟ 93

س220/ ما هي بعض مزايا استخدام CNN (شبكة عصبية تلافيفية) بدلاً من DNN (شبكة عصبية كثيفة) في مهمة تصنيف الصور؟ 93

س221/ كيف يمكنك تحديد حجم المرشح (الفلتر) عند إجراء عملية التفاف في شبكة CNN؟..... 93

س222/ عند تصميم CNN، هل يمكننا معرفة عدد الطبقات التلافيفية التي يجب أن نستخدمها؟ 94

س223/هل من الجيد استخدام CNN لتصنيف إشارة أحادية البعد؟ 94

س224/ما هي شبكة Deconvolutional؟..... 94

س225/صف بنية الشبكة العصبية التلافيفية النموذجية (CNN)؟ 95

س226/ما هي خوارزمية التعلم العميق الأفضل لاكتشاف الوجه face detection؟ 95

س227/ لماذا تعمل الشبكة العصبية التلافيفية (CNN) بشكل أفضل مع بيانات الصورة؟ 96

س228/ في CNN، إذا كان حجم الإدخال 5 × 5 وحجم الفلتر هو 7 × 7 ، فما هو حجم الإخراج؟ 96

س229/ما هي الطبقات المختلفة للشبكة العصبية التلافيفية CNN؟..... 96

س230/كيف يمكنك تحويل طبقة كثيفة Dense Layer لشبكة CNN إلى طبقة تلافيفية كاملة Fully Convolutional Layer (FCN)؟ 97

س231/ما هو التجميع Poolingعلى CNN وكيف يعمل؟..... 97

س232/ما فائدة استخدام نوى صغيرة small kernels مثل3x3 من استخدام عدد قليل من النوى الكبيرة large kernels؟..... 97

س233/كيف تختار عمق الشبكة العصبية؟..... 98

س234/ماذا تفهم بمصطلحات الدُفعات Batches والتكرارات Iterations والفترات Epochs في تدريب نموذج الشبكة العصبية؟ 99

س235/ماذا تقصد بإعطاء "epoch = 1" في الشبكة العصبية؟ 99

س236/ لنفترض أن عليك بناء بُنية شبكة عصبية؛ كيف ستقرر عدد الخلايا العصبية neurons والطبقات المخفية hidden layers اللازمة للشبكة؟ 99

س237/ ماذا تقصد بالفلترة **filtering** والخطوة **stride** والحشو **padding** في الشبكة العصبية التلافيفية **CNN**؟ 100

س238/ ما هي المعلمات الفائقة **hyperparameters** المختلفة المستخدمة في الشبكات العصبية التلافيفية **CNN** أثناء تدريب النموذج؟ 101

س239/ ماذا تقصد بنموذج **MobileNet**؟ 102

س240/ كيف تقسم مجموعة البيانات الخاصة بك إلى بيانات اختبار وتدريب؟ 103

س241/ ما هي **RNN** في التعلم العميق؟ 103

س242/ ما هي الخطوات المتبعة في عمل شبكة **LSTM**؟ 103

س243/ ما الفرق بين الشبكات العصبية التلافيفية (**CNN**) والشبكات العصبية المتكررة (**RNN**) وفي أي الحالات تستخدم كل واحدة منها؟ 103

س244/ ما هو تلاشي الانحدار **vanishing gradient** عند استخدام **RNNs**؟ 104

س245/ ما الهدف من استخدام **LSTM**؟ 104

س246/ كيف تكون بنية المحولات **transformers** أفضل من **RNNs** في التعلم العميق؟ 104

س247/ كيف سيتم تنفيذ التسوية بالدفعات **Batch Normalization** في **RNN**؟..... 105

س248/ قارن بين الشبكة العصبية ذات التغذية الأمامية **FNN** والشبكة العصبية المتكررة **RNN**؟ 105

س249/ ما الفرق بين **Recurrent Neural Networks** و**Recursive Neural Networks**؟ 105

س250/ كم عدد الأبعاد التي يجب أن تحتويها مدخلات طبقة **RNN**؟ ماذا يمثل كل بعد؟ ماذا عن نواتجها؟ 106

س251/ ما هي أنواع الشبكات العصبية المتكررة (**RNN**) التي تعرفها؟..... 107

س252/ كيف يختلف الانتشار الخلفي للشبكة العصبية المتكررة **RNN** عن باقي الشبكات العصبية؟ 108

س253/ لماذا تعمل **RNNs** بشكل أفضل مع البيانات المتسلسلة **sequential data**؟ . 108

س254/ لماذا تعمل **RNNs** بشكل أفضل مع البيانات النصية؟ 109

س255/ هل يمكنك إضافة تنظيم **L2** إلى شبكة عصبية متكررة **RNN** للتغلب على مشكلة تلاشي التدرج؟..... 109

س256/ ما هي تطبيقات الشبكة العصبية المتكررة (**RNN**)؟ 109

س257/ ما هي **GRU** وما اهميتها؟ 109

س258/ اشرح أهمية **GRU**؟..... 110

س259/ أيهما أفضل **LSTM** أم **GRU**؟..... 110

س260/اشرح أهمية LSTM؟ 110
س261/كيف تحل LSTM تحدي تلاشي الانحدار vanishing gradient؟..... 110
س262/لماذا GRU أسرع مقارنة بـ LSTM؟..... 110
س263/ما هي شبكات الخصومة التوليدية (GANs)؟ 111
س264/لماذا تحظى شبكات الخصومة التوليدية (GAN) بشعبية كبيرة؟..... 111
س265/ما هي الفكرة وراء شبكات GAN؟..... 111
س266/ما هي مكونات شبكة الخصومة العامة General Adversarial Network؟ ... 111
س267/ما هي طرق تقليل الأبعاد dimensionality reduction التي تعرفها وكيف تقارن مع بعضها البعض؟ 112
س268/ لماذا نحتاج إلى المشفرات التلقائية autoencoders عندما تكون هناك بالفعل تقنيات فعالة لتقليل الأبعاد dimensionality reduction مثل تحليل المكونات الرئيسية PCA؟ 113
س269/هل يمكن استخدام المشفرات التلقائية autoencoders لتوليد الميزات؟ إذا كانت الإجابة نعم، فكيف؟..... 113
س270/ما هي المشفرات التلقائي العميقة deep autoencoders؟..... 114
س271/أين يتم استخدام المشفرات التلقائية autoencoders؟ 115
س272/ماهي انواع المشفرات التلقائية autoencoders؟..... 115
س273/ما الفرق بين Autoencoder وVariational Autoencoder؟ 115
س274/قارن بين PCA وAutoencoders؟..... 116
س275/ما الفرق بين Variational Autoencoders وGAN؟ 117
س276/ ما هو المشفر التلقائي Autoencoder؟ تفاصيل حول المشفر Encoder وفك التشفير Decoder وعنق الزجاجة Bottleneck؟ 117
س277/ اشرح عن Sparse Autoencoder؟..... 118
س278/ هل يمكن استخدام المشفرات التلقائية لتوليد بيانات التدريب؟ إذا كانت الإجابة نعم، فكيف؟..... 119
س279/اشرح عن Denoising Autoencoder؟..... 119
س280/اشرح عن Convolutional Autoencoder (CAE)؟..... 120
س281/كيف تضبط المعلمات الفائقة Hyperparameter في Autoencoders؟..... 120
س282/ما هو تسوية البيانات data normalization في التعلم العميق؟ 121
س283/ ما الفرق بين Batch Normalization وInstance Normalization و Layer Normalization؟ 121

س284/ ما الفرق بين GAN وautoencoders؟ 122

س285/ ماذا تفهم بنقل التعلم transfer learning؟ اذكر بعض نماذج التعلم الشائعة الاستخدام؟ 123

س286/ ماذا تقصد بنموذج VGG16؟ وكيف نستخدمه في تصنيف الصور؟ 123

س287/ صف بعض المعلمات الفائقة hyperparameters لنقل التعلم transfer learning؟ 123

س288/ ما هو الضبط الدقيق fine-tuning وكيف يختلف عن نقل التعلم transfer learning؟ 124

س289/ ما هو نقل التعلم transfer learning وكيف تستخدم نماذج مدربة مسبقًا pre-trained models لنقل التعلم الى شبكتك العصبية؟ 124

س290/ لأي مشكلة معينة، كيف تقرر ما إذا كان عليك استخدام نقل التعلم transfer learning أو الضبط الدقيق fine-tuning؟ 125

س291/ ما هي طرق الضبط الدقيق Fine Tuning لنماذج التعلم العميق؟ 125

س292/ ما هي تطبيقات نقل التعلم transfer learning في التعلم العميق؟ 125

س293/ ما هي أهمية وجود شبكات عصبية متبقيةresidual neural networks ؟ 126

س294/ ماذا تعرف عن نقل التعلم transfer learning؟ 126

س295/ ما هو معنى الضبط الزائد overfitting؟ 127

س296/ كيف يمكنك التغلب على الضبط الزائد overfitting؟ 127

س297/ ما هي الأساليب المستخدمة بشكل شائع للتعامل مع فرط التعلم (overfitting) في التعلم العميق؟ 128

س298/ هل يمكنك مشاركة بعض الحيل أو التقنيات التي تستخدمها لمواجهة الضبط الزائد overfitting لتلائم نموذج التعلم العميق والحصول على تعميم generalization أفضل؟ 128

س299/ كيفية منع فرط التعلم overfitting في التعلم العميق؟ 129

س300/ ما هو الضبط الناقص underfitting وكيف يمكن منعه؟ 129

س301/ كيف تعرف إذا كان نموذج يعاني من الضبط الزائد overfitting؟ 129

س302/ ما هي الاختلافات بين overfitting وunderfitting؟ 130

س303/ عرف الضبط الزائد (فرط التعلم) overfitting؟ كيف نضمن أننا لا نفرط في تعليم نموذج؟ 130

س304/ اذكر سبب أهمية هندسة الميزات feature engineering في بناء النموذج واذكر بعض التقنيات المستخدمة في هندسة الميزات؟ 131

س305/ ما هو التوزيع الطبيعي normal distribution؟ 131

س306/ ما المقصود بتحجيم الميزات Feature Scaling؟ 131

س307/ كيف تختار الميزات في الشبكة العصبية؟ 132

س308/ ما هو الفرق بين اختيار الميزات feature selection وطرق هندسة الميزات feature engineering؟ 132

س309/ هل هناك فرق بين تعلم الميزات feature learning واستخراج الميزات feature extraction؟ 133

س310/ هل نحتاج إلى استخراج الميزات feature extraction في التعلم العميق؟ 133

س311/ ما هي بعض التقنيات المستخدمة لأخذ العينات sampling؟ ما هي الميزة الرئيسية لأخذ العينات؟ 133

س312/ ما هي متجهات الميزات feature vectors؟ 133

س313/ ما هي طرق تحديد الميزة Feature Selection Methods المستخدمة لتحديد المتغيرات الصحيحة؟ 134

س314/ ما الفرق بين اختيار الميزة Feature Selection واستخراج الميزة Feature Extraction ومن الذي يأتي اولاً؟ 134

س315/ كيف ستتعامل مع مجموعة بيانات غير متوازنة imbalanced dataset؟ 134

س316/ كيف تتعامل مع التصنيف الثنائي غير المتوازن unbalanced binary classification؟ 135

س317/ كيف تحقق التوازن في القوة عند التعامل مع مجموعات البيانات غير المتوازنة imbalanced datasets في التعلم العميق؟ 135

س318/ هل ستقوم بإزالة المتغيرات المرتبطة correlated variables أولاً؟ لماذا؟ ... 135

س319/ ما هي الطرق المتاحة لفحص القيم المتطرفة Outliers؟..... 136

س320/ ما هي القيم المتطرفة outliers؟ اذكر ثلاث طرق للتعامل مع القيم المتطرفة. 136

س321/ ما هي بعض الطرق التي يمكنني من خلالها جعل نموذجي أكثر قوة بالنسبة للقيم المتطرفة outliers؟ 137

س322/ كيف ستتعامل مع القيم المفقودة missing values في البيانات؟..... 137

س323/ اشرح كيف يمكننا التقاط العلاقة بين المتغيرات المستمرة continuous والمتغيرات الفئوية categorical؟ 137

س324/ ما هي خوارزميات التعلم الآلي التي يمكن استخدامها لإدخال القيم المفقودة لكل من المتغيرات الفئوية والمستمرة؟..... 137

س325/ تخيل أنك حصلت على مجموعة بيانات تتكون من متغيرات بها أكثر من 30٪ من القيم المفقودة missing values. لنفترض أنه من بين 50 متغيرًا، هناك 16 متغيرًا بها قيم مفقودة، وهي أعلى من 30٪. كيف ستتعامل معهم؟..... 138

س326/ كيف تتعامل مع البيانات المفقودة **missing** أو التالفة **corrupted** في مجموعة البيانات **dataset**؟ 138

س327/ ماذا تقصد بـ **Imputation**؟..... 138

س328/ ما هي مشاكل جودة البيانات **Data Quality** الشائعة؟ 138

س329/ ما هي الخطوات التي يجب اتباعها لاستخدام خوارزمية الانحدار التدريجي **gradient descent**؟ 139

س330/ ما هو انفجار الانحدار التدريجي **exploding gradient descent** في التعلم العميق؟ 139

س331/ ما هي أنواع الانحدار التدريجي **gradient descent**؟..... 139

س332/ لماذا يعتبر الانحدار التدريجي بالدفعات الصغيرة **Mini-Batch Gradient Descent** شائعا جداً؟ 140

س333/ ماذا تفهم عن اقتصاص التدرج **Gradient Clipping**؟..... 140

س334/ ما هي أنواع الانتشار الخلفي **variants of Backpropagation**؟ 140

س335/ صف طريقتين للتعامل مع مشكلة تلاشي الانحدار **vanishing gradient** في الشبكة العصبية؟ 140

س336/ ما هي مشكلة تلاشي التدرج **Vanishing Gradient** في الشبكات العصبية الاصطناعية؟ 141

س337/ قارن بين الانحدار التدريجي الدفعي **Batch Gradient Descent** والانحدار التدريجي العشوائي **Stochastic Gradient Descent** والانحدار التدريجي للدفعة الصغيرة **Mini Batch gradient descent**؟..... 142

س338/ كيف تحل مشكلة انفجار الانحدار **exploding gradient problem**؟..... 142

س339/ هل من الضروري خلط **shuffle** بيانات التدريب عند استخدام الانحدار التدريجي الدفعي؟..... 143

س340/ عند استخدام الانحدار التدريجي ذو الدفعات الصغيرة، ما سبب أهمية خلط البيانات بشكل عشوائي؟ 143

س341/ كيف تعرف ما إذا كان نموذجك يعاني من مشكلة انفجار الانحدار **exploding gradient**؟ 143

س342/ كيف تعرف ما إذا كان نموذجك يعاني من مشكلة تلاشي الانحدار **vanishing gradients**؟ 143

س343/ اشرح الانحدار التدريجي الدفعي **Batch Gradient Descent**؟ 144

س344/ ما هو الفرق بين الانحدار التدريجي العشوائي (**SGD**) والانحدار التدريجي (**GD**)؟ . 144

س345/ ما هو الانحدار التدريجي **gradient descent**؟ 144

س346/ ما الأسباب التي تجعل الانحدار التدريجي للدفعات الصغيرة **mini-batch gradient** مفيدًا جدًا؟ 145

س347/ما هو الانتشار الأمامي **forward propagation**؟ 145

س348/ما هو الانتشار الخلفي **backpropagation**؟ 145

س349/ هل يمكنك تدريب شبكة عصبية دون استخدام الانتشار الخلفي **backpropagation**؟ إذا كانت الإجابة بنعم، فما التقنية التي ستستخدمها لتحقيق ذلك؟ 146

س350/ لماذا لا نرى مشكلة انفجار الانحدار أو تلاشي الانحدار في الشبكات العصبية ذات التغذية الأمامية **FNN**؟ 146

س351/ اشرح الانتشار الأمامي **forwarding propagation** والانتشار الخلفي **backpropagation** في سياق التعلم العميق؟..... 146

س352/اشرح تسوية البيانات **Data Normalization**؟ ما هي الحاجة لذلك؟..... 147

س353/ما هي التقنيات المختلفة لتحقيق تسوية البيانات **data normalization**؟ 147

س354/ هل توجد أي مشاكل عند استخدام التسوية بالدفعات **Batch Normalization** في الشبكات العصبية العميقة؟ 147

س355/لماذا يجب أن نستخدم التسوية بالدفعات **Batch Normalization**؟ 148

س356/ ما هو التنظيم **regularization** وأين يمكن أن يكون مفيدًا؟ ما هو مثال على استخدام التنظيم في نموذج؟..... 148

س357/متى يجب تفضيل انحدار **ridge** على **lasso**؟ 148

س358/اشرح الفرق بين **Lasso** و**Ridge**؟ 149

س359/ما الفرق بين التنظيم **regularization** والتسوية **normalization**؟ 149

س360/اشرح الفرق بين التسوية **Normalization** والتوحيد **Standardization**؟ 149

س361/ما هي مزايا استخدام **Batchnorm**؟ 150

س362/ما هي فوائد استخدام التسوية بالدفعات **batch normalization** عند تدريب شبكة عصبية؟ 150

س363/ما هي إعادة تحجيم البيانات **Data Rescaling** وكيف يتم ذلك؟ 150

س364/كيف يتم توحيد البيانات **Data Standardization**؟..... 150

س365/لماذا التسوية **normalization** مهمة؟..... 151

س366/ما هي المعلمات الفائقة **hyperparameters** في التعلم العميق؟..... 151

س367/كيف يمكن تدريب المعلمات الفائقة في الشبكات العصبية؟ 151

س368/ هل يمكنك تسمية بعض المعلمات الفائقة **hyperparameters** المستخدمة لتدريب الشبكة العصبية؟..... 151

س370/ ما هي المعلمات الفائقة لنموذج الانحدار اللوجستي؟ 153
س371/ ما هي دوال التنشيط activation functions؟ 153
س372/ كيف تختار دالة التنشيط Activation Function لنموذج التعلم العميق؟ 153
س373/ ماذا يحدث إذا استخدمنا دالة التنشيط ReLU ثم sigmoid كطبقة نهائية؟ .. 153
س374/ ماذا يحدث إذا لم تستخدم أي دوال تنشيط في الشبكة العصبية؟ 154
س375/ ماهي دالة التنشيط Activation Function؟ 154
س376/ ما هو الفرق بين دالة التنشيط الخطية ودالة التنشيط غير الخطية؟ 155
س377/ ما هي خوارزميات التحسين optimization algorithms المختلفة التي نستخدمها بشكل عام في الشبكة العصبية؟ 155
س378/ ماذا تقصد بـ RMS Prop؟ 155
س379/ ما هو الفرق بين Adagrad و Adadelta و Adam؟ 155
س380/ محسن RMSProp و Adam يضبطان التدرجات؟ هل هذا يعني أنهم يؤدون gradient clipping؟ 156
س381/ لماذا نستخدم بشكل عام دالة Softmax غير الخطية كعملية أخيرة في الشبكة؟ 156
س382/ ما الفرق بين دوال Sigmoid و Softmax؟ 156
س283/ أيهما تعتقد أنه أقوى: شبكات عصبية من طبقتين بدون أي دالة تنشيط أو شجرة قرار من طبقتين؟ 157
س384/ ما هي أهمية استخدام دالة التنشيط Non-linear Activation function غير الخطية؟ 157
س385/ لماذا من المهم إدخال اللاخطية non-linearities في الشبكة العصبية؟ 158
س386/ لماذا لا يُفضل استخدام Sigmoid أو Tanh كدالة تنشيط في الطبقة المخفية للشبكة العصبية؟ 158
س387/ لماذا تُستخدم دالة Leaky ReLU في التعلم العميق؟ 158
س388/ ما هي دوال التنشيط التي استخدمتها حتى الآن في مشاريعك وكيف تختار واحدة. 159
س389/ ما مدى فائدة الحذف العشوائي dropout في نماذج التعلم العميق؟ هل يسرع أم يبطئ من عملية التدريب ولماذا؟ 159
س390/ اشرح لماذا يعمل الحذف العشوائي dropout في الشبكة العصبية كمنظم regularizer؟ 159
س391/ ما الفرق بين Dropout وDrop Connect؟ 159
س392/ هل يمكننا استخدام الحذف العشوائي dropout في مجموعة الاختبار test set ؟ 160

س393/ ما معنى الحذف العشوائي dropout في التعلم العميق؟ 160
س394/ ما هو التوقف المبكر Early stopping في التعلم العميق؟..... 160
س395/ ما هو التنظيم Regularization وما نوع المشاكل التي يحلها التنظيم؟..... 160
س396/ ما هو التنظيم Regularization وما نوع المشاكل التي يحلها التنظيم؟..... 161
س397/ لماذا يتسبب تنظيم L1 في تباين المعلمات بينما لا يؤدي تنظيم L2 إلى ذلك؟ 161
س398/ صف كيف يمكن تفسير تنظيم L2 كنوع من تناقص الوزن weight decay ؟ .. 162
س399/ إذا كان لدينا خطأ تحيز كبير high bias error ماذا يعني ذلك؟ كيف نعالجها؟ . 162
س400/ كيف يمكننا معرفة ما إذا كانت بياناتك تعاني من انحياز منخفض low bias وتباين كبير high variance؟..... 162
س401/ لنفترض أنك وجدت أن نموذجك يعاني من تباين كبير high variance. ما الخوارزمية التي تعتقد أنها يمكنها التعامل مع هذا الموقف ولماذا؟..... 162
س402/ اشرح سبب أهمية عملية التهيئة للأوزان والتحيزات للشبكات العصبية Neural Networks (NN)؟ 163
س403/ كيف تحصل الشبكات العصبية على القيم المثلى للأوزان والتحيز؟..... 163
س404/ ما هي أنواع التحيزات biases التي يمكن أن تحدث أثناء أخذ العينات sampling؟. 164
س405/ ما هو survivorship bias؟..... 164
س406/ لماذا نحتاج إلى selection bias؟ 164
س407/ ما هو Cluster Sampling؟ 164
س408/ ما هو Systematic Sampling؟ 165
س409/ ما هي طرق أخذ العينات المختلفة sampling methods؟..... 165
س411/ ماذا يجب أن تفعل عندما يعاني نموذجك من انحياز منخفض low bias وتباين كبير high variance؟ 165
س412/ ما المقصود بالتحيز Bias والتباين Variance وماذا تقصد بموازنة التحيز والتباين Bias-Variance Tradeoff؟..... 165
س413/ هل التباين الكبير في البيانات جيد أم سيئ؟..... 166
س414/ قارن بين التحيز bias والتباين variance فيما يتعلق بنماذج التعلم العميق وكيف يمكنك تحقيق التوازن بين الاثنين؟..... 166
س415/ ما هو دور الأوزان والتحيز في الشبكة العصبية؟..... 167
س416/ ما الذي يجعل التهيئة الصفرية zero initialization ليست عملية تهيئة جيدة للوزن؟..... 168
س417/ اشرح معنى مصطلح تهيئة الوزن weight initialization في الشبكات العصبية؟ 168

س418/ هل يمكنك تدريب نموذج الشبكة العصبية عن طريق تهيئة جميع التحيزات biases على أنها 0؟ 169

س419/ ما هو تأثير نموذج مع معدل التعلم المحدد بشكل غير صحيح على الأوزان؟... 169

س420/ ما هي الأساليب المستخدمة بشكل شائع لضبط معدل التعلم learning rate؟169

س421/ هل تجربة معدلات التعلم التالية: 0.2، 0.1، ... ، 0.5 استراتيجية جيدة لتحسين معدل التعلم؟ 169

س422/ ماذا تفهم من خلال معدل التعلم learning rate في نموذج الشبكة العصبية؟ ماذا يحدث إذا كان معدل التعلم مرتفعًا جدًا أو منخفضًا جدًا؟ 170

س423/ ما هي بعض الاختلافات التي قد تتوقعها في نموذج يقلل من الخطأ التربيعي squared error، مقابل نموذج يقلل من الخطأ المطلق absolute error؟ في أي الحالات يكون كل مقياس خطأ مناسبًا؟ 170

س424/ ما هو الفرق بين الخطأ التربيعي squared error والخطأ المطلق absolute error؟ 171

س425/ كيف تختار دالة الخطأ Loss Function لنموذج التعلم العميق؟..... 171

س426/ لماذا يُفضل الانتروبيا المتقاطعة cross-entropy كدالة تكلفة لمشاكل التصنيف متعدد الفئات multi-class classification؟..... 172

س427/ ما هي دالة الخطأ Loss Function المفضلة للتصنيف متعدد الفئات؟ 172

س428/ اشرح دالة الخطأ cross-entropy؟..... 172

س429/ الفرق بين الخطأ error والخطأ المتبقي residual error؟..... 172

س430/ ما المقصود بخطأ التدريب Training Loss وخطأ التحقق من الصحة Validation Loss؟..... 172

س431/ عند تدريب شبكة عصبية، لاحظ أن الخطأ لا يقل في الفترات الأولى. ما هي الأسباب الأصلية لذلك؟ 173

س432/ ما فائدة دالة الخطأ loss function؟ 173

س433/ على أي نوع من المشاكل يمكن تطبيق دالة خطأ الانتروبيا Cross-Entropy؟ . 173

س434/ إن معالجة المتغير الفئوي كمتغير مستمر من شأنه أن يؤدي إلى نموذج تنبؤي أفضل؟..... 174

س435/ ماذا تقصد بالمتغير الوهمي Dummy Variable؟ أين يتم استخدامه في التعلم الآلي؟..... 174

س436/ نحن نعلم أن one-hot encoding يزيد من أبعاد مجموعة البيانات، لكن label encoding لا يفعل ذلك. كيف؟ 174

س437/ اشرح one-hot encoding وlabel encoding. كيف تؤثر على أبعاد مجموعة البيانات المحددة؟..... 175

س438/ كيف يؤثر الترميز one-hot والترميز label على أبعاد مجموعة البيانات؟..... 175
س439/ ما المقصود بلعنة الأبعاد curse of dimensionality؟ 175
س440/ ما هي زيادة البيانات data augmentation؟ أعط أمثلة. 175
س441/ ما هو المقصود بـ "لعنة الأبعاد curse of dimensionality"؟ كيف يمكننا حل هذه المشكلة؟..... 176
س442/ هل من الممكن استخدام الشبكة العصبية كأداة لتقليل الأبعاد dimensionality reduction؟..... 176
س443/ اشرح تقليل الأبعاد dimension reduction في التعلم الآلي؟ 176
س444/ كيف تفلت الشبكة العصبية العميقة / تقاوم لعنة الأبعاد Curse of Dimensionality؟ 176
س445/ ما هي مكتبات تصوير ورسم البيانات Data Visualization التي تستخدمها؟.. 177
س446/ ما هو تحويل Box-Cox Transformation؟..... 177
س447/ ماذا تفهم بالرسم البياني الحسابي computational graph؟ 177
س448/ ما هي الطرق الجماعية Ensemble methods وكيف أنها مفيدة في التعلم العميق؟ 177
س449/ ما الفرق بين أخطاء Type1 و Type 2؟..... 178
س450/ ماذا تقصد بمنحنى AUC؟ 178
س451/ لماذا نستخدم دالة التلخيص summary؟ 178
س452/ كيف يرتبط الاستدعاء Recall بالمعدلات الإيجابية الحقيقية True positive؟ 179
س453/ كيف تختلف AUC عن ROC؟ 179
س454/ اشرح كيف يعمل منحنى ROC؟ 179
س455/ عرف الدقة precision والاستدعاء recall؟ 179
س456/ ما هي F1 Score؟ كيف يمكنك أن تستفيد منها؟..... 180
س457/ كيف تجد RMSE و MSE في نموذج الانحدار الخطي؟ 180
س458/ صِف الحالة الذي ستستخدم فيه MSE كمقياس للجودة؟ 181
س459/ أعط بعض الأمثلة عن الإيجابية الخاطئة FP، السلبية الخاطئة FN، الإيجابية الحقيقية TP، السلبية الحقيقية TN؟..... 181
س460/ اشرح TP وTN و FP و FN في مصفوفة الارتباك Confusion Matrix بمثال؟ .. 182
س461/ ما هي مصفوفة الارتباك Confusion Matrix؟..... 182
س462/ لماذا نحتاج مجموعات بيانات التحقق من الصحة validation والاختبار test؟... 183
س463/ ما هو SVM في التعلم الآلي؟ 183

س464/اشرح الفرق بين KNN وK-mean Clustering؟ 183
س465/ما هي النوى (kernels) المختلفة الموجودة في SVM؟ 183
س466/هل يمكنك ذكر بعض مزايا وعيوب أشجار القرار decision trees؟ 184
س467/هل من الممكن استخدام KNN لمعالجة الصور؟ 184
س468/ما هو نايف بايز Naive Bayes؟ 185
س469/ما هي أهمية جاما Gamma والتنظيم Regularization في SVM؟ 185
س470/ كيف تحدد عدد المجموعات number of clusters في خوارزمية التجميع clustering algorithm؟ 185
س471/ما هي الطريقة الافتراضية لتقسيم أشجار القرار decision trees؟ 185
س472/ما هي اهم المعلمات الفائقة لـ SVM؟ 185
س473/ما هي مقاييس المسافة التي يمكن استخدامها في KNN؟ 185
س474/ما هي التقنية الجماعية ensemble technique التي تستخدمها Random Forest؟ 186
س475/ما هي فوائد التقليم pruning في خوارزمية شجرة القرار؟ 186
س476/أيهما أفضل، خوارزمية نايف بايز أم أشجار القرار؟ 186
س477/ما هي مزايا استخدام نايف بايز للتصنيف؟ 186
س478/ما هي متجهات الدعم Support Vectors في SVM؟ 186
س479/ما هو الفرق بين Gini Impurity وEntropy في شجرة القرار؟ 187
س480/كيف يتم تقليم شجرة القرار؟ 187
س481/كيف يمكنك اختيار k لـ k-means؟ 187
س482/ما هي الاختلافات بين تحليلات univariate وbivariate وmultivariate؟ 187
س483/ما هي p-value؟ 188
س484/ماذا تعني قيمة p-value P حول البيانات الإحصائية statistical data؟ 188
س485/ماذا تعني كلمة "Naive" في Naive Bayes؟ 188
س486/اشرح خطوات بناء شجرة القرار decision tree؟ 188
س487/كيف تبني نموذج غابة عشوائي random forest model؟ 189
س488/ما هو uniform distribution وskewed Distribution؟ 189
س489/ كيف يمكنك التعامل مع الأنواع المختلفة من seasonality في نمذجة السلاسل الزمنية؟ 190
س490/في التصميم التجريبي، هل من الضروري عمل التوزيع العشوائي randomization؟ إذا كانت الإجابة نعم، فلماذا؟ 190

س491/ ما هو الانتروبيا **Entropy** واكتساب المعلومات **Information Gain** في خوارزمية شجرة القرار؟ 190
س492/ متى ستستخدم **SVM** ومتى تستخدم **Random Forest**؟..... 191
س493/ ما هي المعرفة الرياضية الأساسية وراء **Naïve Bayes**؟..... 192
س494/ متى تستخدم **Random Forest** ومتى تستخدم **XGBoost**؟..... 192
س495/ ما هو الهامش **margin**، النواة **kernels**، التنظيم **Regularization** في **SVM**؟ 192
س496/ ما هي الطريقة المختلفة لتقسيم الشجرة في شجرة القرار؟..... 192
س497/ ما هو ضعف خوارزمية شجرة القرار؟ 192
س498/ ما هي **hyperplane** في **SVM**؟..... 193
س499/ الغابة العشوائية بها **1000** شجرة، خطأ تدريب: **0.0** وخطأ التحقق **20.00**، ما المشكلة هنا؟ 193
س500/ ما هي نظرية بايز **Bayes's Theorem** وكيف يتم استخدامها في التعلم الآلي؟ . 193
س501/ ما هي الحاجة إلى إضافة العشوائية **randomness** في عملية تهيئة الوزن؟ 193
س502/ كيف يؤثر قطع الاتصال العشوائي **Randomized Connection Dropping** على نموذج التعلم العميق؟..... 194
س503/ هل نايف بايز **Naïve Bayes** سيئة؟ إذا كانت الإجابة بنعم، فما هي الجوانب..... 194
س504/ أيهما أفضل – غابة عشوائية **random forest** أم أشجار قرارات متعددة **multiple decision trees**؟ 194
س505/ لماذا أداء **XGBoost** أفضل من **SVM**؟ 194
س506/ ما هي أفكارك حول استخدام **GPT 3** لعملك؟..... 195
س507/ ما الرؤية الحاسوبية **Computer Vision** وما علاقته بالذكاء الاصطناعي وماهي تطبيقاته؟ 195
س508/ ما هو الفرق بين اكتشاف الوجه **Face Detection** والتعرف عليه **Face Recognition**؟ 196
س509/ ما هو **YOLOv3**؟..... 196
س510/ هل الرؤية الحاسوبية تستخدم التعلم العميق؟..... 197
س511/ هل **OpenCV** تعلم الآلي أم تعلم عميق؟ 197
س512/ ما هي الآثار الأخلاقية لاستخدام التعلم العميق؟..... 197

س1/ اشرح مفهوم التعلم الآلي Machine Learning (ML) وافترض أنك تشرح ذلك لطفل يبلغ من العمر 5 سنوات؟

ج / نعم، السؤال نفسه هو الجواب.

التعلم الآلي هو بالضبط نفس الطريقة التي يمارس بها الأطفال أنشطتهم اليومية، والطريقة التي يمشون بها أو ينامون، وما إلى ذلك. ومن الحقائق الشائعة أن الأطفال لا يستطيعون المشي بعيدًا ويسقطون ثم يستيقظون مرة أخرى ثم يحاولون. هذا هو الشيء نفسه عندما يتعلق الأمر بالتعلم الآلي، فالأمر كله يتعلق بكيفية عمل الخوارزمية وفي نفس الوقت إعادة تحديد كل مرة للتأكد من أن النتيجة النهائية مثالية قدر الإمكان.

س2/ كيف أبدأ مسيرتي المهنية في التعلم الآلي؟

ج / لا يوجد دليل ثابت أو نهائي يمكنك من خلاله بدء حياتك المهنية في التعلم الآلي. تتمثل الخطوة الأولى في فهم المبادئ الأساسية للموضوع وتعلم بعض المفاهيم الأساسية مثل الخوارزميات وهياكل البيانات وقدرات البرمجة وحساب التفاضل والتكامل والجبر الخطي والإحصاء. لتحليل البيانات بشكل أفضل، يجب أن يكون لديك فهم واضح لإحصاءات التعلم الآلي. ستكون الخطوة التالية هي الالتحاق بدورة ML أو قراءة أفضل الكتب للتعلم الآلي. يمكنك أيضًا العمل في مشاريع للحصول على خبرة عملية.

س3/ ما مدى صعوبة التعلم الآلي؟

ج / التعلم الآلي هو مفهوم واسع يحتوي على الكثير من الجوانب المختلفة. مع التوجيه الصحيح والعمل الجاد المتسق، قد لا يكون من الصعب جدًا التعلم. إنه بالتأكيد يتطلب الكثير من الوقت والجهد، ولكن إذا كنت مهتمًا بالموضوع وترغب في التعلم، فلن يكون الأمر صعبًا للغاية.

س4/ كيف يمكنك الدخول الى مجال التعلم الآلي؟

ج / الطريقة الأكثر شيوعًا للدخول في مهنة التعلم الآلي هي اكتساب المهارات اللازمة. تعلم لغات البرمجة مثل **C** و **C ++** و **Python** و **Java**. اكتساب المعرفة الأساسية حول خوارزميات التعلم الآلي المختلفة، والمعرفة الرياضية حول حساب التفاضل والتكامل والإحصاء. سيساعدك هذا على قطع شوط طويل.

س5/ ما هي أفضل طريقة لتعلم التعلم الآلي؟

ج / يمكن اعتبار أي طريقة تناسب أسلوبك في التعلم أفضل طريقة للتعلم. قد يستمتع الأشخاص المختلفون بطرق مختلفة. تتمثل بعض الطرق الشائعة في اتباع أساسيات دورة التعلم الآلي مجانًا، ومشاهدة مقاطع فيديو **YouTube**، وقراءة المدونات ذات الموضوعات ذات الصلة، وقراءة الكتب التي يمكن أن تساعدك على التعلم الذاتي.

س6/ ما هو التعلم الآلي للمبتدئين؟

ج / سيتألف التعلم الآلي للمبتدئين من المفاهيم الأساسية مثل أنواع التعلم الآلي (التعلم الخاضع للإشراف، وغير الخاضع للإشراف، والتعلم المعزز). كل نوع من هذه الأنواع من **ML** لها خوارزميات ومكتبات مختلفة داخلها، مثل التصنيف والانحدار. هناك العديد من خوارزميات التصنيف وخوارزميات الانحدار مثل الانحدار الخطي. سيكون هذا أول شيء ستتعلمه قبل المضي قدمًا في المفاهيم الأخرى.

س7/ هل يتطلب التعلم الآلي البرمجة؟

ج / البرمجة هي جزء من التعلم الآلي. من المهم معرفة لغات البرمجة مثل بايثون.

س8/ ما الدرجة والشهادة العلمية التي تحتاجها للتعلم الآلي؟

ج / ستبحث معظم شركات التوظيف عن درجة الماجستير أو الدكتوراه في المجال ذي الصلة. يشمل مجال الدراسة علوم الكمبيوتر أو الرياضيات. لكن امتلاك المهارات اللازمة حتى بدون الدرجة يمكن أن يساعدك في الحصول على وظيفة **ML** أيضًا.

س9/ كيف يكون التعلم العميق أفضل من التعلم الآلي؟

ج/ يعد التعلم الآلي قويًا بطريقة تكفي لحل معظم المشكلات. ومع ذلك، فإن التعلم العميق له اليد العليا عندما يتعلق الأمر بالعمل مع البيانات التي تحتوي على عدد كبير من الأبعاد. باستخدام البيانات الكبيرة الحجم، يمكن لنموذج التعلم العميق العمل معها بسهولة لأنها مصممة للتعامل مع ذلك.

س10/ ما هي الأنواع المختلفة للشبكات العصبية العميقة؟

ج / هناك أنواع مختلفة للشبكات العصبية العميقة: -

- Convolutional Neural Networks (CNNs)
- Long Short-Term Memory Networks (LSTMs)
- Recurrent Neural Networks (RNNs)

- Generative Adversarial Networks (GANs)
- Radial Basis Function Networks (RBFNs)
- Multilayer Perceptrons (MLPs)
- Self-Organizing Maps (SOMs)
- Deep Belief Networks (DBNs)
- Restricted Boltzmann Machines (RBMs)

س11/ ما الفرق بين التعلم الآلي Machine Learning (ML) والتعلم العميق Deep Learning (DL)؟

ج/ يشكل التعلم الآلي مجموعة فرعية من الذكاء الاصطناعي، حيث نستخدم الإحصائيات والخوارزميات لتدريب الآلات بالبيانات، وبالتالي مساعدتها على التحسن من خلال التجربة.

التعلم العميق هو جزء من التعلم الآلي، والذي يتضمن محاكاة الدماغ البشري من حيث الهياكل تسمى الخلايا العصبية، وبالتالي تكوين الشبكات العصبية.

	التعلم الآلي	التعلم العميق
تبعيات البيانات	يعمل بشكل أفضل على مجموعات البيانات الصغيرة والمتوسطة	يعمل بشكل أفضل لمجموعات البيانات الكبيرة
تبعيات الأجهزة	العمل على الآلات المنخفضة الجودة	يتطلب آلة قوية ، ويفضل أن يكون مع GPU
التفسير	من السهل تفسير الخوارزميات	صعب التفسير
وقت التنفيذ	من بضع دقائق إلى ساعات	قد يستغرق الأمر ما يصل إلى أسبوع
هندسة الميزات	تحتاج إلى فهم الميزات التي تمثل البيانات	لا حاجة لفهم أفضل ميزة تمثل البيانات

س12/ ما هي بعض التطبيقات الأكثر استخدامًا في التعلم العميق؟

ج/ يستخدم التعلم العميق في مجموعة متنوعة من المجالات اليوم. الأكثر استخدامًا هي كما يلي:

- ❖ تحليل المشاعر Sentiment Analysis
- ❖ رؤية الكمبيوتر Computer Vision
- ❖ توليد النص التلقائي Automatic Text Generator
- ❖ كشف الكائن Object Detection
- ❖ معالجة اللغات الطبيعية Natural Language Processing (NLP)

- التعرف على الصور Image Detection

س13/ ما الفرق بين القيمة المتوسطة mean value والقيمة المتوقعة expected value؟

ج/ إنها ليست مختلفة ولكن المصطلحات تستخدم في سياقات مختلفة. يشار إلى القيمة المتوسطة **mean value** عمومًا عند الحديث عن توزيع احتمالي أو عينة من السكان بينما تتم الإشارة إلى القيمة المتوقعة **expected value** عمومًا في سياق متغير عشوائي.

- لأخذ العينات البيانات **For Sampling Data**

القيمة المتوسطة هي القيمة الوحيدة التي تأتي من بيانات أخذ العينات **Sampling**.

القيمة المتوقعة هي متوسط جميع الوسائل، أي القيمة المبنية من عينات متعددة. القيمة المتوقعة هي متوسط المجتمع.

- للتوزيعات **For Distributions**

القيمة المتوسطة والقيمة المتوقعة هي نفسها بغض النظر عن التوزيع، بشرط أن يكون التوزيع في نفس المجتمع.

س14/ ما هي بعض قيود التعلم العميق؟

ج/ هناك بعض قيود التعلم العميق كما هو مذكور أدناه:

- تتطلب الشبكات في التعلم العميق كمية هائلة من البيانات للتدريب الجيد.
- يمكن أن تكون مفاهيم التعلم العميق معقدة للتنفيذ في بعض الأحيان.
- من الصعب تحقيق قدر كبير من كفاءة النموذج في كثير من الحالات.

س15/ ما هي وظائف التعلم الخاضع للإشراف Supervised Learning؟

ج/

1. التصنيف **Classification**
2. التعرف على الكلام **Speech Recognition**
3. الانحدار **Regression**
4. توقع السلاسل الزمنية **Predict Time Series**

س16/ ما هي وظائف التعلم غير الخاضع للإشراف Unsupervised Learning؟

ج/

1. البحث عن مجموعات البيانات **Finding clusters of the data**
2. إيجاد تمثيلات منخفضة الأبعاد للبيانات **Finding low-dimensional representations of the data**
3. العثور على اتجاهات مثيرة للاهتمام في البيانات **Finding interesting directions in data**
4. البحث عن ملاحظات جديدة / تنظيف قاعدة البيانات **Finding novel observations/ database cleaning**
5. إيجاد إحداثيات وارتباطات مثيرة للاهتمام **Finding interesting coordinates and correlations**

س17/ ما هي بعض الأمثلة على خوارزميات التعلم الخاضعة للاشراف supervised learning algorithms في التعلم العميق؟

ج/ هناك ثلاث خوارزميات تعلم رئيسية خاضعة للإشراف في التعلم العميق:

- الشبكات العصبية الاصطناعية Artificial neural networks
- الشبكات العصبية التلافيفية Convolutional neural networks
- الشبكات العصبية المتكررة Recurrent neural networks

س18/ ما هي بعض الأمثلة على خوارزميات التعلم غير الخاضع للاشراف unsupervised learning algorithms في التعلم العميق؟

ج/ هناك ثلاث خوارزميات تعلم رئيسية غير خاضع للإشراف في التعلم العميق:

- المشفرات التلقائية Autoencoders
- آلات بولتزمان Boltzmann machines
- خرائط التنظيم الذاتي Self-organizing maps

س19/ كيف تختار الصيغة المناسبة appropriate formula لحل مشاكل التصنيف؟

ج/ يعد اختيار المقاييس والصيغة الصحيحة لتصنيف بياناتك أمرًا في غاية الأهمية لفهم النموذج وتحسينه. استخدم إجراء تقييم النموذج للاختيار من بين أنواع النماذج المختلفة والميزات ومعلمات الضبط **tuning parameters**. قم بتدريب واختبار هذه النماذج على نفس مجموعة البيانات، أو قم باختبار النماذج أو تحقق من صحة النماذج من خلال مقارنة متوسط النتائج مع نتائج تجزئة الاختبار.

س20/ ما الفرق بين مجموعة بيانات التدريب Training ومجموعة بيانات الاختبار Testing ومجموعة بيانات التحقق Validation؟ ما هي النسبة العامة لهذه المجموعات؟

ج/

- مجموعة بيانات التدريب **Training Dataset**

عينة البيانات المستخدمة لتناسب النموذج.

مجموعة البيانات الفعلية التي نستخدمها لتدريب النموذج. النموذج يرى ويتعلم من هذه البيانات.

- مجموعة بيانات التحقق **Validation Dataset**

عينة البيانات المستخدمة لتقديم تقييم غير متحيز لنموذج مناسب لمجموعة بيانات التدريب أثناء ضبط معلمات النموذج الفائقة **hyperparameters**. يصبح التقييم أكثر تحيزًا حيث يتم دمج المهارة في مجموعة بيانات التحقق في تكوين النموذج.

- مجموعة بيانات الاختبار **Test Dataset**

عينة البيانات المستخدمة لتقديم تقييم غير متحيز لنموذج نهائي ملائم لمجموعة بيانات التدريب.

بشكل عام، يتم تقسيم مجموعة بيانات التدريب ومجموعة بيانات التحقق ومجموعة بيانات الاختبار بنسبة **60**٪ و**20**٪ و**20**٪ على التوالي. في بعض الأحيان قد تكون **80**٪ و**20**٪ لمجموعات بيانات التدريب والاختبار على التوالي.

س21/ لماذا الشبكات العميقة Deep Networks أفضل من الشبكات الضحلة Shallow Networks؟

ج/ تحتوي كل شبكة عصبية على طبقة مخفية مع طبقات الإدخال والإخراج. تُعرف الشبكات العصبية التي تستخدم طبقة مخفية واحدة بالشبكات العصبية الضحلة بينما يُشار إلى الشبكات التي تستخدم طبقات مخفية متعددة باسم الشبكات العصبية العميقة. كل من الشبكات الضحلة والعميقة قادرة على التكيف مع أي وظيفة ولكن الشبكات الضحلة تتطلب الكثير من المعلمات، على عكس الشبكات العميقة التي يمكن أن تناسب الوظائف حتى مع عدد محدود من المعلمات بسبب طبقات متعددة. تُفضل الشبكات العميقة اليوم على الشبكات الضحلة لأنه في كل طبقة يتعلم النموذج تمثيلًا جديدًا ومجرّدًا للمدخلات. كما أنها أكثر كفاءة من حيث عدد المعلمات والحسابات مقارنة بالشبكات الضحلة.

يمكن لكل من الشبكات العصبية العميقة والضحلة تقريب قيم الوظيفة. لكن الشبكة العصبية العميقة أكثر كفاءة لأنها تتعلم شيئًا جديدًا في كل طبقة. تحتوي الشبكة العصبية الضحلة على طبقة مخفية واحدة فقط. لكن الشبكة العصبية العميقة لديها العديد من الطبقات المخفية التي تخلق تمثيلاً أعمق وقدرة حسابية.

س22/ ما هي الاختلافات بين المتوسط Mean والوسيط Median والمنوال Mode؟ ما مدى فائدة هذه في التعامل مع القيم المفقودة في مجموعة البيانات المحددة؟

ج/

- المتوسط **Mean**

المتوسط هو متوسط مجموعة البيانات. هي نسبة مجموع المشاهدات الإجمالية إلى إجمالي عدد المشاهدات.

- الوسيط **Median**

إنها القيمة المتوسطة لمجموعة البيانات. إذا كان العدد الإجمالي للملاحظات في مجموعة البيانات فرديًا في العدد ، فإن الوسيط هو القيمة المتوسطة أو الملاحظة. إذا كان العدد الإجمالي للملاحظات في مجموعة البيانات عددًا زوجيًا ، فسيتم إعطاء الوسيط بمتوسط القيمتين الأوسطتين لمجموعة البيانات.

- المنوال **Mode**

المنوال هو الملاحظة أو القيمة الأكثر تكرارًا في مجموعة البيانات بأكملها.

إذا تم توزيع البيانات بشكل طبيعي **Normally distributed**، فيمكن احتساب القيم المفقودة أو استبدالها بمتوسط جميع ملاحظات مجموعة البيانات.

إذا كانت البيانات منحرفة **skewed**، فمن الأفضل احتساب أو استبدال القيم المفقودة بواسطة الوسيط لجميع ملاحظات مجموعة البيانات.

عندما تكون البيانات مرتبطة بالتردد **frequency**، يتم استخدام المنوال **Mode** لاستبدال أو تحديد القيم المفقودة في مجموعة البيانات المعنية.

س23/ ما هي Backend المختلفة المتوفرة في Keras؟

ج/

- TensorFlow
- Theano
- CNTK

س24/ ما هي بعض أطر عمل framework التعلم العميق أو الأدوات التي تُستخدم في التعلم العميق؟

ج/ أفضل أطر عمل التعلم framework العميق الموجودة اليوم هي:

- TensorFlow
- Keras
- PyTorch
- Caffe2
- CNTK
- MXNet
- Theano

س25/ ما هي الموترات tensors؟

ج/ الموترات tensors عبارة عن مصفوفات متعددة الأبعاد في التعلم العميق تُستخدم لتمثيل البيانات. يمثلون البيانات ذات الأبعاد الأعلى. نظرًا لطبيعة لغات البرمجة عالية المستوى، يمكن فهم بنية التنسرات بسهولة واستخدامها على نطاق واسع.

س26/ ما هي العناصر القابلة للبرمجة في TensorFlow؟

ج/ في TensorFlow، يمكن للمستخدمين برمجة ثلاثة عناصر:

- Constants
- Variables
- Placeholders

س27/ ما هي بعض مزايا استخدام TensorFlow؟

ج/ يتمتع TensorFlow بالعديد من المزايا، ومنها ما يلي:

- قدر كبير من المرونة واستقلالية النظام الأساسي.
- التدريب باستخدام CPU او GPU.
- يدعم التمايز التلقائي ومميزاته
- يعالج الخيوط threads والحساب غير المتزامن asynchronous computation بسهولة
- مفتوح المصدر.
- لديها مجتمع كبير.

س28/ ماذا تقصد بـ Tensorboard؟

ج/ **Tensorboard** هي الواجهة المستخدمة للتصوير والرسم البياني والأدوات الأخرى لفهم النموذج وتصحيحه وتحسينه. تزودنا **Tensorboard** بمجموعة من تطبيقات الويب التي تساعدنا في فحص وفهم عمليات تشغيل **TensorFlow** والرسوم البيانية. حاليًا، يوفر خمسة أنواع من التصورات: الكميات القياسية **scalar** والصور والصوت والمدرجات التكرارية **histograms** والرسوم البيانية **graphs**.

س29/ هل تعتقد أن التعلم العميق يؤدي بشكل أفضل من التعلم الآلي؟ إذا كان الأمر كذلك لماذا؟

ج/ يؤدي التعلم العميق إلى تنفيذ تقنيات أخرى إذا كان حجم البيانات كبيرًا. ولكن مع حجم البيانات الصغير، يفضل استخدام خوارزميات التعلم الآلي التقليدية. تحتاج تقنيات التعلم العميق إلى بنية تحتية متطورة للتدريب في وقت معقول.

س30/ ما هي الرياضيات المستخدمة في التعلم الآلي والعميق؟

ج/ سوف تحتاج إلى معرفة المفاهيم الإحصائية statistical concepts والجبر الخطي linear algebra والاحتمالية probability وحساب التفاضل والتكامل متعدد المتغيرات Multivariate Calculus والتحسين Optimization. عندما تتعمق في المفاهيم الأكثر تعمقًا للتعلم العميق، ستحتاج إلى مزيد من المعرفة فيما يتعلق بهذه الموضوعات.

س31/ هل يمكنني دراسة التعلم العميق قبل تعلم الآلة؟

ج/ نعم، بالتأكيد يمكنك ذلك، ولكن البدء بالتعلم الآلي يعطيك فهم أعمق لكيفية عمل الخوارزميات الخاصة بال.

س32/ اذكر مزايا وعيوب استخدام الشبكات العصبية Neural Networks؟

ج/

مزايا الشبكات العصبية:

- الكفاءة
- التعلم المستمر
- تعدد المهام
- تطبيقات واسعة

مساوئ الشبكات العصبية:

- الاعتماد على الأجهزة

- ❖ خوارزميات معقدة
- ❖ طبيعة الصندوق الأسود
- ❖ النتائج التقريبية
- ❖ الاعتماد على البيانات

س33/ لقد قمت ببناء نموذج DL وأثناء التدريب لاحظت أنه بعد عدد معين من الفترات، تتناقص الدقة. ما هي المشكلة وكيف يتم حلها؟

ج / يشير الانخفاض في دقة نموذج التعلم العميق بعد بضع فترات إلى أن النموذج يتعلم من خصائص مجموعة البيانات ولا يأخذ في الاعتبار الميزات **features**. يشار إلى هذا باسم **الضبط الزائد overfitting** لنموذج التعلم العميق. يمكنك إما استخدام التوقف المبكر **early stopping** أو الحذف العشوائي **dropout** لإصلاح هذه المشكلة. التوقف المبكر لأن العبارة تعني التوقف عن تدريب نموذج التعلم العميق أكثر في اللحظة التي تلاحظ انخفاضًا في عدم دقة النموذج. الحذف العشوائي هو تقنية يتم فيها إسقاط عدد قليل من العقد أو طبقات الإخراج بحيث يكون للعقد المتبقية أوزان مختلفة.

س34/ ما هي أهم الانتقادات والقيود المفروضة على التعلم العميق؟

ج /

- ❖ لا يعمل على عينات البيانات الصغيرة.
- ❖ يتطلب الأمر قدرًا كبيرًا من الخبرة الفنية والرياضية لإعداد واحد جيدًا من الصفر (أو بناء وضبط بشكل صحيح في R / Python).
- ❖ إنه صندوق أسود black box ما لم تبدأ في فتح الطبقات لترى كيف تتم معالجة البيانات - فلن يخبرك كيف ترتبط x بـ y وحجم هذه العلاقة.

س35/ ما هي المتغيرات الفئوية categorical variables؟

ج / المتغير الفئوي (يسمى أحيانًا المتغير الاسمي **nominal variable**) هو متغير يحتوي على فئتين أو أكثر، ولكن لا يوجد ترتيب جوهري للفئات. على سبيل المثال، المتغير الثنائي (مثل نعم / لا سؤال) هو متغير فئوي له فئتان (نعم أو لا) ولا يوجد ترتيب جوهري للفئات. لون الشعر هو أيضًا متغير قاطع له عدد من الفئات (أشقر، بني، بني، أحمر ، إلخ) ومرة أخرى ، لا توجد طريقة متفق عليها لترتيبها من الأعلى إلى الأدنى. المتغير الاسمي البحت هو المتغير الذي يسمح لك ببساطة بتعيين الفئات ولكن لا يمكنك ترتيب الفئات بوضوح. إذا كان للمتغير ترتيب واضح ، فسيكون هذا المتغير متغيرًا ترتيبيًا.

س36/ ما هي المقاييس التي يمكن استخدامها لقياس ارتباط correlation البيانات الفئوية categorical data؟

ج / يمكن استخدام اختبار **Chi Square** للقيام بذلك. يعطي مقياس الارتباط بين المتنبئين الفئوية.

س37/ ما الخوارزميات التي يمكن استخدامها لاختيار المتغير المهم important variable؟

ج / يمكن استخدام مخططات **Random Forest** و**Xgboost** ومخطط الرسوم البيانية ذات الأهمية المتغيرة للاختيار المتغير المهم.

س38/ كيف تختار الميزات المهمة important features في مجموعة البيانات المحددة given data set؟

ج/

- في الانحدار اللوجستي، يمكننا استخدام **()step** التي تعطي درجة **AIC** من مجموعة الميزات
- في شجرة القرار، يمكننا استخدام اكتساب المعلومات **information gain** (الذي يستخدم الانتروبيا داخليًا)
- في الغابات العشوائية، يمكننا استخدام **varImpPlot**

س39/ افترض أنك تعمل على مجموعة بيانات، واشرح كيف ستختار المتغيرات المهمة؟

ج / فيما يلي بعض الطرق التي يمكن استخدامها لتحديد المتغيرات المهمة:

- ❖ استخدام طريقة انحدار اللاسو **Lasso Regression**.
- ❖ باستخدام **Random Forest** ، ارسم مخططًا ذا أهمية متغيرة.
- ❖ باستخدام الانحدار الخطي **Linear regression**.

س40/ أثناء العمل على مجموعة بيانات، كيف يمكنك اختيار المتغيرات المهمة important variables؟

ج/ يمكنك استخدام الطرق التالية لاختيار المتغيرات المهمة:

- ❖ قم بإزالة المتغيرات المرتبطة قبل اختيار المتغيرات المهمة
- ❖ استخدم الانحدار الخطي وحدد المتغيرات التي تعتمد على تلك القيم p.

- استخدم التحديد الخلفي Backward والأمامي Forward والتحديد التدريجي Stepwise
- استخدم Xgboost وRandom Forest ورسم مخطط الأهمية المتغيرة.
- قم بقياس كسب المعلومات information gain لمجموعة معينة من الميزات وحدد أفضل الميزات وفقًا لذلك.

س41/ لماذا يحظى التعلم العميق بشعبية كبيرة ومطلوب هذه الأيام؟

ج / إنه مفيد للكثير من تحليلات واحتياجات البيانات الضخمة - رؤية الكمبيوتر Computer Vision، البرمجة اللغوية العصبية NLP... كما أنه يعمل بشكل جيد جدًا مع زيادة حجم البيانات، بينما تستقر العديد من الخوارزميات في مرحلة ما.

س42/ ما هي أفضل الكتب عن التعلم العميق؟

ج/ فيما يلي بعض الكتب الجيدة عن التعلم العميق:

- Deep Learning by Ian Goodfellow
- Neural Networks and Deep Learning by Michael Nielsen
- Deep Learning with Python by Francois Chollet
- TensorFlow 1.x Deep Learning Cookbook by Antonio Gulli, Amita Kapoor
- Deep Learning: A Practitioner's Approach by Adam Gibson, Josh Patterson
- Grokking Deep Learning by Andrew W. Trask
- Machine Learning Yearning by Andrew Ng

س43/ كيف يمكنك تغيير الشبكة العصبية من التوقع regression إلى التصنيف classification؟

ج / يمكن بسهولة تغيير الشبكة العصبية المصممة لمشكلة الانحدار إلى التصنيف.

يتطلب تغييرين في الكود:

- تغيير في طبقة الإخراج output layer.
- تغيير في دالة الخسارة loss function.

س44/ هناك العديد من خوارزميات التعلم الآلي حتى الآن. إذا أعطيت مجموعة بيانات dataset، فكيف يمكن للمرء تحديد الخوارزمية التي سيتم استخدامها لذلك؟

ج/ يتعين علينا اختيار خوارزمية التعلم الآلي المناسبة اعتماداً على بيان المشكلة ومجموعة البيانات ولا يوجد نموذج أفضل خوارزمية مقارنة بآخر.

- إذا كانت مشكلة انحدار regression، فاستخدم الانحدار الخطي، وأشجار القرار، والغابة العشوائية، و KNN، وما إلى ذلك.
- إذا كانت مشكلة تصنيف classification، فاستخدم الانحدار اللوجستي، والغابات العشوائية، و XGboost ، و AdaBoost ، و SVM ، وما إلى ذلك.
- إذا كان التعلم غير خاضع للإشراف، فاستخدم خوارزميات التجميع clustering مثل خوارزمية K-mean.

جرب هذه النماذج بناءً على المشكلة واستخدم مقاييس الأداء للتحقق من صحة النموذج الخاص بك ومعرفة النموذج المناسب لبيان مشكلتك. ولكن، إذا كنت في حيرة من أمرك بشأن الخيار الذي تختاره، فانتقل إلى نموذج Random Forest.

س45/ قم بتسمية بعض مكتبات التعلم الآلي لأغراض مختلفة؟

ج/

الغرض	المكتبة
الحساب العلمي Scientific Computation	Numpy
البيانات المجدولة Tabular Data	Pandas
نمذجة البيانات والمعالجة المسبقة Data Modelling & Preprocessing	Scikit Learn
معالجة النص Text processing	Regular Expressions, NLTK
التعلم العميق Deep Learning	Tensorflow, Pytorch

س46/ كيف يمكنك تطوير نموذج لتحديد الانتحال والسرقة الفكرية plagiarism؟

ج/ اتبع الخطوات أدناه لتطوير نموذج يحدد السرقة الفكرية plagiarism:

❖ قم بترميز Tokenise المستند.

- استخدم مكتبة **NLTK** في **Python** لإزالة كلمات الإيقاف **stopwords** من البيانات.
- قم بإنشاء **LDA** أو **SDA** للمستند ثم استخدم مكتبة **GenSim** لتحديد الكلمات الأكثر صلة، سطراً بسطر.
- استخدم **Google Search API** للبحث عن هذه الكلمات.

س47/ ما هي مراحل بناء نموذج في التعلم الآلي؟

ج/ لبناء نموذج في التعلم الآلي، عليك اتباع بعض الخطوات:

1. فهم نموذج العمل Understand the business model
2. اقتناء البيانات Data acquisitions
3. تنظيف البيانات Data cleaning
4. تحليل البيانات الاستكشافية Exploratory data analysis
5. استخدام خوارزميات التعلم الآلي لعمل نموذج
6. استخدام مجموعة بيانات غير معروفة unknown dataset للتحقق من دقة النموذج

س48/ هل يمكنك إخبارنا بكيفية تصميم فلتر للبريد الإلكتروني العشوائي Email Spam Filter؟

ج/

- فهم نموذج العمل: حاول فهم السمات ذات الصلة بالبريد العشوائي
- اقتناء البيانات: اجمع البريد العشوائي لقراءة النمط المخفي منها
- تنظيف البيانات: تنظيف البيانات غير المهيكلة أو شبه المهيكلة
- تحليل البيانات الاستكشافية: استخدم المفاهيم الإحصائية لفهم البيانات مثل المنتشرة، المتطرفة، إلخ.
- استخدام خوارزميات التعلم الآلي لعمل نموذج: يمكنك استخدام خوارزميات ساذجة أو بعض الخوارزميات الأخرى أيضًا
- استخدم مجموعة بيانات غير معروفة للتحقق من دقة النموذج.

س49/ ما هو التعلم الآلي الاستقرائي inductive machine learning؟

ج/ يدور التعلم الآلي الاستقرائي حول عملية التعلم من خلال الأمثلة الحية.

س50/ ما هي تقنيات الخوارزمية المختلفة في التعلم الآلي؟

ج/ الأنواع المختلفة من التقنيات في التعلم الآلي هي:

- Supervised Learning
- Unsupervised Learning
- Semi-supervised Learning
- Reinforcement Learning
- Transduction
- Learning to Learn

س51/ ما هو الفرق بين نمذجة البيانات Data Modeling وتصميم قواعد البيانات Database Design؟

ج / نموذج البيانات data model هو نموذج مفاهيمي يوضح الكيانات المختلفة التي يتم الحصول على البيانات منها والعلاقات بينها. تصميم قاعدة البيانات Database design، من ناحية أخرى، هو عملية بناء مخطط بناءً على كيفية إنشاء قاعدة البيانات.

س52/ما هو قانون الأعداد الكبيرة Law of Large Numbers؟

ج / ينص قانون الاحتمال هذا على أنه للاقتراب من النتيجة المتوقعة، يجب إجراء تجربة عددًا كبيرًا من المرات، كل منها مستقل عن الآخر، ثم حساب متوسط النتيجة.

س53/ هل يمكنك التمييز بين البيانات ذات التنسيق الطويل Long-Format Data والبيانات ذات التنسيق الواسع Wide-Format Data؟

ج / هناك طريقتان مختلفتان يمكن من خلالهما كتابة مجموعة البيانات. تشير البيانات ذات التنسيق الطويل Long-Format Data إلى تكرار قيم البيانات في العمود الأول. البيانات ذات التنسيق الواسع Wide-Format Data يعني عدم وجود قيم تتكرر في العمود الأول.

س54/ما هي دوال الخسارة Loss Function ودوال التكلفة Cost Function؟ اشرح الفرق الرئيسي بينهما؟

ج / تشير دالات التكلفة والخسارة تقريبًا إلى نفس المعنى. ولكن، تنطبق دالة الخسارة بشكل أساسي على مجموعة تدريب واحدة مقارنة بدالة التكلفة التي تتعامل مع عقوبة لعدد من مجموعات التدريب أو الدُفعة الكاملة. كما يطلق عليه أحيانًا دالة الخطأ Error Function.

س55/ ما هي الخوارزميات الخمس الشائعة لتعلم الآلة؟

ج/

- Decision Trees
- Neural Networks (back propagation)
- Probabilistic Networks
- Nearest Neighbor
- Support Vector Machines (SVM)

س56/ ماذا تقصد بالميزات Features والتسميات Labels في مجموعة البيانات؟

ج/

الميزة Feature

الميزات **Features** هي متغيرات فردية مستقلة تعمل كمدخلات في النظام. تستخدم نماذج التنبؤ هذه الميزات لعمل تنبؤات. يمكن أيضًا استخراج الميزات الجديدة من الميزات القديمة باستخدام طريقة تُعرف باسم "هندسة الميزات". لتبسيط الأمر، يمكنك اعتبار عمود واحد من مجموعة البيانات الخاصة بك ميزة واحدة. تسمى الميزات أيضًا السمات **attributes**. وعدد الميزات أبعاد **dimensions**.

التسمية Label

التسميات **Label** هي الإخراج النهائي أو الإخراج الهدف. يمكن اعتبارها أيضًا فئات الإخراج. نحصل على الملصقات (التسميات) كمخرجات عند تزويدنا بالميزات كمدخلات.

← Features →					Label
Position	Experience	Skill	Country	City	Salary ($)
Developer	0	1	USA	New York	103100
Developer	1	1	USA	New York	104900
Developer	2	1	USA	New York	106800
Developer	3	1	USA	New York	108700
Developer	4	1	USA	New York	110400
Developer	5	1	USA	New York	112300
Developer	6	1	USA	New York	114200
Developer	7	1	USA	New York	116100
Developer	8	1	USA	New York	117800
Developer	9	1	USA	New York	119700
Developer	10	1	USA	New York	121600

س57/ ماذا تقصد بالمتغيرات المستقلة independent variable والتابعة dependent variable؟

ج/

المتغير المستقل **independent variable** هو متغير يمثل الكمية المستخدمة في التجربة. المتغير المستقل (المعروف أحيانًا باسم المتغير الذي تم التلاعب به **manipulated variable**) هو المتغير الذي لا يتأثر تغييره بأي متغير آخر في التجربة.

غالبًا ما يكون **X** هو المتغير المستخدم لتمثيل المتغير المستقل في المعادلة.

المتغير التابع **dependent variable** هو متغير يمثل كمية تعتمد قيمتها على كيفية معالجة المتغير المستقل.

المتغير المستقل هو ما تغيره، والمتغير التابع هو الذي يتغير بسبب ذلك.

غالبًا ما يكون **Y** هو المتغير المستخدم لتمثيل المتغير التابع في المعادلة.

س58/ ماذا تقصد بالضوضاء Noise في مجموعة بيانات Dataset معينة وكيف يمكنك إزالة الضوضاء في مجموعة البيانات؟

ج/ **Noise** عبارة عن عناصر بيانات أو ميزات أو سجلات غير مرغوب فيها لا تساعد في شرح الميزة نفسها أو العلاقة بين الميزة والهدف. غالبًا ما يتسبب **Noise** في فقدان الخوارزميات لأنماط البيانات. البيانات الصاخبة **Noisy data** هي بيانات لا معنى لها. تم استخدام المصطلح كمرادف للبيانات الفاسدة **corrupt data**. ومع ذلك، فإن معناها يشمل أي بيانات لا يمكن للآلات فهمها وتفسيرها بشكل صحيح، مثل النص غير المنظم **unstructured text**. يمكن وصف أي بيانات تم استلامها أو تخزينها أو تغييرها بطريقة لا يمكن للبرنامج قراءتها أو استخدامها على أنها بيانات صاخبة.

طرق اكتشاف وإزالة الضوضاء **Noise** في مجموعة البيانات:

- K-fold validation
- Density-based anomaly detection
- Clustering-based anomaly detection
- SVM-based anomaly detection
- Autoencoder-based anomaly detection
- Manual method

س59/ ما هي عملية تنفيذ الانحدار الخطي linear regression؟

ج/ يتكون تحليل الانحدار الخطي من أكثر من مجرد تركيب خط خطي عبر سحابة من نقاط البيانات. تتكون من 3 مراحل:

1. تحليل ارتباط correlation واتجاه البيانات directionality ،
2. تقدير النموذج ، أي ملاءمة الخط fitting the line ،
3. تقييم صلاحية وفائدة النموذج.

س60/ كيف تشرح الانحدار الخطي linear regression لشخص غير تقني؟

ج/ يعني الانحدار على أساس بعض البيانات أنك ستتوقع متغيرًا.

على سبيل المثال، إذا كان لديك بيانات عن الطول والوزن بشكل عام، فهي مرتبطة بشكل خطي. هذا يعني أنه إذا قمت برسم بيانات الطول والوزن للعديد من الأشخاص في مستوى x-y، فسيكون لديك بنية خطية أكثر أو أقل.

لذلك إذا كان لديك خط مستقيم في المستوى x-y ، فإنه يتكون من شيئين INTERCEPT (الطول من الأصل إلى حيث يقطع الخط المحور y) والميل (الميل مع المحور الأفقي).

ثم باستخدام بعض الإجراءات الإحصائية (المربعات الصغرى بشكل أساسي) يمكننا تقدير الميل والتقاطع.

ثم يتم استخدام التقديرات لتشكيل معادلة الانحدار.

على سبيل المثال، إذا كانت نتيجتك توفر تقديرًا للتقاطع هو 5 وكان تقدير الميل 0.4، فستكون معادلتك y = 5 + 0.4x

هذه المعادلة مثل الآلة. إذا قمت بإدخال قيم x، فستحصل على ناتج من المتوقع y.

قد لا تكون هذه القيمة المتوقعة مساوية للقيمة الفعلية ولكنها ستمنحك فكرة عن النتائج المستقبلية.

س61/ استنادًا إلى مجموعة البيانات، كيف ستعرف الخوارزمية التي يجب تطبيقها؟

ج/

- إذا كانت مشكلة متعلقة بالتصنيف ، فيمكننا استخدام الانحدار اللوجستي، وأشجار القرار ، إلخ . . .

- إذا كانت مشكلة متعلقة بالانحدار ، فيمكننا استخدام الانحدار الخطي. إذا كان يعتمد على التجميع ، فيمكننا استخدام **KNN**.
- يمكننا أيضًا تطبيق **XGB** و **RF** للحصول على دقة أفضل.

س62/ أين تستخدم R & Python؟

ج / يمكن استخدام **R** كلما كانت البيانات مهيكلة. لغة **Python** فعالة في التعامل مع البيانات غير المهيكلة. لا يمكن لـ **R** التعامل مع البيانات ذات الحجم الكبير. استخدام **Python backend** مع **Theano / tensor** من السهل تنفيذه بسرعة مقارنةً بـ **R**.

س63/ ما هو اختيار النموذج Model Selection في التعلم الآلي؟

ج / تُعرف عملية اختيار النماذج Model Selection من بين النماذج الرياضية المختلفة، والتي تُستخدم لوصف مجموعة البيانات نفسها، باسم اختيار النموذج. يتم تطبيق اختيار النموذج على مجالات الإحصاء والتعلم الآلي واستخراج البيانات.

س64/ ما هي المراحل الثلاث لبناء الفرضيات أو النموذج في التعلم الآلي؟

ج / المراحل الثلاث لبناء النموذج في التعلم الآلي هي:

- بناء نموذج Model building
- اختبار النموذج Model testing
- تطبيق النموذج Applying the model

س65/ كيف يتم استخدام الخوارزمية الجينية في التعلم الآلي؟

ج / حالات استخدام الخوارزميات الجينية Genetic Algorithm (GA) في التعلم الآلي كالاتي:

- اختيار الميزة Feature Selection
- ضبط نموذج المعلمات الفائقة Model Hyperparameter Tuning
- تحسين خط أنابيب التعلم الآلي Machine Learning Pipeline Optimization

س66/ كيف يمكنك تحسين بُنية مصنف التعلم العميق باستخدام الخوارزميات الجينية؟

س67/ ما هي الفروق بين الارتباط correlation والتغاير covariance؟

ج/ الارتباط correlation هو مفهوم رياضي يستخدم في الإحصاء ونظرية الاحتمالات لقياس وتقدير ومقارنة عينات البيانات المأخوذة من مجموعات سكانية مختلفة. بعبارات أبسط، يساعد الارتباط في إنشاء علاقة كمية بين متغيرين.

التغاير covariance هو أيضا مفهوم رياضي. إنها طريقة أبسط للوصول إلى ارتباط بين متغيرين. يساعد التغاير بشكل أساسي في تحديد التغيير أو التأثير الذي يحدثه أحد المتغيرات على الآخر.

س68/ عدد الأنواع المختلفة من المخططات Plots التي نستخدمها بشكل عام في التعلم الآلي والتعلم العميق؟

ج/ هناك مخططات Plots مختلفة نستخدمها في التعلم الآلي والتي يمكن تصورها باستخدام بايثون منها:

- مخطط مبعثر Scatter plot
- مخطط صندوقي Box plot
- مخطط شريطي Bar chart
- مخطط خطي Line plot
- مخطط المدرج التكراري Histogram

س69/ ما هي الافتراضات المطلوبة للانحدار الخطي Linear Regression؟

ج/ هناك أربع افتراضات رئيسية:

1. هناك علاقة خطية بين المتغيرات التابعة وعوامل الانحدار ، مما يعني أن النموذج الذي تقوم بإنشائه يناسب البيانات بالفعل.
2. يتم توزيع الأخطاء أو بقايا البيانات بشكل طبيعي ومستقلة عن بعضها البعض.
3. يوجد حد أدنى من العلاقة الخطية المتعددة بين المتغيرات التفسيرية explanatory variables.
4. Homoscedasticity : التباين حول خط الانحدار - هو نفسه لجميع قيم متغير التوقع predictor variable.

س70/ ماهي Sensitivityو Specificity؟

ج/الحساسية **Sensitivity** تعني "نسبة الإيجابيات الحقيقية المصنفة بشكل صحيح" بعبارة أخرى "إيجابية حقيقية **True Positive** "

تعني الخصوصية **Specificity** "نسبة السلبيات الحقيقية المصنفة بشكل صحيح" بعبارة أخرى "سلبية حقيقية **True Negative** "

س71/ ما هي مقاييس الأداء التي يمكن استخدامها لتقدير كفاءة نموذج الانحدار الخطي Linear regression؟

ج/ مقياس الأداء المستخدم في هذه الحالة هو:

- Mean Squared Error (MSE)
- R2 score
- Adjusted R2 score
- Mean Absolute Error (MAE)

س72/ ما هو مقياس الأداء الأفضل R2 أو R2 المعدل؟

ج/ **R2** المعدل هو الافضل لأن أداء المتنبئين يؤثر عليه. **R2** مستقل عن المتنبئين ويظهر تحسنًا في الأداء من خلال الزيادة في حالة زيادة عدد المتنبئين.

س73/ ما هو R2؟ ما هي بعض المقاييس الأخرى التي يمكن أن تكون أفضل من R2 ولماذا؟

ج/ **R2** يمثل حسن قياس الملاءمة. يفسر التباين من خلال الانحدار / التباين الكلي، كلما زاد عدد المتنبئين الذين تضيفهم، يصبح **R ^ 2** أعلى.

ومن ثم استخدم **R ^ 2** المعدلة التي تعدل درجات الحرية أو مقاييس خطأ التدريب.

س74/ كيف تقرر ما إذا كان نموذج الانحدار الخطي الخاص بك يناسب البيانات؟

ج/ يعتمد قرار ما إذا كان النموذج يناسب البيانات على المقاييس، ويعرف أيضًا باسم جودة الملاءمة **goodness of fit**. يمكن استخدام الكثير من القياسات لتحديد مدى جودة الملاءمة، على سبيل المثال:

- Pearson's chi-squared test
- (adjusted) R squared
- Cook's distance
- Mean squared error, mean average error

- Residual error plot

س75/ متى تستخدم الانحدار اللوجستي ومتى تستخدم الانحدار الخطي؟

ج/

- إذا كنت تتعامل مع مشكلة تصنيف مثل (نعم / لا، احتيال / عدم احتيال، رياضة/موسيقى/رقص)، فاستخدم الانحدار اللوجستي.
- إذا كنت تتعامل مع قيم مستمرة/منفصلة، فانتقل إلى الانحدار الخطي.

س76/ هل يمكن استخدام الانحدار اللوجستي logistic regression للفئات التي تزيد عن 2؟

ج/ لا، لا يمكن استخدام الانحدار اللوجستي للفئات التي تزيد عن **2** نظرًا لأنه مصنف ثنائي.

س77/ ما هو نموذج التصويت voting model؟

ج / نموذج التصويت **voting model** هو نموذج تجميعي يجمع بين العديد من المصنفات ولكن لإنتاج النتيجة النهائية، في حالة وجود نموذج قائم على التصنيف، يأخذ في الاعتبار تصنيف نقطة بيانات معينة لجميع النماذج ويختار الأكثر تصديقًا/ تصويتًا/تم إنشاؤه من جميع الفئات المحددة في العمود الهدف.

س78/ هل طرق الانحدار التدريجي تتقارب converge دائمًا مع نقاط متشابهة؟

ج / لا، لا تتقارب طرق الانحدار التدريجي دائمًا إلى نفس النقطة لأنها تتقارب مع حد أدنى محلي local minimum أو نقطة local optima في بعض الحالات. يعتمد ذلك كثيرًا على البيانات التي يتعامل معها الفرد والقيم الأولية لمعلمة التعلم.

س79/ كيف تتعامل مع عينات بيانات قليلة جدًا؟ هل من الممكن بناء نموذج منها؟

ج / إذا كان هناك عدد قليل جدًا من عينات البيانات، فيمكننا الاستفادة من **oversampling** لإنتاج نقاط بيانات جديدة. بهذه الطريقة، يمكننا الحصول على نقاط بيانات جديدة.

يعتقد العديد من الباحثين والممارسين أن البيانات الصغيرة هي مستقبل علم البيانات. ليس من المجدي أن يكون لديك مجموعات بيانات ضخمة لكل نوع من المشاكل. اتبع هذه الإرشادات للتغلب على تحديات مجموعة البيانات الصغيرة:

- افهم أساسيات الإحصاء لتعرف نوع المشكلات التي يمكن أن تتوقعها عند التعامل مع عدد صغير من الملاحظات.
- تعرف على الاستراتيجيات الرئيسية لتجنب الضبط الزائد والحصول على نتائج دقيقة من البيانات الصغيرة.
- إجراء جميع خطوات تنظيف البيانات وتحليلها بكفاءة (على سبيل المثال ، استخدام **Tidyverse** في لغة **R** أو أدوات **Python** لعلوم البيانات).
- كن على دراية بحدود النموذج عند استخلاص النتائج من توقعاته.

س80/ ماهي المقاييس التي تُستخدم لتحليل الاتجاه المركزي للبيانات central tendency of data؟

ج / **mean**، **median**، **mode** هو المقاييس الإحصائية الثلاثة التي تساعدنا على تحليل الاتجاه المركزي **central tendency** للبيانات. نستخدم هذه الإجراءات للعثور على القيمة المركزية للبيانات لتلخيص مجموعة البيانات بأكملها.

س81/ كيف يمكنك بناء data pipeline؟

ج /خطوط أنابيب البيانات **data pipeline** هي جوهر مهندسي التعلم الآلي، الذين يأخذون نماذج علوم البيانات ويكتشفون طرقًا لتوسيع نطاقها وأتمتتها إذا كنت معتادًا على أدوات بناء الأنظمة الأساسية وخطوط أنابيب البيانات حيث يمكننا استضافة خطوط الأنابيب والنماذج.

س82/ كيف تعتقد أن الحوسبة الكمومية quantum computing ستؤثر على التعلم الآلي؟

ج / بعض أجزاء التعلم الآلي، لا سيما المشكلات القائمة على الرسم البياني والخوارزميات التي تعتمد على الحلول التوافقية. بمجرد أن تتوسع أجهزة الكمبيوتر الكمومية، ستكون هذه المشكلات أسهل بكثير على الآلات الكمومية من تلك التقليدية. إذا كنت تعمل في تحليلات الشبكات الاجتماعية، والبائع المتجول Travelling salesman، ومشكلات أخرى محتملة التكلفة.

س83/ ما هي الخطوات الضرورية والمتضمنة في مشروع التعلم الآلي؟

ج / هناك العديد من الخطوات الأساسية التي يجب أن نتبعها لتحقيق نموذج عمل جيد أثناء القيام بمشروع التعلم الآلي. قد تشمل هذه الخطوات ضبط المعلمات parameter tuning، وإعداد البيانات data preparation، وجمع البيانات data collection، وتدريب النموذج training the model، وتقييم النموذج model evaluation، والتنبؤ prediction، إلخ.

س84/ كيف يرتبط علم البيانات والتعلم الآلي ببعضهما البعض؟

ج / علم البيانات وتعلم الآلة هما مصطلحان مرتبطان ارتباطًا وثيقًا ولكن غالبًا ما يساء فهمهما. كلاهما يتعامل مع البيانات. ومع ذلك، هناك بعض الفروق الأساسية التي توضح لنا كيف يختلف كل منهما عن الآخر.

علم البيانات هو مجال واسع يتعامل مع كميات كبيرة من البيانات ويسمح لنا باستخلاص رؤى من هذه البيانات الضخمة. تهتم العملية الكاملة لعلم البيانات بخطوات متعددة متضمنة في استخلاص رؤى من البيانات المتاحة. تتضمن هذه العملية خطوات حاسمة مثل جمع البيانات وتحليل البيانات ومعالجة البيانات وتصور البيانات وما إلى ذلك.

من ناحية أخرى، يمكن اعتبار التعلم الآلي مجالًا فرعيًا لعلوم البيانات. إنه يتعامل أيضًا مع البيانات، ولكن هنا، نحن نركز فقط على تعلم كيفية تحويل البيانات المعالجة إلى نموذج وظيفي، والذي يمكن استخدامه لتعيين المدخلات إلى المخرجات، على سبيل المثال، النموذج الذي يمكن أن يتوقع صورة كمدخل وإخبار لنا إذا كانت تلك الصورة تحتوي على زهرة كمخرج.

باختصار، يتعامل علم البيانات مع جمع البيانات ومعالجتها ، وأخيرًا ، استخلاص رؤى منها. يسمى مجال علم البيانات الذي يتعامل مع بناء النماذج باستخدام الخوارزميات التعلم الآلي. لذلك، يعد التعلم الآلي جزءًا لا يتجزأ من علوم البيانات.

س85/ ما هي السلاسل الزمنية Time series؟

ج / بيانات السلاسل الزمنية Time series data، والتي يشار إليها أيضًا باسم البيانات ذات الطابع الزمني Time-stamped data، هي سلسلة من نقاط البيانات المفهرسة بالترتيب الزمني. الطابع الزمني هو البيانات التي يتم جمعها في فترات زمنية time interval مختلفة.

تتكون نقاط البيانات هذه عادةً من قياسات متتالية يتم إجراؤها من نفس المصدر خلال فترة زمنية وتستخدم لتتبع التغيير بمرور الوقت.

س86/ ما هي المهارات المهمة التي يجب ان تمتلكها في بايثون فيما يتعلق بتحليل البيانات data analysis؟

ج / فيما يلي بعض المهارات المهمة التي يجب امتلاكها والتي ستكون مفيدة عند إجراء تحليل البيانات باستخدام Python.

- ❖ فهم جيد لأنواع البيانات المضمنة وخاصة القوائم والقواميس والمجموعات، والصفوف.
- ❖ التمكن من مصفوفات NumPy ذات الأبعاد N.

- التمكن من إطارات بيانات الباندا **Pandas dataframes**.
- القدرة على إجراء عمليات متجه ومصفوفة العناصر على مصفوفات **NumPy**.
- مع العلم أنه يجب عليك استخدام توزيع **Anaconda** ومدير حزمة **conda**.
- الإلمام بـ **Scikit-Learn** و**Scikit-Learn Cheat Sheet**
- القدرة على كتابة قوائم شاملة **list comprehensions** فعالة بدلاً من حلقات **for** التقليدية.
- القدرة على كتابة دوال صغيرة ونظيفة (مهمة لأي مطور) ، ويفضل أن تكون دوال نقية لا تغير الكائنات.
- معرفة كيفية تحديد أداء برنامج نصي من **Python** وكيفية تحسين الاختناقات **bottlenecks**.

س87/ أيهما أفضل لتحليلات النص - Python أم R؟

ج / يمكن استخدام كل من Python وR لتحليل النص. يأتي R مع العديد من المكتبات المضمنة لتحليل النص، كما هو الحال في Python.

ترجع اختلافاتهم إلى طبيعة البيانات التي تتم دراستها. تعد Python أفضل عند العمل بكميات ضخمة من البيانات. يتمتع R بدعم أفضل للبيانات غير المهيكلة.

س88/ اشرح تحليل السلاسل الزمنية Time Series Analysis؟

ج / تحليل السلاسل الزمنية هو شكل من أشكال تحليل البيانات التي تبحث في قيم البيانات التي تم جمعها في تسلسل معين. يدرس كلاهما البيانات التي تم جمعها بمرور الوقت، والعوامل في النقاط الزمنية المختلفة التي تم فيها جمع البيانات.

س89/ يتم إعطاؤك مجموعة بيانات عن اكتشاف السرطان. لقد قمت ببناء نموذج تصنيف وحققت دقة تصل إلى 96%. لماذا لا تكون سعيدًا بأداء نموذجك؟ ماذا يمكنك أن تفعل بهذا الشأن؟

ج / يؤدي اكتشاف السرطان إلى بيانات غير متوازنة imbalanced data. في مجموعة البيانات غير المتوازنة، لا ينبغي أن تستند الدقة كمقياس للأداء. من المهم التركيز على نسبة 4%، والتي تمثل المرضى الذين تم تشخيصهم بشكل خاطئ. يعد التشخيص المبكر أمرًا بالغ الأهمية عندما يتعلق الأمر باكتشاف السرطان، ويمكن أن يحسن بشكل كبير من تشخيص المريض.

ومن ثم، لتقييم أداء النموذج، يجب أن نستخدم الحساسية Sensitivity (المعدل الإيجابي الحقيقي)، والخصوصية Specificity (المعدل السلبي الحقيقي)، F1-scoreg لتحديد أداء فئة المصنف.

س90/ في الشبكة العصبية، ماذا لو تمت تهيئة جميع الأوزان بنفس القيمة؟

ج/ تخيل الآن أنك تقوم بتهيئة جميع الأوزان بنفس القيمة (على سبيل المثال صفر أو واحد). في هذه الحالة، ستحصل كل وحدة مخفية على نفس الإشارة بالضبط. على سبيل المثال إذا تمت تهيئة جميع الأوزان على 1، تحصل كل وحدة على إشارة مساوية لمجموع المدخلات (ومخرجات ((sigmoid(sum(inputs)). إذا كانت جميع الأوزان عبارة عن أصفار، وهو أسوأ من ذلك، فلن تحصل كل وحدة مخفية على إشارة صفرية. بغض النظر عن المدخلات - إذا كانت جميع الأوزان متماثلة، فستكون جميع الوحدات في الطبقة المخفية هي نفسها أيضًا.

هذه هي المشكلة الرئيسية في التناظر symmetry وسبب وجوب تهيئة الأوزان بشكل عشوائي (أو على الأقل بقيم مختلفة). لاحظ أن هذه المشكلة تؤثر على كافة البنى التي تستخدم كل اتصال لكل منهما.

بعبارات أبسط،إذا كانت جميع الخلايا العصبية لها نفس قيمة الأوزان، فستحصل كل وحدة مخفية على نفس الإشارة بالضبط. في حين أن هذا قد يعمل أثناء الانتشار الأمامي، فإن مشتق دالة التكلفة أثناء الانتشار الخلفي سيكون هو نفسه في كل مرة.

باختصار، لا يوجد تعلم يحدث بواسطة الشبكة! ماذا تسمي ظاهرة عدم قدرة النموذج على تعلم أي أنماط من البيانات؟ نعم، انه الضبط الناقص underfitting.

لذلك، إذا كانت جميع الأوزان لها نفس القيمة الأولية، فقد يؤدي ذلك إلى الضبط الناقص.

س91/ لماذا يجب عليك تحديث خوارزمية تعلم الآلة بانتظام؟ كم مرة يجب تحديثها؟

ج/ من المهم الاستمرار في تحديث (تعديل) خوارزميات التعلم الآلي بانتظام. يعتمد تكرار تحديثها على حالة استخدام الأعمال. على سبيل المثال، تحتاج خوارزميات الكشف عن الاحتيال fraud detection إلى التحديث بانتظام. ولكن إذا كنت بحاجة إلى دراسة بيانات التصنيع باستخدام التعلم الآلي، فستحتاج هذه النماذج إلى التحديث بشكل أقل انتظامًا.

س92/ إذا كنت نموذج التدريب يعطي دقة 90٪ ونموذج الاختبار يعطي دقة 60٪؟ ثم ما المشكلة التي تواجهها؟

ج/ الضبط الزائد overfitting.

يمكن تقليل الضبط الزائد بعدة طرق مثل (تقليم شجار القرار، إزالة بعض المعلومات المتوفرة في مجموعة البيانات).

س93/ ما هو الفرق بين الارتباط correlation والسببية causality؟

ج/

- الارتباط **correlation** هو المقياس الذي يساعدنا على فهم العلاقة بين متغيرين أو أكثر.
- يمثل السببية **causality** تلك العلاقة السببية بين حدثين. ومن المعروف أيضًا أنه يمثل السبب والنتيجة.
- السببية تعني وجود ارتباط ولكن الارتباط لا يعني بالضرورة السببية.

س94/ كيف يتم نشر التعلم الآلي في سيناريوهات العالم الحقيقي؟

ج / فيما يلي بعض السيناريوهات التي يجد فيها التعلم الآلي تطبيقات في العالم الحقيقي:

- التجارة الإلكترونية **Ecommerce**: فهم اضطراب العميل ونشر الإعلانات المستهدفة وتجديد النشاط التسويقي.
- محرك البحث **Search engine**: ترتيب الصفحات حسب التفضيلات الشخصية للباحث
- المالية **Finance**: تقييم فرص الاستثمار والمخاطر، واكتشاف المعاملات الاحتيالية
- الرعاية الطبية **Medicare**: تصميم الأدوية اعتمادًا على تاريخ المريض واحتياجاته
- الروبوتات **Robotics**: التعلم الآلي للتعامل مع المواقف الخارجة عن المألوف
- وسائل التواصل الاجتماعي **Social media**: فهم العلاقات والتوصية بالصلات
- استخراج المعلومات **Extraction of information**: صياغة الأسئلة للحصول على إجابات من قواعد البيانات عبر الويب.

س95/ ما هي بعض الانتقادات الموجهة للشبكات العصبية؟

ج/

- تتطلب الشبكات العصبية الكثير من البيانات لتدريبها **too much data to train**. قد تتطلب شبكة التصنيف آلاف الأمثلة في فئة واحدة حتى تتمكن من التعرف عليها في بيانات غير مرئية. نتيجة لذلك ، في بعض الأحيان لا يكون من الممكن إنشاء نماذج الشبكات العصبية الاصطناعية **ANN** للتطبيقات الهامشية **fringe applications**.

- ❖ الشبكات العصبية غير قابلة للتفسير **not interpretable**. يحتاج المستخدم إلى إدخال البيانات في الشبكة ويقوم بإخراج المخرجات المطلوبة،لكن العمل الذي يدخل في معالجة المدخلات وإعطاء المخرجات لا يمكن فهمه للبشر.
- ❖ الطاقة المطلوبة لتدريب الشبكة العصبية عالية للغاية مقارنة بكمية الطاقة التي يستخدمها الدماغ البشري (حوالي **20** واط) للقيام بنفس الأشياء تقريبًا مثل تصنيف الصور.

س96/ هل يمكنك استخدام التعلم الآلي لتحليل السلاسل الزمنية؟

ج/ نعم، يمكن استخدامه ولكن ذلك يعتمد على التطبيقات.

س97/ ما هي تطبيقات التعلم الآلي الخاضع للإشراف؟

ج/ فيما يلي تطبيقات التعلم الآلي الخاضع للإشراف:

- كشف الاحتيال **Fraud Identification**:

التعلم الخاضع للإشراف يدرب النموذج لتحديد الأنماط المشبوهة؛ يمكننا تحديد حالات الاحتيال الممكنة.

- الرعاية صحية **Healthcare**:

من خلال تقديم صور عن مرض ما ، يمكن للتعلم الآلي الخاضع للإشراف تدريب النموذج على اكتشاف ما إذا كان الشخص مصابًا بالمرض أم لا.

- تحديد البريد الإلكتروني العشوائي **Email spam identification**:

نقوم بتدريب النموذج من خلال البيانات السابقة التي تحتوي على رسائل بريد إلكتروني تم تصنيفها على أنها بريد عشوائي أو ليست بريد عشوائي. يتم توفير هذه البيانات المسمى كمدخلات للنموذج.

- تحليل المشاعر **Sentiment Analysis**:

يتعلق ذلك بعملية استخدام الخوارزميات لتعدين المستندات وتحديد ما إذا كانت سلبية أو محايدة أو إيجابية في المشاعر.

س99/ ما هي خوارزمية التعلم الآلي البارامترية Parametric وغير البارامترية Non-parametric؟ وأهميتها؟

ج / الخوارزمية التي لا تضع افتراضات قوية هي خوارزمية غير بارامترية **Non-parametric** وهي حرة في التعلم من بيانات التدريب. الخوارزمية التي تضع افتراضات قوية هي خوارزمية بارامترية **Parametric** وتتضمن:

- ❖ تحديد شكل الدالة.
- ❖ تعلم معاملات الدالة من بيانات التدريب.

س100/ اعط أمثلة على خوارزمية التعلم الآلي البارامترية وخوارزمية التعلم الآلي غير البارامترية؟

ج/

- خوارزمية التعلم الآلي البارامترية: الانحدار الخطي، الانحدار اللوجستي.
- خوارزمية التعلم الآلي غير البارامترية: أشجار القرار، SVM ، الشبكة العصبية.

س101/ هل تعلم الآلة علم أم فن؟

س102/ ما هي النماذج التمييزية discriminative والتوليدية generative؟

ج/ لفهم هذه المصطلحات بشكل أفضل، دعنا نفكر في مثال. لنفترض أن هناك شخصًا لديه طفلان - الطفل "أ" والطفل "ب". يتعلم الطفل "أ" ويفهم كل شيء بعمق بينما لا يستطيع الطفل "ب" سوى معرفة الاختلافات بين ما يراه. في أحد الأيام، أخذهم ذلك الشخص إلى حديقة الحيوانات حيث رأوا غزالًا وأسدًا. بعد القدوم من حديقة الحيوانات، أطلعهم الشخص على حيوان وسألهم عما هو عليه. قام الطفل "أ" برسم صور كل من الحيوانات التي رآها في حديقة الحيوان. قارن الصور وأجاب "الحيوان غزال" بناءً على أقرب تطابق للصورة. عندما يتعلم الطفل "ب" الأشياء بناءً على الاختلافات فقط، لذلك أجاب بسهولة: "الحيوان غزال".

في التعلم الآلي، نسمي الطفل "أ" نموذجًا توليديًا والطفل "ب" نموذجًا مميزًا. لتوضيح الأمر أكثر، يتعلم النموذج التوليدي التوزيع الاحتمالي المشترك **p(x,y)**. يتنبأ بالاحتمال الشرطي باستخدام نظرية بايز. في حين أن النموذج التمييزي يتنبأ بتوزيع الاحتمال الشرطي **p (y|x)**. يتم استخدام كلا النموذجين في مشاكل التعلم تحت الإشراف.

س103/ ما هي تطبيقات التعلم الآلي؟

ج/

- سيارات ذاتية القيادة **Self-Driving Cars**
- تصنيف الصور **Image Classification**
- تصنيف النص **Text Classification**
- محرك البحث **Search Engine**
- الخدمات المصرفية **Banking**
- مجال الرعاية الصحية **Healthcare Domain**

س104/ ما الفرق بين التعلم الآلي الخاضع للإشراف supervised والتعلم غير الخاضع للإشراف unsupervised؟

ج/

التعلم الآلي الخاضع للإشراف	التعلم غير الخاضع للإشراف
إذا تعلمت الخوارزمية شيئًا ما من بيانات التدريب بحيث يمكن تطبيق المعرفة على بيانات الاختبار، فيشار إليها بالتعلم الخاضع للإشراف.	إذا لم تتعلم الخوارزمية أي شيء مسبقًا بسبب عدم وجود متغير استجابة أو بيانات تدريب، فيتم الإشارة إليها على أنها تعلم غير خاضع للإشراف.
يتم استخدامه بشكل رئيسي لعمل تنبؤات لمتغير تابع **dependent variable**.	يتم استخدامه بشكل أساسي لإجراء التحليل وتجميع نقاط البيانات المتشابهة معًا.
التصنيف والانحدار أمثلة على التعلم الخاضع للإشراف.	يعد التجميع (التكتل) وتقليل الأبعاد أمثلة على التعلم غير الخاضع للإشراف.

س105/ قارن بين التصنيف classification والانحدار regression في التعلم الآلي؟

ج/ في التعلم الآلي، هناك أنواع مختلفة من مشاكل التنبؤ بناءً على التعلم الخاضع للإشراف والتعلم غير الخاضع للإشراف. هم التصنيف **classification** والانحدار **regression** والتكتل **clustering** والارتباط **association**. هنا، سنناقش التصنيف والانحدار.

التصنيف **classification**: في التصنيف، يتم إنشاء نموذج التعلم الآلي الذي يساعد في تمييز البيانات في فئات منفصلة. يتم تصنيف البيانات وتصنيفها بناءً على معلمات الإدخال.

على سبيل المثال، يجب إجراء تنبؤات بشأن العملاء الخارجين لمنتج معين بناءً على بعض البيانات المسجلة. إما أن العملاء سوف يخرجون أو لا يفعلون. لذلك، تسميات هذا ستكون "نعم" و "لا".

الانحدار **regression**: هو عملية إنشاء نموذج لتمييز البيانات إلى قيم حقيقية مستمرة، بدلاً من استخدام الفئات أو القيم المنفصلة. يمكنه أيضًا تحديد حركة التوزيع اعتمادًا على البيانات التاريخية. يتم استخدامه للتنبؤ بحدوث حدث اعتمادًا على درجة ارتباط المتغيرات.

على سبيل المثال، يعتمد التنبؤ بالظروف الجوية على عوامل مثل درجة الحرارة وضغط الهواء والإشعاع الشمسي والارتفاع والمسافة من البحر. العلاقة بين هذه العوامل تساعد في التنبؤ بحالة الطقس.

س106/ كيف يتم تحديد مدى ملاءمة خوارزمية التعلم الآلي لمشكلة معينة؟

ج/ لتحديد خوارزمية التعلم الآلي لمشكلة معينة، يجب اتباع الخطوات التالية:

الخطوة 1: تصنيف المشكلة **Problem classification**: يعتمد تصنيف المشكلة على تصنيف المدخلات والمخرجات:

- تصنيف المدخلات **Classifying the input**: يعتمد تصنيف المدخلات على ما إذا كانت هناك بيانات مصنفة (التعلم الخاضع للإشراف) أو غير مصنفة (التعلم غير الخاضع للإشراف)، أو ما إذا كان يجب إنشاء نموذج يتفاعل مع البيئة ويحسن نفسه (التعلم المعزز).
- تصنيف المخرجات **Classifying the output**: إذا كان ناتج النموذج مطلوبًا كفئة، فيجب استخدام بعض تقنيات التصنيف.

إذا كان الناتج رقمًا، فيجب استخدام تقنيات الانحدار **regression**؛ إذا كان الناتج عبارة عن مجموعة مختلفة من المدخلات، فيجب استخدام تقنيات التجميع **clustering**.

الخطوة 2: فحص الخوارزميات الموجودة في متناول اليد **Checking the algorithms in hand**: بعد تصنيف المشكلة، يجب مراعاة الخوارزميات المتاحة التي يمكن نشرها لحل المشكلة المصنفة.

الخطوة 3: تنفيذ الخوارزميات **Implementing the algorithms**: إذا كانت هناك عدة خوارزميات متاحة، فسيتم تنفيذها جميعًا. أخيرًا، تم تحديد الخوارزمية التي تقدم أفضل أداء.

س107/ اشرح مصطلح التعلم الآلي شبه الخاضع للإشراف Semi-Supervised Machine Learning؟

ج/ يُعرَّف التعلم شبه الخاضع للإشراف **Semi-Supervised Machine Learning** بأنه نهج للتعلم الآلي يجمع بين كمية أقل من البيانات المصنفة وكمية هائلة من البيانات غير المصنفة أثناء عملية التدريب. يقع بين التعلم غير الخاضع للإشراف والتعلم الخاضع للإشراف.

س108/ ما الفرق بين التنقيب في البيانات Data Mining والتعلم الآلي Machine Learning؟

ج/ يتعلق التنقيب في البيانات Data Mining بالعمل على بيانات غير منظمة ثم استخراجها إلى مستوى يتم فيه تحديد أنماط مثيرة للاهتمام وغير معروفة.

التعلم الآلي Machine Learning هو عملية أو دراسة ما إذا كان يرتبط ارتباطًا وثيقًا بتصميم وتطوير الخوارزميات التي توفر قدرة للآلات على التعلم.

التنقيب في البيانات	التعلم الآلي
يقدم خوارزميات من البيانات وكذلك من الخبرة السابقة.	يستخرج معلومات مفيدة من كمية كبيرة من البيانات.
يتم استخدامه لفهم تدفق البيانات.	يعلم أجهزة الكمبيوتر أن تتعلم وتفهم من تدفق البيانات.
لديها قواعد بيانات ضخمة مع بيانات غير منظمة.	لديها البيانات الموجودة وكذلك الخوارزميات.
يتطلب تدخل بشري فيه.	لا حاجة للجهد البشري المطلوب بعد التصميم
يتم تطوير النماذج باستخدام تقنية التنقيب في البيانات	يمكن استخدام خوارزمية التعلم الآلي في شجرة القرار والشبكات العصبية وبعض أجزاء الذكاء الاصطناعي الأخرى
إنه أكثر من البحث باستخدام طرق مثل التعلم الآلي.	إنه تعلم ذاتيًا ويدرب النظام على القيام بمهام ذكية.

س109/ ما الفرق بين TensorFlow وKerasg؟

ج/

- **TensorFlow** عبارة عن منصة مفتوحة المصدر للتعلم الآلي. إنها بيئة شاملة ومرنة من الأدوات والمكتبات والموارد الأخرى التي توفر مهام سير عمل مع واجهات برمجة تطبيقات عالية المستوى. يوفر إطار العمل مستويات مختلفة من المفاهيم لتختار منها بناء نماذج التعلم الآلي ونشرها.

  يقدم **TensorFlow** مستويات متعددة من التجريد لبناء النماذج وتدريبها.

  يتيح لك **TensorFlow** تدريب نموذجك ونشره بسهولة، بغض النظر عن اللغة أو النظام الأساسي الذي تستخدمه.

  يمنحك **TensorFlow** المرونة والتحكم مع ميزات مثل **Keras Functional API** و**Model Sub classing API** لإنشاء طبولوجيا معقدة.

- **Keras**، هي مكتبة شبكات عصبية عالية المستوى تعمل في الجزء العلوي من **TensorFlow** و **CNTK** و **Theano**. يتيح استخدام **Keras** في التعلم العميق إنشاء نماذج أولية سهلة وسريعة بالإضافة إلى التشغيل بسهولة على وحدة المعالجة المركزية **CPU** ووحدة معالجة الرسوميات **GPU**.

يمكن كتابة إطار العمل هذا في كود **Python** الذي يسهل تصحيحه ويسمح بسهولة التوسعة.

يحتوي **Keras** على واجهة بسيطة ومتسقة تم تحسينها لحالات الاستخدام الشائعة والتي توفر ملاحظات واضحة وقابلة للتنفيذ لأخطاء المستخدم.

تُبنى نماذج **Keras** من خلال ربط كتل البناء القابلة للتكوين معًا، مع قيود قليلة.

بمساعدة **Keras**، يمكنك بسهولة كتابة قوالب بناء مخصصة لأفكار وأبحاث جديدة.

تقدم **Keras** واجهات برمجة تطبيقات متسقة وبسيطة تساعد في تقليل عدد إجراءات المستخدم المطلوبة لحالات الاستخدام الشائعة، كما أنها توفر ملاحظات واضحة وقابلة للتنفيذ عند حدوث خطأ من المستخدم.

س110/ ماذا يفهم المرء بمصطلح علم البيانات Data Science (DS)؟

ج/ علم البيانات Data Science (DS) هو عملية استخدام تقنيات رياضية وحسابية متنوعة لاستخراج رؤى ذات مغزى **meaningful insights** من مجموعات البيانات.

س111/ كيف يمكنك جمع وتحليل البيانات لاستخدام وسائل التواصل الاجتماعي للتنبؤ بالطقس؟

ج / يمكننا جمع بيانات الوسائط الاجتماعية باستخدام **twitter** و **Facebook** و **instagram API**. بعد ذلك، على سبيل المثال، بالنسبة إلى **Twitter** ، يمكننا إنشاء ميزات من كل تغريدة ، على سبيل المثال تاريخ التغريد وعدد المفضلات وإعادة التغريد وبالطبع الميزات التي تم إنشاؤها من محتوى **Twitter** نفسه. ثم استخدم نموذج السلاسل الزمنية المتعددة المتغيرات للتنبؤ بالطقس.

س112/ هل المزيد من البيانات دائمًا أفضل؟

ج/

احصائياً:

يعتمد ذلك على جودة بياناتك، على سبيل المثال، إذا كانت بياناتك متحيزة، فإن مجرد الحصول على المزيد من البيانات لن يساعد.

هذا يعتمد على النموذج الخاص بك. إذا كان نموذجك يعاني من انحياز كبير، فلن يؤدي الحصول على المزيد من البيانات إلى تحسين نتائج الاختبار الخاصة بك إلى أبعد من نقطة. ستحتاج إلى إضافة المزيد من الميزات وما إلى ذلك.

عملياً:

هناك أيضًا مفاضلة بين الحصول على المزيد من البيانات والتخزين الإضافي، والقوة الحسابية، والذاكرة التي تتطلبها. ومن ثم، فكر دائمًا في تكلفة الحصول على المزيد من البيانات.

س113/ كيف يساعد التعلم العميق علماء البيانات data scientists؟

ج / يعد التعلم العميق عنصرًا مهمًا في علم البيانات، والذي يتضمن الإحصاء والنمذجة التنبؤية. إنه مفيد للغاية لعلماء البيانات المكلفين بجمع وتحليل وتفسير كميات كبيرة من البيانات؛ التعلم العميق يجعل هذه العملية أسرع وأسهل.

س114/ ما هي مزايا رسم بياناتك قبل إجراء التحليل؟

ج / **90**٪ من المعلومات التي تعالجها دماغنا هي معلومات بصرية، ويعالج دماغنا آلاف المرات أسرع من النصوص. ثلاث أسباب بسيطة لماذا يجب عليك رسم بياناتك أولاً قبل القيام بأي شيء آخر.:

- فهم النمط المفقود.
- تحديد القيم المتطرفة.
- توليد فرضيات ذات مغزى.

س115/ كيف يمكنك اختيار المصنف classifier بناءً على حجم مجموعة التدريب؟

ج /

- إذا كانت مجموعة التدريب صغيرة الحجم، أو نماذج عالية التحيز أو منخفضة التباين، على سبيل المثال، فإن **Naive Bayes** تميل إلى الأداء بشكل أفضل لأنها أقل عرضة للضبط الزائد **overfitting**.
- إذا كانت مجموعة التدريب كبيرة الحجم أو منخفضة التحيز أو نماذج عالية التباين، على سبيل المثال، الانحدار اللوجستي، تميل إلى الأداء بشكل أفضل لأنها يمكن أن تعكس علاقات أكثر تعقيدًا.

س116/ متى يجب استخدام التصنيف classification على حساب الانحدار regression؟

ج/ في التعلم الخاضع للإشراف، لدينا مجموعات بيانات وقائمة بالنتائج. أنواع النتائج التي ساعدنا في تصنيفها إلى تصنيف وانحدار. بالنسبة لمشاكل الانحدار، تكون النتائج عادةً بأرقام حقيقية بينما تكون النتائج بالنسبة لمشاكل التصنيف فئات classes أو أصناف categories. لذلك، يمكننا القول إننا سنستخدم الانحدار إذا كانت المخرجات بأرقام حقيقية وسنذهب مع التصنيفات إذا كانت المخرجات في شكل فئات أو اصناف.

س117/ كيف ستفرق بين مشكلة التصنيف متعدد الفئات multi-class classification ومتعدد التسميات multi-label classification؟

ج / في مشكلة التصنيف متعدد الفئات multi-class، تحتوي مهمة التصنيف على أكثر من فئتين حصريتين بشكل متبادل بينما في مشكلة التصنيف متعددة التسميات (العلامات) multi-label، يكون لكل تسمية مهمة تصنيف مختلفة، ومع ذلك، فإن المهام مرتبطة بطريقة ما. على سبيل المثال، تصنيف مجموعة من صور الحيوانات التي قد تكون قطط أو كلاب أو دببة هي مشكلة تصنيف متعددة الفئات تفترض أن كل عينة لها تسمية واحدة فقط مما يعني أنه يمكن تصنيف الصورة على أنها قطة أو كلب ولكن ليس كلاهما في آن واحد. تخيل الآن أنك تريد معالجة الصورة أدناه. يجب تصنيف الصورة الموضحة أدناه على أنها قطة وكلب لأن الصورة تظهر كلا الحيوانات. في مشكلة التصنيف متعدد العلامات، يتم تعيين مجموعة من العلامات لكل عينة ولا تكون الفئات متعارضة. لذلك، يمكن أن ينتمي النمط إلى فئة واحدة أو أكثر في مشكلة تصنيف متعددة التسميات.

س118/ ما الخوارزميات المستخدمة في التصنيف الثنائي Binary classification؟

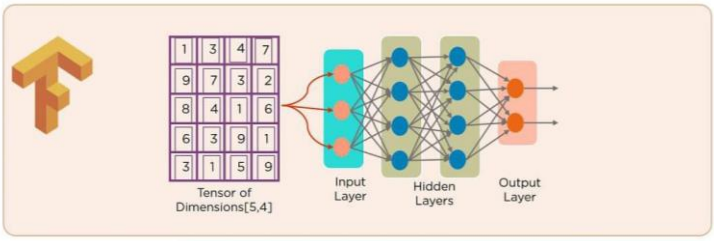
ج / الانحدار اللوجستي، KNN، Random Forest، CART، C50 هي خوارزميات قليلة يمكنها إجراء التصنيف الثنائي Binary classification.

س119/ ما هي الخوارزميات المستخدمة لتصنيف متعدد الحدود Multinomial classification؟

ج / تستخدم Naïve Bayes، Random Forest على نطاق واسع لتصنيف متعدد الحدود Multinomial classification.

س120/ ماذا تقصد بـ Tensor في TensorFlow؟

ج / الموتر **Tensor** هو كائن رياضي يتم تمثيله كمصفوفات ذات أبعاد أعلى. تسمى مصفوفات البيانات هذه بأبعاد وتصنيفات مختلفة يتم تغذيتها كمدخلات للشبكة العصبية باسم "**Tensors**".



س121/ ما الذي يجعل TensorFlow المكتبة الأكثر تفضيلاً في التعلم العميق؟

ج / يوفر Tensorflow كلاً من واجهات برمجة تطبيقات C ++ وPython، مما يسهل العمل عليه ولديه وقت تجميع أسرع مقارنة بمكتبات التعلم العميق الأخرى مثل Keras و Torch. يدعم Tensorflow كلاً من وحدات المعالجة CPU و GPU.

س122/ ما هي خوارزمية imputation المختلفة المتاحة للتعامل مع البيانات المفقودة؟

ج / خوارزمية **imputation** تعني "استبدال القيم الفارغة ببعض القيم" منها:

- Mean imputation
- Median Imputation
- MICE
- miss forest Amelia

س123/ ما هي الخوارزمية التي يمكن استخدامها في احتساب القيمة value imputation في كل من فئات البيانات الفئوية والمستمرة؟

ج / يمكن استخدام خوارزمية KNN لحساب القيمة المفقودة لكل من المتغيرات الفئوية والمستمرة.

س124/ لماذا يتم استخدام Python لتنظيف البيانات في علم البيانات؟

ج / يتعين على علماء البيانات تنظيف مجموعات البيانات الضخمة وتحويلها بالشكل الذي يمكنهم العمل به. من المهم التعامل مع البيانات الزائدة عن الحاجة للحصول على نتائج أفضل عن طريق إزالة القيم المتطرفة غير المنطقية

nonsensical outliers، والسجلات المشوهة malformed records، والقيم المفقودة missing values، والتنسيق غير المتسق inconsistent formatting، وما إلى ذلك.

تُستخدم مكتبات Python مثل Matplotlib و Pandas و Numpy و Keras و SciPy على نطاق واسع لتنظيف البيانات وتحليلها. تُستخدم هذه المكتبات لتحميل البيانات وتنظيفها وإجراء تحليل فعال. على سبيل المثال، يحتوي ملف CSV المسمى "الطالب" على معلومات حول طلاب المعهد مثل أسمائهم، والمعيار، والعنوان، ورقم الهاتف، والدرجات، والعلامات، وما إلى ذلك.

س125/ لماذا يعد تنظيف البيانات data cleaning أمرًا بالغ الأهمية؟ كيف تنظف البيانات؟

ج/ يعد تنظيف البيانات data cleaning من مصادر متعددة أمرًا مرهقًا لتحويلها إلى تنسيق يمكن لمحللي البيانات أو العلماء العمل معه. مع زيادة عدد مصادر البيانات، يزداد الوقت المستغرق لتنظيف البيانات بشكل كبير بسبب عدد المصادر وحجم البيانات المتولدة في هذه المصادر. قد يستغرق الأمر ما يصل إلى 80٪ من الوقت لتنظيف البيانات، مما يجعلها جزءًا مهمًا من مهمة التحليل.

س126/ما هو التحويل الثنائي البيانات Binarization؟ كيفية التحويل الثنائي؟

ج/يُعرف تحويل البيانات إلى قيم ثنائية على أساس قيم العتبة باسم التحويل الثنائي للبيانات Binarization. يتم تعيين القيم الأقل من الحد إلى 0 ويتم تعيين القيم الأكبر من الحد على 1. هذه العملية مفيدة عندما يتعين تنفيذ هندسة الميزات feature engineering. يمكن أيضًا استخدام هذا لإضافة ميزات فريدة unique features. يمكن تحويل البيانات إلى صيغة ثنائية باستخدام مكتبة Scikit-Learn.

س127/ما هي المكتبات الأكثر شعبية المستخدمة في علم البيانات؟

ج / فيما يلي المكتبات الشائعة المستخدمة لاستخراج البيانات data extraction، والتنظيف cleaning، والتصوير visualization، ونشر نماذج علم البيانات:

- TensorFlow: يدعم الحوسبة المتوازية مع إدارة مكتبة لا تشوبها شائبة مدعومة من Google.
- SciPy: يستخدم بشكل أساسي لحل المعادلات التفاضلية ، والبرمجة متعددة الأبعاد ، ومعالجة البيانات ، والتصور من خلال الرسوم البيانية والمخططات.

- Pandas: تُستخدم لتنفيذ إمكانيات ETL (استخراج مجموعات البيانات وتحويلها وتحميلها) في تطبيقات الأعمال.
- Matplotlib: كونها مجانية ومفتوحة المصدر ، يمكن استخدامها كبديل لـ MATLAB ، مما يؤدي إلى أداء أفضل واستهلاك أقل للذاكرة.
- PyTorch: الأفضل للمشاريع التي تتضمن خوارزميات التعلم الآلي والشبكات العصبية العميقة.

س128/ في علم البيانات، لماذا يتم استخدام Python لتنظيف البيانات؟

ج/ يقضي علماء البيانات وقتًا طويلاً في تنظيف مجموعات البيانات والحصول عليها بالشكل الذي يمكنهم العمل به. من المهارات الأساسية لعلماء البيانات أن يكونوا قادرين على العمل مع البيانات الفوضوية messy data، أو القيم المفقودة missing values، أو البيانات غير المتسقة inconsistent، أو الضوضاء noise، أو البيانات غير المنطقية nonsensical. للعمل بسلاسة، يوفر Python وحدة مدمجة Pandas.

س129/ ما هو تصوير البيانات Data Visualization؟

ج / تصوير البيانات Data Visualization هو عملية تحويل رؤى البيانات الرقمية والنصية إلى تنسيق مرئي. يتم استخدام الرسوم البيانية والمخططات والجداول والمساعدات الأخرى لجعل تصوير البيانات ممكنًا.

س130/ ما هي المكتبات في Python المستخدمة في تحليل البيانات Data Analysis والحسابات العلمية Scientific Computations؟

ج/

- SciPy
- Pandas
- Matplotlib
- NumPy
- SciKit
- Seaborn

س131/ ماذا تقصد بـ Matplotlib؟

ج /**Matplotlib** هي مكتبة رسوميات مذهلة في **Python** لرسومات المصفوفات ثنائية الأبعاد. **Matplotlib** عبارة عن مكتبة رسوميات بيانات متعددة المنصات مبنية على مصفوفات **NumPy** ومصممة للعمل مع مكدس **SciPy** الأوسع.

س132/ ما هي الاختلافات بين البيانات المصنفة Labelled وغير المصنفة Unlabeled؟

ج/

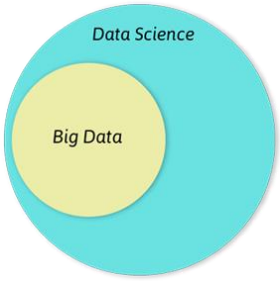
البيانات المصنفة **labelled Data** هي مجموعة من العينات التي تم تمييزها بتسمية (علامة) واحدة أو أكثر. عادةً ما يأخذ التصنيف مجموعة من البيانات غير المسماة ويوسع كل جزء من تلك البيانات غير المسماة بعلامات مفيدة وغنية بالمعلومات.

البيانات غير المصنفة **Unlabeled data** هي وصف لأجزاء من البيانات التي لم يتم تمييزها بعلامات تحدد الخصائص أو التصنيفات. تُستخدم البيانات غير المصنفة عادةً في أشكال مختلفة من التعلم الآلي.

س133/ كيف يتم استخدام Hadoop في علم البيانات؟

ج / **Hadoop** هو نعمة لعلماء البيانات. دعونا نلقي نظرة على كيفية مساعدة **Hadoop** في زيادة إنتاجية علماء البيانات. يتمتع **Hadoop** بقدرة فريدة حيث يمكن تخزين جميع البيانات واسترجاعها من مكان واحد. وبهذه الطريقة يمكن تحقيق ما يلي:

- ❖ استكشاف مجموعات البيانات الكبيرة
- ❖ القدرة على استخراج مجموعات البيانات الكبيرة
- ❖ إعداد البيانات على نطاق واسع
- ❖ تسريع الابتكار المدفوع بالبيانات



س134/ قارن بين تحليلات البيانات Data Analytics وعلم البيانات Data Science؟

ج / يستخدم علم البيانات **Data Science** الرؤى المستخرجة من البيانات لحل مشاكل عمل محددة. تعد تحليلات البيانات **data analytics** ممارسة أكثر استكشافية للكشف عن الارتباطات والأنماط الموجودة في مجموعة البيانات

تحليلات البيانات	علم البيانات

تحليلات البيانات هي مجموعة فرعية من علم البيانات.	علم البيانات عبارة عن تقنية واسعة تتضمن مجموعات فرعية مختلفة مثل تحليلات البيانات، والتنقيب في البيانات، وتصوير البيانات، وما إلى ذلك.
الهدف من علم البيانات هو اكتشاف رؤى ذات مغزى من مجموعات البيانات الضخمة واشتقاق أفضل الحلول الممكنة لحل مشكلات الأعمال.	الهدف من تحليلات البيانات هو توضيح التفاصيل الدقيقة للرؤى المسترجعة.
يتطلب معرفة بلغات البرمجة المتقدمة.	يتطلب فقط لغات البرمجة الأساسية.
تتمثل مهمة عالم البيانات في تقديم تصورات ثاقبة للبيانات من البيانات الأولية التي يسهل فهمها.	تتمثل مهمة محلل البيانات في تحليل البيانات من أجل اتخاذ القرارات.

س135/ كيف تقيم نموذج الانحدار اللوجستي logistic regression model؟

ج/ يعتمد ذلك على نوع نموذج الانحدار اللوجستي الذي قمت بتشغيله ولماذا قمت بتشغيله.

في بعض الأحيان تقوم بتشغيل الانحدار اللوجستي كأداة تصنيف. في هذه الحالة، يمكنك إلقاء نظرة على مصفوفة الارتباك وAUC وما إلى ذلك. في أحيان أخرى تقوم بتشغيل الانحدار اللوجستي لنفس الأغراض مثل الانحدار الخطي - للنظر في العلاقة بين متغير تابع ومتغير واحد أو أكثر من المتغيرات المستقلة -. هناك يمكنك إلقاء نظرة على نسب الأرجحية وأخطائها القياسية؛ يمكنك أيضًا إلقاء نظرة على العديد من المقاييس R ^ 2، على الرغم من عدم وجود أي منها مثالي.

س136/ لماذا يتم استخدام تحويل فورييه Fourier transform في التعلم العميق؟

ج/تحويل فورييه عبارة عن حزمة فعالة تستخدم لتحليل وإدارة كميات كبيرة من البيانات الموجودة في قاعدة البيانات. يمكن أن تأخذ في الوقت الحقيقي مجموعة البيانات ومعالجتها بسرعة. يضمن ذلك الحفاظ على الكفاءة العالية ويجعل النموذج أيضًا أكثر انفتاحًا لمعالجة مجموعة متنوعة من الإشارات.

س137/ ما هي بعض الأمثلة على خوارزميات التعلم الخاضعة للأشراف supervised learning algorithms في التعلم العميق؟

ج/ هناك ثلاث خوارزميات تعلم رئيسية خاضعة للإشراف في التعلم العميق:

- الشبكات العصبية الاصطناعية Artificial neural networks
- الشبكات العصبية التلافيفية Convolutional neural networks
- الشبكات العصبية المتكررة Recurrent neural networks

س138/ لماذا يتحسن أداء التعلم العميق حيث يتم تغذية المزيد من البيانات إليه؟

ج/

- تتمثل إحدى أفضل فوائد التعلم العميق في قدرته على أداء الاستخراج التلقائي للميزات من البيانات الأولية.
- عندما يزداد عدد البيانات التي يتم تغذيتها في خوارزمية التعلم ، سيكون هناك المزيد من الحالات المتطورة التي تؤخذ في الاعتبار ، وبالتالي ستتعلم الخوارزمية اتخاذ القرارات الصحيحة في تلك الحالات المتطورة.

س139/ بالنظر إلى وجود العديد من خوارزميات التعلم العميق، كيف ستحدد نموذج التعلم العميق التي يجب استخدامها لمجموعة البيانات؟

ج/ هنالك أربع خطوات يجب أن تفهمها عند اختيار نموذج التعلم العميق:

1) فهم مجال المشكلة Understanding the Problem Domain
2) إيجاد الدقة "الصحيحة" Finding the "Right" Accuracy
3) معرفة بياناتك Knowing Your Data
4) اختيار المعمارية Picking the Architecture

س140/ لماذا تعتبر وحدات معالجة الرسوميات GPU مهمة لتنفيذ نماذج التعلم العميق؟

ج/ عندما نحاول بناء أي نموذج للشبكة العصبية، فإن مرحلة التدريب النموذجية هي أكثر الوظائف استهلاكا للموارد. يشتمل كل تكرار للتدريب النموذجي على آلاف (أو حتى أكثر) من عمليات ضرب المصفوفة التي تحدث. إذا كان هناك أقل من حوالي مائة الف معلمة في نموذج الشبكة العصبية، فلن يستغرق التدريب أكثر من بضع دقائق (أو بضع ساعات على الأكثر). ولكن عندما يكون لدينا الملايين من المعلمات، فهذا هو الوقت الذي من المحتمل أن تستسلم فيه أجهزة الكمبيوتر الكبيرة لدينا. هذا هو المكان الذي تظهر فيه وحدات معالجة الرسوميات GPU في الصورة. وحدات معالجة الرسوميات ليست سوى وحدات المعالجة المركزية (CPU) ولكن مع المزيد من ALUs (وحدات المنطق الحسابي) أكثر من وحدات المعالجة المركزية (CPU) العادية والتي تم تصميمها خصيصًا لهذا النوع من الحسابات الرياضية الثقيلة.

س141/ هل هناك فرق بين الشبكات العصبية والتعلم العميق؟

ج/

	الشبكات العصبية	التعلم العميق
التعريف	الشبكة العصبية هي نموذج من الخلايا العصبية مستوحى من الدماغ البشري. يتكون من العديد من الخلايا العصبية التي ترتبط ببعضها البعض.	تتميز الشبكات العصبية للتعلم العميق عن الشبكات العصبية على أساس عمقها أو عدد الطبقات المخفية.
الوقت والدقة	يستغرق تدريبهم وقتًا أقل بشكل عام. لديهم دقة أقل من أنظمة التعلم العميق	يستغرق تدريبهم مزيدًا من الوقت بشكل عام. لديهم دقة أعلى من الشبكات العصبية
الأداء	يعطي أداء منخفض مقارنة بشبكات التعلم العميق.	يعطي أداء عالي مقارنة بالشبكات العصبية.
تفسير المهام	يتم تفسير المهام بشكل سيئ من قبل الشبكة العصبية.	تدرك شبكة التعلم العميق المهام بشكل أكثر فاعلية.
التطبيقات	تجعل القدرة على نمذجة العمليات غير الخطية الشبكات العصبية أدوات ممتازة لمعالجة مجموعة متنوعة من القضايا ، بما في ذلك التصنيف والتعرف على الأنماط والتنبؤ والتحليل والتجميع واتخاذ القرار والتعلم الآلي والتعلم العميق والمزيد.	يمكن استخدام نماذج التعلم العميق في مجموعة متنوعة من الصناعات ، بما في ذلك التعرف على الأنماط والتعرف على الكلام ومعالجة اللغة الطبيعية وألعاب الكمبيوتر والسيارات ذاتية القيادة وتصفية الشبكات الاجتماعية والمزيد.

س142/ دقة النموذج Model accuracy أم أداء النموذج Model performance؟ أيهما تفضل ولماذا؟

ج/ هذا سؤال خادع، يجب أولاً الحصول على فكرة واضحة، ما هو أداء النموذج؟ إذا كان الأداء يعني السرعة، فهذا يعتمد على طبيعة التطبيق، أي تطبيق متعلق بسيناريو الوقت الفعلي سيحتاج إلى سرعة عالية كميزة مهمة. مثال: ستفقد أفضل نتائج البحث فضلها إذا لم تظهر نتائج طلب البحث بسرعة.

إذا تم التلميح إلى الأداء في سبب عدم كون الدقة accuracy هي المعيار الأكثر أهمية بالنسبة لأي مجموعة بيانات غير متوازنة، أكثر من الدقة، ستكون F1-score أكثر اهمية وفي حالة عدم توازن البيانات، فستكون الدقة Precisionوالاستدعاء Recall أكثر أهمية من البقية.

س143/ كيف يمكنك تحسين معلمات النموذج أثناء بناء النموذج؟

ج / تحسين معلمات النموذج هو عملية للعثور على أفضل القيم التي تتخذها معلمات النموذج. يمكن ضبط معلمات النموذج باستخدام خوارزمية بحث الشبكة Grid Search أو البحث العشوائي Random Search.

س144/ ما هي هياكل البيانات الشائعة المستخدمة في التعلم العميق؟

ج/ ينتقل التعلم العميق مباشرة من أبسط هياكل البيانات مثل القوائم List إلى تلك المعقدة مثل الرسوم البيانية الحسابية Computation Graphs.

فيما يلي أكثرها شيوعًا:

- List
- Matrix
- Dataframe
- Tensors
- Computation Graphs

س145/ ما هي لغة البرمجة أو التقنيات التي يجب أن يمتلكها شخص ما أو يتعلمها ليكون مهندس التعلم العميق DL؟

ج / اختيار لغة البرمجة هو المهمة الأولى التي تضعك على مسار التعلم العميق. اللغات الشائعة المفضلة لـ DL هي Python و R.

س146/ ما هي بعض المشاكل الشائعة التي يواجهها مهندسو التعلم العميق؟

ج/

- ❖ نوعية رديئة للبيانات.
- ❖ سوء تجهيز بيانات التدريب.
- ❖ الضبط الزائد لبيانات التدريب.
- ❖ التعلم الآلي هو عملية معقدة.
- ❖ نقص بيانات التدريب.
- ❖ التنفيذ البطيء.
- ❖ العيوب في الخوارزمية عندما تزداد البيانات.

س147/ ما هي المكتبة المفضلة في التعلم العميق ولماذا؟

ج / كيراس Keras هي مكتبة مفتوحة المصدر تعمل بكفاءة على وحدة المعالجة المركزية ووحدة معالجة الرسومات. يتم استخدامه للتعلم العميق، خاصة للشبكات العصبية

س148/ ما هو الاستخدام النهائي للتعلم العميق في وقتنا الحالي وكيف يساعد علماء البيانات؟

ج / يستخدم التعلم العميق لعدد من الحالات بما في ذلك التعرف على اللغة، والسيارات ذاتية القيادة، وإنشاء النصوص، وتحرير الفيديو والصور والمزيد. ومع ذلك، فإن أهم استخدام للتعلم العميق ربما يكون في مجال رؤية الكمبيوتر حيث يتم تغذية أجهزة الكمبيوتر بالبيانات ذات الصلة لتعلم اكتشاف الأشياء، واستعادة الصور وتجزئتها، والتشخيص الطبي، ومراقبة المحاصيل والثروة الحيوانية، وأكثر من ذلك. يستخدم العلماء التعلم العميق عبر الصناعات لأتمتة المهام القائمة على المراقبة والمتكررة لتحسين الإنتاجية والدقة.

س149/ قارن بين التعلم الآلي Machine Learning والبيانات الضخمة Big Data؟

ج/

الميزة	التعلم الالي	البيانات الضخمة
استخدام البيانات	التكنولوجيا التي تساعد في الحد من التدخل البشري	البحث عن البيانات، خاصة عند العمل مع البيانات الضخمة.
العمليات	تساعد البيانات الموجودة في تعليم الآلة لما يمكن القيام به بشكل أكبر	أنماط التصميم مع التحليلات على البيانات الموجودة من حيث اتخاذ القرار.
التعرف على الأنماط	على غرار البيانات الكبيرة، تساعد البيانات الموجودة في التعرف على الأنماط.	يساعد تحليل التسلسل والتصنيف في التعرف على الأنماط.
حجم البيانات	أفضل أداء أثناء العمل بمجموعات البيانات الصغيرة.	تساعد مجموعات البيانات في فهم وحل المشكلات المرتبطة بأحجام البيانات الكبيرة.

س150/ يتم تقديم مجموعة بيانات لك حول كشف الاحتيال fraud detection في المرافق utilities. لقد قمت ببناء نموذج مصنف وحقق أداء بنسبة 98.5٪. هل هذا نموذج جيد؟ إذا كانت الإجابة بنعم، فبرر ذلك. إذا لم يكن كذلك، فماذا يمكنك أن تفعل حيال ذلك؟

ج/ مجموعة البيانات حول كشف الاحتيال في المرافق ليست متوازنة imbalanced بما فيه الكفاية. في مثل هذه المجموعة من البيانات، لا يمكن أن تكون درجة الدقة accuracy مقياسًا للأداء حيث يمكن فقط توقع تسمية فئة الأغلبية بشكل صحيح ولكن في هذه الحالة، فإن نقطة اهتمامنا هي التنبؤ بتسمية الأقلية minority label. لكن غالبًا ما يتم التعامل مع الأقليات على أنها ضوضاء ويتم تجاهلها. لذلك، هناك احتمال كبير لسوء تصنيف misclassification تسمية الأقلية مقارنةً بتسمية الأغلبية majority label. لتقييم أداء النموذج في حالة مجموعات البيانات غير المتوازنة، يجب أن نستخدم الحساسية Sensitivity (المعدل الإيجابي الحقيقي) أو الخصوصية Specificity (المعدل السلبي الحقيقي) لتحديد الأداء لنموذج التصنيف.
إذا لم يكن أداء فئة الأقلية جيدًا، فيمكننا القيام بما يلي:

- يمكننا استخدام under sampling أو over sampling لموازنة البيانات.
- يمكننا تغيير قيمة عتبة التنبؤ prediction threshold.
- يمكننا تعيين أوزان للتسميات بحيث تحصل تسميات فئات الأقلية على أوزان أكبر.
- يمكننا الكشف عن الحالات الشاذة anomalies.

س151/ كيف تشرح الانحدار اللوجستي لاقتصادي وطبيب وعالم أحياء؟

ج/ يعد الانحدار اللوجستي أحد أبسط خوارزميات التعلم الآلي. يتم استخدامه للتنبؤ بالعلاقة بين متغير تابع فئوي ومتغيرين مستقلين أو أكثر. يتم إعطاء الصيغة الرياضية بواسطة

$$Y = \frac{e^{a+bX}}{1+e^{a+bX}}$$

حيث X هي المتغير المستقل، و a ، و b هي المعاملات ، و Y هي المتغير التابع الذي يمكن أن يأخذ قيمًا فئوية.

س152/ ما هو الفرق بين محلل البيانات Data Analyst وعالم البيانات Data Scientist؟

ج/

عادةً ما يعمل محللو البيانات مع البيانات المنظمة لحل مشاكل الأعمال الملموسة باستخدام أدوات مثل لغة البرمجة **SQL** أو **R** أو **Python**، وبرامج تصوير البيانات، والتحليل الإحصائي.

قد تتضمن المهام الشائعة لمحلل البيانات ما يلي:

- التعاون مع القادة التنظيميين لتحديد الاحتياجات المعلوماتية
- الحصول على البيانات من المصادر الأولية والثانوية
- تنظيف وإعادة تنظيم البيانات للتحليل
- تحليل مجموعات البيانات لتحديد الاتجاهات والأنماط التي يمكن ترجمتها إلى رؤى قابلة للتنفيذ
- تقديم النتائج بطريقة سهلة الفهم لإبلاغ القرارات التي تعتمد على البيانات

غالبًا ما يتعامل علماء البيانات مع المجهول **unknown** باستخدام تقنيات بيانات أكثر تقدمًا لعمل تنبؤات حول المستقبل. قد يقومون بأتمتة خوارزميات التعلم الآلي الخاصة بهم أو تصميم عمليات النمذجة التنبؤية التي يمكنها التعامل مع البيانات المهيكلة وغير المهيكلة. يعتبر هذا الدور بشكل عام إصدارًا أكثر تقدمًا من محلل البيانات. قد تتضمن بعض المهام اليومية لعالم البيانات ما يلي:

- جمع البيانات الخام وتنظيفها ومعالجتها
- تصميم النماذج التنبؤية وخوارزميات التعلم الآلي لتعدين مجموعات البيانات الضخمة
- تطوير الأدوات والعمليات لمراقبة دقة البيانات وتحليلها
- أدوات تصوير البيانات ولوحات المعلومات والتقارير
- كتابة برامج لأتمتة جمع البيانات ومعالجتها

س153/ ما هي الخطوات المختلفة التي ينطوي عليها مشروع تحليلات البياناتanalytics project ؟

ج/

- فهم مشكلة العمل وتحويلها إلى مشكلة تحليل البيانات.
- استخدام تقنيات تحليل البيانات الاستكشافية exploratory data analysis لفهم مجموعة البيانات المحددة.
- بمساعدة اختيار الميزات feature selection وأساليب هندسة الميزات feature engineering، يتم إعداد مجموعة بيانات التدريب والاختبار.
- استكشاف خوارزميات التعلم الآلي / التعلم العميق واستخدم واحدة لبناء نموذج تدريب.

- تغذية مجموعة بيانات التدريب إلى النموذج وتحسين أداء النموذج من خلال تحليل المعلمات الإحصائية المختلفة.
- اختبار أداء النموذج باستخدام مجموعة بيانات الاختبار.
- نشر النموذج ، إذا لزم الأمر ، ومراقبة أداء النموذج.

س154/ ماذا تعني البرمجة اللغوية العصبية NLP؟

ج / البرمجة اللغوية العصبية (NLP) هي اختصار لـ "Natural Language Processing". إنه يتعامل مع دراسة كيفية تعلم أجهزة الكمبيوتر كمية هائلة من البيانات النصية من خلال البرمجة. بعض الأمثلة الشائعة على البرمجة اللغوية العصبية هي: التشذيب Stemming، وتحليل المشاعر Sentimental Analysis، التقطيع Tokenization، وإزالة كلمات التوقف removal of stop words، وما إلى ذلك.

س155/ كيف تتعامل مع مجموعة بيانات تفتقد إلى أكثر من 30 في المائة من قيمها؟

ج / ستعتمد الطريقة على حجم مجموعة البيانات.

- إذا كانت مجموعة البيانات كبيرة، فإن أسرع طريقة هي إزالة الصفوف التي تحتوي على القيم المفقودة. نظرًا لأن مجموعة البيانات كبيرة، فلن يؤثر ذلك على قدرة النموذج على تحقيق النتائج.
- إذا كانت مجموعة البيانات صغيرة، فليس من العملي إزالة القيم ببساطة. في هذه الحالة، من الأفضل حساب المتوسط أو الوضع الخاص بهذه الميزة المعينة وإدخال تلك القيمة حيث لا توجد إدخالات مفقودة.

هناك طريقة أخرى تتمثل في استخدام خوارزمية التعلم الآلي للتنبؤ بالقيم المفقودة. يمكن أن يؤدي هذا إلى نتائج دقيقة ما لم تكن هناك إدخالات ذات تباين كبير جدًا عن بقية مجموعة البيانات.

س156/ هل يمكنك بناء نماذج تعلم عميقة يعتمد فقط على الانحدار الخطي؟

ج / نعم، من الممكن بالتأكيد بناء شبكات عميقة باستخدام دالة خطية كدالة تنشيط لكل طبقة إذا تم تمثيل المشكلة بمعادلة خطية. ومع ذلك، فإن المشكلة التي تتكون من تكوين دوال خطية هي دالة خطية ولا يوجد شيء غير عادي يمكن تحقيقه من خلال تنفيذ شبكة عميقة لأن إضافة المزيد من العقد إلى الشبكة لن يزيد من القدرة التنبؤية لنموذج التعلم الآلي.

س157/ في التعلم الآلي، ما عدد الكلاسات التي يمكن ان تستخدم الانحدار اللوجستي Logistic Regression؟

ج / لا يمكن استخدام الانحدار اللوجستي لأكثر من فئتين. الانحدار اللوجستي، افتراضيًا، مصنف ثنائي. ومع ذلك، في الحالات التي تحتاج فيها مشاكل التصنيف متعدد الفئات إلى حل، يمكن تمديد العدد الافتراضي للفئات، أي الانحدار اللوجستي متعدد الحدود multinomial logistic regression.

س158/ أخبرني عن حالة قد تكون فيها التقنيات الجماعية ensemble techniques مفيدة؟

ج / تستخدم التقنيات الجماعية ensemble techniques مجموعة من خوارزميات التعلم لتحسين الأداء التنبئي. عادةً ما تقلل من فرط التعلم في النماذج وتجعل النموذج أكثر قوة (من غير المرجح أن تتأثر بالتغييرات الصغيرة في بيانات التدريب).

يمكنك سرد بعض الأمثلة عن طرق التجميع (التعبئة bagging، التعزيز boosting، طريقة "bucket of models") وإثبات كيف يمكنها زيادة القدرة التنبؤية.

س159/ ما هو المقصود بالتعلم الجماعي Ensemble Learning؟

ج /لحل برنامج حسابي معين، يتم إنشاء نماذج متعددة مثل المصنفات أو الخبراء بشكل استراتيجي ودمجها. تُعرف هذه العملية باسم التعلم الجماعي Ensemble Learning.

س160/ أخبرني عن حالة قد تكون فيها التقنيات الجماعية ensemble techniques مفيدة؟

ج /تجمع التقنية الجماعية ensemble techniques بين عدة نماذج في نموذج تنبؤي واحد لتقليل التباين وتحسين النتائج. تنقسم طريقة التجميع إلى مجموعتين:

- الطريقة المتسلسلة: يتم إنشاء المتعلمين الأساسيين بالتتابع.
- الطريقة الموازية: يتم إنشاء المتعلمين الأساسيين بشكل متوازي.

التقنيات الجماعية هي:

- تكييس Bagging
- التكديس Stacking
- التعزيز Boosting

السيناريو: لنفترض أنك تريد شراء زوج جديد من سماعات الرأس. ماذا ستفعل؟ كونك مستهلكًا واعيًا، أولاً، ستقوم بإجراء بحث حول الشركة التي تقدم أفضل

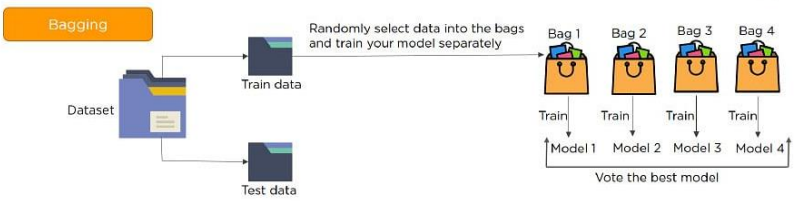
سماعات الرأس وأيضًا أخذ بعض الاقتراحات من أصدقائك. باختصار، ستتخذ قرارات مستنيرة بعد إجراء بحث شامل عن العمل.

س161/ ما معنى bagging وboosting في التعلم العميق؟

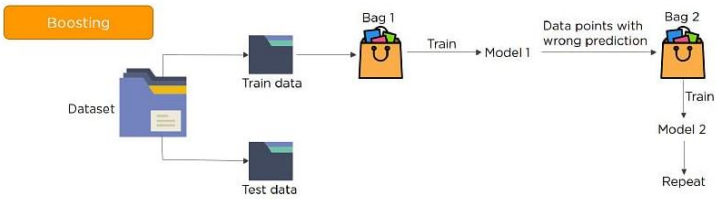
ج/

- التعبئة bagging هو مفهوم تقسيم مجموعة البيانات ووضعها بشكل عشوائي في bags لتدريب النموذج.
- التعزيز boosting هو السيناريو الذي يتم فيه استخدام نقاط البيانات غير الصحيحة لإجبار النموذج على إنتاج مخرجات خاطئة. يستخدم هذا لإعادة تدريب النموذج وزيادة الدقة.

باستخدام التعبئة bagging، نأخذ مجموعة بيانات ونقسمها إلى بيانات تدريب وبيانات اختبار. ثم نختار البيانات بشكل عشوائي لوضعها في الأكياس وتدريب النموذج بشكل منفصل.



مع التعزيز Boosting، يكون التركيز على اختيار نقاط البيانات التي تعطي مخرجات خاطئة لتحسين الدقة.



س162/ ما هو التعلم الشامل end-to-end learning؟ أعط بعض مزاياها

ج/ إنها عملية تعلم عميقة حيث يحصل النموذج على بيانات أولية كمدخلات ويتم تدريب جميع الأجزاء المختلفة في وقت واحد لإنتاج النتيجة المرجوة بدون مهام وسيطة. تتمثل ميزة التعلم الشامل end-to-end learning في أنه لا توجد حاجة للقيام بهندسة ميزات ضمنيًا مما يؤدي عادةً إلى تحيز أقل. من الأمثلة الجيدة التي يمكنك اقتباسها في محتوى التعلم الشامل السيارات ذاتية القيادة driverless cars.

يستخدمون المدخلات المقدمة من الإنسان كدليل ويتم تدريبهم على تعلم ومعالجة المعلومات تلقائيًا باستخدام CNN لإكمال المهام.

س163/ لماذا يُشار أحيانًا إلى خوارزميات التعلم القائم على المثيل instance-based learning باسم خوارزميات التعلم الكسول Lazy learning؟

ج / التعلم الكسول Lazy learning هو أسلوب تعلم آلي يتم فيه تأجيل إجراءات الاستقراء induction والتعميم generalization حتى يتم الانتهاء من التصنيف. تُعرف خوارزمية التعلم القائم على المثيل instance-based learning أيضًا باسم خوارزمية التعلم البطيء slow learning بسبب نفس الخاصية.

س164/ متى يفضل التعلم متعدد المهام multi-task learning؟

ج / يعد التعلم متعدد المهام multi-task learning باستخدام الشبكات العصبية العميقة مجالًا فرعيًا يتم فيه تعلم العديد من المهام بواسطة نموذج مشترك. هذا يقلل من الضبط الزائد overfitting، ويعزز كفاءة البيانات، ويسرع عملية التعلم باستخدام المعلومات المساعدة. يعد التعلم متعدد المهام مفيدًا عندما يكون هناك كمية صغيرة من البيانات لأي مهمة معينة ويمكننا الاستفادة من تدريب نموذج التعلم العميق على مجموعة بيانات كبيرة.

س165/ ما هو التعلم المعزز reinforcement learning؟

ج / التعلم المعزز reinforcement learning هو آلية تعلم حول كيفية تعيين الحالات situations إلى الإجراءات actions. يجب أن تساعدك النتيجة النهائية على زيادة إشارة المكافأة الثنائية binary reward signal. في هذه الطريقة، لا يتم إخبار المتعلم بالإجراء الذي يجب اتخاذه ولكن بدلاً من ذلك يجب أن يكتشف الإجراء الذي يقدم مكافأة قصوى maximum reward. حيث أن هذه الطريقة تعتمد على آلية المكافأة / الجزاء reward/penalty.

س166/ اشرح كيف يعمل نظام التوصية recommender system؟

ج / يستخدم نظام التوصية السلوك التاريخي للتنبؤ بكيفية قيام المستخدم بتصنيف عنصر معين. على سبيل المثال، توصي Netflix المستخدمين بالعروض التلفزيونية والأفلام من خلال تحليل الوسائط التي صنفها المستخدمون في الماضي، واستخدامها للتوصية بالوسائط الجديدة التي قد تنال إعجابهم.

س167/ ما هي أنظمة التوصية recommender systems؟

ج/ غالبًا ما يكون فهم سلوك المستهلك هو الهدف الأساسي للعديد من الشركات. على سبيل المثال، ضع في اعتبارك حالة أمازون. إذا بحث أحد المستخدمين عن فئة منتج على موقعه على الويب، فإن التحدي الرئيسي الذي يواجه خوارزميات الواجهة الخلفية في أمازون هو التوصل إلى اقتراحات من المحتمل أن تحفز المستخدمين على إجراء عملية شراء. ومثل هذه الخوارزميات هي قلب أنظمة التوصية recommender systems. تهدف هذه الأنظمة إلى تحليل سلوك العملاء وتقييم ولعهم بالمنتجات المختلفة. بصرف النظر عن Amazon، يتم استخدام أنظمة التوصية أيضًا بواسطة Netflix وYoutubeg و Flipkart وما إلى ذلك.

س168/ اشرح التصفية على أساس المحتوى content-based filtering في أنظمة التوصية recommender systems؟

ج / التصفية القائمة على المحتوى content-based filtering هي إحدى التقنيات المستخدمة لبناء أنظمة التوصية recommender systems. في هذه التقنية، يتم إنشاء التوصيات من خلال الاستفادة من خصائص المحتوى الذي يهتم به المستخدم.

على سبيل المثال، إذا كان المستخدم يشاهد أفلامًا تنتمي إلى نوع الحركة والغموض ومنحها تقييمات جيدة، فهذا مؤشر واضح على أن المستخدم يحب أفلامًا من هذا النوع. إذا تم عرض أفلام من نوع مشابه للتوصيات، فهناك احتمال أكبر بأن المستخدم سيحب هذه التوصيات أيضًا.

بمعنى آخر، هنا، يتم أخذ محتوى الفيلم في الاعتبار عند إنشاء توصيات للمستخدمين.

س169/ ما هو التصفية التعاونية collaborative filtering؟

ج/التصفية التعاونية collaborative filtering هي شكل من أشكال تصفية المحتوى التي تستخدم أوجه التشابه بين مختلف المستخدمين لتقديم التوصيات.

س170/ ما نوع نظام التوصية recommendation system الذي تستخدمه أمازون للتوصية بعناصر مماثلة؟

ج / تستخدم أمازون Amazon خوارزمية تصفية تعاونية collaborative filtering للتوصية بعناصر مماثلة.

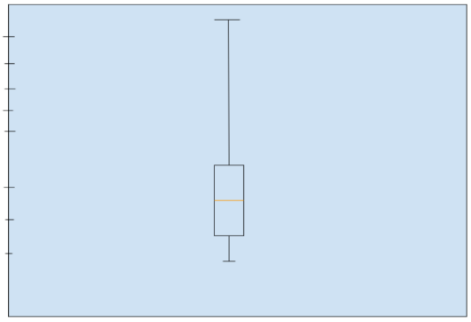
س171/ في Google، إذا قمت بكتابة "How are" فإنه يمنحك التوصية على أنها "How do you do" / "How are you" ، فهذا يعتمد على ماذا؟

ج / يأتي هذا النوع من محركات التوصية **recommendation engine** من التصفية التعاونية **collaborative filtering**.

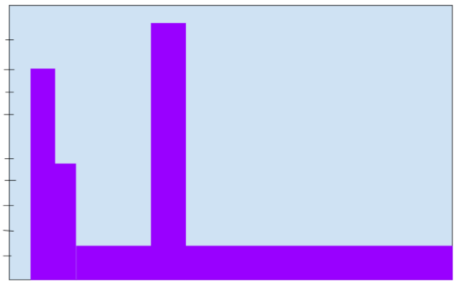
س172/ ما هو الفرق بين box plot وhistogram؟

ج /يُشار إلى تكرار قيم ميزة معينة بصريًا بواسطة كل من المخططات الصندوقية **box plot** والمدرج التكراري و**histogram**.

غالبًا ما يتم استخدام **Boxplots** في مقارنة العديد من مجموعات البيانات وبالمقارنة مع **Histograms**، تأخذ مساحة أقل وتحتوي على تفاصيل أقل. تُستخدم **Histograms** لمعرفة وفهم توزيع الاحتمالات الكامن وراء مجموعة البيانات



يشير الرسم البياني أعلاه إلى **box plot** لمجموعة بيانات.



يشير الرسم البياني أعلاه إلى **Histogram** لمجموعة بيانات.

س173/ لماذا يتم استخدام لغة R في تصوير البيانات Data Visualization؟

ج / يستخدم R على نطاق واسع في تصوير البيانات Data Visualization للأسباب التالية:

- يمكننا إنشاء أي نوع من الرسوم البيانية تقريبًا باستخدام R.
- يحتوي R على مكتبات متعددة مثل lattice ، و ggplot2 و leaflet ، وما إلى ذلك ، والعديد من الدوال المدمجة أيضًا.
- من الأسهل تخصيص الرسومات في R مقارنة ببايثون.
- يستخدم R في هندسة الميزات وفي تحليل البيانات الاستكشافية exploratory data analysis أيضًا.

س174/ ما هو XGBoost؟

ج /XGBoost هو نوع جديد من خوارزمية التعزيز boosting التي تستفيد من التعزيز وتصميم الأجهزة والعقوبات النموذجية لإنشاء خوارزمية تعزيز دقيقة للغاية وسريعة جدًا (Python، R ... الحزم المتوفرة هذه الأيام). يجعل تعزيز بديل قابل للتطبيق للغابات العشوائية للاستخدام في التطبيقات التنبؤية السريعة.

س175/ ما هو التجميع Clustering في التعلم الآلي؟

ج / عندما تحاول التعرف على شيء ما، مثل الموسيقى، فقد يكون أحد الأساليب هو البحث عن مجموعات أو مجموعات ذات معنى. يمكنك تنظيم الموسيقى حسب النوع، بينما قد ينظم صديقك الموسيقى حسب العقد. تساعدك الطريقة التي تختارها لتجميع العناصر على فهم المزيد عنها كمقاطع موسيقية فردية. قد تجد أن لديك تقاربًا عميقًا مع موسيقى البانك روك وتقسم هذا النوع إلى أساليب مختلفة أو موسيقى من مواقع مختلفة. من ناحية أخرى، قد ينظر صديقك إلى الموسيقى من الثمانينيات ويكون قادرًا على فهم كيف تأثرت الموسيقى عبر الأنواع في ذلك الوقت بالمناخ الاجتماعي والسياسي. في كلتا الحالتين، تعلمت أنت وصديقك شيئًا مثيرًا للاهتمام حول الموسيقى، على الرغم من اتباعكما نهجًا مختلفًا.

في التعلم الآلي أيضًا، غالبًا ما نقوم بتجميع الأمثلة كخطوة أولى لفهم موضوع (مجموعة بيانات) في نظام التعلم الآلي. يسمى تجميع الأمثلة غير المسماة التجميع Clustering.

نظرًا لأن الأمثلة غير مسماة unlabeled، يعتمد التجميع على التعلم الآلي غير الخاضع للإشراف. إذا تم تصنيف الأمثلة، فسيصبح التجميع تصنيفًا.

س176/ لماذا يعتبر الانحدار اللوجستي نوعًا من تقنيات التصنيف وليس الانحدار؟

ج/ الانحدار اللوجستي Logistic Regression هو أسلوب تصنيف يستخدم في التعلم الآلي. يستخدم دالة لوجستية لنمذجة المتغير التابع. المتغير التابع ثنائي التفرع بطبيعته، أي يمكن أن يكون هناك فئتان محتملتان فقط (على سبيل المثال: إما أن يكون السرطان خبيثًا أم لا).

س177/ اشرح الانحدار اللوجستي Logistic Regression؟

ج/ هو شكل من أشكال التحليل التنبئي. يتم استخدامه لإيجاد العلاقات الموجودة بين متغير ثنائي تابع dependent binary variable ومتغير واحد أو أكثر من المتغيرات المستقلة independent variables عن طريق استخدام معادلة الانحدار اللوجستي.

س178/ ما هو الفرق بين مجموعة التحقق من الصحة validation set ومجموعة الاختبار Test set؟

ج/ تعتبر مجموعة التحقق من الصحة validation set في الغالب جزءًا من مجموعة التدريب حيث يتم استخدامها لتحديد المعلمات التي تساعدك على تجنب الضبط الزائد للنموذج الذي يتم بناؤه.

بينما يتم استخدام مجموعة الاختبار Test set لاختبار أو تقييم أداء نموذج التعلم الآلي المُدرَّب.

س179/ ما هي تقنية cross-validation (CV) الذي ستستخدمه في مجموعة بيانات السلاسل الزمنية time series dataset؟

ج/ يتم استخدام cross-validation لضبط المعلمات الفائقة وإنتاج قياسات أداء النموذج. باستخدام بيانات السلاسل الزمنية، لا يمكننا استخدام تقنية cross-validation التقليدية لسببين رئيسيين وهما كالتالي -

- التبعيات الزمنية Temporal dependencies
- الاختيار التعسفي لمجموعة الاختبار Arbitrary Choice of Test Set

بالنسبة لبيانات السلاسل الزمنية، نستخدم CV المتداخل الذي يوفر تقديرًا غير متحيز تقريبًا للخطأ الحقيقي. تتكون CV المتداخلة من حلقة داخلية لضبط المعلمة وحلقة خارجية لتقدير الخطأ.

س180/ هل من الممكن اختبار احتمال تحسين دقة النموذج بدون تقنيات التحقق المتبادل cross-validation؟ إذا كانت الإجابة بنعم، رجاء التوضيح؟

ج / نعم، من الممكن اختبار احتمالية تحسين دقة النموذج بدون تقنيات التحقق المتبادل **cross-validation**. يمكننا القيام بذلك عن طريق تشغيل نموذج تعلم آلي لـ **n** عدد التكرارات، وتسجيل الدقة. ارسم كل الدقة وأزل **5**٪ من قيم الاحتمالات المنخفضة **low probability**. قم بقياس القطع الأيسر [المنخفض] واليمين [العالي] المقطوع. مع نسبة **95**٪ المتبقية من **confidence**، يمكننا القول أن النموذج يمكن أن يكون منخفضًا أو مرتفعًا [كما هو مذكور ضمن نقاط القطع].

س181/ ما هو k-fold cross-validation؟

ج / التحقق المتقاطع cross-validation هو أسلوب يستخدم لتقدير فعالية نموذج التعلم الآلي. المعلمة k هي عبارة عن عدد المجموعات التي يمكن تقسيم مجموعة البيانات إليها.

تبدأ العملية مع خلط مجموعة البيانات بأكملها بطريقة عشوائية. ثم يتم تقسيمها إلى مجموعات k، والمعروفة أيضًا باسم الطيات folds. يتم تطبيق الإجراء التالي على كل طية فريدة **unique fold**:

1. قم بتعيين طية واحدة كطية اختبار وطيات k-1 المتبقية كمجموعة اختبار.
2. ابدأ تدريب النموذج على مجموعة التدريب. لكل عملية تكرار للتحقق المتبادل ، قم بتدريب نموذج جديد مستقل عن النماذج المستخدمة في التكرارات السابقة.
3. تحقق من صحة النموذج في مجموعة الاختبار واحفظ نتيجة كل تكرار.
4. متوسط النتائج من كل تكرار للحصول على النتيجة النهائية.

س182/ عدد تقنيات التحقق المتبادل cross validation؟

ج / هناك بعض الطرق الشائعة المستخدمة للتحقق المتبادل cross validation. هذه الطرق موضحة أدناه:

1. Validation Set Approach
2. Leave-P-out cross-validation
3. Leave one out cross-validation
4. K-fold cross-validation

5. Stratified k-fold cross-validation

س183/ ما هو Cross-validation في التعلم الآلي؟

ج / التحقق المتقاطع Cross-validation هو أسلوب للتحقق من الصحة لتقييم كيفية تعميم نتائج التحليل الإحصائي لمجموعة بيانات مستقلة. تُستخدم هذه الطريقة في الخلفيات التي يتم فيها توقع الهدف، ويحتاج المرء إلى تقدير مدى دقة إنجاز النموذج.

س184/ ما هي الحدود الدنيا المحلية local minima والحد الأدنى العالمية Global Minima في الانحدار؟

ج/

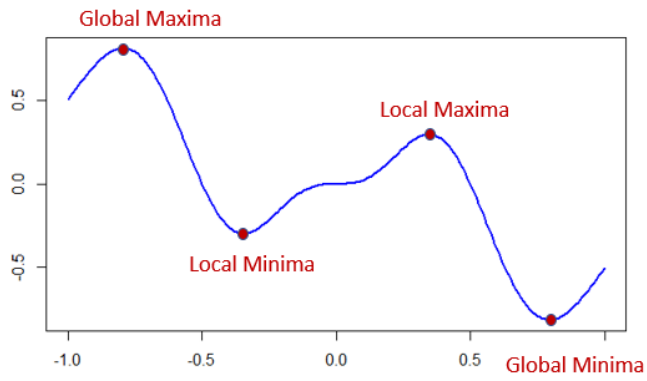
- الحد الأدنى المحلي **local minima**:

تسمى النقطة في المنحنى الأدنى عند مقارنتها بالنقاط السابقة والنقاط التالية بالحد الأدنى المحلي.

- الحد الأدنى العالمي **Global Minima**:

النقطة الدنيا في المنحنى عند مقارنتها بجميع النقاط في المنحنى تسمى **Global Minima**.

بالنسبة للمنحنى، يمكن أن يكون هناك أكثر من حد أدنى محلي واحد، ولكنه يحتوي على حد أدنى عالمي واحد فقط. في الانحدار التدريجي **gradient descent**، نستخدم هذه الحدود الدنيا المحلية والعالمية لتقليل دوال الخسارة (الخطأ).



س185/ ما هي الانتروبية Entropy في التعلم الآلي؟

ج / يقيس الانتروبيا Entropy في التعلم الآلي العشوائية في البيانات التي يجب معالجتها. كلما زاد الانتروبيا في البيانات المقدمة، أصبح من الصعب استخلاص أي استنتاج مفيد من البيانات. على سبيل المثال، دعونا نأخذ تقليب عملة معدنية. نتيجة هذا الفعل عشوائية لأنها لا تحبذ الرؤوس أو الذيل. هنا، لا يمكن توقع نتيجة أي عدد من الرميات بسهولة حيث لا توجد علاقة محددة بين فعل التقليب والنتائج المحتملة.

س186/ ما هو PCA في التعلم الآلي؟

ج/يعد تحليل المكونات الرئيسية Principal component analysis(PCA) أحد أكثر خوارزميات التعلم الآلي غير الخاضعة للإشراف استخدامًا عبر مجموعة متنوعة من التطبيقات: تحليل البيانات الاستكشافية، وتقليل الأبعاد، وضغط المعلومات، وإزالة الضوضاء، وغير ذلك الكثير!

س187/ هل rotation ضروري في PCA؟

ج/نعم، التدوير rotation ضروري بالتأكيد لأنه يزيد الفروق بين التباين الذي تلتقطه المكونات.

س188/ ماذا يحدث إذا لم يتم تدوير المكونات في PCA؟

ج/إنه تأثير مباشر. إذا لم يتم تدوير المكونات، فسوف تتضاءل في النهاية وسيتعين على المرء استخدام الكثير من المكونات المختلفة لشرح تباين مجموعة البيانات.

س189/ ما هو تحليل التمايز الخطي Linear discriminant analysis (LDA)؟

ج / LDA هي خوارزمية نمذجة تنبؤية لتصنيف متعدد الفئات. سيحسب LDA الاتجاهات التي ستمثل المحاور axes التي تزيد من الفصل بين الفئات.

يتم استخدام LDA بشكل أساسي هنا لتقليل عدد الميزات إلى رقم أكثر قابلية للإدارة قبل التصنيف. كل من الأبعاد الجديدة عبارة عن مجموعة خطية من قيم البكسل، والتي تشكل قالبًا.

س190/ ما هي العلاقة بين PCA وLDA؟

ج/يتم استخدام كلتا الطريقتين لتقليل الأبعاد dimensionality reduction. PCA غير خاضع للإشراف بينما LDA خاضع للأشراف.

س191/ يتم إعطاؤك مجموعة بيانات تدريب بها الكثير من الأعمدة والصفوف. كيف تقلل من أبعاد هذه البيانات؟

ج/

- سيساعدنا تحليل المكونات الرئيسية (PCA) هنا والذي يمكن أن يفسر الحد الأقصى للتباين في مجموعة البيانات.
- يمكننا أيضًا التحقق من العلاقة المشتركة للبيانات العددية وإزالة مشكلة العلاقة الخطية المتعددة (إن وجدت) وإزالة بعض الأعمدة التي قد لا تؤثر على النموذج.
- يمكننا إنشاء مجموعة بيانات متعددة وتنفيذها على دفعات.

س192/ كيف يختلف PCA عن LDA؟

ج/

- **PCA** غير خاضع للإشراف. **LDA** غير خاضع للإشراف.
- **PCA** يأخذ في الاعتبار التباين **variance**. يأخذ **LDA** في الاعتبار توزيع الفئات **distribution of classes**.

س193/ ما هي عيوب النموذج الخطي linear model؟

ج/ ثلاثة عيوب للنموذج الخطي هي:

- افتراض الخطية للأخطاء.
- لا يمكنك استخدام هذا النموذج للنتائج الثنائية أو العد
- هناك الكثير من مشاكل فرط التعلم (الضبط الزائد) التي لا يمكن حلها

س194/ يتم منحك مجموعة بيانات عن كشف الاحتيال fraud detection. حقق نموذج التصنيف دقة 95٪ هل هو جيد؟

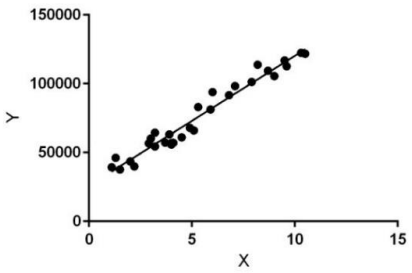
ج/ دقة **95**٪ جيدة. ولكن قد يتعين علينا التحقق من العناصر التالية:

- ما هي مجموعة البيانات الخاصة بمشكلة التصنيف.
- هل الحساسية **Sensitivity** والخصوصية **Specificity** مقبولة.
- إذا كانت هناك حالات أقل سلبية فقط ، ولم يتم تصنيف جميع الحالات السلبية بشكل صحيح ، فقد تكون هناك مشكلة.
- بالإضافة إلى ذلك ، يتعلق الأمر باكتشاف الاحتيال ، وبالتالي يجب توخي الحذر هنا في التنبؤ (أي عدم التنبؤ بالخطأ بالاحتيال كمريض غير احتيالي.

س195/ ما هو الانحدار الخطي Linear Regression في التعلم الآلي؟

ج/ الانحدار الخطي Linear Regression هو خوارزمية تعلم آلي تعتمد على التعلم الخاضع للإشراف. ينفذ مهمة الانحدار. نماذج الانحدار قيمة تنبؤ مستهدفة بناءً على متغيرات مستقلة. يستخدم في الغالب لاكتشاف العلاقة بين المتغيرات والتنبؤ. تختلف نماذج الانحدار المختلفة بناءً على - نوع العلاقة بين المتغيرات التابعة والمستقلة التي يدرسونها، وعدد المتغيرات المستقلة التي يتم استخدامها.

ينفذ الانحدار الخطي مهمة التنبؤ بقيمة متغير تابع (y) بناءً على متغير مستقل معين (x). لذلك، تكتشف تقنية الانحدار هذه علاقة خطية بين x (المدخلات) و y (المخرجات). ومن ثم، فإن الاسم هو الانحدار الخطي.



في الشكل أعلاه، X (المدخلات) هي خبرة العمل و Y (المخرجات) هي راتب الشخص. خط الانحدار هو أفضل خط ملائم لنموذجنا.

س196/ ماذا تقصد بالتعلم العميق Deep learning ولماذا أصبح شائعًا الآن؟

ج/ التعلم العميق Deep learning ليس سوى نموذج للتعلم الآلي الذي أظهر وعدًا لا يصدق في السنوات الأخيرة. هذا بسبب حقيقة أن التعلم العميق يظهر تشابهًا كبيرًا مع عمل الدماغ البشري.

الآن على الرغم من أن التعلم العميق كان موجودًا منذ سنوات عديدة، إلا أن الاختراقات الكبرى من هذه التقنيات جاءت في السنوات الأخيرة فقط. هذا لسببين رئيسيين:

- الزيادة في كمية البيانات التي يتم إنشاؤها من خلال المصادر المختلفة النمو في موارد الأجهزة المطلوبة لتشغيل هذه النماذج.
- تعد وحدات معالجة الرسومات أسرع عدة مرات وتساعدنا في بناء نماذج تعلم أكبر وأعمق في وقت أقل نسبيًا مما كنا نحتاجه سابقًا.

س197/ كيف يمكننا استخدام مجموعة بيانات بدون المتغير المستهدف في خوارزميات التعلم الخاضعة للإشراف؟

ج / أدخل مجموعة البيانات في خوارزمية التجميع **clustering**، وقم بإنشاء مجموعات مثالية، وقم بتسمية أرقام الكتلة (الكلاستر) على أنها المتغير المستهدف الجديد **new target**. الآن، تحتوي مجموعة البيانات على متغيرات مستقلة **independent** ومستهدفة **target** موجودة. هذا يضمن أن مجموعة البيانات جاهزة للاستخدام في خوارزميات التعلم الخاضع للإشراف.

س198/ هل يمكنك تسمية بعض هياكل البيانات data structures التي يشيع استخدامها في التعلم العميق؟

ج / يمكنك التحدث عن الرسوم البيانية الحسابية **computational graphs** والموترات **tensors** والمصفوفات **matrices** وإطارات البيانات **data frames** والقوائم **lists**.

س199/ كيف يتم استخدام تحويل فورييه Fourier Transform لصالح التعلم العميق؟

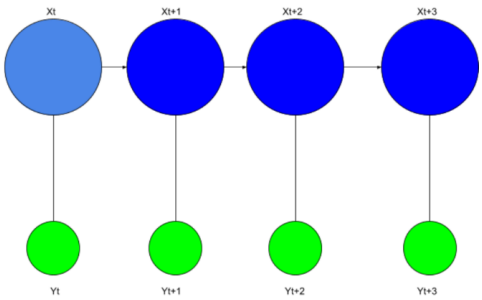
ج / تعمل دالة تحويل فورييه **Fourier transform** على تحليل مجموعات البيانات الكبيرة وصيانتها وإدارتها بكفاءة. يمكنك استخدامه لإنشاء بيانات مصفوفة في الوقت الفعلي تساعد في معالجة إشارات متعددة.

س200/ ما هي سلاسل ماركوف Markov chains؟

ج / تحدد سلاسل ماركوف Markov Chains أن الاحتمال المستقبلي للحالة يعتمد فقط على حالتها الحالية.

تنتمي سلاسل ماركوف إلى فئة نوع العملية العشوائية.

يوضح الرسم البياني أدناه نموذجًا تدريجيًا لسلاسل ماركوف التي يعتمد ناتجها على حالتها الحالية.



مثال ممتاز على سلاسل ماركوف هو نظام التوصية بالكلمات **word recommendation**. في هذا النظام، يتعرف النموذج على الكلمة التالية ويوصي بها بناءً على الكلمة السابقة مباشرة وليس أي شيء قبل ذلك. تأخذ سلاسل ماركوف الفقرات السابقة التي كانت مشابهة لمجموعات بيانات التدريب وتقوم بإنشاء التوصيات للفقرات الحالية وفقًا للكلمة السابقة.

س201/ اشرح فوائد استخدام الإحصائيات statistics بواسطة علماء البيانات Data Scientists؟

ج/ تساعد الإحصائيات **statistics** عالم البيانات في تكوين فكرة أفضل عن توقعات العميل. باستخدام الأسلوب الإحصائي، يمكن لعلماء البيانات الحصول على المعرفة فيما يتعلق باهتمام المستهلك، والسلوك، والمشاركة، والاحتفاظ، وما إلى ذلك. كما أنه يساعدك على بناء نماذج بيانات قوية للتحقق من صحة بعض الاستدلالات والتنبؤات.

س202/ ما هي آلة بولتزمان Boltzmann machine؟

ج/آلة بولتزمان **Boltzmann machine** هي نوع من الشبكات العصبية المتكررة التي تستخدم قرارات ثنائية، إلى جانب التحيزات، لتعمل. يمكن ربط هذه الشبكات العصبية معًا لإنشاء شبكات اعتقاد عميقة deep belief networks، وهي معقدة للغاية وتستخدم لحل أكثر المشكلات تعقيدًا.

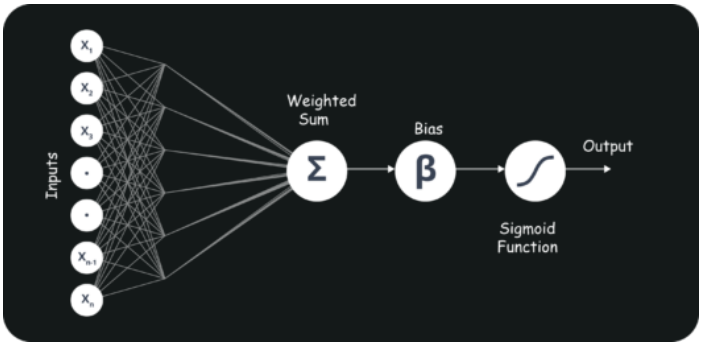
س203/ ما هو البيرسيبترون perceptron؟

ج/ يشبه البيرسيبترون perceptron العصبون الفعلي في دماغ الإنسان. يتلقى مدخلات من كيانات مختلفة ويطبق دوال على هذه المدخلات، مما يحولها إلى مخرجات.

يستخدم المدرك بشكل أساسي لإجراء التصنيف الثنائي حيث يرى المدخلات، ويحسب الدوال بناءً على أوزان المدخلات، ويخرج التحويل المطلوب.

س204/ صف هيكل الشبكات العصبية الاصطناعية ANN؟

ج / تعمل الشبكات العصبية الاصطناعية **ANN** على نفس مبدأ الشبكة العصبية البيولوجية. وهو يتألف من مدخلات يتم معالجتها باستخدام مجموع الاوزان والتحيز بمساعدة دوال التنشيط.



س205/ ما هو البيرسيبترون Perceptron؟ وكيف يعمل؟

ج/

- تم تطوير Perceptron في الخمسينيات والستينيات من قبل العالم فرانك روزنبلات ، مستوحى من أعمال سابقة لوارن ماكولوتش ووالتر بيتس.
- تعد أداة Perceptron واحدة من أبسط وحدات ANN (الشبكة العصبية الاصطناعية) التي تقوم بحسابات معينة لاكتشاف الميزات أو ذكاء الأعمال في بيانات الإدخال.
- يعتمد Perceptron على خلية عصبية اصطناعية تسمى وحدة منطق العتبة (TLU)
- المدخلات والمخرجات عبارة عن أرقام وليست قيمًا ثنائية وكل وصلة إدخال مرتبطة بوزن.
- يحسب TLU مجموعًا مرجحًا لمدخلاته: z = w1 x1 + w2 x2 + â(¯ +) wn xn = wT x ، ثم يطبق دالة خطوة على هذا المجموع ويخرج النتيجة: hw (x) hw(x) = step(z), where z = wT x.
- يمكن استخدام TLU واحد لتصنيف ثنائي خطي بسيط .

س206/ ما هي الخطوات المتبعة في تدريب perceptron في التعلم العميق؟

ج/هناك خمس خطوات رئيسية تحدد تدريب perceptron:

- تهيئة العتبات والأوزان.
- تقديم المدخلات.
- حساب النواتج.
- تحديث الأوزان في كل خطوة.
- تكرار الخطوات من 2 إلى 4.

س207/ ما هي مزايا استخدام Perceptron متعدد الطبقات على Perceptron أحادي الطبقة؟

ج/ يستخدم البيرسيبترون متعدد الطبقات MLP أسلوب التعلم الخاضع للإشراف يسمى الانتشار الخلفي backpropagation للتدريب. طبقاته المتعددة وتنشيطه غير الخطي يميز MLP عن البيرسيبترون الخطي. يمكنه تمييز البيانات غير القابلة للفصل خطيًا.

س208/ ما هي محددات استخدام البيرسبترون Perceptron؟

ج/ نموذج البيرسبترون perceptron له محددات (قيود) على النحو التالي:

- يمكن أن يكون إخراج البيرسبترون رقمًا ثنائيًا (0 أو 1) بسبب دالة hard limit transfer.
- يمكن استخدام البيرسبترون فقط لتصنيف مجموعات متجهات الإدخال القابلة للفصل خطيًا. إذا كانت متجهات الإدخال غير خطية، فليس من السهل تصنيفها بشكل صحيح.

س209/ قارن بين البيرسيبترون أحادي الطبقة single layer perceptron والبيرسيبترون متعدد الطبقات multi-layer Perceptron؟

ج/

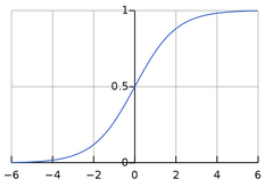
البيرسيبترون أحادي الطبقة	البيرسيبترون متعدد الطبقات
لا يمكن تصنيف نقاط البيانات غير الخطية	يمكن تصنيف البيانات غير الخطية
يأخذ كمية محدودة من المعلمات	يأخذ الكثير من المعلمات
كفاءة أقل مع البيانات الكبيرة	كفاءة عالية مع مجموعات البيانات الكبيرة

س210/ ما هو الفرق بين Perceptron والانحدار اللوجستي Logistic Regression؟

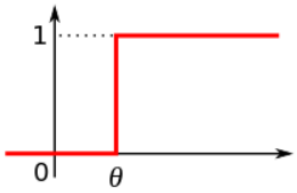
ج/ تعد البيرسيبترون متعدد الطبقات (MLP) واحدة من أكثر الشبكات العصبية الأساسية التي نستخدمها في التصنيف. بالنسبة لمشكلة التصنيف الثنائي، نعلم أن الناتج يمكن أن يكون إما 0 أو 1. وهذا يشبه تمامًا الانحدار اللوجستي البسيط، حيث

نستخدم دالة لوجيت logit function لتوليد احتمال بين 0 و 1 ، فما الفرق بين الاثنين؟

ببساطة، هذا هو مجرد الاختلاف في دالة العتبة! threshold function عندما نقيد نموذج الانحدار اللوجستي ليعطينا إما 1 بالضبط أو 0 بالضبط، نحصل على نموذج Perceptron:



Logistic regression - Logit function



Perceptron step function

س211/ ما هو البيرسيبترون متعدد الطبقات Multilayer perceptron (MLP)؟

ج / كما هو الحال في الشبكات العصبية، تحتوي MLPs على طبقة إدخال وطبقة مخفية وطبقة إخراج. لها نفس البنية مثل طبقة واحدة من البيرسيبترون مع طبقة مخفية واحدة أو أكثر. يمكن لمُحسِّن الطبقة الواحدة تصنيف الفئات الخطية القابلة للفصل فقط بإخراج ثنائي (0،1)، ولكن يمكن لـ MLP تصنيف الفئات غير الخطية.

س212/ ماهي CNN؟

ج/ CNNs هي شبكات عصبية تلافيفية convolutional neural networks تُستخدم لإجراء تحليل على الصور والمرئيات. يمكن لهذه الفئات من الشبكات العصبية إدخال صورة متعددة القنوات والعمل عليها بسهولة.

س213/ ما هي الطبقات المختلفة الموجودة في CNN؟

ج/فيما يلي أربع طبقات رئيسية تشكل شبكة عصبية تلافيفية CNN:

- الالتفاف (Convolution): هذه طبقات تتكون من مدخلات تسمى المرشحات (الفلاتر) التي تُستخدم كمعلمات لتدريب الشبكة.
- ReLu: يتم استخدامه كدالة تنشيط ويتم استخدامه دائمًا مع طبقة الالتفاف.

- التجميع (Pooling): التجميع هو مفهوم تقليص (اختزال) كيانات البيانات المعقدة التي تتشكل بعد الالتفاف ويستخدم بشكل أساسي للحفاظ على حجم الصورة بعد الانكماش.
- الاتصال (Connectedness): يستخدم هذا للتأكد من أن جميع الطبقات في الشبكة العصبية متصلة بالكامل ويمكن حساب التنشيط باستخدام التحيز بسهولة.

س214/ ما معنى valid padding وsame padding في CNN؟

ج/

- ❖ Valid padding: يتم استخدامها عندما لا تكون هناك حاجة للحشو padding. سيكون لمصفوفة الإخراج الأبعاد (n – f + 1) X (n – f + 1) بعد الالتفاف.
- ❖ Same padding: هنا ، تتم إضافة عناصر الحشو في جميع أنحاء مصفوفة الإخراج. سيكون لها نفس أبعاد مصفوفة الإدخال.

س215/ قم بتسمية بعض بُنى CNNs التي تعرفها؟

ج/

- ❖ ResNet
- ❖ InceptionV3
- ❖ MobileNet

س216/ لماذا يُفضل CNN على ANN لمهام تصنيف الصور Image Classification على الرغم من أنه من الممكن حل تصنيف الصور باستخدام ANN؟

ج/ تتمثل إحدى المشكلات الشائعة في استخدام ANN لتصنيف الصور في أن تفاعل ANN يتفاعل بشكل مختلف مع إدخال الصور وإصداراتها التي تم تغييرها. دعنا نفكر في مثال بسيط حيث توجد صورة لكلب في الجزء العلوي الأيسر من الصورة وفي صورة أخرى، توجد صورة لكلب في أسفل اليمين. ستفترض ANN أن الكلب سيظهر دائمًا في هذا القسم من أي صورة، ولكن هذا ليس هو الحال. تتطلب ANN نقاط بيانات محددة مما يعني أنه إذا كنت تقوم ببناء نموذج تعلم عميق للتمييز بين القطط والكلاب، فيجب توفير طول الأذنين وعرض الأنف وغيرها من الميزات كنقاط بيانات أثناء استخدام CNN لتصنيف الصور المكاني يتم استخراج الميزات من الصور المدخلة. عندما يكون هناك الآلاف من الميزات التي سيتم استخراجها،

فإن CNN هي الخيار الأفضل لأنها تجمع الميزات من تلقاء نفسها، على عكس ANN حيث يجب قياس كل ميزة فردية.

يصبح تدريب نموذج الشبكة العصبية ثقيلًا من الناحية الحسابية (يتطلب تخزينًا إضافيًا وقدرة معالجة) مع زيادة عدد الطبقات والمعلمات. يمكن أن يكون ضبط العدد المتزايد من المعلمات مهمة شاقة مع ANN، على عكس CNN حيث يتم تقليل وقت ضبط المعلمات مما يجعلها خيارًا مثاليًا لمشاكل تصنيف الصور.

س217/ كيفية إصلاح Constant validation accuracy في تدريب نموذج CNN؟

ج/ تعد دقة التحقق المستمرة Constant validation accuracy مشكلة شائعة عند تدريب أي شبكة عصبية لأن الشبكة تتذكر العينة فقط وتؤدي إلى مشكلة الضبط الزائد overfitting. يعني overfitting أن نموذج الشبكة العصبية يعمل بشكل رائع في عينة التدريب ولكن أداء النموذج يغرق في مجموعة التحقق من الصحة. فيما يلي بعض النصائح لمحاولة إصلاح دقة التحقق المستمرة في CNN: -

- يُنصح دائمًا بتقسيم مجموعة البيانات إلى مجموعة تدريب training وتحقق validation واختبار test.
- عند العمل مع القليل من البيانات، يمكن حل هذه المشكلة عن طريق تغيير معلمات الشبكة العصبية عن طريق التجربة والخطأ trial and error.
- زيادة حجم مجموعة بيانات التدريب.
- استخدم التسوية بالدفعات batch normalization.
- التنظيم Regularization.
- تقليل تعقيد الشبكة Reduce the network complexity.

س218/ لماذا نستخدم التلافيف convolutions للصور بدلاً من استخدام طبقات متصلة بالكامل fully connected layers؟

ج / تعمل كل نواة التفاف convolution kernel في CNN مثل كاشف الميزات feature detector الخاص بها ولها ترجمة مضمنة جزئيًا في التباين. يتيح استخدام التلافيف convolutions للفرد الاحتفاظ بالمعلومات المكانية spatial information من الصورة وتشفيرها والاستفادة منها، على عكس الطبقات المتصلة بالكامل fully connected layers التي لا تحتوي على أي معلومات مكانية نسبية.

س219/ ما فائدة استخدام التجميع الأقصى max-pooling في تصنيف الشبكات العصبية التلافيفية CNN؟

ج/ تصبح خرائط الميزات feature maps أصغر بعد التجميع الأقصى max-pooling في CNN وبالتالي تساعد في تقليل الحساب وأيضًا توفير مزيد من الترجمة في التباين. أيضًا، لا نفقد الكثير من المعلومات الدلالية لأننا نتخذ أقصى قدر من التنشيط.

س220/ ما هي بعض مزايا استخدام CNN (شبكة عصبية تلافيفية) بدلاً من DNN (شبكة عصبية كثيفة) في مهمة تصنيف الصور؟

ج/

- عدد المعلمات في الشبكة العصبية التلافيفية أصغر بكثير من عدد الشبكة العصبية الكثيفة. ومن ثم، فإن شبكة CNN أقل احتمالاً أن تتفوق.
- تتيح لك CNN إلقاء نظرة على أوزان عامل التصفية وتصور ما تعلمته الشبكة. لذا، فهذا يعطي فهمًا أفضل للنموذج.
- تدرب CNN النماذج بطريقة هرمية، أي أنها تتعلم الأنماط من خلال شرح الأنماط المعقدة باستخدام أنماط أبسط.

س221/ كيف يمكنك تحديد حجم المرشح (الفلتر) عند إجراء عملية التفاف في شبكة CNN؟

ج/

- أثناء إجراء عملية الالتفاف convolution في CNN، تكتشف المرشحات filters الأنماط المكانية مثل الحواف في الصور عن طريق اكتشاف التغييرات في شدة قيم الصور.
- لا توجد إجابة محددة لعدد المرشحات أو أفضل عدد من المرشحات التي يمكن للمرء استخدامها.
- لتحديد حجم المرشح filter size، أود أن أقول إنه يعتمد بشدة على نوع وتعقيد بيانات الصورة.
- يتم تعلم عدد لا بأس به من الميزات من التجربة بعد العمل المتكرر مع أنواع مماثلة من مجموعات البيانات.
- بشكل عام، كلما زاد عدد الميزات التي تريد التقاطها في صورة ما، زاد عدد المرشحات المطلوبة في شبكة CNN. عدد المرشحات عبارة عن معلمة فائقة hyperparameter يمكن ضبطها لاحقًا.

س222/ عند تصميم CNN، هل يمكننا معرفة عدد الطبقات التلافيفية التي يجب أن نستخدمها؟

ج/

- أثناء تصميم شبكة CNN، فإن الطبقات التلافيفية Convolutional layers هي الطبقات التي يتم فيها تطبيق المرشحات filters على الصورة الأصلية، أو على خرائط المعالم feature maps الأخرى في شبكة CNN العميقة.
- كلما زادت الطبقات التلافيفية، كان ذلك أفضل لأن كل طبقة تلافيفية تقلل من عدد ميزات الإدخال إلى الطبقات المتصلة بالكامل fully connected layers، على الرغم من أنه بعد طبقتين أو ثلاث طبقات يصبح كسب الدقة صغيرًا نوعًا ما، لذا عليك أن تقرر ما إذا كان تركيزك الرئيسي هو دقة التعميم generalisation accuracy أو وقت التدريب training time.
- تختلف جميع مهام التعرف على الصور، لذا فإن أفضل طريقة هي محاولة زيادة عدد الطبقات التلافيفية واحدة تلو الأخرى حتى تشعر بالرضا عن النتيجة.

س223/ هل من الجيد استخدام CNN لتصنيف إشارة أحادية البعد؟

ج/ نعم، يمكن استخدام CNN لتصنيف إشارة أحادية البعد.

س224/ ما هي شبكة Deconvolutional؟

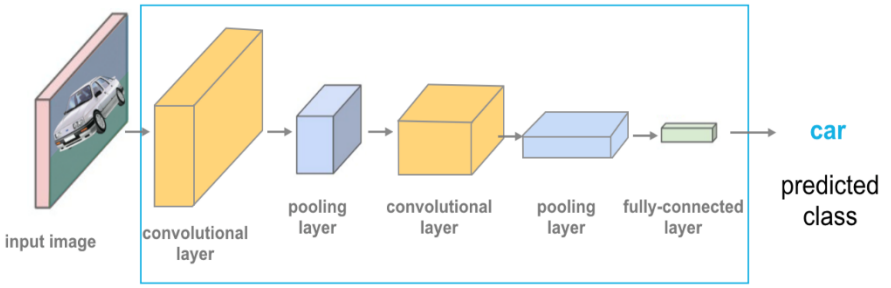
ج / شبكة Deconvolutional هي شبكات عصبية تلافيفية (CNN) تعمل في عملية معكوسة. شبكات Deconvolutional، المعروفة أيضًا باسم deconvolutional neural networks، متشابهة جدًا في طبيعتها مع شبكات CNN التي تعمل في الاتجاه المعاكس ولكنها تطبيق متميز للذكاء الاصطناعي (AI).

تسعى شبكات Deconvolutional إلى العثور على الميزات أو الإشارات المفقودة التي ربما لم يتم اعتبارها في السابق مهمة لمهمة الشبكة العصبية التلافيفية CNN. قد يتم فقد الإشارة بسبب التفافها مع إشارات أخرى. يمكن استخدام deconvolution للإشارات في كل من تخليق الصور وتحليلها image synthesis and analysis.

س225/ صف بنية الشبكة العصبية التلافيفية النموذجية (CNN)؟

ج / في بُنية CNN النموذجية، ترتبط بضع طبقات تلافيفية بأسلوب متسلسل cascade style.

- ❖ تتبع كل طبقة تلافيفية convolutional layer طبقة وحدة خطية مصححة (ReLU)، ثم طبقة تجميع pooling layer ، ثم طبقة تلافيفية واحدة أو أكثر (+ ReLU) ، ثم طبقة تجميع أخرى.
- ❖ الناتج من كل طبقة التفاف هو مجموعة من الكائنات تسمى خرائط المعالم feature maps، يتم إنشاؤها بواسطة مرشح kernel واحد.
- ❖ تُستخدم خرائط المعالم لتحديد مدخلات جديدة للطبقة التالية.
- ❖ في النهاية، هناك طبقة واحدة أو أكثر متصلة بالكامل fully connected layers.
- ❖ اعتمادًا إلى حد كبير على نوع المشكلة، قد تكون الشبكة عميقة.



س226/ ما هي خوارزمية التعلم العميق الأفضل لاكتشاف الوجه face detection؟

ج / هناك العديد من خوارزميات التعلم الآلي المتاحة لاكتشاف الوجه ولكن أفضلها هي تلك التي تتضمن شبكات CNN والتعلم العميق. بعض الخوارزميات البارزة لاكتشاف الوجه مذكورة أدناه FaceNet Probablisit Face Embedding ArcFace Cosface Spherface

س227/ لماذا تعمل الشبكة العصبية التلافيفية (CNN) بشكل أفضل مع بيانات الصورة؟

ج/ تعد الشبكة العصبية التلافيفية CNN أفضل من شبكة التغذية الأمامية FNN نظرًا لأن CNN تتميز بمشاركة المعلمات parameter sharing وتقليل الأبعاد dimensionality reduction. بسبب مشاركة المعلمات في CNN، يتم تقليل عدد المعلمات وبالتالي تنخفض الحسابات أيضًا. الحدس الرئيسي هو أن التعلم من جزء واحد من الصورة مفيد أيضًا في جزء آخر من الصورة. بسبب تقليل الأبعاد في CNN، يتم تقليل القوة الحسابية اللازمة.

س228/ في CNN، إذا كان حجم الإدخال 5 × 5 وحجم الفلتر هو 7 × 7، فما هو حجم الإخراج؟

ج/ هذه إجابة بديهية جدا. كما رأينا أعلاه، نقوم بالالتفاف على "x" خطوة واحدة في كل مرة، إلى اليمين ، وفي النهاية ، حصلنا على Z بأبعاد 2 × 2 ، من أجل X بأبعاد 3 × 3.

وبالتالي، لجعل حجم الإدخال مشابهًا لحجم المرشح، فإننا نستخدم الحشو padding: إضافة 0 ثانية إلى مصفوفة الإدخال بحيث يصبح حجمها الجديد 7 × 7. وبالتالي، فإن حجم الإخراج سيستخدم الصيغة:

Dimension of image = (n, n) = 5 X 5

Dimension of filter = (f,f) = 7 X 7

Padding = 1 (adding 1 pixel with value 0 all around the edges)

Dimension of output will be (n+2p-f+1) X (n+2p-f+1) = 1 X 1

س229/ ما هي الطبقات المختلفة للشبكة العصبية التلافيفية CNN؟

ج/ هناك أربع طبقات في CNN:

- طبقة تلافيفية Convolutional Layer: الطبقة التي تنفذ عملية تلافيفية، مما يؤدي إلى إنشاء عدة نوافذ صور أصغر لتجاوز البيانات.
- طبقة ReLU: إنها تجلب اللاخطية للشبكة وتحول جميع البيكسلات السلبية إلى الصفر. الإخراج هو خريطة معالم مصححة.
- طبقة التجميع Pooling Layer: التجميع هو عملية أخذ عينات لأسفل تقلل من أبعاد خريطة المعالم.

- ❖ طبقة متصلة بالكامل Fully Connected Layer: تتعرف هذه الطبقة على الكائنات الموجودة في الصورة وتصنفها.

س230/ كيف يمكنك تحويل طبقة كثيفة Dense Layer لشبكة CNN إلى طبقة تلافيفية كاملة Fully Convolutional Layer (FCN)؟

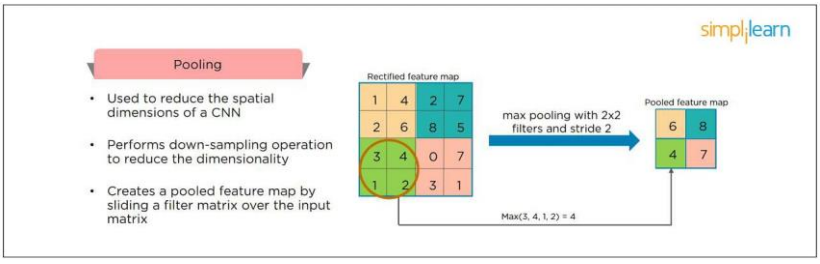
ج / إذا كان لديك CNN مع بعض الطبقات الكثيفة في الأعلى، فيمكنك تحويل هذه الطبقات الكثيفة إلى طبقات تلافيفية لإنشاء FCN بالطريقة التالية:

استبدل الطبقة الأقل كثافة بطبقة تلافيفية بحجم نواة يساوي حجم إدخال الطبقة، مع مرشح واحد لكل خلية عصبية في الطبقة الكثيفة ، واستخدم **valid padding**.

- ❖ بشكل عام ، يجب أن تكون الخطوة **1 stride** ، ولكن يمكنك ضبطها على قيمة أعلى إذا كنت تريد ذلك.
- ❖ يجب أن تكون دالة التنشيط مماثلة للطبقة الكثيفة.
- ❖ يجب تحويل الطبقات الكثيفة الأخرى بنفس الطريقة ، ولكن باستخدام مرشحات **1 × 1**.
- ❖ من الممكن في الواقع تحويل شبكة **CNN** المدربة بهذه الطريقة عن طريق إعادة تشكيل مصفوفات أوزان الطبقات الكثيفة بشكل مناسب.

س231/ ما هو التجميع Pooling على CNN وكيف يعمل؟

ج / يستخدم التجميع **Pooling** لتقليل الأبعاد المكانية لشبكة **CNN**. يقوم بإجراء عمليات اختزال للعينات لتقليل الأبعاد وإنشاء خريطة معالم مجمعة عن طريق تحريك مصفوفة مرشح فوق مصفوفة الإدخال.



س232/ ما فائدة استخدام نوى صغيرة small kernels مثل 3x3 من استخدام عدد قليل من النوى الكبيرة large kernels؟

ج / تتيح لك النوى الأصغر استخدام المزيد من المرشحات حتى تتمكن من استخدام عدد أكبر من دوال التنشيط والسماح لـ **CNN** بالتعرف على دالة تعيين أكثر تميزًا.

أيضًا، تلتقط النوى الأصغر سياقًا مكانيًا **spatial context** أكثر وتستخدم عددًا أقل من العمليات الحسابية والمعلمات مما يجعلها خيارًا أفضل على تلك الكبيرة.

س233/ كيف تختار عمق الشبكة العصبية؟

ج / يعد تحديد عدد الخلايا العصبية في الطبقات المخفية (عمق الشبكة العصبية) جزءًا مهمًا جدًا من تحديد البنية الشاملة لشبكتك العصبية. على الرغم من أن هذه الطبقات لا تتفاعل بشكل مباشر مع البيئة الخارجية، إلا أن لها تأثيرًا هائلاً على الناتج النهائي. يجب مراعاة كل من عدد الطبقات المخفية وعدد الخلايا العصبية في كل من هذه الطبقات المخفية بعناية.

سيؤدي استخدام عدد قليل جدًا من الخلايا العصبية في الطبقات المخفية إلى شيء يسمى الضبط الناقص **underfitting**. يحدث الضبط الناقص عندما يكون هناك عدد قليل جدًا من الخلايا العصبية في الطبقات المخفية لاكتشاف الإشارات بشكل مناسب في مجموعة بيانات معقدة.

يمكن أن يؤدي استخدام عدد كبير جدًا من الخلايا العصبية في الطبقات المخفية إلى العديد من المشكلات. أولاً، قد يؤدي وجود عدد كبير جدًا من الخلايا العصبية في الطبقات المخفية إلى الضبط الزائد **overfitting** يحدث الضبط الزائد عندما تمتلك الشبكة العصبية قدرًا كبيرًا من القدرة على معالجة المعلومات بحيث لا تكون الكمية المحدودة من المعلومات الموجودة في مجموعة التدريب كافية لتدريب جميع الخلايا العصبية في الطبقات المخفية. يمكن أن تحدث مشكلة ثانية حتى عندما تكون بيانات التدريب كافية. يمكن أن يزيد عدد الخلايا العصبية في الطبقات المخفية بشكل مفرط من الوقت الذي يستغرقه تدريب الشبكة. يمكن أن يزداد مقدار وقت التدريب إلى الحد الذي يستحيل معه تدريب الشبكة العصبية بشكل مناسب. من الواضح أنه يجب التوصل إلى حل وسط بين عدد كبير جدًا من الخلايا العصبية في الطبقات المخفية وعدد قليل جدًا منها.

هناك العديد من الاساليب لتحديد العدد الصحيح للخلايا العصبية لاستخدامها في الطبقات المخفية، مثل ما يلي:

- يجب أن يكون عدد الخلايا العصبية المخفية بين حجم طبقة الإدخال وحجم طبقة الإخراج.
- يجب أن يكون عدد الخلايا العصبية المخفية 3/2 حجم طبقة الإدخال، بالإضافة إلى حجم طبقة الإخراج.
- يجب أن يكون عدد الخلايا العصبية المخفية أقل من ضعف حجم طبقة الإدخال.

توفر لك هذه القواعد الثلاث نقطة بداية يجب عليك مراعاتها. في النهاية، فإن اختيار بنية لشبكتك العصبية سيعتمد على التجربة والخطأ.

س234/ ماذا تفهم بمصطلحات الدُفعات Batches والتكرارات Iterations والفترات Epochs في تدريب نموذج الشبكة العصبية؟

ج/

- تشير الفترة Epoch إلى التكرار حيث يتم تمرير مجموعة البيانات الكاملة للأمام وللخلف عبر الشبكة العصبية مرة واحدة فقط.
- لا يمكن تمرير مجموعة البيانات الكاملة إلى الشبكة مرة واحدة، لذلك يتم تقسيم مجموعة البيانات إلى أجزاء. يشار إلى هذا باسم الدُفعة Batch.
- يُشار إلى إجمالي عدد الدُفعات number of batches اللازمة لإكمال فترة واحدة باسم التكرار iteration. على سبيل المثال، إذا كان لديك 60000 صف بيانات وكان حجم الدُفعة 1000، فسيتم تشغيل 60 تكرارًا لكل فترة.

س235/ ماذا تقصد بإعطاء "epoch = 1" في الشبكة العصبية؟

ج/ وهذا يعني أن عبور مجموعة البيانات مرة واحدة " traversing the data set one time " في الشبكة العصبية.

س236/ لنفترض أن عليك بناء بُنية شبكة عصبية؛ كيف ستقرر عدد الخلايا العصبية neurons والطبقات المخفية hidden layers اللازمة للشبكة؟

ج/ بالنظر إلى مشكلة العمل، لا توجد قاعدة صارمة وسريعة لتحديد العدد الدقيق للخلايا العصبية والطبقات المخفية المطلوبة لبناء بنية شبكة عصبية. يكمن الحجم الأمثل للطبقة المخفية في الشبكة العصبية بين حجم طبقات الإخراج وحجم المدخلات. ومع ذلك، إليك بعض الأساليب الشائعة التي تتمتع بميزة تحقيق بداية رائعة لبناء بنية شبكة عصبية: -

- ❖ لمعالجة أي مشكلة نمذجة تنبؤية محددة في العالم الحقيقي، فإن أفضل طريقة هي البدء بتجربة منهجية تقريبية ومعرفة ما هو الأفضل لأي مجموعة بيانات معينة استنادًا إلى الخبرة السابقة في العمل مع الشبكات العصبية على مشاكل مماثلة في العالم الحقيقي. بناءً على فهم أي مجال مشكلة معين وخبرة الفرد في العمل مع الشبكات العصبية، يمكن للمرء اختيار تكوين الشبكة. دائمًا ما يكون عدد الطبقات والخلايا العصبية المستخدمة في مشكلات مماثلة طريقة رائعة لبدء اختبار تكوين الشبكة العصبية.
- ❖ يُنصح دائمًا، في البداية، ببُنية شبكة عصبية بسيطة ثم المضي قدمًا لتعزيز تعقيد الشبكة العصبية.

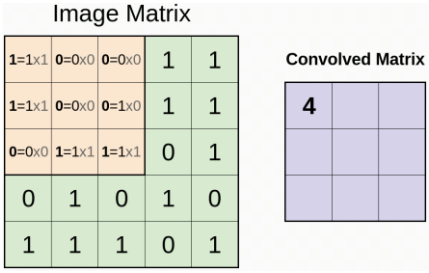
- حاول العمل مع أعماق مختلفة من الشبكات وقم بتكوين شبكات عصبية عميقة فقط لتحدي مشاكل النمذجة التنبؤية حيث يمكن أن يكون العمق مفيدًا.

س237/ ماذا تقصد بالفلترة filtering والخطوة stride والحشو padding في الشبكة العصبية التلافيفية CNN؟

ج/

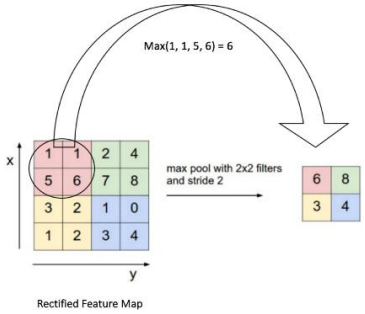
الفلتر Filter

يتم تمثيل الفلتر (المرشح) بواسطة متجه للأوزان نقوم بلف المدخلات. تعمل كل طبقة شبكة كمرشح لوجود ميزات أو أنماط محددة موجودة في الصورة الأصلية. لاكتشاف مرشح، من غير المناسب مكان وجود هذه الميزة أو النمط المحدد في الصورة الأصلية. تم تصميم هذه المرشحات خصيصًا لاكتشاف ما إذا كانت الصورة تحتوي على أي من هذه الخصائص أم لا. يتم إزاحة هذا المرشح عدة مرات ويتم تطبيقه في مواضع مختلفة للصورة حتى تتم تغطية الصورة بأكملها بالتفصيل.



الخطوات Strides

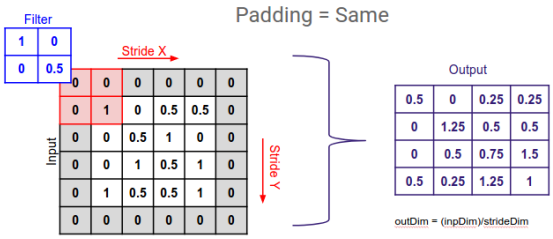
الخطوة **Stride** هو عدد وحدات البكسل التي تم إزاحتها عبر مصفوفة الإدخال بالكامل. عندما تكون الخطوة **1** ، فإننا ننقل المرشحات إلى **1** بكسل في المرة الواحدة. عندما تكون الخطوة **2**، فإننا ننقل المرشحات إلى **2** بكسل في المرة الواحدة وهكذا.



الحشو Padding

إنها طبقة إضافية يمكننا إضافتها إلى حدود الصورة للحصول على معلومات أكثر دقة عن الصورة. أحيانًا لا يتناسب المرشح تمامًا مع الصورة المدخلة. ثم لدينا خياران:

قم بتثبيت الصورة بالأصفار وهو ما يسمى أيضًا بالحشو الصفري **zero padding** بحيث تلائم الصورة.



قم بإسقاط جزء الصورة حيث لا يتناسب المرشح مع الصورة أو الصورة. يسمى هذا بالحشو الصحيح **valid padding** الذي يحتفظ فقط بجزء صالح من الصورة ويزيل الآخر.

س238/ ما هي المعلمات الفائقة hyperparameters المختلفة المستخدمة في الشبكات العصبية التلافيفية CNN أثناء تدريب النموذج؟

ج/

يعد ضبط المعلمات الفائقة للشبكة العصبية العميقة أمرًا صعبًا لأنه بطيء في تدريب شبكة عصبية عميقة وهناك العديد من المعلمات لتكوينها.

- ❖ معدل التعلم **Learning rate**
  - يتحكم معدل التعلم في تحديث الوزن في خوارزمية التحسين. من خلال تحديد معدل التعلم، وخفض معدل التعلم تدريجيًا، أو الأساليب القائمة على الزخم **momentum** أو معدلات التعلم التكيفي **adaptive learning rates**، اعتمادًا على اختيارنا للمحسن مثل **SGD** أو **Adam** أو **Adagrad** أو **AdaDelta** أو **RMSProp**.
- ❖ عدد الفترات **Number of epochs**
  - عدد الفترات هو عدد المرات التي يمكن أن تمر فيها مجموعة التدريب بأكملها عبر الشبكة العصبية. قم بزيادة عدد الفترات حتى نرى فجوة صغيرة بين خطأ الاختبار وخطأ التدريب.

- ❖ حجم الدفعة **Batch size**
  - يفضل استخدام الدفعة الصغيرة بشكل عام في عملية التعلم الخاصة بـ **convnet**. يعد النطاق من **16** إلى **128** خيارًا جيدًا للاختبار. **Convnet** حساسة لحجم الدفعة.
- ❖ دالة التنشيط **Activation function**
  - تقدم دالة التنشيط اللاخطية للنموذج. المعدل يعمل بشكل جيد مع **convnet**. بدائل إجراء ضبط **hyperparameter** هي **sigmoid** **tanh** ودوال التنشيط الأخرى حسب المهمة.
- ❖ عدد الطبقات والوحدات المخفية **Number of hidden layers and units**
  - من الجيد إضافة المزيد من الطبقات حتى يتوقف خطأ الاختبار عن التحسن. المفاضلة مكلفة حسابيًا لتدريب الشبكة. قد يؤدي وجود عدد صغير من الوحدات إلى **overfitting** في حين أن وجود المزيد من الوحدات عادة لا يضر بالتنظيم المناسب.
- ❖ تهيئة الوزن **Weight initialization**
  - تهيئة الأوزان بأرقام عشوائية صغيرة لمنع الخلايا العصبية الميتة **dead neurons**، ولكن ليست صغيرة جدًا لتجنب الانحدار الصفري **zero gradient**. بشكل عام، يعمل التوزيع الموحد بشكل جيد.
- ❖ الحذف العشوائي للتسوية **Dropout for regularization**
  - الحذف العشوائي **Dropout** هو أسلوب تنظيم أفضل لتجنب الضبط الزائد **overfitting** في الشبكات العصبية العميقة. هذه الطريقة تحذف ببساطة الوحدات في الشبكة العصبية وفقًا للاحتمال المطلوب. تعتبر القيمة الافتراضية **0.5** اختيارًا جيدًا للاختبار باستخدامها.
- ❖ بحث الشبكة أو البحث العشوائي **Grid search or randomized search**
  - يعد الضبط اليدوي للمعامل الفائق متعباً وغير معقول أيضًا. **Grid search** تبحث بدقة في جميع مجموعات المعلمات عن القيم المعطاة. **randomized search** تأخذ عينة لعدد معين من المرشحين من مساحة المعلمة مع توزيع محدد.

س239/ ماذا تقصد بنموذج MobileNet؟

ج/ **MobileNet** هو نوع من الشبكات العصبية التلافيفية **CNN** المصممة لتطبيقات الرؤية المتنقلة والمدمجة. وهي تستند إلى بنية مبسطة تستخدم تلافيفات قابلة للفصل في العمق لبناء شبكات عصبية عميقة خفيفة الوزن يمكن أن يكون لها زمن انتقال منخفض للأجهزة المحمولة والأجهزة المدمجة.

س240/ كيف تقسم مجموعة البيانات الخاصة بك إلى بيانات اختبار وتدريب؟

ج/ من الناحية المثالية، يجب ألا تختبر بيانات التدريب. قد يسبب نموذجك الضبط الزائد overfitting لمجموعة التدريب وبالتالي سيفشل في البيانات الجديدة. لا يمكن أن تضمن الدقة الجيدة في مجموعة بيانات التدريب نجاح نموذجك على البيانات غير المرئية. هذا هو السبب في أنه يوصى بالحفاظ على بيانات التدريب منفصلة عن بيانات الاختبار. الفكرة الأساسية هي استخدام مجموعة الاختبار كبيانات غير مرئية.

بعد تدريب بياناتك على مجموعة التدريب، يجب عليك اختبار النموذج الخاص بك على مجموعة الاختبار. إذا كان نموذجك يعمل جيدًا في مجموعة الاختبار، فيمكنك أن تكون أكثر ثقة بشأن نموذجك.

اما عن كيفية تقسيم مجموعة البيانات الى بيانات تدريب واختبار فهي كالاتي:

نسبة الانقسام الأكثر شيوعًا هي 20:80. هذا هو 80٪ من مجموعة البيانات تذهب إلى مجموعة التدريب و20 ٪ من مجموعة البيانات تذهب إلى مجموعة الاختبار. قبل تقسيم البيانات، تأكد من أن مجموعة البيانات كبيرة بما يكفي. يعمل قسم التدريب/الاختبار بشكل جيد مع مجموعات البيانات الكبيرة.

س241/ ما هي RNN في التعلم العميق؟

ج/ RNNs هي اختصار للشبكات العصبية المتكررة recurrent neural networks، والتي تشكل نوعًا شائعًا من الشبكات العصبية الاصطناعية. يتم استخدامها لمعالجة تسلسل البيانات والنصوص والجينوم والكتابة اليدوية والمزيد. تستفيد RNNs من الانتشار الخلفي لمتطلبات التدريب.

س242/ ما هي الخطوات المتبعة في عمل شبكة LSTM؟

ج/ هناك ثلاث خطوات رئيسية متضمنة في عمل شبكة LSTM:

- تلتقط الشبكة المعلومات التي يجب أن تتذكرها وتحدد ما يجب نسيانه.
- يتم تحديث قيم حالة الخلية بناءً على الخطوة 1.
- تحسب الشبكة وتحلل أي جزء من الحالة الحالية يجب أن يصل إلى المخرجات.

س243/ ما الفرق بين الشبكات العصبية التلافيفية (CNN) والشبكات العصبية المتكررة (RNN) وفي أي الحالات تستخدم كل واحدة منها؟

ج/ تطبق CNN التفافًا على البيانات قبل استخدامها في طبقات متصلة بالكامل.

- يتم استخدامها بشكل أفضل في الحالات التي تريد فيها الثبات الموضعي، أي أنك تريد التقاط الميزات بغض النظر عن مكانها في عينة الإدخال.
- فكر في صورة بها كل أنواع الحيوانات. إذا قمت بتطبيق شبكة عصبية تلافيفية لتصنيف ما إذا كان هناك قطة في الصورة، فإنها ستحدد القطة بغض النظر عن موضع القطة في الصورة (في الأعلى أو الأسفل أو اليسار أو اليمين). هذا مفيد جدا لتصنيف الصور.

RNN هي شبكات عصبية تحافظ على الحالة بين عينات الإدخال. يتذكرون عينات الإدخال السابقة ويستخدمونها للمساعدة في تصنيف عينة الإدخال الحالية.

- تكون مفيدة للغاية عندما يكون ترتيب بياناتك مهمًا. لذلك على سبيل المثال في الكلام (تساعد الكلمات السابقة في تحديد الكلمة الحالية) والفيديو (يتم ترتيب الإطارات) وكذلك معالجة النصوص.
- بشكل عام، تعتبر المشكلات المتعلقة ببيانات السلاسل الزمنية (البيانات ذات الطابع الزمني عليها) مرشحة جيدة لحلها بشكل جيد مع الشبكات العصبية المتكررة.

س244/ ما هو تلاشي الانحدار vanishing gradient عند استخدام RNNs؟

ج/ تلاشي الانحدار vanishing gradient هو سيناريو يحدث عندما نستخدم RNNs. نظرًا لأن RNNs تستخدم الانتشار الخلفي backpropagation، فإن التدرجات في كل خطوة على الطريق سوف تميل إلى أن تصبح أصغر عندما تمر الشبكة عبر التكرارات العكسية. هذا يعادل تعلم النموذج ببطء شديد، مما يتسبب في مشاكل الكفاءة في الشبكة.

س245/ ما الهدف من استخدام LSTM؟

ج/ LSTM تعني ذاكرة طويلة قصيرة المدى long short-term memory. إنه نوع من RNN يستخدم لتسلسل سلسلة من البيانات. وهو يتألف من سلاسل التغذية الراجعة التي تمنحه القدرة على الأداء مثل كيان حسابي للأغراض العامة.

س246/ كيف تكون بنية المحولات transformers أفضل من RNNs في التعلم العميق؟

ج/ باستخدام المعالجة المتسلسلة، واجه المبرمجون:

- استخدام قوة معالجة عالية.
- صعوبة التنفيذ الموازي.

تسبب هذا في ظهور بنية المحولات transformers. هنا، توجد آلية تسمى آلية الانتباه attention mechanism، والتي تُستخدم لمطابقة جميع التبعيات dependencies بين الجمل، وبالتالي إحراز تقدم كبير في حالة نماذج البرمجة اللغوية العصبية NLP.

س247/ كيف سيتم تنفيذ التسوية بالدفعات Batch Normalization في RNN؟

ج/ لا يمكن استخدام التسوية بالدُفعات Batch Normalization في RNN لأنه يتم حساب الإحصائيات لكل دفعة، وبالتالي لن يأخذ التسوية بالدُفعات في الاعتبار الجزء المتكرر من الشبكة العصبية. يمكن أن يكون البديل لهذا هو تسوية الطبقة layer normalization في RNN أو إعادة معاملات طبقة LSTM التي تسمح باستخدام التسوية بالدُفعات.

س248/ قارن بين الشبكة العصبية ذات التغذية الأمامية FNN والشبكة العصبية المتكررة RNN؟

ج/ في RNN، سيتم تغذية إخراج الحالة السابقة كمدخلات الحالة التالية (الخطوة الزمنية time step). ليس هذا هو الحال مع شبكة التغذية الأمامية FNN التي تتعامل مع إدخال بطول ثابت ومخرج بطول ثابت.

لا توجد تغذية راجعة (حلقات)؛ أي أن ناتج أي طبقة لا يؤثر على نفس الطبقة. تميل شبكات التغذية الأمامية FNN إلى أن تكون شبكات مباشرة تربط المدخلات بالمخرجات.

س249/ ما الفرق بين Recurrent Neural Networks و Recursive Neural Networks؟

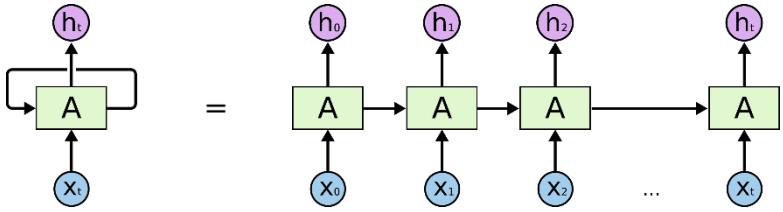
ج/

- ❖ الشبكات العصبية المتكررة Recurrent Neural Networks:

يتم استخدامه للمدخلات المتسلسلة حيث يكون عامل الوقت هو عامل التمييز الرئيسي بين عناصر التسلسل، ولهذا السبب يتم استخدامه بشكل شائع في السلاسل الزمنية.

عندما نفتح الشبكة، في كل خطوة زمنية، فإنها تقبل إدخال المستخدم في تلك الخطوة الزمنية وإخراج الطبقة المخفية التي تم حسابها في الخطوة الزمنية السابقة.

يتم تقاسم الأوزان (وتبقى الأبعاد ثابتة) على طول طول التسلسل length of the sequence.

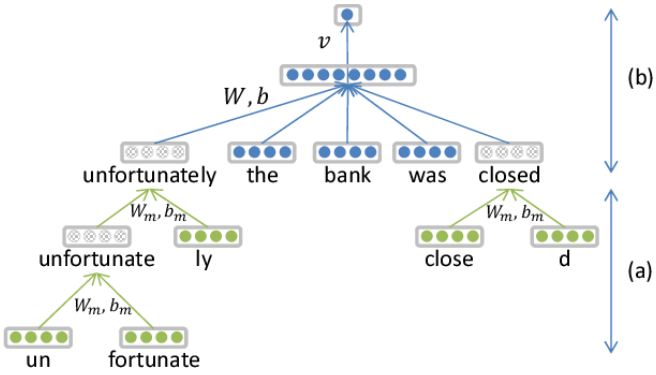


❖ الشبكة العصبية العودية Recursive neural network:

يشبه إلى حد كبير شبكة هرمية حيث لا يوجد حقًا جانب زمني لتسلسل الإدخال ولكن يجب معالجة الإدخال بشكل هرمي بطريقة شجرة.

يستخدم بشكل كبير في البرمجة اللغوية العصبية NLP، حيث تكون طريقة تعلم شجرة تحليل جملة ما تأخذ بشكل متكرر ناتج العملية التي يتم إجراؤها على جزء أصغر من النص.

هنا، يتم تقاسم الأوزان (وتبقى الأبعاد ثابتة) في كل عقدة.



س250/ كم عدد الأبعاد التي يجب أن تحتويها مدخلات طبقة RNN؟ ماذا يمثل كل بعد؟ ماذا عن نواتجها؟

ج/ يجب أن تحتوي طبقة RNN على مدخلات ثلاثية الأبعاد:

- البعد الأول هو أبعاد الدُفعة (حجمها هو حجم الدُفعة).
- البعد الثاني يمثل الوقت (حجمه هو عدد الخطوات الزمنية)،
- والبعد الثالث يحمل المدخلات في كل خطوة زمنية (حجمه هو عدد ميزات الإدخال لكل خطوة زمنية).

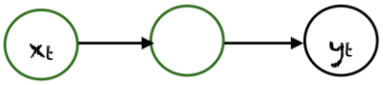
على سبيل المثال، إذا كنت تريد معالجة دفعة تحتوي على 5 سلاسل زمنية من 10 خطوات زمنية لكل منها، مع قيمتين لكل خطوة زمنية (على سبيل المثال، درجة الحرارة وسرعة الرياح)، فسيكون الشكل [5، 10، 2].

النواتج أيضًا ثلاثية الأبعاد، بنفس البعدين الأولين، لكن البعد الأخير يساوي عدد الخلايا العصبية. على سبيل المثال، إذا كانت طبقة RNN تحتوي على 32 خلية عصبية تعالج الدفعة التي ناقشناها للتو، فسيكون الناتج على شكل [5، 10، 32].

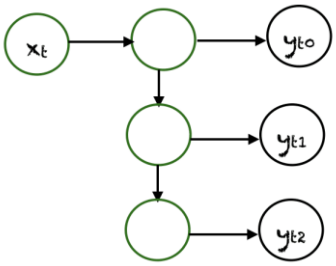
س251/ ما هي أنواع الشبكات العصبية المتكررة (RNN) التي تعرفها؟

ج / هناك أنواع مختلفة من الشبكات العصبية المتكررة RNN ببنيات مختلفة. بعض الأمثلة هي:

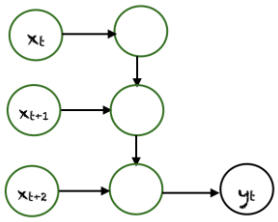
- واحد لواحد One to one: يوجد هنا زوج واحد (xt، yt). تستخدم الشبكات العصبية التقليدية بُنية واحد لواحد.



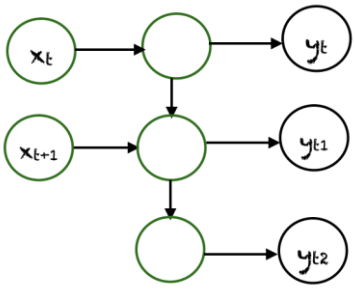
- واحد إلى متعدد One to many: في شبكة إلى عدة شبكات، يمكن أن ينتج عن الإدخال الفردي في xt مخرجات متعددة، على سبيل المثال، yt0، yt1، yt2 يعد إنشاء الموسيقى منطقة مثال، حيث يتم استخدام شبكة إلى عدة شبكات.



- متعدد لواحد Many to one: في هذه الحالة، ينتج عن العديد من المدخلات من خطوات زمنية مختلفة مخرجات واحدة. على سبيل المثال، يمكن أن تنتج xt و 1 + xt و 2 + xt ناتجًا واحدًا. يتم استخدام هذه الشبكات في تحليل المشاعر sentiment analysis أو اكتشاف المشاعر emotion detection، حيث يعتمد تصنيف الفئة على سلسلة من الكلمات.



❖ متعدد لمتعدد Many to many: هناك العديد من احتمالات متعدد لمتعدد. يتم عرض مثال أدناه، حيث ينتج عن مدخلين ثلاث مخرجات. يتم تطبيق العديد من الشبكات في الترجمة الآلية machine translation، على سبيل المثال، أنظمة الترجمة من الإنجليزية إلى الفرنسية أو العكس.



س252/ كيف يختلف الانتشار الخلفي للشبكة العصبية المتكررة RNN عن باقي الشبكات العصبية؟

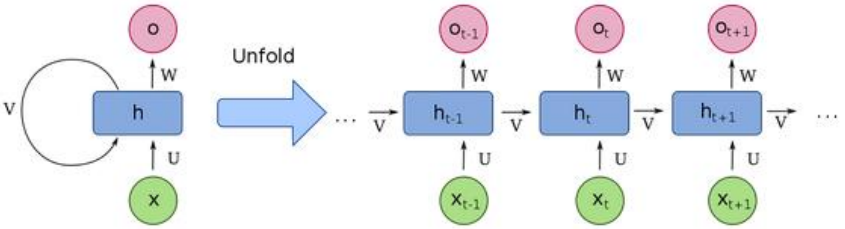
ج/ تعالج RNN بشكل أساسي المتسلسلات sequences خطوة واحدة في كل مرة، لذلك أثناء الانتشار الخلفي backpropagation، تتدفق التدرجات gradients للخلف عبر خطوات زمنية time steps. هذا يسمى الانتشار الخلفي عبر الزمن backpropagation through time.

س253/ لماذا تعمل RNNs بشكل أفضل مع البيانات المتسلسلة sequential data؟

ج / تُستخدم الشبكات العصبية المتكررة RNNs بشكل شائع عندما نتعامل مع البيانات المتسلسلة sequential data. والسبب هو أن النموذج يستخدم طبقات تعطي النموذج ذاكرة قصيرة المدى. باستخدام هذه الذاكرة، يمكنها التنبؤ بالبيانات التالية بشكل أكثر دقة.

س254/ لماذا تعمل RNNs بشكل أفضل مع البيانات النصية؟

ج / المكون الرئيسي الذي يميز الشبكات العصبية المتكررة (RNN) عن النماذج الأخرى هو إضافة حلقة في كل عقدة. تجلب هذه الحلقة آلية التكرار recurrence في RNNs. في الشبكة العصبية الاصطناعية الأساسية (ANN)، يتم إعطاء كل مدخلات نفس الوزن ويتم تغذيتها للشبكة في نفس الوقت. لذلك، بالنسبة إلى جملة مثل "رأيت الفيلم وكرهته"، سيكون من الصعب التقاط المعلومات التي تربط "هو" بـ "الفيلم".



تهدف إضافة حلقة إلى الاحتفاظ بمعلومات العقدة السابقة للعقدة التالية، وما إلى ذلك. هذا هو السبب في أن RNNs أفضل بكثير بالنسبة للبيانات المتسلسلة، وبما أن البيانات النصية هي أيضًا متسلسلة بطبيعتها، فهي تعد تحسينًا على شبكات ANN.

س255/ هل يمكنك إضافة تنظيم L2 إلى شبكة عصبية متكررة RNN للتغلب على مشكلة تلاشي التدرج؟

ج / يمكن أن يؤدي هذا في الواقع إلى تفاقم مشكلة تلاشي التدرج لأن تنظيم L2 ستقلص الأوزان نحو الصفر.

س256/ ما هي تطبيقات الشبكة العصبية المتكررة (RNN)؟

ج / يمكن استخدام RNN لتحليل المشاعر **sentiment analysis**، واستخراج النص **text mining**، وتعليق الصور **image captioning**. يمكن للشبكات العصبية المتكررة أيضًا معالجة مشاكل السلاسل الزمنية مثل التنبؤ بأسعار الأسهم في شهر أو ربع.

س257/ ما هي GRU وما اهميتها؟

ج / تعد **Gated Recurrent Unit (GRU)** جزءاً من نموذج محدد للشبكة العصبية المتكررة التي تهدف إلى استخدام الاتصالات من خلال سلسلة من العقد لأداء مهام التعلم الآلي المرتبطة بالذاكرة والتكتل ، على سبيل المثال ، في التعرف على الكلام.

تساعد الوحدات المتكررة ذات البوابات على ضبط أوزان إدخال الشبكة العصبية لحل مشكلة التدرج المتلاشي التي تعد مشكلة شائعة في الشبكات العصبية المتكررة.

س258/ اشرح أهمية GRU؟

ج/ كتحسين لهيكل الشبكة العصبية المتكررة العامة، تحتوي **GRU** على ما يسمى بوابة التحديث وبوابة إعادة الضبط. باستخدام هذين المتجهين، يقوم النموذج بتحسين المخرجات من خلال التحكم في تدفق المعلومات من خلال النموذج. مثل الأنواع الأخرى من نماذج الشبكات المتكررة، يمكن للنماذج التي تحتوي على **GRU** الاحتفاظ بالمعلومات على مدى فترة زمنية - وهذا هو السبب في أن إحدى أبسط الطرق لوصف هذه الأنواع من التقنيات هي أنها نوع من الشبكات العصبية "تتمحور حول الذاكرة". على النقيض من ذلك، غالبًا ما لا تمتلك الأنواع الأخرى من الشبكات العصبية التي لا تحتوي على **GRU** القدرة على الاحتفاظ بالمعلومات.

بالإضافة إلى التعرف على الكلام، يمكن استخدام نماذج الشبكة العصبية التي تستخدم **GRU** للبحث في الجينوم البشري وتحليل خط اليد وغير ذلك الكثير. يتم استخدام بعض هذه الشبكات المبتكرة في تحليل سوق الأوراق المالية والعمل الحكومي. يستفيد الكثير منهم من قدرة المحاكاة للآلات على تذكر المعلومات.

س259/ أيهما أفضل LSTM أم GRU؟

ج / يعمل LSTM جيدًا مع المشكلات التي تكون فيها الدقة بالغة الأهمية والمتسلسلة كبيرة، بينما إذا كنت تريد استهلاكًا أقل للذاكرة وعمليات أسرع ، فاختر GRU

س260/ اشرح أهمية LSTM؟

ج/ تم اختراع LSTM خصيصًا لتجنب مشكلة تلاشي التدرج vanishing gradient.

س261/ كيف تحل LSTM تحدي تلاشي الانحدار vanishing gradient؟

ج / تحل LSTMs المشكلة باستخدام بنية تدرج مضافة فريدة تتضمن الوصول المباشر إلى عمليات تفعيل بوابة النسيان forget gate ، مما يمكّن الشبكة من تشجيع السلوك المطلوب من تدرج الخطأ باستخدام تحديث البوابات المتكرر في كل خطوة زمنية من عملية التعلم.

س262/ لماذا GRU أسرع مقارنة بـ LSTM؟

ج / من خلال عمل كلتا الطبقتين، أي LSTM وGRU، تستخدم GRU معلمة تدريب أقل وبالتالي تستخدم ذاكرة أقل وتنفذ بشكل أسرع من LSTM بينما

LSTM أكثر دقة على مجموعة بيانات أكبر. يمكن للمرء أن يختار LSTM إذا كنت تتعامل مع تسلسلات كبيرة وتشعر بالقلق، يتم استخدام GRU عندما يكون لديك استهلاك أقل للذاكرة وتريد نتائج أسرع.

س263/ ما هي شبكات الخصومة التوليدية (GANs)؟

ج / تُستخدم شبكات الخصومة التوليدية GAN لتحقيق النمذجة التوليدية في التعلم العميق. إنها مهمة غير خاضعة للإشراف تتضمن اكتشاف أنماط في بيانات الإدخال لتوليد المخرجات.

يتم استخدام المولد generator لإنشاء أمثلة جديدة، بينما يتم استخدام المميز discriminator لتصنيف الأمثلة التي تم إنشاؤها بواسطة المولد.

س264/ لماذا تحظى شبكات الخصومة التوليدية (GAN) بشعبية كبيرة؟

ج / تُستخدم شبكات الخصومة التوليدية Generative adversarial (GAN) networks لأغراض متنوعة. في حالة العمل مع الصور، لديهم قدر كبير من العمل الفعال.

- إنشاء الفن (Creation of art): تُستخدم شبكات GAN لإنشاء الصور الفنية والرسومات واللوحات.
- تحسين الصورة (Image enhancement): يتم استخدامها لتحسين دقة الصور المدخلة بشكل كبير.
- ترجمة الصور (Image translation): تُستخدم أيضًا لتغيير جوانب معينة، مثل النهار إلى الليل ومن الصيف إلى الشتاء، في الصور بسهولة.

س265/ ما هي الفكرة وراء شبكات GAN؟

ج / تُستخدم شبكات الخصومة التوليدية (Generative adversarial networks) لتحقيق النمذجة التوليدية في التعلم العميق. إنها مهمة غير خاضعة للإشراف تتضمن اكتشاف أنماط في بيانات الإدخال لتوليد المخرجات.

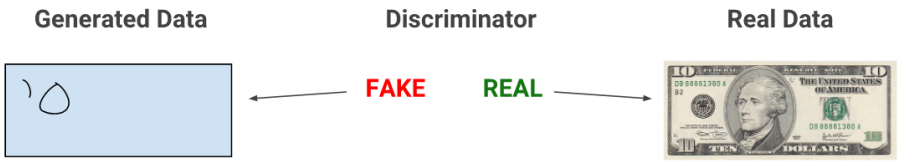
يتم استخدام المولد generator لإنشاء أمثلة جديدة، بينما يتم استخدام المميز discriminator لتصنيف الأمثلة التي تم إنشاؤها بواسطة المولد.

س266/ ما هي مكونات شبكة الخصومة العامة General Adversarial Network؟

ج/تتكون شبكة الخصومة التوليدية (GAN) من جزأين:

- يتعلم المولد generator توليد بيانات معقولة. تصبح الأمثلة التي تم إنشاؤها أمثلة تدريب سلبية للمميز.
- يتعلم المميز discriminator التمييز بين البيانات المزيفة للمولد والبيانات الحقيقية. يعاقب المُميِّز المولد على إنتاج نتائج غير معقولة.

عندما يبدأ التدريب، يُنتج المولد بيانات مزيفة بشكل واضح ، وسرعان ما يتعلم المميّز أن يقول إنها مزيفة:



مع تقدم التدريب، يقترب المولد من إنتاج مخرجات يمكن أن تخدع أداة التمييز:

أخيرًا، إذا سارت عملية تدريب المولدات بشكل جيد، فسيزداد المميز سوءاً في معرفة الفرق بين الحقيقي والمزيف. يبدأ في تصنيف البيانات المزيفة على أنها حقيقية،وتقل دقتها.

س267/ ما هي طرق تقليل الأبعاد dimensionality reduction التي تعرفها وكيف تقارن مع بعضها البعض؟

ج/ الطرق الشائعة لتقليل الأبعاد:

- Principal Component Analysis (PCA)
- Backward Elimination
- Forward Selection
- Score comparison
- Missing Value Ratio
- Low Variance Filter
- High Correlation Filter
- Random Forest
- Factor Analysis

- Autoencoder

س268/ لماذا نحتاج إلى المشفرات التلقائية autoencoders عندما تكون هناك بالفعل تقنيات فعالة لتقليل الأبعاد dimensionality reduction مثل تحليل المكونات الرئيسية PCA؟

ج/ تعد لعنة الأبعاد Curse of Dimensionality (المشكلات التي تظهر عند العمل باستخدام البيانات عالية الأبعاد) مشكلة شائعة عند العمل على التعلم الآلي أو مشروعات التعلم العميق. تسبب مشكلة لعنة الابعاد الكثير من الصعوبات أثناء تدريب نموذج لأنها تتطلب تدريب الكثير من المعلمات على مجموعة بيانات نادرة تؤدي إلى مشكلات مثل الضبط الزائد overfitting، وأوقات التدريب الكبيرة large training times، والتعميم السيئ poor generalization. يتم استخدام PCA وautoencoders لمعالجة هذه المشكلات. PCA هي تقنية غير خاضعة للإشراف حيث يتم عرض البيانات الفعلية في اتجاه التباين العالي بينما تعد autoencoders شبكات عصبية تستخدم لضغط البيانات في مساحة كامنة منخفضة الأبعاد ثم محاولة إعادة بناء البيانات الفعلية عالية الأبعاد.

تكون PCA أو autoencoders فعالة فقط عندما يكون للميزات علاقة ما ببعضها البعض. قاعدة الإبهام العامة general thumb rule بين اختيار PCA وAutoencoders هي حجم البيانات. تعمل Autoencoders بشكل رائع مع مجموعات البيانات الأكبر حجمًا، كما أن PCA تعمل بشكل جيد لمجموعات البيانات الأصغر. يُفضل عادة Autoencoders عندما تكون هناك حاجة لنمذجة غير الخطية والعلاقات المعقدة نسبيًا. يمكن لـ Autoencoders تشفير الكثير من المعلومات بأبعاد أقل عندما يكون هناك انحناء curvature في بنية خافتة منخفضة low dim structure أو غير خطية non-linearity، مما يجعلها خيارًا أفضل على PCA في مثل هذه السيناريوهات.

يُفضل عادةً استخدام Autoencoders لتحديد البيانات الشاذة data anomalies في البيانات بدلاً من تقليل البيانات. يمكن تحديد نقاط البيانات الشاذة باستخدام خطأ إعادة الإعمار reconstruction error، PCA ليست جيدة لإعادة بناء البيانات خاصة عندما تكون هناك علاقات غير خطية.

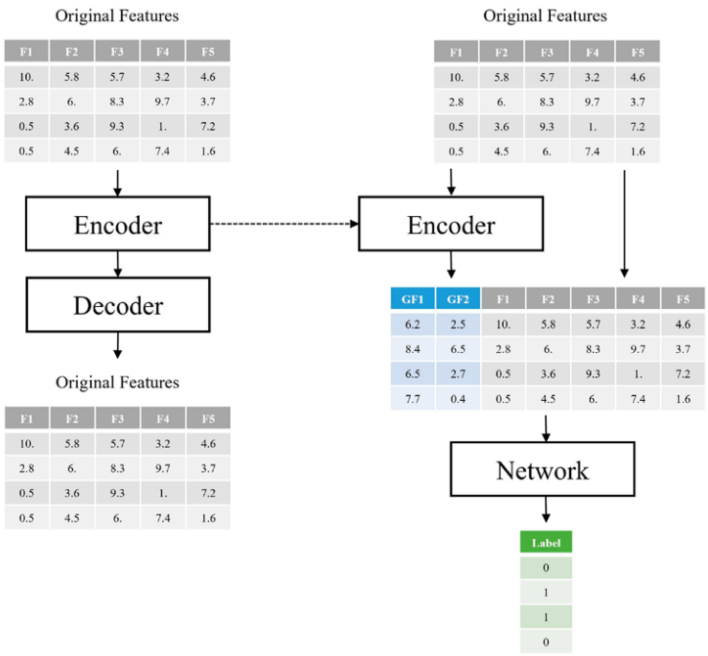
س269/ هل يمكن استخدام المشفرات التلقائية autoencoders لتوليد الميزات؟ إذا كانت الإجابة نعم، فكيف؟

ج/ نعم، يمكن استخدام مكون التشفير في المشفر التلقائي لإنشاء ميزات جديدة لنموذج ليتم التعامل معه ومعالجته عند تنفيذ مهمة ما. نظرًا لأن المشفر قد

تعلم أن يأخذ إدخالًا قياسيًا وضغطه بحيث تحتوي كل ميزة في التمثيل المشفر على المعلومات الأكثر أهمية من الإدخال القياسي، يمكن استغلال هذه الميزات المشفرة لإنشاء ميزات والمساعدة في التنبؤ بنموذج آخر.

الغرض من إنشاء الميزات هو إنشاء ميزات أو إضافتها بدلاً من استبدالها. وبالتالي، عند استخدام المشفرات التلقائية لتوليد الميزات، يوفر المشفر مجموعة واحدة من الميزات المشفرة الغنية بالمعلومات والتي يتم أخذها في الاعتبار جنبًا إلى جنب مع مجموعة الميزات الأصلية.

لاستخدام المشفر التلقائي لإنشاء الميزات، يتم ربط إخراج المشفر (أو دمجه من خلال آلية أخرى) مع بيانات الإدخال الأصلية، بحيث يمكن لبقية الشبكة النظر في كلا مجموعتي الميزات ومعالجتهما.



س270/ ما هي المشفرات التلقائي العميقة deep autoencoders؟

ج / تعد المشفرات التلقائي العميقة deep autoencoders امتدادًا للمشفرات التلقائية العادية. هنا، تكون الطبقة الأولى مسؤولة عن تنفيذ دالة من الدرجة الأولى للمدخلات. ستعتني الطبقة الثانية بدوال الدرجة الثانية، وتستمر.

عادةً ما يكون المشفر التلقائي العميق عبارة عن مزيج من شبكتين متماثلتين أو أكثر من شبكات الاعتقاد العميقة deep-belief networks حيث:

- تتكون الطبقات الخمس الأولى الضحلة من جزء التشفير.
- الطبقات الأخرى تهتم بجزء فك التشفير.

س271/ أين يتم استخدام المشفرات التلقائية autoencoders؟

ج/ تتمتع استخدام المشفرات التلقائية بمجموعة متنوعة من الاستخدامات في العالم الحقيقي. فيما يلي بعض من أشهرها:

- إضافة اللون إلى الصور بالأبيض والأسود.
- إزالة الضوضاء من الصور.
- تقليل الأبعاد Dimensionality reduction.
- إزالة الميزات والتباين.

س272/ ماهي انواع المشفرات التلقائية autoencoders؟

ج/هناك أربعة أنواع رئيسية من المشفرات التلقائية:

- Denoising autoencoder.
- Sparse Autoencoder.
- Deep Autoencoder.
- Contractive Autoencoder.
- Undercomplete Autoencoder.
- Convolutional Autoencoder.
- Variational Autoencoder.

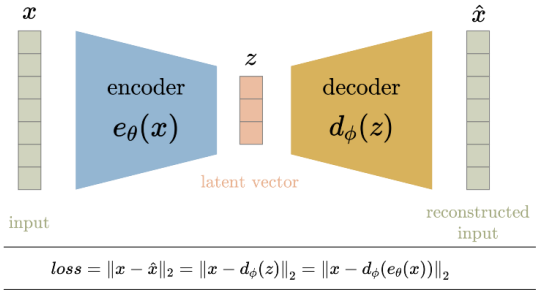
س273/ ما الفرق بين Autoencoderg Variational Autoencoder؟

ج/

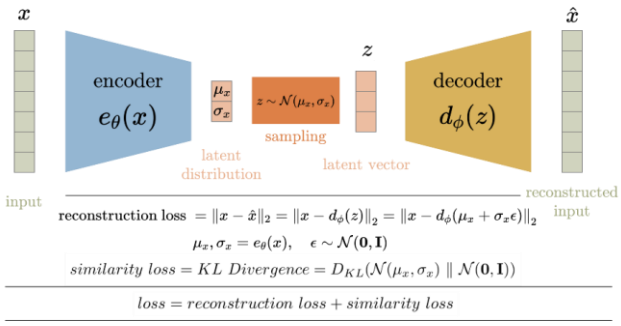
- يتكون المشفر التلقائي autoencoder من جزأين، وحدة التشفير encoder ووحدة فك التشفير decoder. يقوم المشفر بضغط البيانات من مساحة ذات أبعاد أعلى إلى مساحة ذات أبعاد أقل (تسمى أيضًا المساحة الكامنة latent space)، بينما يقوم وحدة فك التشفير بالعكس، أي تحويل المساحة الكامنة مرة أخرى إلى مساحة ذات أبعاد أعلى. أحد عيوب هذه المساحة الكامنة هو عدم تنظيمها: قد تكون هناك أجزاء من المساحة الكامنة لا تتوافق مع أي نقطة بيانات من البيانات الأصلية، لذلك نقول إن المساحة الكامنة تفتقر إلى القدرة التوليدية lacks the generative capability. بمجرد تدريب الشبكة، وإزالة بيانات التدريب، ليس لدينا طريقة لمعرفة ما

إذا كان الناتج الذي تم إنشاؤه بواسطة وحدة فك التشفير من ناقل كامن تم اختياره بشكل عشوائي صالحًا أم لا.

- الاستخدامان الرئيسيان للمشفر التلقائي autoencoder هما ضغط البيانات إلى بعدين (أو ثلاثة) بحيث يمكن رسمها بيانيًا وضغط الصور أو المستندات وفك ضغطها، مما يزيل الضوضاء في البيانات.



- يعالج المشفر التلقائي المتغير variational encoder (VAE) مشكلة المساحة الكامنة غير المنتظمة من خلال فرض قيود على هذا الفضاء الكامن: يجبره على أن يكون توزيعًا عاديًا. لذا بدلاً من إخراج المتجهات في الفضاء الكامن، فإن مشفر VAE ينتج المتوسط والانحراف المعياري لكل متغير كامن. ثم يتم أخذ عينات من المتجه الكامن من هذا المتوسط والانحراف المعياري الذي يتم تغذيته بعد ذلك إلى مفكك الشفرة لإعادة بناء المدخلات.
- يتمثل الاستخدام الرئيسي الوحيد لبرنامج المشفر التلقائي المتغير في إنشاء بيانات جديدة مرتبطة ببيانات المصدر الأصلية. الآن من الصعب تحديد ما هي البيانات الإضافية المفيدة بالضبط. يعد المشفر التلقائي المتغير نظامًا توليديًا ويخدم غرضًا مشابهًا لشبكة الخصومة التوليدية GAN.



س274/ قارن بين PCA وAutoencoders؟

ج/

- ❖ PCA هو في الأساس تحول خطي و Autoencoders قادرة على نمذجة الدوال المعقدة غير الخطية.
- ❖ ميزات PCA غير مرتبطة خطيًا تمامًا مع بعضها البعض نظرًا لأن الميزات عبارة عن إسقاطات على الأساس المتعامد. لكن الميزات المشفرة تلقائيًا قد يكون لها ارتباطات نظرًا لأنها مدربة للتو على إعادة بناء دقيقة.
- ❖ PCA أسرع وأرخص من الناحية الحسابية من Autoencoders.
- ❖ يشبه Autoencoders ذو الطبقة الواحدة مع دالة التنشيط الخطي إلى حد كبير PCA.
- ❖ إن Autoencoders عرضة للضبط الزائد بسبب العدد الكبير من المعلمات. (على الرغم من أن التنظيم regularization والتصميم الدقيق يمكن أن يتجنبوا ذلك)

س275/ ما الفرق بين Variational Autoencoders وGAN؟

ج/

- على عكس شبكة الخصومة التوليدية (GAN)، فإن المشفرات التلقائية المتغيرة (VAE) قابلة للمقارنة بمعنى أنه يمكنك بسهولة التقييم بين اثنين من المشفرات التلقائية المتغيرة VAE من خلال النظر في دالة الخسارة أو الحدود الدنيا التي تحققها. في الوقت الحاضر لا يمكن القيام بذلك لشبكات GAN.
- لكن المشكلة مع VAEs هي أنهم يبسطون المهمة الموضوعية objective task لأنهم ملزمون بالعمل في مساحة كامنة.

س276/ ما هو المشفر التلقائي Autoencoder؟ تفاصيل حول المشفر Encoder وفك التشفير Decoder وعنق الزجاجة Bottleneck؟

ج / المشفر التلقائي **Autoencoder** هو شبكة عصبية تحتوي على ثلاث طبقات، طبقة إدخال، وطبقة مخفية أو طبقة تشفير **encoding layer**، وطبقة فك تشفير **decoding layer**.

المشفر التلقائي هو نوع من نموذج التعلم الآلي الأساسي الذي يكون فيه الإخراج هو نفسه الإدخال. إنه ينتمي إلى عائلة الشبكة العصبية حيث يعمل عن طريق ضغط المدخلات في تمثيل الفضاء الكامن وإعادة بناء أو استخراج الإخراج من تمثيل الفضاء الكامن.

الشبكة العصبية للمشفر التلقائي هي خوارزمية تعلم الآلة غير الخاضعة للإشراف والتي تطبق الانتشار الخلفي، وتعيين القيم المستهدفة لتكون مساوية للمدخلات.

تُستخدم المشفرات التلقائية لتقليل حجم مدخلاتنا إلى تمثيل أصغر. إذا احتاج أي شخص إلى البيانات الأصلية، فيمكنه إعادة بنائها من البيانات المضغوطة.

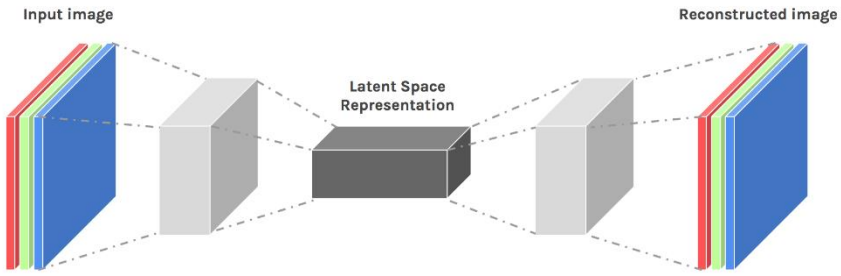
يتكون المشفر التلقائي من ثلاث طبقات:

- المشفر **Encoder**: يقوم هذا الجزء من الشبكة بضغط الإدخال في تمثيل الفضاء الكامن. تقوم طبقة المشفر بتشفير الصورة المدخلة كتمثيل مضغوط بأبعاد مخفضة. الصورة المضغوطة هي النسخة المشوهة من الصورة الأصلية.
- شفرة **Code**:يمثل هذا الجزء من الشبكة المدخلات المضغوطة التي يتم تغذيتها إلى وحدة فك التشفير.
- مفكك الشفرة **Decoder**: تقوم هذه الطبقة بفك تشفير الصورة المشفرة إلى البعد الأصلي. الصورة التي تم فك تشفيرها هي إعادة بناء للصورة الأصلية مع فقدان البيانات ويتم إعادة بنائها من تمثيل الفضاء الكامن.

عنق الزجاجة **Bottleneck**:

تُعرف الطبقة بين المشفر ووحدة فك التشفير، أي الرمز أيضًا باسم **Bottleneck**. هذه طريقة جيدة التصميم لتحديد جوانب البيانات المرصودة التي تحتوي على المعلومات ذات الصلة والجوانب التي يمكن رفضها. يقوم بذلك عن طريق موازنة معيارين:

- انضغاط التمثيل **Compactness of representation**، يقاس بانضغاطية **compressibility**.
- يحتفظ ببعض المتغيرات السلوكية ذات الصلة من المدخلات.



س277/ اشرح عن Sparse Autoencoder؟

ج / في المشفرات التلقائية المتناثرة **Sparse Autoencoder** ذات المنفذ المتناثر **sparsity enforcer** الذي يوجه شبكة طبقة واحدة لتعلم قاموس الكود الذي يقلل من الخطأ في إعادة إنتاج المدخلات مع تقييد عدد كلمات الكود لإعادة البناء.

يتكون المشفر التلقائي المتناثر من طبقة مخفية واحدة متصلة بموجه الإدخال بواسطة مصفوفة وزن تشكل خطوة التشفير **encoding**. تُخرج الطبقة المخفية إلى متجه إعادة البناء، باستخدام مصفوفة وزن مرتبطة لتشكيل وحدة فك التشفير **encoding**.

تُستخدم المشفرات التلقائية المتفرقة لتعلم ميزات من مهمة أخرى ، مثل التصنيف. يجب أن يستجيب المشفر التلقائي المنتظم الذي يجب أن يكون متفرقًا للسمات الإحصائية الفريدة لمجموعة البيانات المدربة ، بدلاً من العمل ببساطة كدالة هوية **identity function**.

س278/ هل يمكن استخدام المشفرات التلقائية لتوليد بيانات التدريب؟ إذا كانت الإجابة نعم، فكيف؟

ج / نعم، يمكنك استخدام المشفرات التلقائية لإنشاء بيانات تدريب ولكنها ستكون مماثلة لبيانات التدريب الأصلية ولكن بجودة أقل.

أفضل طريقة لإنشاء بيانات تدريب هي استخدام شبكات الخصومة التوليدية GAN حيث يتم تدريب إحدى الشبكات على إنشاء بيانات تدريب ذات مظهر حقيقي (المولد) ويتم تدريب الأخرى على التمييز بين البيانات ذات المظهر الاصطناعي (المُميِّز). كلاهما يحاول أن يتفوق على الآخر وفي نهاية عملية التدريب يمكنك استخدام شبكة المولدات لغرضك.

س279/ اشرح عن Denoising Autoencoder؟

ج / بشكل عام، يتم استخدام المشفر التلقائي لاختيار الميزات واستخراج الميزات. عندما يكون عدد الطبقات المخفية أكبر من طبقات الإدخال، فإن الناتج يكون مساويًا للإدخال. لتصحيح هذه المشكلة، نستخدم **Denoising Autoencoder**.

يؤدي **Denoising Autoencoder** إلى إتلاف البيانات عن طريق جعل بعض قيم الإدخال بشكل عشوائي إلى الصفر. بشكل عام، يتم تعيين **50**٪ من عقد الإدخال على صفر. أثناء حساب دالة الخسارة، يتعين علينا مقارنة قيم الإخراج مع المدخلات الأصلية، وليس مع القيمة التالفة. من خلال القيام بذلك يمكننا القضاء على مخاطر تعلم دالة الهوية.

يساعد تقليل الضوضاء المشفرات التلقائية على تعلم تمثيل الفضاء الكامن **latent space** الموجود في البيانات. إنه يضمن تمثيلًا جيدًا يمكن اشتقاقه بقوة من إدخال تالف وسيكون مفيدًا لاستعادة المدخلات النظيفة المقابلة.

س280/ اشرح عن Convolutional Autoencoder (CAE)؟

ج / تستخدم المشفرات التلقائية التلافيفية **CAE** عامل التلافيف لاستغلال هذه الملاحظة. بدلاً من هندسة المرشحات التلافيفية يدويًا، نترك النموذج يتعرف على المرشحات المثلى التي تقلل من خطأ إعادة البناء. يمكن بعد ذلك استخدام هذه المرشحات في أي مهمة تتعلق بالرؤية الحاسوبية **computer vision**. تعد المشفرات التلقائية التلافيفية من أحدث الأدوات للتعلم غير الخاضع للإشراف للفلاتر التلافيفية. بمجرد التعرف على هذه المرشحات، يمكن تطبيقها على أي إدخال لاستخراج الميزات. يمكن استخدام هذه الميزات للقيام بأي مهمة تتطلب تمثيلًا مضغوطًا للإدخال، مثل التصنيف.

CAEs هي نوع من الشبكات العصبية التلافيفية (**CNNs**). يتمثل الاختلاف الرئيسي بين التفسير الشائع لكل من **CNN** و **CAE** في أن المصممين يتم تدريبهم من البداية إلى النهاية لتعلم المرشحات والجمع بين الميزات بهدف تصنيف مدخلاتهم. يتم تدريب الأخير فقط على تعلم المرشحات القادرة على استخراج الميزات التي يمكن استخدامها لإعادة بناء الإدخال.

نظرًا لطبيعتها التلافيفية ، يتكيف **CAE** بشكل جيد مع الصور الدقيقة عالية الأبعاد حيث أن عدد المعلمات المطلوبة لإنتاج خريطة التنشيط هو نفسه دائمًا ، بغض النظر عن حجم الإدخال. لذلك، فإن **CAEs** عبارة عن مستخلصات ميزة للأغراض العامة تختلف عن المشفرات التلقائية **AEs** التي تتجاهل تمامًا بنية الصورة ثنائية الأبعاد. في الواقع، في **AEs** ، يجب أن تكون الصورة غير مقيدة في متجه واحد ويجب بناء الشبكة وفقًا للقيود المفروضة على عدد المدخلات.

س281/ كيف تضبط المعلمات الفائقة Hyperparameter في Autoencoders؟

ج / أثناء تدريب المشفر التلقائي **autoencoder**، سنقوم بضبط المعلمة الفائقة **Hyperparameter** من أجل الحصول على المخرجات المطلوبة.

- حجم الكود **Code size**: يمثل عدد العقد في الطبقة الوسطى. ينتج عن الحجم الأصغر مزيد من الضغط.
- عدد الطبقات **Number of layers**: يمكن أن يتكون المشفر التلقائي من العديد من الطبقات كما نريد.
- عدد العقد لكل طبقة **Number of nodes per layer**: يتناقص عدد العقد لكل طبقة مع كل طبقة لاحقة من المشفر **encoder**، ويزداد مرة أخرى في وحدة فك التشفير **decoder**. فك التشفير متماثل مع المشفر من حيث بنية الطبقة.

- دالة الخسارة Loss function: نستخدم متوسط الخطأ التربيعي MSE أو الانتروبيا الثنائية المتقاطعة لـ Autoencoder. إذا كانت قيم الإدخال في النطاق [0، 1] فإننا نستخدم الانتروبيا المتقاطعة cross-entropy، وإلا فإننا نستخدم متوسط الخطأ التربيعي.

س282/ ما هو تسوية البيانات data normalization في التعلم العميق؟

ج/تسوية البيانات data normalization هو خطوة معالجة مسبقة تُستخدم لتجديد البيانات في نطاق معين. هذا يضمن أن الشبكة يمكن أن تتعلم بشكل فعال لأنها تتمتع بتقارب أفضل عند إجراء الانتشار الخلفيbackpropagation.

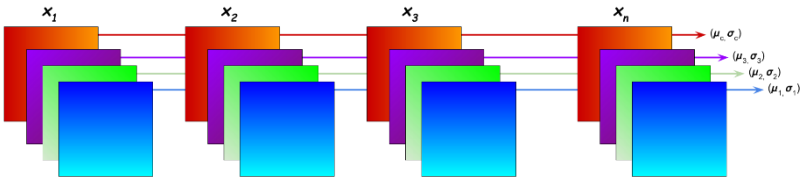
س283/ ما الفرق بين Instance Normalization وBatch Normalization وLayer Normalization؟

ج / دعنا نضع بعض الرموز: سنفترض أن التنشيط في أي طبقة سيكون بأبعاد NxCxHxW، حيث:

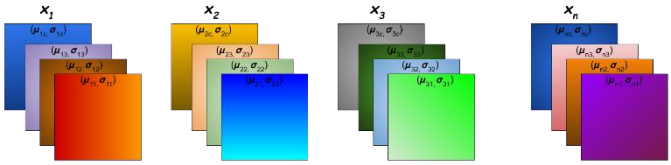
- N = حجم الدفعة ،
- C = عدد القنوات (المرشحات) في تلك الطبقة ،
- H = ارتفاع كل خريطة تنشيط
- W = عرض كل خريطة تنشيط

بشكل عام، يتطلب تسوية التنشيطات تحويل وقياس التنشيطات بالمتوسط (μ) والانحراف المعياري (σ) على التوالي. تختلف تسوية الدُفعات Batch Normalization وتسوية المثيل Instance Normalization وتسوية الطبقة Layer Normalization في طريقة حساب هذه الإحصائيات.

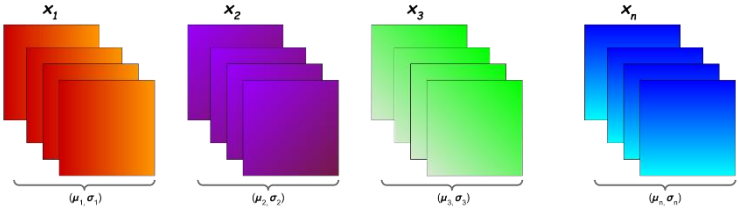
- في تسوية الدُفعات Batch Normalization، يتم حساب المتوسط والتباين لكل قناة فردية عبر جميع العينات والأبعاد المكانية H و W.



- في تسوية المثيل Instance Normalization، يتم حساب المتوسط والتباين لكل قناة فردية لكل عينة فردية عبر كل من الأبعاد المكانية H و W.



❖ في تسوية الطبقة Layer Normalization، يتم حساب المتوسط والتباين لكل عينة فردية عبر جميع القنوات والأبعاد المكانية H و W.



س284/ ما الفرق بين GAN وautoencoders؟

ج/

❖ تتمثل مهمة المشفرات التلقائية autoencoders في تعلم شبكة تشفير وفك تشفير في نفس الوقت. هذا يعني أن مُدخلًا (على سبيل المثال، صورة) يتم إعطاؤه إلى المشفر encoder، والذي يحاول تقليل الإدخال إلى نموذج مشفر مضغوط بشدة، والذي يتم تغذيته بعد ذلك إلى وحدة فك التشفير decoder.

تتعلم الشبكة هذا التشفير / فك التشفير لأن مقياس الخسارة يزداد مع الاختلاف بين صورة الإدخال والإخراج - في كل تكرار، يصبح المشفر أفضل قليلاً في العثور على شكل مضغوط فعال لمعلومات الإدخال، ويحصل وحدة فك التشفير قليلاً أفضل في إعادة بناء المدخلات من النموذج المشفر.

للتلخيص، يتعلم المشفر التلقائي تمثيل بعض معلومات الإدخال بكفاءة عالية، وبالتالي كيفية إعادة بناء المدخلات من شكله المضغوط.

❖ في شبكات الخصومة التوليدية GAN، لدينا "مولد generator " وظيفته أخذ بعض إشارات الضوضاء وتحويلها إلى مساحة مستهدفة معينة (مرة أخرى، الصور هي مثال شائع). المكون الآخر (الخصم the adversary) هو "المُميِّز discriminator "، وتتمثل مهمته في تمييز الصور الحقيقية المستمدة من المساحة المستهدفة المطلوبة عن الصور المزيفة التي تم

إنشاؤها بواسطة المولد. في هذه الحالة، يتم تدريب الشبكة على مرحلتين متعاقبتين، ولكل منهما خسارة مختلفة.

باختصار، تستخدم GAN حلقة تغذية مرتدة معادية لتعلم كيفية إنشاء بعض المعلومات "التي تبدو حقيقية".

س285/ ماذا تفهم بنقل التعلم transfer learning؟ اذكر بعض نماذج التعلم الشائعة الاستخدام؟

ج/ نقل التعلم transfer learning هو عملية نقل التعلم من نموذج إلى نموذج آخر دون الحاجة إلى تدريبه من البداية. يأخذ أجزاء مهمة من نموذج مدرب مسبقًا ويطبقها لحل مشاكل التعلم الآلي الجديدة ولكن المتشابهة.

بعض نماذج التعلم الشائعة هي:

- VGG-16
- BERT
- GTP-3
- Inception V3
- XCeption

س286/ ماذا تقصد بنموذج VGG16؟ وكيف نستخدمه في تصنيف الصور؟

ج/ **VGG-16** عبارة عن شبكة عصبية تلافيفية بعمق **16** طبقة. يمكنك تحميل نسخة مسبقة التدريب من الشبكة تم تدريبها على أكثر من مليون صورة من قاعدة بيانات **ImageNet**. يمكن للشبكة التي تم تدريبها مسبقًا تصنيف الصور إلى **1000** فئة كائن، مثل لوحة المفاتيح والماوس والقلم الرصاص والعديد من الحيوانات. نتيجة لذلك، تعلمت الشبكة تمثيلات غنية بالميزات لمجموعة واسعة من الصور. يبلغ حجم إدخال الصور للشبكة **224 × 224**

س287/ صف بعض المعلمات الفائقة hyperparameters لنقل التعلم transfer learning؟

ج / المعلمات الفائقة التي يجب ضبطها هي عدد الخلايا العصبية number of neurons ودالة التنشيط activation function والمحسن optimizer ومعدل التعلم learning rate وحجم الدفعة batch size والفترات epochs.

س288/ ما هو الضبط الدقيق fine-tuning وكيف يختلف عن نقل التعلم transfer learning؟

ج / في نقل التعلم transfer learning، يظل جزء استخراج الميزة دون تغيير ويتم إعادة تدريب طبقة التنبؤ فقط عن طريق تغيير الأوزان بناءً على التطبيق. على العكس من ذلك في الضبط الدقيق fine-tuning، يمكن إعادة تدريب طبقة التنبؤ إلى جانب مرحلة استخراج الميزة مما يجعل العملية مرنة.

س289/ ما هو نقل التعلم transfer learning وكيف تستخدم نماذج مدربة مسبقًا pre-trained models لنقل التعلم الى شبكتك العصبية؟

ج /ما هو هدفنا عندما نقوم بتدريب شبكة عصبية؟ نرغب في تحديد الأوزان الصحيحة للشبكة من خلال التكرارات المتعددة للأمام والخلف. من خلال استخدام النماذج المدربة مسبقًا والتي تم تدريبها مسبقًا على مجموعات البيانات الكبيرة، يمكننا استخدام الأوزان والبنية التي تم الحصول عليها وتطبيق التعلم على بيان المشكلة بشكل مباشر. هذا هو المعروف باسم نقل التعلم. نحن "ننقل التعلم transfer the learning" للنموذج المدرب مسبقًا إلى بيان المشكلة المحدد لدينا.

يجب أن تكون حذرًا للغاية أثناء اختيار الطراز المدرب مسبقًا الذي يجب أن تستخدمه في حالتك. إذا كانت عبارة المشكلة التي لدينا مختلفة تمامًا عن تلك التي تم تدريب النموذج المدرب مسبقًا عليها - فإن التنبؤ الذي سنحصل عليه سيكون غير دقيق للغاية. على سبيل المثال، قد يعمل النموذج الذي تم تدريبه مسبقًا على التعرف على الكلام بشكل مروع إذا حاولنا استخدامه لتحديد الكائنات التي تستخدمها.

نحن محظوظون لأن العديد من البنى المدربة مسبقًا متاحة لنا مباشرة في مكتبة Keras. تم استخدام مجموعة بيانات Imagenet على نطاق واسع لبناء هياكل مختلفة نظرًا لأنها كبيرة بما يكفي (1.2 مليون صورة) لإنشاء نموذج عام. بيان المشكلة هو تدريب نموذج يمكنه تصنيف الصور بشكل صحيح إلى 1000 فئة كائن منفصلة. تمثل فئات الصور التي يبلغ عددها 1000 فئة فئات الكائنات التي نواجهها في حياتنا اليومية، مثل أنواع الكلاب والقطط والأشياء المنزلية المختلفة وأنواع المركبات وما إلى ذلك.

تُظهر هذه الشبكات المُدرَّبة مسبقًا قدرة قوية على التعميم على الصور خارج مجموعة بيانات ImageNet عبر تعلم النقل. نقوم بإجراء تعديلات على النموذج الموجود مسبقًا عن طريق ضبط النموذج. نظرًا لأننا نفترض أن الشبكة المدربة مسبقًا قد تم تدريبها جيدًا، فلن نرغب في تعديل الأوزان في وقت مبكر جدًا وبالكثير. أثناء التعديل، نستخدم بشكل عام معدل تعلم أصغر من المعدل المستخدم لتدريب النموذج في البداية.

س290/ لأي مشكلة معينة، كيف تقرر ما إذا كان عليك استخدام نقل التعلم transfer learning أو الضبط الدقيق fine-tuning؟

ج/ نقل التعلم transfer learning هو طريقة تُستخدم عند تطوير نموذج لمهمة واحدة يُعاد استخدامها للعمل في مهمة ثانية. يعد الضبط الدقيق fine-tuning أحد الأساليب لتحقيق نقل التعلم. في نقل التعلم، نقوم بتدريب النموذج باستخدام مجموعة بيانات وبعد أن نقوم بتدريب النموذج نفسه باستخدام مجموعة بيانات أخرى لها توزيع مختلف للفئات. في الضبط الدقيق، أحد أساليب نقل التعلم، لدينا مجموعة بيانات، ونقوم بتقسيم 80-20 ونستخدم 80٪ منها في التدريب. ثم نقوم بتدريب نفس النموذج مع 20٪ المتبقية. عادة، نقوم بتغيير معدل التعلم إلى معدل أصغر، لذلك لا يكون له تأثير كبير على الأوزان المعدلة بالفعل. لتحديد الطريقة التي تختارها، يجب على المرء أن يجرب أولاً باستخدام نقل التعلم لأنه سهل وسريع، وإذا لم يكن كافياً للغرض، فاستخدم الضبط الدقيق.

س291/ ما هي طرق الضبط الدقيق Fine Tuning لنماذج التعلم العميق؟

ج/ طرق الضبط الدقيق لنماذج التعلم العميق:

- استخراج الميزات **Feature extraction:** يمكننا استخدام نموذج مدرب مسبقًا كآلية لاستخراج الميزات. ما يمكننا القيام به هو أنه يمكننا إزالة طبقة المخرجات (الطبقة التي تعطي احتمالات التواجد في كل فئة من الفئات **1000**) ثم استخدام الشبكة بالكامل كمستخرج ميزة ثابتة لمجموعة البيانات الجديدة.
- استخدام بنية النموذج المدرب مسبقًا **Use the Architecture of the pre-trained model:** ما يمكننا فعله هو استخدام بنية النموذج بينما نقوم بتهيئة جميع الأوزان بشكل عشوائي وتدريب النموذج وفقًا لمجموعة البيانات الخاصة بنا مرة أخرى.
- تدريب بعض الطبقات بينما تقوم بتجميد البعض الآخر **Train some layers while freeze others:** هناك طريقة أخرى لاستخدام نموذج تم تدريبه مسبقًا وهي التدريب جزئيًا. ما يمكننا القيام به هو الاحتفاظ بأوزان الطبقات الأولية للنموذج مجمدة بينما نقوم بإعادة تدريب الطبقات العليا فقط. يمكننا تجربة واختبار عدد الطبقات التي سيتم تجميدها وعدد الطبقات التي سيتم تدريبها.

س292/ ما هي تطبيقات نقل التعلم transfer learning في التعلم العميق؟

ج/ هناك بعض الأمثلة على تطبيقات نقل التعلم:

التقنيات:

- التعرف على الصور Image Recognition
- معالجة اللغة الطبيعية Natural Language Processing(NLP)
- التعرف على الكلام Speech Recognition

الصناعات:

- القيادة الذاتية Autonomous Driving
- الألعاب Gaming
- رعاية صحية Healthcare
- تصفية البريد العشوائي Spam filtering

س293/ ما هي أهمية وجود شبكات عصبية متبقية residual neural networks؟

ج/ تحل الشبكات العصبية المتبقية residual neural networks مشكلة التدهور degradation problem عن طريق الاختصارات shortcuts أو تخطي الاتصالات skip connections، عن طريق قصر الدائرة الضحلة للطبقات العميقة short circuiting shallow layers to deep layers. يمكننا تكديس الكتل المتبقية أكثر فأكثر، دون تدهور في الأداء. هذا يتيح بناء شبكات عميقة للغاية.

تعتبر الشبكة العصبية المتبقية (ResNet) بلا شك علامة فارقة في التعلم العميق. تم تجهيز ResNet بوصلات مختصرة بين الطبقات، وتعرض تدريبات فعالة باستخدام خوارزميات بسيطة من الدرجة الأولى.

س294/ ماذا تعرف عن نقل التعلم transfer learning؟

ج/ أنت تعرف كيفية ركوب الدراجة bicycle، لذلك سيكون من السهل عليك تعلم قيادة الدراجة الهوائية bike. هذا هو نقل التعلم. لديك بعض المهارة ويمكنك تعلم مهارة جديدة تتعلق بها دون الحاجة إلى تعلمها من الصفر. نقل التعلم هو عملية يمكن من خلالها نقل التعلم من نموذج إلى آخر دون الحاجة إلى جعل النموذج يتعلم كل شيء من البداية. يمكن استخدام الميزات والأوزان لتدريب النموذج الجديد الذي يوفر إمكانية إعادة الاستخدام. يعمل نقل التعلم بشكل جيد في تدريب النموذج بسهولة عندما تكون البيانات محدودة.

س295/ ما هو معنى الضبط الزائد overfitting؟

ج/يعد الضبط الزائد (فرط التعلم) overfitting مشكلة شائعة جدًا عند العمل مع التعلم العميق. إنه سيناريو حيث تقوم خوارزمية التعلم العميق بمطاردة البيانات بقوة للحصول على بعض المعلومات الصحيحة.

هذا يجعل نموذج التعلم العميق يلتقط الضوضاء بدلاً من البيانات المفيدة، مما يتسبب في تباين كبير للغاية وانحياز منخفض. هذا يجعل النموذج أقل دقة، وهذا تأثير غير مرغوب فيه يمكن منعه.

س296/ كيف يمكنك التغلب على الضبط الزائد overfitting؟

ج/يمكننا التغلب على الضبط الزائد **overfitting** باستخدام واحد أو أكثر من التقنيات التالية:

- تبسيط النموذج **Simplifying the model**: يمكننا تقليل الضبط الزائد للنموذج عن طريق تقليل تعقيد النموذج. يمكننا إما إزالة الطبقات أو تقليل عدد الخلايا العصبية في حالة نموذج التعلم العميق، أو تفضيل نموذج متعدد الحدود ذي الترتيب الأقل في حالة الانحدار.
- استخدام التنظيم **Use Regularization**: التنظيم هو الأسلوب الشائع المستخدم لإزالة تعقيد النموذج عن طريق إضافة عقوبة إلى دالة الخسارة. هناك نوعان من تقنيات التنظيم وهما **L1** و **L2**. تعاقب **L1** على مجموع القيم المطلقة للوزن بينما تعاقب **L2** على مجموع القيم التربيعية للوزن. عندما تكون البيانات معقدة للغاية بحيث لا يمكن نمذجتها ، يُفضل استخدام تقنية **L2** ويكون **L1** أفضل إذا كانت البيانات المراد نمذجتها بسيطة للغاية. ومع ذلك ، يفضل **L2** بشكل أكثر شيوعًا.
- زيادة البيانات **Data Augmentation**: زيادة البيانات ليست سوى إنشاء المزيد من عينات البيانات باستخدام مجموعة البيانات الموجودة. على سبيل المثال ، في حالة الشبكة العصبية التلافيفية ، فإن إنتاج صور جديدة عن طريق التقليب **flipping** والتدوير **rotation** والقياس **scaling** وتغيير سطوع **brightness** مجموعة الصور الحالية يساعد في زيادة حجم مجموعة البيانات وتقليل الضبط الزائد.
- التوقف المبكر **Early Stopping**: التوقف المبكر هو أسلوب تنظيم يحدد النقطة التي تؤدي منها بيانات التدريب إلى خطأ التعميم **generalization error** ويبدأ في الضبط الزائد. تتوقف الخوارزمية عن تدريب النموذج في تلك المرحلة.
- تقليل الميزات **Feature reduction**: إذا كان لدينا عدد صغير من عينات البيانات مع عدد كبير من الميزات ، فيمكننا منع الضبط الزائد من خلال

اختيار أهم الميزات فقط. يمكننا استخدام تقنيات مختلفة لهذا مثل اختبار F ، والحذف الأمامي **Forward elimination** ، والحذف الخلفي **Backward elimination**.

- ❖ الحذف العشوائي **Dropouts**: في حالة الشبكات العصبية، يمكننا أيضًا تعطيل نسبة من الخلايا العصبية في كل طبقة بشكل عشوائي. تسمى هذه التقنية بالحذف العشوائي وهي شكل من أشكال التنظيم. ومع ذلك، عندما نستخدم تقنية الحذف العشوائي، يتعين علينا تدريب البيانات لمزيد من الفترات.

س297/ ما هي الأساليب المستخدمة بشكل شائع للتعامل مع فرط التعلم (overfitting) في التعلم العميق؟

ج/

1. Hold-out
2. Cross-validation
3. Data augmentation
4. Feature selection
5. L1 / L2 regularization
6. Remove layers / number of units per layer
7. Dropout
8. Early stopping

س298/ هل يمكنك مشاركة بعض الحيل أو التقنيات التي تستخدمها لمواجهة الضبط الزائد overfitting لتلائم نموذج التعلم العميق والحصول على تعميم generalization أفضل؟

ج/ يتم تعريف الضبط الزائد **overfitting** عندما يكون أداء النموذج جيدًا على بيانات التدريب (تحيز منخفض) وأداء سيئ / ضعيف في بيانات الاختبار (تباين مرتفع). باختصار، لقد تعلم النموذج عبر نمط معين من البيانات وليس مفيدًا لأي بيانات أخرى. يمكن الكشف عن الضبط الزائد عن طريق التحقق من مقاييس الأداء مثل خطأ ودقة نموذج معين. هناك العديد من النصائح والتقنيات التي يمكن للمرء استخدامها لتقليل ملاءمة نموذج التعلم العميق:

- ❖ زيادة حجم بيانات التدريب.
- ❖ تقليل عدد الطبقات في الطبقة المخفية، سيؤدي ذلك إلى تقليل سعة الشبكات.
- ❖ تطبيق التنظيم **regularization** .

- أضف طبقات الحذف العشوائي Add dropout layers.
- التوقف المبكر Early stopping - حاول إيقاف التدريب قبل زيادة فقدان التحقق validation loss من الصحة.
- استفد من زيادة البيانات data augmentation.

س299/ كيفية منع فرط التعلم overfitting في التعلم العميق؟

ج / يمكنك تقليل overfitting من خلال تدريب الشبكة على المزيد من الأمثلة أو تقليل overfitting عن طريق تغيير تعقيد الشبكة.

تتمثل فائدة الشبكات العصبية العميقة جدًا في أن أداءها يستمر في التحسن حيث يتم تغذيتها بمجموعات بيانات أكبر وأكبر. إن النموذج الذي يحتوي على عدد لا حصر له من الأمثلة سوف يستقر في النهاية من حيث قدرة الشبكة على التعلم.

س300/ ما هو الضبط الناقص underfitting وكيف يمكن منعه؟

ج / يعد الضبط الناقص underfitting مشكلة عندما يكون لدينا خطأ منخفض في كل من مجموعة التدريب ومجموعة الاختبار. قليل من الخوارزميات تعمل بشكل أفضل للتفسيرات لكنها تفشل في الحصول على تنبؤات أفضل.

س301/ كيف تعرف إذا كان نموذج يعاني من الضبط الزائد overfitting؟

ج / يكاد يكون من المستحيل الكشف عن الضبط الزائد قبل اختبار البيانات. يمكن أن يساعد في معالجة السمة المتأصلة في الضبط الزائد، وهي عدم القدرة على تعميم مجموعات البيانات. وبالتالي، يمكن فصل البيانات إلى مجموعات فرعية مختلفة لتسهيل التدريب والاختبار. يتم تقسيم البيانات إلى قسمين رئيسيين، أي مجموعة اختبار ومجموعة تدريب.

تمثل مجموعة التدريب غالبية البيانات المتاحة (حوالي 80٪)، وهي تدرب النموذج. تمثل مجموعة الاختبار جزءًا صغيرًا من مجموعة البيانات (حوالي 20٪)، ويتم استخدامها لاختبار دقة البيانات التي لم تتفاعل معها من قبل. من خلال تقسيم مجموعة البيانات، يمكننا فحص أداء النموذج في كل مجموعة من البيانات لاكتشاف الضبط الزائد عند حدوثه، وكذلك معرفة كيفية عمل عملية التدريب.

يمكن قياس الأداء باستخدام النسبة المئوية للدقة التي لوحظت في كلتا مجموعتي البيانات لاستنتاج وجود الضبط الزائد. إذا كان أداء النموذج في مجموعة التدريب أفضل من أداءه في مجموعة الاختبار، فهذا يعني أن النموذج من المحتمل يعاني من الضبط الزائد.

س302/ ما هي الاختلافات بين overfitting وunderfitting؟

ج / في الإحصاء والتعلم الآلي، تتمثل إحدى المهام الأكثر شيوعًا في ملاءمة نموذج لمجموعة من بيانات التدريب، حتى تكون قادرًا على عمل تنبؤات موثوقة بشأن البيانات العامة غير المدربة.

في الضبط الزائد **overfitting**، يصف النموذج الإحصائي خطأ عشوائي أو ضوضاء بدلاً من العلاقة الأساسية. يحدث الضبط الزائد عندما يكون النموذج شديد التعقيد، مثل وجود عدد كبير جدًا من المعلمات بالنسبة إلى عدد الملاحظات. النموذج الذي فيه ضبط زائد، لديه أداء تنبؤي ضعيف، لأنه يبالغ في رد فعله للتقلبات الطفيفة في بيانات التدريب.

يحدث الضبط الناقص **underfitting** عندما لا يتمكن النموذج الإحصائي أو خوارزمية التعلم الآلي من التقاط الاتجاه الأساسي للبيانات. قد يحدث الضبط الناقص، على سبيل المثال، عند ملاءمة نموذج خطي للبيانات غير الخطية. مثل هذا النموذج أيضًا سيكون له أداء تنبؤي ضعيف.

س303/ عرف الضبط الزائد (فرط التعلم) overfitting؟ كيف نضمن أننا لا نفرط في تعليم نموذج؟

ج / يحدث الضبط الزائد **overfitting** عندما يبحث النموذج في بيانات التدريب للتأثير بشكل كبير على أداء النموذج في أحدث البيانات. يشير هذا إلى أننا نسجل الاضطراب في بيانات التدريب، ونتعلم المفاهيم حسب النموذج. المشكلة هي أن المفاهيم التي لا تستخدم بيانات الاختبار تؤثر سلبًا على قدرة النموذج على تصنيف البيانات الجديدة؛ لذلك، إنقاص دقة بيانات الاختبار.

لتجنب فرط التعلم، يتعين علينا تطبيق الطرق التالية:

- نجمع المزيد من البيانات حتى نتمكن من تدريب النموذج على عينات متنوعة.
- يمكننا تجنب فرط التعلم باستخدام طرق التجميع ، مثل **Random Forest**. وفقًا لفكرة التعبئة **bagging** ، نستخدمها لتقليل التغيير في التوقعات من خلال الانضمام إلى نتيجة أشجار القرار المتعددة في عينات مختلفة من مجموعة البيانات.
- من خلال تحديد الخوارزمية الصحيحة ، يمكننا تجنب فرط التعلم **overfitting**.

س304/ اذكر سبب أهمية هندسة الميزات feature engineering في بناء النموذج واذكر بعض التقنيات المستخدمة في هندسة الميزات؟

ج/هندسة الميزات feature engineering هي "فن" صياغة ميزات مفيدة من البيانات الحالية بعد الهدف المراد تعلمه ونموذج التعلم الآلي المستخدم. إنه ينطوي على تحويل البيانات إلى نماذج ترتبط بشكل أفضل بالهدف الأساسي المراد تعلمه. عند القيام بذلك بشكل صحيح، يمكن لهندسة الميزات زيادة قيمة البيانات الحالية وتحسين أداء نماذج التعلم الآلي الخاصة بك. من ناحية أخرى، قد يتطلب استخدام الميزات السيئة إنشاء نماذج أكثر تعقيدًا لتحقيق نفس المستوى من الأداء.

تقنيات هندسة الميزات التالية التي يجب أن تعرفها:

1. Imputation
    A. Categorical Imputation
    B. Numerical Imputation
1. Discretization
2. Categorical Encoding
3. Feature Splitting
4. Handling Outliers
5. Variable Transformations
6. Scaling
7. Creating Features

س305/ ما هو التوزيع الطبيعي normal distribution؟

ج / التوزيع الطبيعي normal distribution هو توزيع احتمالي حيث تكون القيم متماثلة على جانبي متوسط البيانات. هذا يعني أن القيم الأقرب إلى المتوسط أكثر شيوعًا من القيم البعيدة عنه.

س306/ ما المقصود بتحجيم الميزات Feature Scaling؟

ج/يمكن أن يكون هي تحجيم الميزات من خلال التوحيد القياسي **standardization** (أو تسوية درجة Z) خطوة مهمة في المعالجة المسبقة للعديد من خوارزميات التعلم الآلي. يتضمن التوحيد إعادة قياس الميزات بحيث يكون لها خصائص التوزيع الطبيعي القياسي بمتوسط صفر وانحراف معياري بواحد.

س307/ كيف تختار الميزات في الشبكة العصبية؟

ج / يعتبر الارتباط القوي جدًا **very strong correlation** بين الميزة الجديدة والميزة الحالية علامة جيدة إلى حد ما على أن الميزة الجديدة توفر القليل من المعلومات الجديدة.

من المحتمل أن يكون الارتباط المنخفض **low correlation** بين الميزة الجديدة والميزات الحالية هو المفضل **preferable**.

يعتبر الارتباط الخطي القوي بين الميزة الجديدة والمتغير المتوقع علامة جيدة على أن الميزة الجديدة ستكون ذات قيمة، ولكن عدم وجود ارتباط كبير ليس بالضرورة علامة على ميزة ضعيفة، لأن الشبكات العصبية لا تقتصر على المجموعات الخطية من المتغيرات.

إذا تم إنشاء الميزة الجديدة يدويًا من مجموعة من الميزات الموجودة، ففكر في تركها. يكمن جمال الشبكات العصبية في الحاجة إلى القليل من الهندسة والمعالجة المسبقة - وبدلاً من ذلك يتم تعلم الميزات بواسطة الطبقات الوسيطة.

كلما كان ذلك ممكنًا، يفضل ميزات التعلم على هندستها.

س308/ ما هو الفرق بين اختيار الميزات feature selection وطرق هندسة الميزات feature engineering؟

ج/

اختيار الميزات	هندسة الميزات (المعالم)
طرق اختيار الميزة هي الطرق المستخدمة للحصول على مجموعة فرعية من المتغيرات من مجموعة البيانات المطلوبة لإنشاء نموذج يناسب الاتجاهات في مجموعة البيانات بشكل أفضل.	طرق هندسة المعالم هي الطرق المستخدمة لإنشاء ميزات جديدة من مجموعة البيانات المحددة باستخدام المتغيرات الحالية. تسمح هذه الأساليب بمواءمة الاتجاهات المعقدة بشكل أفضل في مجموعة البيانات.
مثال: الاقتطاع Imputation ، التفكك Discreteziation ، الترميز الفئوي Categorical Encoding ، إلخ.	مثال: (الخوارزميات المستندة إلى القواعد والأشجار ، ونماذج MARS ، وما إلى ذلك) ، وطرق التصفية ، وطرق الالتفاف (حذف الميزة التكرارية ، والخوارزميات الجينية ، وما إلى ذلك)

س309/ هل هناك فرق بين تعلم الميزات feature learning واستخراج الميزات feature extraction؟

ج / استخراج الميزة feature extraction هو تطبيق بعض القواعد /الدوال المحددة مسبقًا على البيانات والحصول على هذه الميزات. تعلم الميزات feature learning أكثر توجهاً نحو المهام، على سبيل المثال تتعلم ميزات للتعرف على الأرقام كهدفك النهائي. كلاهما متشابهان بمعنى أنهما مجرد وسيلة للحصول على أدلة خوارزمية التعلم للقيام بتنبؤاتها.

س310/ هل نحتاج إلى استخراج الميزات feature extraction في التعلم العميق؟

ج / أكبر ميزة في التعلم العميق هي أننا لا نحتاج إلى استخراج الميزات يدويًا من الصورة. تتعلم الشبكة استخراج الميزات أثناء التدريب. أنت فقط تغذي الصورة للشبكة (قيم البكسل).

س311/ ما هي بعض التقنيات المستخدمة لأخذ العينات sampling؟ ما هي الميزة الرئيسية لأخذ العينات؟

ج / فيما يلي تقنيات أخذ العينات sampling شائعة الاستخدام:

- Simple Random Sampling
- Systematic Sampling
- Cluster Sampling
- Purposive Sampling
- Quota Sampling
- Convenience Sampling

أخذ العينات sampling أكثر كفاءة من حيث التكلفة والوقت من دراسة مجموعة بيانات كاملة. يتيح لك تحليل مجموعة فرعية من تلك البيانات، وهو أمر أسهل مع توفير رؤى حول مجموعة البيانات بأكملها.

س312/ ما هي متجهات الميزات feature vectors؟

ج / متجهات المعالم او الميزات feature vectors هي مجموعة من المتغيرات التي تحتوي على قيم تصف خصائص كل ملاحظة في مجموعة البيانات. تعمل هذه المتجهات كمتجهات إدخال لنموذج التعلم الآلي.

س313/ ما هي طرق تحديد الميزة Feature Selection Methods المستخدمة لتحديد المتغيرات الصحيحة؟

ج / فيما يلي بعض الأساليب المستخدمة لاختيار الميزة Feature Selection في تحليل البيانات:

- Pearson's Correlation
- Chi-Square
- Recursive Feature Elimination
- Backward Elimination
- Lasso Regression
- Ridge Regression

س314/ ما الفرق بين اختيار الميزة Feature Selection واستخراج الميزة Feature Extraction ومن الذي يأتي اولاً؟

ج / يتمثل الاختلاف الرئيسي بين اختيار الميزة **Feature Selection** واستخراج الميزة **Feature Extraction** في أن تحديد الميزة يحتفظ بمجموعة فرعية من الميزات الأصلية بينما يؤدي استخراج الميزات إلى إنشاء ميزات جديدة تمامًا.

الخطوة الأولى لاستخراج الميزات ثم تطبيق طريقة اختيار الميزة على هذه الميزات قبل تطبيقها على المصنف.

س315/ كيف ستتعامل مع مجموعة بيانات غير متوازنة imbalanced dataset؟

ج /مجموعة البيانات غير المتوازنة Imbalanced dataset هي مشكلة تصنيف حيث لا يتم توزيع عدد المشاهدات لكل فئة بالتساوي. بالنسبة لبعض الفئات، سيكون هناك عدد كبير من الملاحظات بينما توجد ملاحظات أقل بالنسبة للآخرين. يمكننا حل هذه المشكلة عن طريق:

- جمع المزيد من البيانات حتى الاختلالات (غير المتوازنه) في مجموعة البيانات.
- أعادة تشكيل مجموعة البيانات لتصحيح الاختلالات.
- تجربة خوارزمية مختلفة تمامًا في مجموعة البيانات الخاصة بك.

س316/ كيف تتعامل مع التصنيف الثنائي غير المتوازن unbalanced binary classification؟

ج / أثناء إجراء التصنيف الثنائي، إذا كانت مجموعة البيانات غير متوازنة، فلا يمكن التنبؤ بدقة النموذج بشكل صحيح باستخدام **R2 score** فقط. على سبيل المثال، إذا كانت البيانات التي تنتمي إلى أحد الفئتين أقل من حيث الكمية مقارنة بالفئة الأخرى، فستأخذ الدقة التقليدية نسبة صغيرة جدًا من الفئة الأصغر. إذا كانت نسبة **5**٪ فقط من الأمثلة تنتمي إلى فئة أصغر، وكان النموذج يصنف جميع المخرجات التي تنتمي إلى الفئة الأخرى، فستظل الدقة حوالي **95**٪. لكن هذا سيكون خطأ. للتعامل مع هذا، يمكننا القيام بما يلي:

- استخدم طرقًا أخرى لحساب أداء النموذج مثل الدقة / الاسترجاع ، و **F1 score**، وما إلى ذلك.
- أعد أخذ عينات البيانات باستخدام تقنيات مثل اختزال العينة **undersampling** (تقليل حجم العينة للفئة الأكبر) ، والاختزال الزائد **oversampling** (زيادة حجم العينة للفئة الأصغر باستخدام التكرار ، و **SMOTE**، وتقنيات أخرى من هذا القبيل.
- استخدام التحقق المتقاطع **K-fold**
- استخدام التعلم الجماعي **ensemble learning** بحيث تأخذ كل شجرة قرار بعين الاعتبار العينة الكاملة للفئة الأصغر ومجموعة فرعية فقط من الفئة الأكبر.

س317/ كيف تحقق التوازن في القوة عند التعامل مع مجموعات البيانات غير المتوازنة imbalanced datasets في التعلم العميق؟

ج / من المستحيل أن يكون لديك مجموعة بيانات واقعية متوازنة تمامًا عند العمل على مشكلات التعلم العميق، لذلك سيكون هناك مستوى معين من عدم التوازن في الفصل داخل البيانات التي يمكن معالجتها إما عن طريق:

- موازنة الوزن **Weight Balancing**
- أخذ العينات **Over and Under Sampling**

س318/ هل ستقوم بإزالة المتغيرات المرتبطة correlated variables أولاً؟ لماذا؟

ج/في موقف أكثر عمومية، عندما يكون لديك متغيرين مستقلين مترابطين للغاية **very highly correlated**، يجب عليك بالتأكيد إزالة أحدهما لأنك واجهت لغزًا متعدد

الخطوط وستكون معاملات الانحدار الخاصة بنموذج الانحدار المتعلقة بالمتغيرين المترابطين للغاية غير موثوقة.

س319/ ما هي الطرق المتاحة لفحص القيم المتطرفة Outliers؟

ج/ يمكننا استخدام الطرق التالية لفحص القيم المتطرفة **Outliers**:

- ❖ النماذج الخطية **Linear models**: يمكن تدريب النماذج الخطية مثل الانحدار اللوجستي لفحص القيم المتطرفة. وبهذه الطريقة، يجمع النموذج الخارجة اللاحقة التي يلتقي بها.
- ❖ مخطط الصندوق **Boxplot**: مخطط الصندوق يصور تخصيص البيانات وإمكانية تغييرها. مخطط الصندوق تشمل الارباع الدنيا والعليا؛ لذلك، فإن الصندوق يمتد بشكل أساسي إلى المدى الرباعي (**IQR**). السبب الرئيسي لاستخدام مخطط الصندوق هو تحديد القيم المتطرفة في البيانات.
- ❖ النماذج المستندة إلى القرب **Proximity-based models**: **K-mean clustering** هو مثال على هذا النوع من النماذج، حيث تشكل نقاط البيانات مجموعات متنوعة أو مجموعات "**k**" بناءً على ميزات مثل المسافة أو التشابه.
- ❖ النماذج الاحتمالية والإحصائية **Probabilistic and Statistical models**: يمكننا استخدام النماذج الإحصائية مثل التوزيع الأسي والتوزيع الطبيعي لتحديد الاختلافات في تخصيص نقاط البيانات. إذا وجدنا أي نقطة بيانات خارج نطاق التوزيع، فيمكننا أن نجعلها خارجية (متطرفة).

س320/ ما هي القيم المتطرفة outliers؟ اذكر ثلاث طرق للتعامل مع القيم المتطرفة.

ج/ يمكنك تجربة ما يلي:

- ❖ جرب نموذجًا مختلفًا. يمكن أن تناسب النماذج غير الخطية البيانات التي تم اكتشافها على أنها قيم متطرفة بواسطة النماذج الخطية. لذلك، تأكد من اختيار النموذج الصحيح.
- ❖ حاول تسوية البيانات Normalization. بهذه الطريقة، يتم سحب نقاط البيانات القصوى إلى نطاق مماثل.
- ❖ يمكنك استخدام الخوارزميات الأقل تأثراً بالقيم المتطرفة؛ مثال على ذلك الغابات العشوائية random forests.

س321/ ما هي بعض الطرق التي يمكنني من خلالها جعل نموذجي أكثر قوة بالنسبة للقيم المتطرفة outliers؟

ج/ يمكن أن يكون لدينا تنظيم مثل **L1** أو **L2** لتقليل التباين (زيادة التحيز).

التغييرات على الخوارزمية:

- استخدم الطرق المستندة إلى الشجرة بدلاً من طرق الانحدار لأنها أكثر مقاومة للقيم المتطرفة. بالنسبة للاختبارات الإحصائية ، استخدم الاختبارات غير البارامترية بدلاً من الاختبارات البارامترية.
- استخدم مقاييس خطأ قوية مثل **MAE** أو **Huber Loss** بدلاً من **MSE**.

التغييرات على البيانات:

- تفكيك البيانات.
- تحويل البيانات (مثل **log**).
- قم فقط بإزالة القيم المتطرفة إذا كنت متأكدًا من أنها حالات شاذة لا تستحق التنبؤ بها.

س322/ كيف ستتعامل مع القيم المفقودة missing values في البيانات؟

ج/ هناك عدة طرق للتعامل مع القيم المفقودة في البيانات المعطاة:

- حذف القيم
- حذف الملاحظة (لا يوصى به دائما)
- استبدال القيمة بالمتوسط والوسيط وطريقة الملاحظة **mode of the observation**
- توقع القيمة مع الانحدار **regression**
- إيجاد قيمة مناسبة مع التجميع **clustering**

س323/ اشرح كيف يمكننا التقاط العلاقة بين المتغيرات المستمرة continuous والمتغيرات الفئوية categorical؟

ج/ نعم، يمكن استخدام تقنية **ANCOVA**. إنها تعني تحليل التغاير. يتم استخدامه لحساب الارتباط بين المتغيرات المستمرة **continuous** والفئوية **categorical**.

س324/ ما هي خوارزميات التعلم الآلي التي يمكن استخدامها لإدخال القيم المفقودة لكل من المتغيرات الفئوية والمستمرة؟

ج / تعمل خوارزميات K-NN بشكل أفضل عندما يتعلق الأمر بإدخال القيم في البيانات الفئوية والمستمرة.

س325/ تخيل أنك حصلت على مجموعة بيانات تتكون من متغيرات بها أكثر من 30٪ من القيم المفقودة missing values. لنفترض أنه من بين 50 متغيرًا، هناك 16 متغيرًا بها قيم مفقودة، وهي أعلى من 30٪. كيف ستتعامل معهم؟

ج/ للتعامل مع القيم المفقودة **missing values** نقوم بما يلي:

- سنحدد فئة مختلفة للقيم المفقودة.
- الآن ، سوف نتحقق من توزيع القيم ، وسنحتفظ بالقيم المفقودة التي تحدد النمط.
- بعد ذلك، سنقوم بشحن هذه القيم إلى فئة أخرى مع استبعاد الآخرين.

س326/ كيف تتعامل مع البيانات المفقودة missing أو التالفة corrupted في مجموعة البيانات dataset؟

ج/ يمكنك العثور على البيانات المفقودة/التالفة في مجموعة البيانات وإما إسقاط تلك الصفوف أو الأعمدة، أو اتخاذ قرار استبدالها بقيمة أخرى.

في Pandas، هناك طريقتان مفيدتان للغاية: isnull() و dropna() والتي ستساعدك في العثور على أعمدة البيانات مع البيانات المفقودة أو التالفة وإسقاط هذه القيم. إذا كنت تريد تعبئة القيم غير الصالحة بقيمة عنصر نائب (على سبيل المثال ، 0) ، يمكنك استخدام طريقة fillna().

س327/ ماذا تقصد بـ Imputation؟

ج/ **Imputation** هو تقنية تُستخدم لاستبدال البيانات المفقودة **missing data** ببعض القيمة البديلة للاحتفاظ بمعظم البيانات / المعلومات الخاصة بمجموعة البيانات. يستبدل القيم المفقودة بالمتوسط **mean** أو وسيط **median** القيم المتبقية. تُستخدم هذه الأساليب لأن إزالة البيانات من مجموعة البيانات في كل مرة غير ممكن ويمكن أن تؤدي إلى تقليل حجم مجموعة البيانات إلى حد كبير، الأمر الذي لا يثير مخاوف بشأن تحيز مجموعة البيانات فحسب، بل يؤدي أيضًا إلى تحليل غير صحيح.

س328/ ما هي مشاكل جودة البيانات Data Quality الشائعة؟

ج/

- قيم مفقودة **Missing Values**
- الضوضاء في مجموعة البيانات **Noise in the Data Set**
- القيم المتطرفة **Outliers**

- مزيج من لغات مختلفة (مثل الإنجليزية والصينية) Mixture of Different Languages
- قيود النطاق Range Constraints

س329/ ما هي الخطوات التي يجب اتباعها لاستخدام خوارزمية الانحدار التدريجي gradient descent؟

ج/ هناك خمس خطوات رئيسية تُستخدم لتهيئة واستخدام خوارزمية الانحدار التدريجي:

- بدء التحيزات والأوزان للشبكة.
- إرسال بيانات الإدخال عبر الشبكة (طبقة الإدخال).
- حساب الفرق (الخطأ) بين القيم المتوقعة والمتوقعة.
- تغيير القيم في الخلايا العصبية لتقليل دالة الخطأ.
- تكرارات متعددة لتحديد أفضل الأوزان للعمل بكفاءة.

س330/ ما هو انفجار الانحدار التدريجي exploding gradient descent في التعلم العميق؟

ج/ انفجار التدرجات Exploding gradients هو مشكلة تسبب سيناريو يجمع التدرجات. يؤدي هذا إلى إنشاء عدد كبير من التحديثات للأوزان في النموذج عند التدريب.

يعتمد عمل الانحدار التدريجي على شرط أن تكون التحديثات صغيرة وخاضعة للتحكم. سيؤثر التحكم في التحديثات بشكل مباشر على كفاءة النموذج.

س331/ ما هي أنواع الانحدار التدريجي gradient descent؟

ج/ هناك ثلاثة أنواع مختلفة من الانحدار التدريجي كما هو موضح أدناه:

- الانحدار التدريجي العشوائي (Stochastic gradient descent): يتم استخدام مثال تدريبي واحد لحساب الانحدار التدريجي وتحديث المعلمات.
- الانحدار التدريجي دفعة واحدة (Batch gradient descent): يُحسب التدرج لمجموعة البيانات بأكملها، ويتم تحديث المعلمات في كل تكرار.
- الانحدار التدريجي بالدفعات الصغيرة (Mini-batch gradient descent): يتم تقسيم العينات إلى مجموعات أصغر حجمًا ثم يتم العمل عليها كما في حالة الانحدار التدريجي العشوائي.

س332/ لماذا يعتبر الانحدار التدريجي بالدفعات الصغيرة Mini-Batch Gradient Descent شائعا جداً؟

ج/ إن الانحدار التدريجي بالدفعات الصغيرة Mini-Batch Gradient Descent شائع على النحو التالي:

- ❖ إنه أكثر كفاءة بالمقارنة مع الانحدار التدريجي العشوائي.
- ❖ يتم التعميم من خلال إيجاد الصغرى المسطحة flat minima.
- ❖ يساعد في تجنب الحدود الدنيا المحلية local minima من خلال السماح بتقريب الانحدار الاشتقاقي لمجموعة البيانات بأكملها.

س333/ ماذا تفهم عن اقتصاص التدرج Gradient Clipping؟

ج/ يتم استخدام Gradient Clipping للتعامل مع مشكلة انفجار التدرج التي تحدث أثناء الانتشار الخلفي backpropagation. يتم فرض قيم التدرج على العنصر إلى حد أدنى أو أقصى قيمة معينة إذا تجاوز التدرج النطاق المتوقع. يوفر Gradient Clipping ثباتًا عدديًا أثناء تدريب شبكة عصبية ولكنه لا يوفر أي تحسينات في الأداء.

س334/ ما هي أنواع الانتشار الخلفي variants of Backpropagation؟

ج/

- ❖ الانحدار التدريجي العشوائي (Stochastic gradient descent): يتم استخدام مثال تدريبي واحد لحساب الانحدار التدريجي وتحديث المعلمات.
- ❖ الانحدار التدريجي دفعة واحدة (Batch gradient descent): يُحسب التدرج لمجموعة البيانات بأكملها، ويتم تحديث المعلمات في كل تكرار.
- ❖ الانحدار التدريجي بالدفعات الصغيرة (Mini-batch gradient descent): يتم تقسيم العينات إلى مجموعات أصغر حجمًا ثم يتم العمل عليها كما في حالة الانحدار التدريجي العشوائي.

س335/ صف طريقتين للتعامل مع مشكلة تلاشي الانحدار vanishing gradient في الشبكة العصبية؟

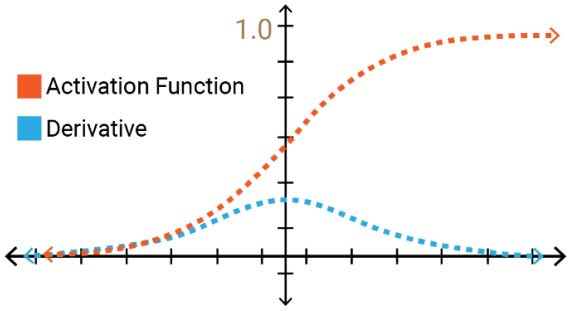
ج/

- ❖ استخدم دالة التنشيط ReLU بدلاً من دالة sigmoid
- ❖ تهيئة الشبكات العصبية باستخدام تهيئة Xavier التي تعمل مع دالة تنشيط tanh.

س336/ ما هي مشكلة تلاشي التدرج Vanishing Gradient في الشبكات العصبية الاصطناعية؟

ج/ تتم مواجهة مشكلة تلاشي التدرج Vanishing Gradient في الشبكات العصبية الاصطناعية باستخدام طرق التعلم القائمة على التدرج gradient-based learning والانتشار الخلفي backpropagation.

في طرق التعلم هذه، تتلقى كل أوزان للشبكات العصبية تحديثًا يتناسب مع المشتق الجزئي لدالة الخطأ فيما يتعلق بالوزن الحالي في كل تكرار للتدريب. في بعض الأحيان عندما تصبح التدرجات صغيرة بشكل متلاشي، فإن هذا يمنع الوزن من تغيير القيمة.



إذا كانت الشبكة العصبية تحتوي على العديد من الطبقات المخفية hidden layers، فإن التدرجات في الطبقات السابقة ستصبح منخفضة جدًا حيث نقوم بضرب مشتقات كل طبقة. نتيجة لذلك، يصبح التعلم في الطبقات السابقة بطيئًا جدًا.

- يمكن أن يتسبب هذا في توقف الشبكة العصبية عن التعلم.
- تحدث مشكلة تلاشي الانحدار التدريجي عند تدريب الشبكات العصبية ذات الطبقات المتعددة لأن التدرج يتضاءل بشكل كبير حيث ينتشر إلى الوراء عبر الشبكة.

تم اقتراح العديد من الإصلاحات والحلول البديلة والتحقيق فيها لإصلاح مشكلة تلاشي الانحدار التدريجي (fix the vanishing gradient problem)، مثل :

- مخططات تهيئة الوزن البديل alternate weight initialization schemes
- تدريب مسبق غير خاضع للإشراف unsupervised pre-training
- التدريب الطبقي layer-wise training
- الاختلافات في الانحدار التدريجي variations on gradient descent

س337/ قارن بين الانحدار التدريجي الدفعي Batch Gradient Descent والانحدار التدريجي العشوائي Stochastic Gradient Descent والانحدار التدريجي للدفعة الصغيرة Mini Batch gradient descent؟

ج/

- الانحدار التدريجي الدفعي Batch Gradient Descent: يعالج جميع أمثلة التدريب لكل تكرار للانحدار التدريجي gradient descent. إذا كان عدد أمثلة التدريب كبيرًا، فإن الانحدار التدريجي للدفعة مكلف للغاية. لذلك، في حالة أمثلة التدريب الكبيرة، نفضل استخدام الانحدار التدريجي العشوائي أو الانحدار التدريجي للدفعة الصغيرة.
- الانحدار التدريجي العشوائي Stochastic Gradient Descent: هذا نوع من الانحدار التدريجي الذي يعالج مثال تدريب واحد لكل تكرار. وبالتالي، يتم تحديث المعلمات حتى بعد تكرار واحد. هذا أسرع بكثير من الانحدار التدريجي الدفعي. ولكن مرة أخرى، عندما يكون عدد أمثلة التدريب كبيرًا، فإنه يعالج مثالًا واحدًا فقط يمكن أن يكون عبئًا إضافيًا للنظام نظرًا لأن عدد التكرارات سيكون كبيرًا جدًا.
- الانحدار التدريجي للدفعة الصغيرة Mini Batch gradient descent: يعمل بشكل أسرع من كل من الانحدار التدريجي الدفعي والانحدار التدريجي العشوائي. هنا أمثلة b حيث تتم معالجة b<m لكل تكرار. لذا، حتى إذا كان عدد أمثلة التدريب كبيرًا، تتم معالجتها على دفعات من أمثلة التدريب b دفعة واحدة. وبالتالي، فهو يعمل مع أمثلة تدريبية أكبر وذلك أيضًا مع عدد أقل من التكرارات.

س338/ كيف تحل مشكلة انفجار الانحدار exploding gradient problem؟

ج / عندما تنمو أوزان النموذج بشكل أسي وتصبح كبيرة بشكل غير متوقع في النهاية عند تدريب النموذج، تحدث مشكلة انفجار الانحدار. في شبكة عصبية بها n طبقات مخفية، يتم ضرب مشتقات n معًا. إذا كانت الأوزان التي يتم ضربها أكبر من 1، فإن الانحدار التدريجي يزيد أضعافًا مضاعفة أكبر من المعتاد وينفجر في النهاية مع انتشار النموذج. الحالة التي تكون فيها قيمة الأوزان أكثر من 1 تجعل المخرجات أكبر بشكل كبير مما يعوق تدريب النموذج ويؤثر على الدقة الكلية للنموذج يشار إليه باسم مشكلة انفجار الانحدار exploding gradients. يعد انفجار التدرجات مشكلة خطيرة لأن النموذج لا يمكنه التعلم من بيانات التدريب الخاصة به مما يؤدي إلى خسارة ضعيفة. يمكن للمرء أن يتعامل مع مشكلة التدرج المتفجر إما عن

طريق اقتصاص التدرج gradient clipping، أو تنظيم الوزن weight regularization، أو باستخدام LSTM.

س339/ هل من الضروري خلط shuffle بيانات التدريب عند استخدام الانحدار التدريجي الدفعي؟

ج / لن يحدث خلط Shuffling مجموعة بيانات التدريب فرقًا كبيرًا لأن الانحدار التدريجي يتم حسابه في كل فترة باستخدام مجموعة بيانات التدريب الكاملة.

س340/ عند استخدام الانحدار التدريجي ذو الدفعات الصغيرة، ما سبب أهمية خلط البيانات بشكل عشوائي؟

ج / يجعل خلط Shuffling الدُفعات الصغيرة التدرجات أكثر تباينًا، مما قد يساعد في التقارب convergence لأنه يزيد من احتمالية الوصول إلى اتجاه جيد.

س341/ كيف تعرف ما إذا كان نموذجك يعاني من مشكلة انفجار الانحدار exploding gradient؟

ج / هناك بعض العلامات الدقيقة التي قد تدل على أنك تعاني من انفجار التدرجات (الانحدارات) أثناء تدريب شبكتك، مثل:

- النموذج غير قادر على الحصول على قوة جذب traction على بيانات التدريب الخاصة بك (على سبيل المثال، خسارة ضعيفة).
- النموذج غير مستقر، مما يؤدي إلى تغييرات كبيرة في الخسارة من التحديث إلى التحديث.
- تذهب خسارة النموذج إلى NaN أثناء التدريب.

إذا كانت لديك هذه الأنواع من المشاكل، فيمكنك البحث بشكل أعمق لمعرفة ما إذا كانت لديك مشكلة في انفجار التدرجات. هناك بعض العلامات الأقل دقة والتي يمكنك استخدامها لتأكيد وجود تدرجات متفجرة:

- سرعان ما تصبح أوزان النموذج كبيرة جدًا أثناء التدريب.
- تذهب أوزان النموذج إلى قيم NaN أثناء التدريب.
- تكون قيم تدرج الخطأ أعلى باستمرار من 1.0 لكل عقدة وطبقة أثناء التدريب.

س342/ كيف تعرف ما إذا كان نموذجك يعاني من مشكلة تلاشي الانحدار vanishing gradients؟

ج/

- ❖ سيتحسن النموذج ببطء شديد خلال مرحلة التدريب ومن الممكن أيضًا أن يتوقف التدريب مبكرًا جدًا، مما يعني أن أي تدريب إضافي لا يؤدي إلى تحسين النموذج.
- ❖ ستشهد الأوزان الأقرب إلى طبقة الإخراج للنموذج المزيد من التغيير في حين أن الطبقات التي تحدث بالقرب من طبقة الإدخال لن تتغير كثيرًا (إن وجدت).
- ❖ تتقلص أوزان النموذج بشكل كبير وتصبح صغيرة جدًا عند تدريب النموذج.
- ❖ تصبح أوزان النموذج 0 في مرحلة التدريب.

س343/ اشرح الانحدار التدريجي الدفعي Batch Gradient Descent؟

ج / في الانحدار التدريجي الدفعي Batch Gradient Descent، يتم أخذ جميع بيانات التدريب في الاعتبار لاتخاذ خطوة واحدة. نأخذ متوسط التدرجات لجميع أمثلة التدريب ثم نستخدم هذا متوسط التدرج لتحديث معلماتنا. إذن فهذه مجرد خطوة واحدة من الانحدار التدريجي في فترة واحدة.

س344/ ما هو الفرق بين الانحدار التدريجي العشوائي (SGD) والانحدار التدريجي (GD)؟

ج / بادئ ذي بدء، يعد كل من الانحدار التدريجي Gradient Descent (GD) والانحدار التدريجي العشوائي Stochastic Gradient Descent (SGD) كلاهما تعلمًا آليًا شائعًا وخوارزميات تحسين التعلم العميق التي تُستخدم لتحديث مجموعة من المعلمات بطريقة تكرارية لتقليل دالة الخطأ. في الانحدار التدريجي لتحديث المعلمات، يجب مراعاة مجموعة البيانات بأكملها لتكرار معين بينما في الانحدار التدريجي العشوائي، يتم إجراء الحساب على عينة تدريب واحدة فقط. على سبيل المثال، إذا كانت مجموعة البيانات تحتوي على 10000 نقطة بيانات، فسيتدرب GD على كل 10000 نقطة بيانات وسيستغرق ذلك وقتًا أطول، بينما من ناحية أخرى، سيكون SGD أسرع بكثير حيث سنقوم بالتدريب على عينة واحدة فقط وتحديث المعلمات. وذلك لأن الانحدار التدريجي العشوائي عادةً ما يتقارب بشكل أسرع من الانحدار التدريجي في مجموعات البيانات الكبيرة، لأن التحديثات تكون أكثر تكرارًا.

س345/ ما هو الانحدار التدريجي gradient descent؟

ج/ الانحدار التدريجي gradient descent هو إجراء تكراري يقلل من دالة التكلفة المحددة بواسطة معلمات النموذج. إنها طريقة تحسين تعتمد على دالة محدبة وتقليص المعلمات بشكل متكرر لمساعدة الدالة المعينة على بلوغ الحد الأدنى المحلي لها. الانحدار التدريجي يقيس التغيير في المعلمة فيما يتعلق بالتغيير في

الخطأ. تخيل شخصًا معصوب العينين على قمة تل ويريد الوصول إلى الارتفاع الأدنى. الأسلوب البسيط الذي يمكنه استخدامه هو الشعور بالأرض في كل اتجاه واتخاذ خطوة في الاتجاه حيث تنخفض الأرض بشكل أسرع. نحتاج هنا إلى مساعدة معدل التعلم learning rate الذي يوضح حجم الخطوة التي نتخذها للوصول إلى الحد الأدنى. يجب اختيار معدل التعلم بحيث لا يكون مرتفعًا جدًا أو منخفضًا جدًا. عندما يكون معدل التعلم المحدد مرتفعًا جدًا، فإنه يميل إلى الارتداد ذهابًا وإيابًا بين الدالة المحدبة للانحدار التدريجي، وعندما يكون منخفضًا جدًا، سنصل إلى الحد الأدنى ببطء شديد.

س346/ ما الأسباب التي تجعل الانحدار التدريجي للدفعات الصغيرة mini-batch gradient مفيدًا جدًا؟

ج/

- يعد الانحدار التدريجي للدفعة الصغيرة فعالاً للغاية مقارنةً بالانحدار التدريج العشوائي **stochastic gradient descent**.
- يتيح لك الوصول إلى التعميم **generalization** من خلال إيجاد الحد الأدنى المسطح **flat minima**.
- يساعد الانحدار التدريجي للدفعة الصغيرة على تجنب الحدود الدنيا المحلية **local minima** للسماح بتقريب التدرج لمجموعة البيانات بأكملها.

س347/ ما هو الانتشار الأمامي forward propagation؟

ج/ الانتشار الأمامي forward propagation هو السيناريو الذي يتم فيه تمرير المدخلات إلى الطبقة المخفية بالأوزان. في كل طبقة مخفية، يتم حساب ناتج دالة التنشيط حتى يمكن معالجة الطبقة التالية. يطلق عليه الانتشار الأمامي حيث تبدأ العملية من طبقة الإدخال وتتحرك نحو طبقة الإخراج النهائية.

س348/ ما هو الانتشار الخلفي backpropagation؟

ج/ يتم استخدام الانتشار الخلفي backpropagation لتقليل دالة التكلفة (الخطأ) من خلال رؤية كيف تتغير القيمة أولاً عندما يتم تعديل الأوزان والتحيزات في الشبكة العصبية. يتم حساب هذا التغيير بسهولة من خلال فهم الانحدار التدريجي في كل طبقة مخفية. يطلق عليه الانتشار الخلفي حيث تبدأ العملية من طبقة الإخراج، وتتحرك للخلف إلى طبقات الإدخال.

س349/ هل يمكنك تدريب شبكة عصبية دون استخدام الانتشار الخلفي backpropagation؟ إذا كانت الإجابة بنعم، فما التقنية التي ستستخدمها لتحقيق ذلك؟

ج/نعم، هناك العديد من خوارزميات التحسين التي لا تتطلب الانتشار الخلفي لتدريب الشبكة العصبية:

- في الشبكة العصبية، الانتشار الخلفي back propagation هو عملية الضبط المتكرر لأوزان الطبقات في الشبكة لتقليل الفرق بين المخرجات الفعلية actual output والمخرجات المرغوبة desired output، أي الخسارة.
- تؤدي هذه الأوزان المعدلة إلى جعل الوحدات المخفية للشبكة العصبية لتمثيل السمات الرئيسية للبيانات. هل هناك أي طرق أخرى لمواصلة العملية بدلاً من الانتشار الخلفي؟
- في الواقع، هناك العديد من خوارزميات التحسين التي لا تتطلب الانتشار الخلفي لتدريب الشبكة العصبية.
- من بينها خوارزميات التحسين التطوري evolutionary optimization وتوجيه الكبسولة لجيف هينتون Jeff Hinton's capsule routing. ومع ذلك، لا تظهر أي من هذه الطرق أداءً تنافسيًا ضد الخوارزميات القائمة على الانتشار الخلفي.

س350/ لماذا لا نرى مشكلة انفجار الانحدار أو تلاشي الانحدار في الشبكات العصبية ذات التغذية الأمامية FNN؟

ج / لان FNN لا تستخدم أسلوب الانتشار الخلفي backpropagation المسبب لمشكلة تلاشي وانفجار الانحدار التدريجي.

س351/ اشرح الانتشار الأمامي forwarding propagation والانتشار الخلفي backpropagation في سياق التعلم العميق؟

ج/

الانتشار الامامي forwarding propagation

كما يوحي الاسم، يتم تغذية بيانات الإدخال في الاتجاه الأمامي عبر الشبكة. تقبل كل طبقة مخفية بيانات الإدخال وتعالجها وفقًا لدالة التنشيط وتمريرها إلى الطبقة التالية.

من أجل توليد بعض المخرجات، يجب تغذية بيانات الإدخال في الاتجاه الأمامي فقط. يجب ألا تتدفق البيانات في الاتجاه العكسي أثناء توليد المخرجات وإلا فإنها

ستشكل دورة ولا يمكن أبدًا إنشاء الناتج. تُعرف تكوينات الشبكة هذه بشبكة التغذية الأمامية. تساعد شبكة التغذية الأمامية في الانتشار الأمامي.

الانتشار الخلفي backpropagation

الانتشار الخلفي هو جوهر تدريب الشبكة العصبية. إنها ممارسة الضبط الدقيق fine-tuning لأوزان الشبكة العصبية بناءً على معدل الخطأ (أي الخسارة) التي تم الحصول عليها في الفترة السابقة (أي التكرار). يضمن الضبط الصحيح للأوزان معدلات خطأ أقل، مما يجعل النموذج موثوقًا من خلال زيادة تعميمه.

لدينا نموذج لا يعطي تنبؤات دقيقة ويعزى ذلك إلى حقيقة أن أوزانه لم يتم ضبطها بعد. نحن أيضا لدينا الخسارة. يدور الانتشار الخلفي حول تغذية هذه الخسارة بشكل عكسي بطريقة يمكننا من خلالها ضبط الأوزان بناءً على أي منها. ستساعدنا دالة التحسين مثل الانحدار التدريجي gradient descent في العثور على الأوزان التي - نأمل - تسفر عن خسارة أقل في التكرار التالي.

س352/ اشرح تسوية البيانات Data Normalization؟ ما هي الحاجة لذلك؟

ج / تسمى عملية توحيد البيانات وإصلاحها بـ "تسوية البيانات Data Normalization". إنها خطوة معالجة مسبقة للتخلص من تكرار البيانات. في كثير من الأحيان، تأتي البيانات وتحصل على نفس المعلومات بتنسيقات مختلفة. في هذه الحالات، يجب إعادة قياس القيم لتلائم نطاقًا معينًا، وتحقيق تقارب أفضل.

س353/ ما هي التقنيات المختلفة لتحقيق تسوية البيانات data normalization؟

ج / هناك أربع تقنيات شائعة لتسوية البيانات **data normalization**:

- scaling to a range
- clipping
- log scaling
- z-score

س354/ هل توجد أي مشاكل عند استخدام التسوية بالدفعات Batch Normalization في الشبكات العصبية العميقة؟

ج / تركز Batch Normalization (BN) على توحيد المدخلات إلى أي طبقة معينة (أي عمليات التنشيط من الطبقات السابقة). يعني توحيد المدخلات أن المدخلات إلى أي طبقة في الشبكة يجب أن يكون لها متوسط صفر تقريبًا وتباين واحد، ولكن كل طبقة لا تحتاج إلى توقع مدخلات بمتوسط صفري وتباين واحد، بدلاً من ذلك، ربما

يكون أداء النموذج أفضل مع بعض الوسائل الأخرى والتباين. ومن ثم تقدم طبقة BN أيضًا معلمتين قابلتين للتعلم γ و β.

في هذه العملية، يحسب BN متوسط الدُفعة المصغرة والتباين) في كل تكرار تدريب، وبالتالي يتطلب أحجام دُفعات أكبر أثناء التدريب بحيث يمكنه تقريب متوسط المحتوى والتباين من الدُفعة المصغرة بشكل فعال. هذا يجعل BN أكثر صعوبة لتدريب الشبكات على تطبيقات مثل اكتشاف الكائنات object detection، والتجزئة الدلالية semantic segmentation، وما إلى ذلك لأنها تعمل عمومًا بدقة عالية للمدخلات (غالبًا ما تصل إلى 1024 × 2048) والتدريب بأحجام دفعات أكبر ليس ممكنًا من الناحية الحسابية. علاوة على ذلك، أثناء وقت الاختبار (أو الاستدلال)، لا تحسب طبقة BN المتوسط والتباين من الدفعة المصغرة لبيانات الاختبار ولكنها تستخدم المتوسط الثابت والتباين المحسوب من بيانات التدريب. هذا يتطلب الحذر أثناء استخدام BN ويدخل تعقيدًا إضافيًا.

س355/ لماذا يجب أن نستخدم التسوية بالدفعات Batch Normalization؟

ج/ التسوية بالدفعات Batch Normalization هو أسلوب لتدريب الشبكات العصبية العميقة جدًا التي تقوم بتوحيد المدخلات إلى طبقة لكل دفعة صغيرة. هذا له تأثير في تثبيت عملية التعلم وتقليل عدد فترات التدريب المطلوبة بشكل كبير لتدريب الشبكات العميقة.

الميزة الأخرى للتسوية بالدفعات هي أنها تعمل بمثابة تسوية regularization. يتم قياس كل دفعة صغيرة mini-batch باستخدام متوسطها وانحرافها المعياري. يقدم هذا بعض الضوضاء لكل طبقة، مما يوفر تأثيرًا تنظيمًا.

س356/ ما هو التنظيم regularization وأين يمكن أن يكون مفيدًا؟ ما هو مثال على استخدام التنظيم في نموذج؟

ج / يعد التنظيم **regularization** مفيدًا لتقليل التباين **variance** في النموذج، مما يعني تجنب فرط التعلم **Overfitting**. على سبيل المثال، يمكننا استخدام تنظيم **L1** في انحدار **Lasso** لمعاقبة **penalize** المعاملات الكبيرة.

س357/ متى يجب تفضيل انحدار ridge على lasso؟

ج / يكون انحدار ridge أكثر ملاءمة عندما تحتوي مجموعة البيانات على عدد أكبر من متغيرات التنبؤ predictor variables من عدد الملاحظات number of observations. ثاني أفضل سيناريو هو عندما يتم اختبار العلاقة الخطية المتعددة multicollinearity في مجموعة.

س358/ اشرح الفرق بين Lasso وRidge؟

ج/

Lasso	Ridge
يُطلق على نموذج الانحدار الذي يستخدم عملية تنظيم المستوى 1 اسم Lasso Regression.	يُطلق على نموذج الانحدار الذي يستخدم عملية تنظيم المستوى 2 اسم Ridge Regression.
يضيف Lasso Regression القيمة المطلقة لحجم المعامل كمصطلح جزائي إلى دالة الخسارة.	يضيف Ridge Regression الحجم التربيعي للمعامل كمصطلح جزائي إلى دالة الخسارة.
يحاول تقدير الوسيط median للبيانات.	يحاول تقدير المتوسط mean للبيانات.

س359/ ما الفرق بين التنظيم regularization والتسوية normalization؟

ج/

- يهدف التنظيم regularization إلى حل مشكلة الضبط الزائد. من خلال إضافة جزء إضافي إلى دالة الخسارة، من المرجح أن تتقارب المعلمات في خوارزميات التعلم مع قيم أصغر، مما قد يقلل بشكل كبير من الضبط الزائد.
- التسوية normalization عادة ما يعيد قياس الميزات إلى $[0,1]$

س360/ اشرح الفرق بين التسوية Normalization والتوحيد Standardization؟

ج/

Standardization	Normalization
تقنية تحويل البيانات بطريقة يتم توزيعها بشكل طبيعي ولها انحراف معياري قدره 1 ومتوسط 0.	تُعرف تقنية تحويل جميع قيم البيانات لتقع بين 1 و 0 بالتسوية. يُعرف هذا أيضًا باسم مقياس min-max.
التوحيد القياسي يعني أن التوزيع الطبيعي القياسي متبوع بالبيانات.	تتم العناية بالبيانات التي تعود إلى النطاق من 0 إلى 1 بواسطة التطبيع.
صيغة التوحيد: $X' = (X - \mu) / \sigma$	صيغة التسوية: $X' = (X - Xmin) / (Xmax - Xmin)$ Xmin : الحد الأدنى لقيمة الميزة، Xmax : القيمة القصوى للميزة.

س361/ ما هي مزايا استخدام Batchnorm؟

ج/ اهم مزايا التسوية بالدفعات batch normalization هي:

- النموذج أقل حساسية لضبط المعلمة الفائقة.
- تصبح معدلات التعلم العالية مقبولة، مما يؤدي إلى تدريب أسرع للنموذج.
- تصبح تهيئة الوزن مهمة سهلة.
- يصبح استخدام دوال التنشيط غير الخطية المختلفة أمرًا ممكنًا.
- تم تبسيط الشبكات العصبية العميقة بسبب التسوية بالدفعات.
- يقدم تنظيمًا معتدلا mild regularisation في الشبكة.

س362/ ما هي فوائد استخدام التسوية بالدفعات batch normalization عند تدريب شبكة عصبية؟

ج/

- يعمل التسوية بالدُفعات batch normalization على تحسين عملية التدريب على الشبكة مما يجعل من السهل بناءها وتدريبها بشكل أسرع على شبكة عصبية عميقة.
- تنظم التسوية بالدُفعات القيم التي تدخل في كل دالة تنشيط مما يجعل دوال التنشيط أكثر قابلية للتطبيق لأن العناصر غير الخطية التي لا يبدو أنها تعمل بشكل جيد تصبح قابلة للتطبيق باستخدام التسوية بالدفعات.
- تسهل التسوية بالدُفعات تهيئة الأوزان وتسمح أيضًا باستخدام معدلات تعليم أعلى مما يؤدي في النهاية إلى زيادة السرعة التي تتدرب بها الشبكة.

س363/ ما هي إعادة تحجيم البيانات Data Rescaling وكيف يتم ذلك؟

ج/ في سيناريوهات العالم الحقيقي، تكون السمات الموجودة في البيانات في نمط متغير. لذا، فإن إعادة تحجيم Rescaling الخصائص إلى مقياس مشترك مفيد للخوارزميات لمعالجة البيانات بكفاءة.

يمكننا إعادة قياس البيانات باستخدام Scikit-Learn عن طريق استخدام الامر MinMaxScaler.

س364/ كيف يتم توحيد البيانات Data Standardization؟

ج/ التوحيد القياسي للبيانات Data Standardization هو الطريقة المستخدمة لإعادة قياس سمات البيانات. من المحتمل أن يكون للسمات قيمة متوسطة تبلغ 0 وقيمة الانحراف المعياري 1. الهدف الرئيسي للتوحيد القياسي هو دفع المتوسط والانحراف المعياري للسمات.

يمكن توحيد البيانات باستخدام Scikit-Learn وذلك باستخدام الامر StandardScaler.

س365/ لماذا التسوية normalization مهمة؟

ج / يمكن أن تحتوي مجموعة البيانات على عمود واحد في النطاق (20000/10000) وقد يحتوي العمود الآخر على بيانات في النطاق (1 ، 2 ، 3). من الواضح أن هذين العمودين في نطاق مختلف ولا يمكنهما تحليل الاتجاه بدقة. لذلك يمكننا تطبيق التسوية normalization هنا باستخدام min-max normalization (أي لتحويله إلى مقياس 0-1).

س366/ ما هي المعلمات الفائقة hyperparameters في التعلم العميق؟

ج/ المعلمات الفائقة hyperparameters هي المتغيرات المستخدمة لتحديد بنية الشبكة العصبية. يتم استخدامها أيضًا لفهم المعلمات، مثل معدل التعلم وعدد الطبقات المخفية، وأكثر من ذلك، الموجودة في الشبكة العصبية.

س367/ كيف يمكن تدريب المعلمات الفائقة في الشبكات العصبية؟

ج/ يمكن تدريب المعلمات الفائقة باستخدام أربع مكونات كما هو موضح أدناه:

- حجم الدُفعة (Batch size): يستخدم للإشارة إلى حجم مقطع الإدخال. يمكن تغيير أحجام الدُفعات وتقطيعها إلى دفعات فرعية بناءً على المتطلبات.
- الفترات (Epochs): تشير الفترة إلى عدد المرات التي تكون فيها بيانات التدريب مرئية للشبكة العصبية حتى تتمكن من التدريب. نظرًا لأن العملية تكرارية، سيختلف عدد الفترات بناءً على البيانات.
- الزخم (Momentum): يستخدم الزخم لفهم الخطوات التالية المتتالية التي تحدث مع تنفيذ البيانات الحالية في متناول اليد. يتم استخدامه لتجنب التذبذبات oscillations عند التدريب.
- معدل التعلم (Learning rate): يُستخدم معدل التعلم كمعامل للإشارة إلى الوقت المطلوب للشبكة لتحديث المعلمات والتعلم.

س368/ هل يمكنك تسمية بعض المعلمات الفائقة hyperparameters المستخدمة لتدريب الشبكة العصبية؟

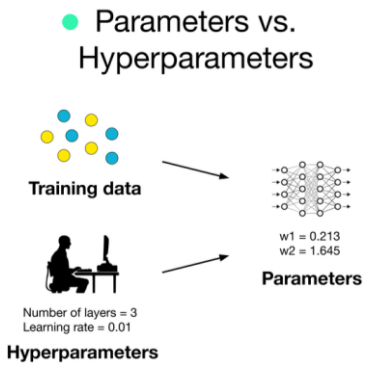
ج / عند تدريب أي شبكة عصبية، يوجد نوعان من المعلمات الفائقة: أحدهما يحدد بنية الشبكة العصبية والآخر يحدد كيفية تدريب الشبكة العصبية. المدرجة هي بعض المعلمات الفائقة التي تم تعيينها قبل تدريب أي شبكة عصبية:

- تهيئة الأوزان Initialization of weights
- تحديد عدد الطبقات المخفية Setting the number of hidden layers
- معدل التعليم Learning Rate
- عدد الفترات Number of epochs
- دوال التنشيط Activation Functions
- حجم الدفعة Batch Size
- الزخم Momentum

س369/ ما هي المعلمات الفائقة hyperparameters وكيف تختلف عن المعلمات parameters؟

- ج/

المعلمات الفائقة hyperparameters	المعلماتparameters
المعلمات الفائقة هي المعلمات المحددة صراحة والتي تتحكم في عملية التدريب.	المعلمات هي نموذج التكوين، وهي داخلية في النموذج.
تعد المعلمات الفائقة ضرورية لتحسين النموذج.	المعلمات ضرورية لعمل التنبؤات.
مستقلة عن مجموعة البيانات.	تعتمد على مجموعة البيانات، والتي تستخدم للتدريب.
يمكن تقدير قيم المعلمات الفائقة عن طريق ضبط المعلمات الفائقة.	يمكن تقدير قيم المعلمات بواسطة خوارزميات التحسين، مثل Gradient Descent.
بعض الأمثلة على معلمات النموذج هي الأوزان في ANN أو متجهات الدعم في المعاملات في الانحدار أو الخطي أو الانحدار اللوجستي.	بعض الأمثلة على معلمات النموذج هي الأوزان في ANN أو متجهات الدعم في SVM أو المعاملات في الانحدار الخطي أو الانحدار اللوجستي.



س370/ ما هي المعلمات الفائقة لنموذج الانحدار اللوجستي؟

ج / عقوبة المصنف **Classifier penalty**، وحلول المصنف **classifier solver**، و**classifier C** هي المعلمات الفائقة القابلة للتدريب لمصنف الانحدار اللوجستي. يمكن تحديد هذه القيم حصريًا في **Grid Search** لضبط المصنف اللوجستي بشكل مفرط.

س371/ ما هي دوال التنشيط activation functions؟

ج/ دوال التنشيط هي كيانات في التعلم العميق تُستخدم لترجمة المدخلات إلى معلمة إخراج قابلة للاستخدام. إنها دالة تحدد ما إذا كانت الخلية العصبية بحاجة إلى التنشيط أم لا عن طريق حساب مجموع الاوزان عليها مع التحيز.

يؤدي استخدام دالة التنشيط إلى جعل إخراج النموذج غير خطي. هناك أنواع عديدة من دوال التنشيط:

- ReLU
- Softmax
- Sigmoid
- Linear
- Tanh

س372/ كيف تختار دالة التنشيط Activation Function لنموذج التعلم العميق؟

ج/

- إذا كان الناتج المراد توقعه حقيقيًا، فمن المنطقي استخدام دالة التنشيط Linear.
- إذا كان الإخراج المتوقع هو احتمال لفئة ثنائية، فيجب استخدام دالة Sigmoid.
- إذا كان الإخراج المتوقع له فئتان، فيمكن استخدام دالة Tanh.
- يمكن استخدام دالة ReLU في العديد من الحالات المختلفة نظرًا لبساطتها الحسابية.

س373/ ماذا يحدث إذا استخدمنا دالة التنشيط ReLU ثم sigmoid كطبقة نهائية؟

ج/ ستتنبأ الشبكة العصبية بفئة واحدة فقط لجميع أنواع المدخلات لأن إخراج دالة التنشيط ReLu دائمًا ما يكون نتيجة غير سلبية.

س374/ ماذا يحدث إذا لم تستخدم أي دوال تنشيط في الشبكة العصبية؟

ج/ الشبكة العصبية بدون دالة التنشيط هي في الأساس مجرد نموذج انحدار خطي. تقوم دالة التنشيط بالتحويل غير الخطي إلى المدخلات مما يجعلها قادرة على التعلم وتنفيذ مهام أكثر تعقيدًا.

س375/ ماهي دالة التنشيط Activation Function؟

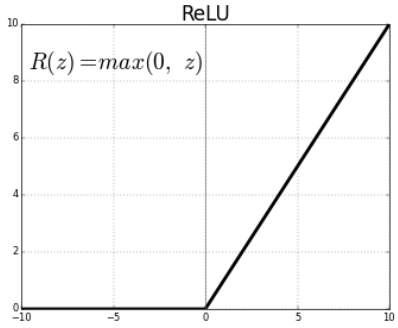
ج/ تطبق دالة التنشيط قاعدة الخطوة (تحويل الناتج العددي إلى +1 أو -1) للتحقق مما إذا كان ناتج دالة الاوزان أكبر من الصفر أم لا.

تحدد دالة التنشيط للعقدة إخراج العقدة بإدخال أو مجموعة من المدخلات إلى العقدة.

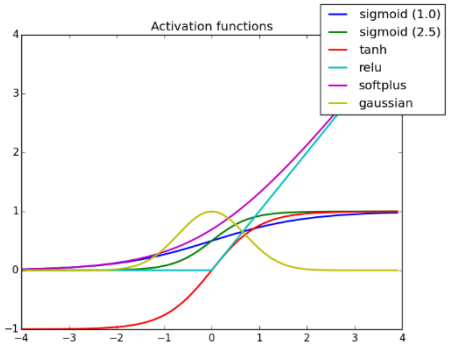
يمكن تقسيم دوال التنشيط إلى ثلاث فئات:

- دوال ريدج Ridge functions
- دوال شعاعية Radial functions
- دوال الطي Fold functions

يظهر نوع من دالة ريدج تسمى الدالة الخطية المصححة (ReLU) أدناه:



يظهر نوع من الدوال الشعاعية تسمى الدالة الغاوسية أدناه:



س376/ ما هو الفرق بين دالة التنشيط الخطية ودالة التنشيط غير الخطية؟

ج / يتم تمثيل دالة التنشيط الخطي بالمعادلة التالية:

$$A = cx$$

تأخذ دالة التنشيط الخطي المدخلات، مضروبة بالأوزان لكل خلية عصبية، وتخلق إشارة خرج تتناسب مع المدخلات.

تستخدم دالة التنشيط غير الخطي دوال التنشيط غير الخطية. يقومون بإنشاء تعيينات معقدة بين مدخلات ومخرجات الشبكات. من المهم إنشاء تعيينات معقدة لأن الصور والصوت والفيديو غير خطية non-linear أو ذات أبعاد عالية high dimensionality.

س377/ ما هي خوارزميات التحسين optimization algorithms المختلفة التي نستخدمها بشكل عام في الشبكة العصبية؟

ج / **Optimizers** عبارة عن خوارزميات أو طرق مستخدمة لتغيير سمات شبكتك العصبية مثل الأوزان ومعدل التعلم من أجل تقليل الأخطاء . هناك أنواع مختلفة من **Optimizers**.

س378/ ماذا تقصد بـ RMS Prop؟

ج / **RMS Prop** هي تقنية تحسين **optimization technique** في الشبكات العصبية. للتعرف على **RMS Prop**، نحتاج إلى معرفة **R prop**. يتم استخدام خوارزمية **R Prop** لتحسين الدفعة الكاملة. يحاول حل مشكلة التدرجات ذات الأحجام المتغيرة.

س379/ ما هو الفرق بين Adagrad و Adadelta و Adam؟

ج/

❖ **Adagrad:**

يقيس **Adagrad** ألفا (معدل التعلم) لكل معلمة وفقًا لتاريخ التدرجات (الخطوات السابقة) لتلك المعلمة والتي تتم بشكل أساسي عن طريق قسمة الانحدار التدريجي الحالي في قاعدة التحديث على مجموع التدرجات السابقة. نتيجة لذلك، ما يحدث هو أنه عندما يكون التدرج كبيرًا جدًا، يتم تقليل ألفا والعكس صحيح.

❖ **AdaDelta:**

Adadelta هو امتداد أكثر قوة لـ **Adagrad** يقوم بتكييف معدلات التعلم بناءً على نافذة متحركة لتحديثات التدرج ، بدلاً من تجميع جميع التدرجات السابقة. بهذه الطريقة ، يستمر **Adadelta** في التعلم حتى بعد إجراء العديد من التحديثات.

❖ **Adam:**

Adam عبارة عن خوارزمية تحسين بديلة للانحدار التدريجي العشوائي لتدريب نماذج التعلم العميق. يجمع Adam بين أفضل خصائص خوارزميات AdaGrad و RMSProp لتوفير خوارزمية تحسين يمكنها التعامل مع التدرجات المتفرقة في المشكلات الصاخبة..

س380/ محسن RMSProp و Adam يضبطان التدرجات؟ هل هذا يعني أنهم يؤدون gradient clipping؟

ج / لا يعني هذا بطبيعته أنهم يقومون باقتصاص التدرج **gradient clipping** لأن اقتصاص التدرج يتضمن إعداد قيم محددة مسبقًا لا يمكن للتدرجات أن تتجاوزها، على عكس **Adam** و**RMS Prop** اللذان يقومان بإجراء تعديلات مضاعفة على التدرجات.

س381/ لماذا نستخدم بشكل عام دالة Softmax غير الخطية كعملية أخيرة في الشبكة؟

ج / تعمل دالة Softmax تقريبًا مثل الطبقة القصوى max layer التي يتم إخراجها إما 0 أو 1 لعقدة إخراج واحدة. من الممكن أيضًا التفاضل في التدريب عن طريق الانحدار التدريجي gradient descent. سيكون جمع كل المخرجات دائمًا مساويًا لـ 1. سيكون للقيمة العالية للإنتاج احتمالية أعلى من غيرها 1.

س382/ ما الفرق بين دوال Sigmoid و Softmax؟

ج /تختلف دوال Sigmoid و Softmax بناءً على استخدامها في تصنيف مهام التعلم الآلي. تستخدم دالة Sigmoid في حالة التصنيف الثنائي binary classification، بينما تستخدم دالة Softmax في حالة التصنيف المتعدد multi-classification.

س283/ أيهما تعتقد أنه أقوى: شبكات عصبية من طبقتين بدون أي دالة تنشيط أو شجرة قرار من طبقتين؟

ج/

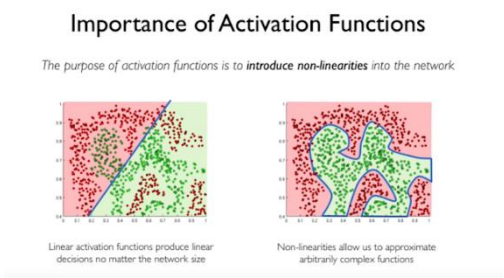
- عندما تقول شبكة عصبية من طبقتين ، فإنها تحتوي أساسًا على طبقة إدخال واحدة وطبقة مخفية واحدة وطبقة إخراج واحدة. تعد دالة التنشيط مهمة أثناء التعامل مع الشبكات العصبية لأنها ضرورية أثناء التعامل مع التعيينات الدالة المعقدة وغير الخطية بين المدخلات ومتغير الاستجابة.
- عندما لا تحتوي الشبكة العصبية المكونة من طبقتين على دالة تنشيط ، فهي مجرد شبكة خطية. الشبكة العصبية بدون دالة التنشيط ستكون ببساطة نموذجًا للانحدار الخطي ، والذي يتمتع بقوة محدودة ولا يؤدي أداءً جيدًا في معظم الأوقات.
- شجرة القرار المكونة من طبقتين هي مجرد شجرة قرارات بعمق 2.
- لذلك ، أثناء المقارنة بين هذين النموذجين ، تكون الشبكة العصبية المكونة من طبقتين (بدون دالة التنشيط) أقوى من شجرة القرار المكونة من طبقتين ، نظرًا لأن الشبكة العصبية المكونة من طبقتين ستأخذ المزيد من السمات في الاعتبار أثناء بناء نموذج وفي حالة شجرة قرار من طبقتين ، سيتم النظر في 2 أو 3 سمات فقط.

س384/ ما هي أهمية استخدام دالة التنشيط Non-linear Activation function غير الخطية؟

ج/ الشبكات العصبية ذات التنشيط الخطي فقط لا تكسب من زيادة عدد الطبقات فيها بسبب حقيقة أن جميع الدوال الخطية تضيف ما يصل إلى دالة خطية واحدة. لذلك، إذا كان هناك العديد من الطبقات في الشبكة ذات دوال التنشيط الخطي فقط، فإن الأمر يبدو كما لو أن هناك طبقة واحدة فقط only one layer.

تسمح دوال التنشيط غير الخطي بتكديس stacking طبقات مختلفة ولن يتم التعامل معها كطبقة واحدة كما هو الحال في طبقة التنشيط الخطي.

اشتقاق دالة خطية هو ثابت (ليس له علاقة بالمدخلات)، لذلك لا يمكن استخدام الانتشار الخلفي backpropagation عندما يتعلق الأمر بالدوال الخطية. تسمح الدوال غير الخطية بالانتشار الخلفي نظرًا لحقيقة أنه يمكن التمييز بينها، ومشتقاتها مرتبطة بالمدخلات.



س385/ لماذا من المهم إدخال اللاخطية non-linearities في الشبكة العصبية؟

ج / بدون اللاخطية non-linearities، ستعمل الشبكة العصبية كـ perceptron بغض النظر عن عدد الطبقات الموجودة مما يجعل الإخراج يعتمد بشكل خطي على المدخلات. بعبارة أخرى، فإن وجود شبكة عصبية بها طبقات n ووحدات مخفية مع دوال التنشيط الخطي يشبه تمامًا وجود شبكة عصبية خطية بدون طبقات مخفية يمكنها فقط العثور على حدود الفصل الخطي. لا يمكن للشبكة العصبية بدون اللاخطية إيجاد الحلول المناسبة وتصنيف البيانات بشكل صحيح للمشكلات المعقدة.

س386/ لماذا لا يُفضل استخدام Sigmoid أو Tanh كدالة تنشيط في الطبقة المخفية للشبكة العصبية؟

ج / مشكلة شائعة في دوال Tanh أو Sigmoid هي أنها تشبع saturate. بمجرد التشبع، لا يمكن لخوارزميات التعلم التكيف مع الأوزان وتحسين أداء النموذج. وبالتالي، فإن دوال Sigmoid أو Tanh تمنع الشبكة العصبية من التعلم بشكل فعال مما يؤدي إلى مشكلة تلاشي الانحدار التدريجي vanishing gradient. يمكن معالجة مشكلة تلاشي الانحدار التدريجي باستخدام دالة التنشيط الخطي المصحح (ReLu) بدلاً من Sigmoid واستخدام تهيئة Xavier.

س387/ لماذا تُستخدم دالة Leaky ReLU في التعلم العميق؟

ج / Leaky ReLU هو نسخة متقدمة من دالة تنشيط ReLU. بشكل عام، تحدد دالة ReLU التدرج ليكون 0 عندما تكون جميع قيم المدخلات أقل من الصفر. هذا يعطل الخلايا العصبية. للتغلب على هذه المشكلة، يتم استخدام دوال تنشيط Leaky ReLU. منحدر صغير جدًا very small slope للقيم السالبة بدلاً من منحدر مستو flat slope.

يستخدم Leaky ReLU ، المعروف أيضًا باسم LReL ، لإدارة دالة للسماح بتمرير قيم سالبة صغيرة الحجم إذا كانت قيمة الإدخال إلى الشبكة أقل من الصفر.

س388/ ما هي دوال التنشيط التي استخدمتها حتى الآن في مشاريعك وكيف تختار واحدة.

ج/دوال التنشيط الأكثر شيوعًا هي:

- Sigmoid
- Tanh
- ReLU
- Softmax

س389/ ما مدى فائدة الحذف العشوائي dropoutفي نماذج التعلم العميق؟ هل يسرع أم يبطئ من عملية التدريب ولماذا؟

ج/تُثبت طريقة الحذف العشوائي dropout غالبًا أنها مفيدة للحالات التي تكون فيها مجموعة البيانات صغيرة، ومن المحتمل أن تزداد الشبكة العصبية العميقة أثناء التدريب. يجب مراعاة العامل الحسابي لمجموعات البيانات الكبيرة، والتي قد تفوق فائدة الحذف العشوائي.

تتضمن طريقة الحذف العشوائي الإزالة العشوائية لطبقة من شبكة عصبية عميقة، مما يسرع عملية التدريب.

س390/ اشرح لماذا يعمل الحذف العشوائي dropout في الشبكة العصبية كمنظم regularizer؟

ج / الحذف العشوائي dropout هو أسلوب تنظيم لتقليل الضبط الزائد overfitting في الشبكات العصبية عن طريق منع التكيفات المشتركة المعقدة co- complex adaptations على بيانات التدريب. إنها طريقة فعالة للغاية لأداء نموذج متوسط مع الشبكات العصبية. يشير مصطلح "dropout" إلى الحذف العشوائي الوحدات (المخفية والمرئية) في الشبكة العصبية.

س391/ ما الفرق بين DropoutوDrop Connect؟

ج / يطبق الحذف العشوائي **Dropout** قناعًا على عمليات التنشيط، بينما يطبق **DropConnect** قناعًا على الأوزان.

أن**DropConnect** هو تعميم لـ **Dropout** حيث يمكن حذف كل اتصال **connection** ، بدلاً من كل وحدة إخراج **output unit** كما في **Dropout** ، مع احتمال **p**.

س392/ هل يمكننا استخدام الحذف العشوائي dropout في مجموعة الاختبار test set ؟

ج / كلا، لا يمكننا استخدام الحذف العشوائي في مجموعة الاختبار وهناك سببان رئيسيان لعدم استخدام الحذف العشوائي dropout لاختبار البيانات:

- الحذف العشوائي يجعل الخلايا العصبية تنتج قيمًا "خاطئة" عن قصد
- نظرًا لتعطيل الخلايا العصبية بشكل عشوائي، سيكون لشبكتك مخرجات مختلفة لكل (تسلسل) تنشيط. هذا يقوض الاتساق consistency.

س393/ ما معنى الحذف العشوائي dropout في التعلم العميق؟

ج/الحذف العشوائي dropout هو أسلوب يستخدم لتجنب الضبط الزائد overfitting في التعلم العميق. إذا كانت قيمة الحذف العشوائي منخفضة جدًا، فسيكون لها تأثير ضئيل على التعلم. إذا كانت عالية جدًا، فيمكن للنموذج أن يتعلم بشكل ناقص، مما يؤدي إلى انخفاض الكفاءة.

س394/ ما هو التوقف المبكر Early stopping في التعلم العميق؟

ج / التوقف المبكر Early stopping في التعلم العميق هو نوع من التنظيم regularization حيث يتم إيقاف التدريب بعد عدة تكرارات.

عند تدريب شبكة كبيرة، ستكون هناك نقطة أثناء التدريب عندما يتوقف النموذج عن التعميم generalizing ويبدأ في تعلم الضوضاء الإحصائية في مجموعة بيانات التدريب. هذا يجعل الشبكات غير قادرة على التنبؤ بالبيانات الجديدة.

إن تحديد التوقف المبكر في الشبكة العصبية سيمنع الشبكة من الضبط الزائد overfitting.

تتمثل إحدى طرق تحديد التوقف المبكر في البدء في تدريب النموذج وإذا بدأ أداء النموذج في التدهور degrade، فقم بإيقاف عملية التدريب.

س395/ ما هو التنظيم Regularization وما نوع المشاكل التي يحلها التنظيم؟

ج / التنظيم Regularization هو في الأساس تقنية تُستخدم لدفع أو تشجيع معاملات نموذج التعلم الآلي نحو الصفر لتقليل مشكلة الضبط الزائد overfitting. الفكرة العامة للتنظيم هي معاقبة النماذج المعقدة عن طريق إضافة عقوبة إضافية لدالة الخسارة من أجل توليد خسارة أكبر. بهذه الطريقة، يمكننا تثبيط النموذج عن تعلم الكثير من التفاصيل ويكون النموذج أكثر عمومية.

هناك طريقتان لتعيين مدة العقوبة الإضافية لدالة الخسارة مما يؤدي إلى نوعين من تقنيات التنظيم:

- ❖ تنظيم L2
- ❖ تنظيم L1

في تنظيم L2، يكون مصطلح العقوبة هو مجموع مربعات حجم معاملات النموذج بينما في تنظيم L1، يكون مجموع القيم المطلقة لمعاملات النموذج.

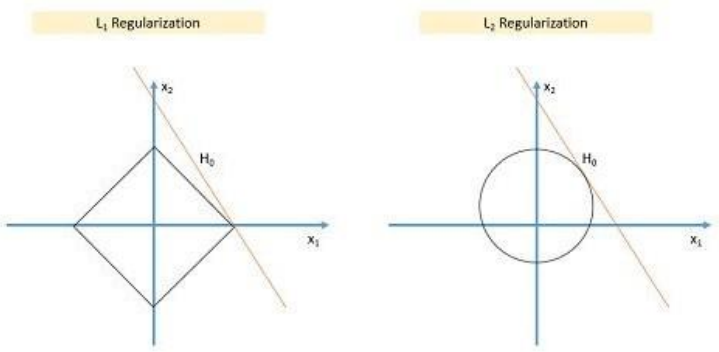
س396/ ما هو التنظيم Regularization وما نوع المشاكل التي يحلها التنظيم؟

ج / إنها عملية إضافة معلمة ضبط إلى نموذج ما للحث على السلاسة من أجل منع الضبط الزائد overfitting أو حل مشكلة غير صحيحة.

إنه يعمل عن طريق إضافة ثابت لمضاعفة متجه الوزن الحالي، حيث يكون الثابت غالبًا إما (Lasso) L1 أو L2 (Ridge)، أو أي معيار. ينتج عن هذا تنبؤات النموذج التي تقلل (تقلص) متوسط دالة الخسارة المحسوبة على مجموعة التدريب المنتظمة.

س397/ لماذا يتسبب تنظيم L1 في تباين المعلمات بينما لا يؤدي تنظيم L2 إلى ذلك؟

ج /يتم استخدام التنظيم Regularizations في الإحصائيات أو في مجال التعلم الآلي لتضمين بعض المعلومات الإضافية من أجل حل مشكلة بطريقة أفضل. تُستخدم تنظيم L1 & L2 عمومًا لإضافة قيود على مشكلات التحسين.



في المثال الموضح أعلاه H0 هي فرضية. إذا لاحظت، في L1، هناك احتمال كبير لضرب الزوايا كحلول بينما في L2، لا يحدث ذلك. لذلك في L1 يتم معاقبة المتغيرات أكثر مقارنة بـ L2 مما يؤدي إلى التناثر sparsity. بمعنى آخر، يتم تربيع الأخطاء في L2، لذلك يرى النموذج خطأ أعلى ويحاول تقليل هذا الخطأ التربيعي.

س398/ صف كيف يمكن تفسير تنظيم L2 كنوع من تناقص الوزن weight decay؟

ج/ غالبًا ما يشار إلى تنظيم L2 باسم تناقص الوزن weight decay لأنه يجعل الأوزان أصغر. تُعرف أيضًا باسم انحدار Ridge وهي تقنية يتم فيها إضافة مجموع المعلمات التربيعية أو أوزان نموذج (مضروبًا في بعض المعامل) إلى دالة الخسارة كمصطلح جزائي يجب تقليله.

س399/ إذا كان لدينا خطأ تحيز كبير high bias error ماذا يعني ذلك؟ كيف نعالجها؟

ج/ يعني خطأ التحيز المرتفع high bias error أن هذا النموذج الذي نستخدمه يتجاهل جميع الاتجاهات المهمة في النموذج وأن النموذج يعاني من مشكلة الضبط الناقص underfitting.

لتقليل الضبط الناقص underfitting:

- ❖ نحن بحاجة إلى زيادة تعقيد النموذج
- ❖ يجب زيادة عدد الميزات

في بعض الأحيان، يعطي أيضًا انطباعًا بأن البيانات صاخبة noisy. ومن ثم يجب إزالة الضوضاء من البيانات بحيث يتم العثور على أهم الإشارات بواسطة النموذج لعمل تنبؤات فعالة.

تؤدي زيادة عدد الفترات إلى زيادة مدة تدريب النموذج. إنه مفيد في تقليل الخطأ.

س400/ كيف يمكننا معرفة ما إذا كانت بياناتك تعاني من انحياز منخفض low bias وتباين كبير high variance؟

ج/ ببساطة، التباين هو التباين في تنبؤ النموذج - إلى أي مدى يمكن لوظيفة ML ضبطها اعتمادًا على مجموعة البيانات المحددة. يأتي التباين من نماذج شديدة التعقيد مع عدد كبير من الميزات. النماذج ذات التحيز العالي سيكون لها تباين منخفض. النماذج ذات التباين العالي سيكون لها انحياز منخفض.

س401/ لنفترض أنك وجدت أن نموذجك يعاني من تباين كبير high variance. ما الخوارزمية التي تعتقد أنها يمكنها التعامل مع هذا الموقف ولماذا؟

ج/

- للتعامل مع مشكلات التباين العالي، يجب أن نستخدم خوارزمية التعبئة **bagging.**
- ستقوم خوارزمية التعبئة بتقسيم البيانات إلى مجموعات فرعية مع أخذ عينات مكررة من البيانات العشوائية.
- بمجرد تقسيم الخوارزمية للبيانات، يمكننا استخدام بيانات عشوائية لإنشاء قواعد باستخدام خوارزمية تدريب معينة.
- بعد ذلك، يمكننا استخدام الاقتراع **polling** للجمع بين تنبؤات النموذج.

س402/ اشرح سبب أهمية عملية التهيئة للأوزان والتحيزات للشبكات العصبية (Neural Networks (NN؟

ج / يمكن أن تكون خطوة التهيئة حاسمة لأداء النموذج، وتتطلب الطريقة الصحيحة.

- يؤدي تهيئة الأوزان إلى الصفر إلى تعلم الشبكة أن الناتج صفر مما يجعل الشبكة لا تتعلم أي شيء.
- يؤدي تهيئة الأوزان بحيث تكون كبيرة جدًا إلى انفجار الانحدار التدريجي.
- يؤدي تهيئة الأوزان لتكون صغيرة جدًا إلى تلاشي الانحدار التدريجي للشبكة.

للعثور على التهيئة المثالية، هناك بعض القواعد الأساسية التي يجب اتباعها:

- يجب أن يكون متوسط mean عمليات التنشيط صفرًا.
- يجب أن يظل تباين variance عمليات التنشيط كما هو عبر كل طبقة.

س403/ كيف تحصل الشبكات العصبية على القيم المثلى للأوزان والتحيز؟

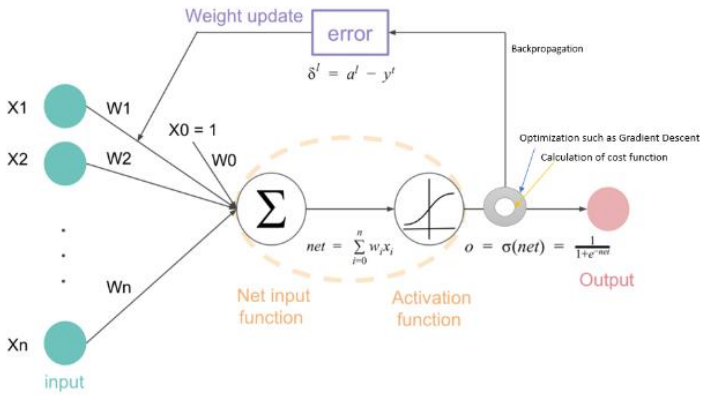
ج / تحصل الشبكات العصبية على الأوزان المثلى وقيم التحيز من خلال تدرج خطأ Error Gradient.

لتقرير ما إذا كنت تريد زيادة أو تقليل الأوزان والانحياز الحاليين، يجب مقارنتها بالقيمة المثلى. تم العثور على هذا من خلال تدرجات الخطأ فيما يتعلق بالأوزان والتحيز:

$$\frac{\partial E}{\partial W}, \frac{\partial E}{\partial b}$$

يتم حساب قيمة التدرج من خوارزمية محددة تسمى الانتشار الخلفي backpropagation.

تستخدم خوارزمية التحسين الانحدار gradient لتحسين قيم الوزن والتحيز.



س404/ ما هي أنواع التحيزات biases التي يمكن أن تحدث أثناء أخذ العينات sampling؟

ج/

1. Selection bias
2. Undercoverage bias
3. Survivorship bias

س405/ ما هو survivorship bias؟

ج/ يحدث survivorship bias عندما يكون هناك الكثير من التركيز على البيانات التي نجت من عملية اختيار معينة، مع تجاهل البيانات التي لم تنجو منها.

س406/ لماذا نحتاج إلى selection bias؟

ج/ يحدث انحياز التحديد selection bias عندما لا تكون بيانات العينة المستخرجة من مجموعة بيانات أكبر ممثلة بشكل كامل. هذا يؤدي إلى استنتاجات خاطئة يتم إجراؤها حول مجموعة البيانات.

س407/ ما هو Cluster Sampling؟

ج/ أخذ العينات العنقودية **Cluster Sampling** هو أسلوب يستخدم عندما يصبح من الصعب دراسة السكان المستهدفين المنتشرين عبر منطقة واسعة ولا يمكن تطبيق أخذ العينات العشوائية البسيطة. العينة العنقودية هي عينة احتمالية حيث تكون كل وحدة أخذ عينات عبارة عن مجموعة أو مجموعة من العناصر.

على سبيل المثال، يريد الباحث مسح الأداء الأكاديمي لطلاب المدارس الثانوية في اليابان. يمكنه تقسيم جميع سكان اليابان إلى مجموعات (مدن) مختلفة. ثم يختار

الباحث عددًا من العناقيد اعتمادًا على بحثه من خلال أخذ عينات عشوائية بسيطة أو منهجية.

س408/ ما هو Systematic Sampling؟

ج/ أخذ العينات المنتظم **Systematic Sampling** هو تقنية إحصائية حيث يتم اختيار العناصر من إطار أخذ العينات المرتب. في أخذ العينات المنتظم، تتقدم القائمة بطريقة دائرية، لذا بمجرد وصولك إلى نهاية القائمة، تتقدم من الأعلى مرة أخرى. أفضل مثال على أخذ العينات المنتظم هو طريقة الاحتمال المتساوي.

س409/ ما هي طرق أخذ العينات المختلفة sampling methods؟

ج/

- Random Sampling
- Systematic Sampling
- Stratified Sampling
- Quota Sampling

س410/ هل يمكن أن يكون لدينا نفس التحيز bias لجميع الخلايا العصبية في الطبقة المخفية؟

ج/ بشكل أساسي، يمكن أن يكون لديك قيمة تحيز مختلفة في كل طبقة أو في كل خلية عصبية أيضًا. ومع ذلك، فمن الأفضل أن يكون لدينا مصفوفة تحيز bias matrix لجميع الخلايا العصبية في الطبقات المخفية أيضًا، والنقطة التي يجب ملاحظتها هي أن كلا الاستراتيجيتين ستعطيك نتائج مختلفة تمامًا.

س411/ ماذا يجب أن تفعل عندما يعاني نموذجك من انحياز منخفض low bias وتباين كبير high variance؟

ج/ يعني التباين العالي high variance أن النموذج الخاص بك يعاني من الضبط الزائد overfitting. يجب تقليل تعقيد نموذجك بطريقة تحصل على موازنة جيدة للتحيز/ التباين bias/variance trade-off, على سبيل المثال عن طريق إزالة الميزات غير ذات الصلة irrelevant features.

س412/ ما المقصود بالتحيز Bias والتباين Variance وماذا تقصد بموازنة التحيز والتباين Bias-Variance Tradeoff؟

ج/ التحيز Bias هو خطأ بسبب افتراضات خاطئة أو مفرطة في التبسيط في خوارزمية التعلم التي تستخدمها. يمكن أن يؤدي ذلك إلى عدم ملاءمة النموذج لبياناتك

underfitting، مما يجعل من الصعب عليه الحصول على دقة تنبؤية عالية ويصعب عليك تعميم معرفتك من مجموعة التدريب إلى مجموعة الاختبار.

التباين Variance هو خطأ بسبب التعقيد الشديد في خوارزمية التعلم التي تستخدمها. يؤدي هذا إلى أن تكون الخوارزمية حساسة للغاية لدرجات عالية من التباين في بيانات التدريب الخاصة بك، مما قد يؤدي إلى زيادة نموذجك في البيانات overfitting. ستحمل الكثير من الضوضاء من بيانات التدريب الخاصة بك لنموذجك ليكون مفيدًا جدًا لبيانات الاختبار الخاصة بك.

يحلل تحلل التباين التحيز بشكل أساسي خطأ التعلم من أي خوارزمية عن طريق إضافة التحيز والتباين وقليلًا من الخطأ غير القابل للاختزال بسبب الضوضاء في مجموعة البيانات الأساسية. بشكل أساسي، إذا جعلت النموذج أكثر تعقيدًا وأضفت المزيد من المتغيرات، فستفقد التحيز ولكنك تحصل على بعض التباين - من أجل الحصول على مقدار الخطأ المخفض على النحو الأمثل، سيتعين عليك موزانة tradeoff التحيز والتباين. لا تريد أي انحياز كبير أو تباين كبير في نموذجك.

س413/ هل التباين الكبير في البيانات جيد أم سيئ؟

ج/ يشير التباين إلى تباين نقطة "تقليل التكلفة" عبر عينات من مجموعة التدريب. إذا كان التباين مرتفعًا، فإن اختيار هذه النقطة لمجموعة تدريب واحدة لن يتم تعميمه جيدًا على البيانات خارج مجموعة التدريب. لذلك لا تريد بالضرورة الوصول إلى هذه النقطة بالضبط.

س414/ قارن بين التحيز bias والتباين variance فيما يتعلق بنماذج التعلم العميق وكيف يمكنك تحقيق التوازن بين الاثنين؟

ج / أثناء فهم التنبؤات، يعد فهم أخطاء التنبؤ هو الأكثر أهمية. هناك نوعان رئيسيان من الأخطاء، قابلة للاختزال reducible وغير قابلة للاختزال irreducible. في الأخطاء القابلة للاختزال، لدينا نوعان، التحيز والتباين. يساعد اكتساب الفهم الصحيح لهذه الأخطاء على بناء نموذج دقيق عن طريق تجنب الضبط الزائد overfitting للنموذج وتثبيته.

من أجل الحصول على التوازن الأمثل optimal balance بين الخطأين، يجب أن يهدف النموذج دائمًا إلى الحفاظ على تحيز منخفض وتباين منخفض. التوازن الأمثل بين التحيز والتباين لن يتناسب أبدًا مع النموذج أو يقل عنه.

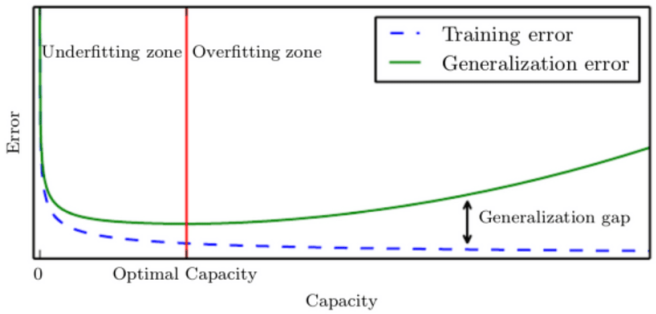
التحيز Bias: في الرسم البياني ادناه، يكون خطأ التدريب (الخط المنقط الأزرق) مرتفعًا في المرحلة الأولية (التحيز العالي) ثم يتناقص بشكل مستدام (انحياز

منخفض) يعني التحيز العالي، والبيانات غير مناسبة، وبالتالي يجب أن تحتوي البيانات على تحيز منخفض لتحقيق نتائج جيدة. من أجل تحقيق انحياز منخفض:

- حاول زيادة عدد التكرارات iterations/الفترات epochs
- جرب شبكة أكبر

التباين Variance: الاختلاف في التعلم العميق ليس سوى الفرق بين خطأ التحقق من الصحة وخطأ التدريب. في الشكل أعلاه، يمكننا أن نرى أن الفجوة بين خطأ التدريب وخطأ التحقق عالية، أي أن التباين مرتفع. هذه هي حالة الضبط الزائد. يجب أن يكون للنموذج تباين منخفض ويمكن تحقيقه من خلال:

- زيادة بيانات التدريب training data
- استخدام التنظيم regularization
- استخدام بنى الشبكات العصبية المختلفة.



س415/ ما هو دور الأوزان والتحيز في الشبكة العصبية؟

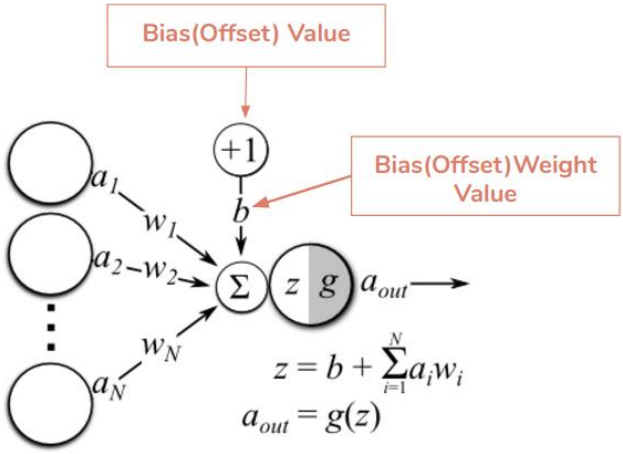
ج / الأوزان والتحيزات (يشار إليها عادةً باسم w و b) هي المعلمات القابلة للتعلم لبعض نماذج التعلم الآلي، بما في ذلك الشبكات العصبية.

الخلايا العصبية هي الوحدات الأساسية للشبكة العصبية. في شبكة ANN، ترتبط كل خلية عصبية في طبقة ببعض أو كل الخلايا العصبية في الطبقة التالية. عندما يتم نقل المدخلات بين الخلايا العصبية، يتم تطبيق الأوزان على المدخلات جنبًا إلى جنب مع التحيز.

$$Y = \sum (weight * input) + bias$$

تتحكم الأوزان في الإشارة (أو قوة الاتصال) بين خليتين عصبيتين. بمعنى آخر، يقرر الوزن مدى تأثير المدخلات على المخرجات.

التحيزات، التي تكون ثابتة، هي مدخلات إضافية في الطبقة التالية والتي سيكون لها دائمًا قيمة 1. لا تتأثر وحدات التحيز بالطبقة السابقة (ليس لديهم أي اتصالات واردة) ولكن لديهم اتصالات صادرة مع خاصة بهم الأوزان. تضمن وحدة التحيز أنه حتى عندما تكون جميع المدخلات أصفارًا، فسيظل هناك تنشيط في الخلايا العصبية.



س416/ ما الذي يجعل التهيئة الصفرية zero initialization ليست عملية تهيئة جيدة للوزن؟

ج/ ان تهيئة الأوزان على الصفر zero initialization في البداية سيؤدي إلى إنتاج جميع الخلايا العصبية لنفس المخرجات ونفس التدرجات عند إجراء الانتشار الخلفي. هذا يعني أن الشبكة لن يكون لديها القدرة على التعلم على الإطلاق بسبب عدم وجود عدم تناسق بين كل من الخلايا العصبية.

س417/ اشرح معنى مصطلح تهيئة الوزن weight initialization في الشبكات العصبية؟

ج/ تهيئة الوزن weight initialization هي إجراء لتعيين أوزان الشبكة العصبية على قيم عشوائية صغيرة تحدد نقطة البداية لتحسين (التعلم أو التدريب) لنموذج الشبكة العصبية.

في كل مرة، يتم تهيئة الشبكة العصبية بمجموعة مختلفة من الأوزان ، مما يؤدي إلى نقطة انطلاق starting point مختلفة لعملية التحسين ، ومن المحتمل أن ينتج عن ذلك مجموعة نهائية مختلفة من الأوزان بخصائص أداء مختلفة.

س418/ هل يمكنك تدريب نموذج الشبكة العصبية عن طريق تهيئة جميع التحيزات biases على أنها 0؟

ج / لا، ليس من الممكن تدريب نموذج عن طريق تهيئة جميع الأوزان إلى 0 لأن الشبكة العصبية لن تتعلم أبدًا أداء مهمة معينة. سيؤدي بدء جميع الأوزان إلى الأصفار إلى بقاء المشتقات كما هي لكل w في $W^{[1]}$ بسبب أي الخلايا العصبية ستتعلم نفس الميزات في كل تكرار. ليس فقط 0، ولكن أي نوع من التهيئة المستمرة للأوزان من المحتمل أن ينتج عنه نتيجة سيئة.

س419/ ما هو تأثير نموذج مع معدل التعلم المحدد بشكل غير صحيح على الأوزان؟

ج / باستخدام الصور كمدخلات، يمكن أن يتسبب معدل التعلم الذي تم تعيينه بشكل غير صحيح في حدوث ميزات مشوشة **noisy features**. إن وجود معدل تعليمي سيئ الاختيار يحدد جودة التنبؤ للنموذج ويمكن أن يؤدي إلى شبكة عصبية غير متقاربة **unconverged neural network**.

س420/ ما هي الأساليب المستخدمة بشكل شائع لضبط معدل التعلم learning rate؟

ج/

- استخدام قيمة معدل التعلم الثابت fixed learning rate لعملية التعلم الكاملة.
- استخدام معدل التعلم مجدول learning rate schedule
- الاستفادة من معدلات التعلم التكيفية adaptive learning rates
- إضافة الزخم momentum إلى معادلة SGD الكلاسيكية.

س421/ هل تجربة معدلات التعلم التالية: 0.1،0.2 ، ... ، 0.5 استراتيجية جيدة لتحسين معدل التعلم؟

ج/لسوء الحظ، لا يمكننا حساب معدل التعلم الأمثل بشكل تحليلي لنموذج معين على مجموعة بيانات معينة. بدلاً من ذلك، يجب اكتشاف معدل تعليمي جيد (أو جيد بما فيه الكفاية) من خلال التجربة والخطأ **trial and error**. نطاق القيم التي يجب مراعاتها لمعدل التعلم أقل من **1.0** وأكبر من **10^-6**. القيمة الافتراضية التقليدية لمعدل التعلم هي **0.1** أو **0.01**، وقد يمثل هذا نقطة بداية جيدة لمشكلتك. يمكن أن يساعد نهج البحث الشبكي **grid search** في تسليط الضوء على ترتيب الاهمية

حيث قد توجد معدلات التعلم الجيدة، وكذلك وصف العلاقة بين معدل التعلم والأداء.

علاوة على ذلك، يمكن للزخم **momentum** تسريع التعلم بشأن تلك المشكلات حيث تحتوي مساحة الوزن عالية الأبعاد التي يتم التنقل فيها من خلال عملية التحسين على هياكل تضلل خوارزمية الانحدار التدريجي **gradient descent**، مثل المناطق المسطحة أو الانحناء الحاد. على الرغم من عدم وجود طريقة واحدة تعمل بشكل أفضل في جميع المشكلات، فقد أثبت **Adaptive Momentum AdaM** أنه قوي على العديد من أنواع هياكل الشبكات العصبية وأنواع المشكلات.

س422/ ماذا تفهم من خلال معدل التعلم learning rate في نموذج الشبكة العصبية؟ ماذا يحدث إذا كان معدل التعلم مرتفعًا جدًا أو منخفضًا جدًا؟

ج/ يعد معدل التعلم **learning rate** أحد أهم المعلمات الفائقة **hyperparameters** القابلة للتكوين المستخدمة في تدريب الشبكة العصبية. تكمن قيمة معدل التعلم بين 0 و1. يعد اختيار معدل التعلم أحد أكثر الجوانب تحديًا لتدريب الشبكة العصبية لأنها المعلمة التي تتحكم في مدى سرعة أو بطء نموذج الشبكة العصبية في التكيف مع مشكلة معينة والتعلم. تعني قيمة معدل التعلم الأعلى أن النموذج يتطلب فترات تدريب قليلة وينتج عنه تغييرات سريعة بينما يشير معدل التعلم الأصغر إلى أن النموذج سيستغرق وقتًا طويلاً للتقارب **converge** أو قد لا يتقارب أبدًا ويتعثر في حل دون المستوى الأمثل. وبالتالي، يُنصح بعدم استخدام معدل تعلم منخفض جدًا أو مرتفع جدًا ولكن بدلاً من ذلك، يجب اكتشاف قيمة معدل تعلم جيدة من خلال التجربة والخطأ.

س423/ ما هي بعض الاختلافات التي قد تتوقعها في نموذج يقلل من الخطأ التربيعي squared error، مقابل نموذج يقلل من الخطأ المطلق absolute error؟ في أي الحالات يكون كل مقياس خطأ مناسبًا؟

ج/ **MSE** أكثر صرامة في وجود القيم المتطرفة. **MAE** هو أكثر قوة بهذا المعنى ، لكنه كذلك أصعب لتلائم النموذج لأنه لا يمكن تحسينه عدديًا. لذلك عندما يكون هناك تباين أقل في النموذج ويكون النموذج سهل التوافق حسابيًا، يجب أن نستخدم **MAE**، وإذا لم يكن الأمر كذلك، فيجب علينا استخدام **MSE**.

MSE: أسهل لحساب التدرج ، **MAE**: البرمجة الخطية اللازمة لحساب التدرج

MAE: أكثر قوة للقيم المتطرفة. إذا كانت عواقب الأخطاء الكبيرة كبيرة، فاستخدم **MSE**

يتوافق **MSE** مع تعظيم احتمالية المتغيرات العشوائية الغاوسية.

س424/ ما هو الفرق بين الخطأ التربيعي squared error والخطأ المطلق absolute error؟

ج/ يتم استخدام كل من متوسط الخطأ التربيعي (MSE) ومتوسط الخطأ المطلق (MAE) في النمذجة التنبؤية. تتميز MSE بخصائص رياضية لطيفة تجعل من السهل حساب الانحدار gradient. ومع ذلك، يتطلب MAE أدوات أكثر تعقيدًا مثل البرمجة الخطية لحساب الانحدار gradient. بسبب المربع، يكون للأخطاء الكبيرة تأثير أكبر نسبيًا على MSE من الخطأ الأصغر. لذلك، يعد MAE أكثر قوة بالنسبة للقيم المتطرفة لأنه لا يستخدم المربع. من ناحية أخرى، يعد MSE أكثر فائدة إذا كنا قلقين بشأن الأخطاء الكبيرة التي تكون عواقبها أكبر بكثير من الأخطاء الأصغر المماثلة. يتوافق MSE أيضًا مع تعظيم احتمالية المتغيرات العشوائية الغاوسية.

س425/ كيف تختار دالة الخطأ Loss Function لنموذج التعلم العميق؟

ج/

- الأهداف الثنائية Binary targets: في هذه الحالة، يتم رسم القيمة المرصودة من -1 إلى 1. تظهر دالة الخطأ لهذه الحالة على النحو التالي:

$$L = log(1 + exp(-y.\hat{y}))$$

حيث $\hat{y}$ هو الناتج المتوقع. y هو الناتج المرصود.

ينفذ هذا النوع من دالة الخطأ طريقة التعلم الآلي المعروفة باسم الانحدار اللوجستي.

- الأهداف الفئوية Categorical targets: إذا كانت y هي احتمالات فئات k، وكانت الفئة r هي فئة الحقيقة الأساسية، يتم تحديد دالة الخسارة لعينة واحد على النحو التالي:

$$L = -log(\hat{y}_r)$$

هذا النوع من دالة الخسارة يطبق الانحدار اللوجستي متعدد الحدود، ويسمى خسارة الانتروبيا المتقاطعة cross-entropy loss.

- الانحدار اللوجستي الثنائي Binary logistic regression مشابه للانحدار اللوجستي متعدد الحدود مع ضبط قيمة k على 2.

س426/ لماذا يُفضل الانتروبيا المتقاطعة cross-entropy كدالة تكلفة لمشاكل التصنيف متعدد الفئات multi-class classification؟

ج/ يمكن استخدام الانتروبيا المتقاطعة cross-entropy لحساب الفرق بين توزيعي الاحتمال. على هذا النحو، يمكننا مطابقة تصنيف أحد الأمثلة على فكرة متغير عشوائي مع توزيع احتمالي على النحو التالي: المتغير العشوائي: المثال الذي نطلب منه تسمية فئة متوقعة.

الخسارة عبر الانتروبيا cross-entropy، أو log loss، تقيس أداء نموذج التصنيف الذي يكون ناتجه قيمة احتمالية بين 0 و1. ويفضل استخدامه في التصنيف.

س427/ ما هي دالة الخطأ Loss Function المفضلة للتصنيف متعدد الفئات؟

ج/ دالة الخطأ Cross-Entropy.

س428/ اشرح دالة الخطأ cross-entropy؟

ج/ يعد دالة الخطأ cross-entropy مقياسًا يستخدم لقياس مدى جودة أداء نموذج التصنيف في التعلم الآلي. يتم قياس الخسارة (أو الخطأ) كرقم بين 0 و 1، مع كون الصفر نموذجًا مثاليًا. الهدف بشكل عام هو جعل النموذج الخاص بك قريبًا من الصفر قدر الإمكان.

غالبًا ما يُعتبر فقدان الانتروبيا المتقاطعة قابلاً للتبادل مع الخسارة اللوجيستية (أو فقدان السجل log loss، ويشار إليه أحيانًا باسم فقدان الانتروبيا الثنائية binary cross entropy loss) ولكن هذا ليس صحيحًا دائمًا.

س429/ الفرق بين الخطأ error والخطأ المتبقي residual error؟

ج/ الخطأ error هو مقياس لمدى انحراف القيمة المرصودة عن القيمة الحقيقية. من ناحية أخرى، يوضح الخطأ المتبقي residual error مدى اختلاف القيمة المرصودة observed value عن القيمة المقدرة estimated value لنقطة بيانات معينة.

س430/ ما المقصود بخطأ التدريب Training Loss وخطأ التحقق من الصحة Validation Loss؟

ج/ خطأ التدريب Training Loss هو مقياس يستخدم لتقييم مدى ملاءمة نموذج التعلم العميق لبيانات التدريب. أي أنه يقيم خطأ النموذج في مجموعة التدريب. لاحظ أن مجموعة التدريب هي جزء من مجموعة البيانات المستخدمة لتدريب

النموذج مبدئيًا. من الناحية الحسابية، يتم حساب خطأ التدريب بأخذ مجموع الأخطاء لكل مثال في مجموعة التدريب.

على العكس من ذلك، يعد خطأ التحقق من الصحة Validation Loss مقياسًا يستخدم لتقييم أداء نموذج التعلم العميق في مجموعة التحقق من الصحة. مجموعة التحقق من الصحة هي جزء من مجموعة البيانات تم وضعه جانبًا للتحقق من أداء النموذج. تشبه خطأ التحقق من الصحة خطأ التدريب ويتم حسابها من مجموع الأخطاء لكل مثال في مجموعة التحقق من الصحة.

س431/ عند تدريب شبكة عصبية، لاحظ أن الخطأ لا يقل في الفترات الأولى. ما هي الأسباب الأصلية لذلك؟

ج/ يمكن أن تكون أسباب ذلك:

1) معدل التعلم learning rate منخفض
2) معلمة التنظيم Regularization parameter عالية
3) عالق في الحدود الدنيا المحلية local minima

س432/ ما فائدة دالة الخطأ loss function؟

ج/ تُستخدم دالة الخطأ (الخسارة) loss function كمقياس للدقة لمعرفة ما إذا كانت الشبكة العصبية قد تعلمت بدقة من بيانات التدريب أم لا. يتم ذلك عن طريق مقارنة مجموعة بيانات التدريب بمجموعة بيانات الاختبار.

دالة الخطأ هي مقياس أساسي لأداء الشبكة العصبية. في التعلم العميق، سيكون لشبكة الأداء الجيد دالة خطأ منخفضة في جميع الأوقات عند التدريب.

س433/ على أي نوع من المشاكل يمكن تطبيق دالة خطأ الانتروبيا Cross-Entropy؟

ج/

- مشاكل التصنيف الثنائي Binary Classification Problems
- مشاكل التصنيف متعدد العلامات Multi-Label Classification Problems
- مشاكل التصنيف متعدد الفئات Multi-Category Classification Problems

س434/ إن معالجة المتغير الفئوي كمتغير مستمر من شأنه أن يؤدي إلى نموذج تنبؤي أفضل؟

ج / يمكن التعامل مع المتنبئ الفئوي باعتباره متنبئًا مستمرًا عندما تكون طبيعة نقاط البيانات التي يمثلها ترتيبية. إذا كان متغير التوقع يحتوي على بيانات ترتيبية، فيمكن معاملته على أنه مستمر ويؤدي تضمينه في النموذج إلى زيادة أداء النموذج.

س435/ماذا تقصد بالمتغير الوهمي Dummy Variable؟ أين يتم استخدامه في التعلم الآلي؟

ج / إذا كان هناك عدد **n** من الفئات في السمة الفئوية، فسيتم إنشاء سمات جديدة **n**. تسمى هذه السمات التي تم إنشاؤها باسم المتغيرات الوهمية **Dummy Variables**. سيتم إنشاء هذه المتغيرات الوهمية بترميز **one-hot** وسيكون لكل سمة قيمة إما **0** أو **1**، مما يمثل وجود أو عدم وجود تلك السمة.

نحن نستخدم المتغيرات الوهمية في الانحدار **Regression** في التعلم الآلي. لتحويل السمة الفئوية **categorical** إلى سمة عددية **numerical**، يمكننا استخدام إجراء ترميز التسمية **label encoding** (يخصص ترميز التسمية عددًا صحيحًا فريدًا لكل فئة من فئات البيانات). لكن هذا الإجراء ليس بمفرده مناسبًا إلى حد كبير، وبالتالي، يتم استخدام ترميز **one hot** في نماذج الانحدار بعد ترميز التسمية. يتيح لنا ذلك إنشاء سمات جديدة وفقًا لعدد الفئات الموجودة في السمة الفئوية. هنا حيث يتم استخدام المتغيرات الوهمية.

س436/ نحن نعلم أن one-hot encoding يزيد من أبعاد مجموعة البيانات، لكن label encoding لا يفعل ذلك. كيف؟

ج / عند استخدام الترميز one-hot، يكون هناك زيادة في أبعاد مجموعة البيانات. سبب الزيادة في الأبعاد هو أن كل فئة في المتغيرات الفئوية تشكل متغيرًا مختلفًا.

مثال: افترض أن هناك متغيرًا "اللون". يحتوي على ثلاثة مستويات فرعية، "أصفر" و "أرجواني" و "برتقالي". لذلك، سينشئ "اللون" الترميز one-hot ثلاث متغيرات مختلفة مثل Color.Yellow و Color.Purple و Color.Orange.

في ترميز label، تحصل الفئات الفرعية لمتغير معين على القيمة 0 و1. لذلك، يتم استخدام ترميز label للمتغيرات الثنائية فقط.

هذا هو السبب في أن الترميز one-hot يزيد من أبعاد البيانات ولا يؤدي ترميز label إلى ذلك.

س437/ اشرح label encoding وone-hot encoding. كيف تؤثر على أبعاد مجموعة البيانات المحددة؟

ج / لا يؤثر label encoding على مجموعة البيانات بأي شكل من الأشكال لأنه في label encoding، لا نقدم سوى تسميات لكل فئة في العمود.

في one-hot encoding، نقوم بإنشاء أعمدة لكل فئة في مجموعة البيانات. وبالتالي، كلما زاد عدد الفئات في العمود، زاد عدد الأعمدة التي تم إنشاؤها بعد one-hot encoding.

س438/ كيف يؤثر الترميز one-hot والترميز label على أبعاد مجموعة البيانات؟

ج/ الترميز one-hot هو تمثيل المتغيرات الفئوية categorical variables كمتجهات ثنائية. ترميز التسمية هو تحويل التسميات / الكلمات إلى شكل رقمي. يزيد استخدام ترميز one-hot من أبعاد مجموعة البيانات. لا يؤثر ترميز التسمية label encoding على أبعاد مجموعة البيانات. ينشئ ترميز one-hot متغيرًا جديدًا لكل مستوى في المتغير بينما، في ترميز التسمية، يتم ترميز مستويات المتغير كـ 1 و0.

س439/ ما المقصود بلعنة الأبعاد curse of dimensionality؟

ج / تشير البيانات عالية الأبعاد High dimensional data إلى البيانات التي تحتوي على عدد كبير من الميزات. بُعد البيانات dimension of data هو عدد الميزات أو السمات في البيانات. يشار إلى المشاكل التي تنشأ أثناء العمل مع البيانات عالية الأبعاد على أنها لعنة الأبعاد curse of dimensionality. هذا يعني بشكل أساسي أن الخطأ يزداد مع زيادة عدد الميزات في البيانات. من الناحية النظرية، يمكن تخزين المزيد من المعلومات في بيانات عالية الأبعاد، ولكن من الناحية العملية، لا يساعد ذلك لأنه يمكن أن يكون له ضوضاء أعلى وتكرار. من الصعب تصميم خوارزميات للبيانات عالية الأبعاد. أيضًا، يزداد وقت التشغيل running time أضعافًا مضاعفة مع بُعد البيانات.

س440/ ما هي زيادة البيانات data augmentation؟ أعط أمثلة.

ج / زيادة البيانات data augmentation هي عملية إنشاء بيانات جديدة من خلال تحسين حجم ونوعية مجموعات البيانات التدريبية لضمان إمكانية بناء نماذج أفضل باستخدامها. هناك تقنيات مختلفة لزيادة البيانات مثل زيادة البيانات الرقمية وزيادة الصورة image augmentation وزيادة البيانات المستند إلى GAN GAN-based augmentation وزيادة النص text augmentation.

س441/ ما هو المقصود بـ "لعنة الأبعاد curse of dimensionality"؟ كيف يمكننا حل هذه المشكلة؟

ج/أثناء تحليل مجموعة البيانات، هناك حالات يزيد فيها عدد المتغيرات أو الأعمدة. ومع ذلك، فنحن مطالبون فقط باستخراج المتغيرات المهمة من المجموعة. على سبيل المثال، ضع في اعتبارك أن هناك ألف ميزة. ومع ذلك، نحتاج فقط إلى استخراج عدد قليل من الميزات المهمة. هذه المشكلة المتمثلة في وجود العديد من الميزات حيث نحتاج فقط إلى القليل منها تسمى "لعنة الأبعاد curse of dimensionality".

هناك خوارزميات مختلفة لتقليل الأبعاد مثل PCA (تحليل المكونات الرئيسية).

س442/ هل من الممكن استخدام الشبكة العصبية كأداة لتقليل الأبعاد dimensionality reduction؟

ج/يمكن بالتأكيد استخدامه. المشفر التلقائي Autoencoders هو ما يتبادر إلى ذهني وأعتقد أنه قد ثبت أنه جيد في ذلك. ولكن من الواضح أنه سيتعين عليك تدريب الشبكة واختيار دالة التكلفة المناسبة (على عكس الطرق غير الخاضعة للإشراف مثل PCA و t-SNE).

س443/ اشرح تقليل الأبعاد dimension reduction في التعلم الآلي؟

ج / بكلمات بسيطة، يشير تقليل الأبعاد dimension reduction إلى تقنية تقليل أبعاد مجموعة ميزات البيانات. عادةً ما تحتوي مجموعات بيانات التعلم الآلي (مجموعة الميزات) على مئات الأعمدة (أي الميزات) أو مجموعة من النقاط، مما يؤدي إلى إنشاء كرة ضخمة في مساحة ثلاثية الأبعاد. من خلال تطبيق تقليل الأبعاد، يمكنك تقليل أو تقليل عدد الأعمدة إلى أعداد قابلة للقياس الكمي، وبالتالي تحويل الكرة ثلاثية الأبعاد إلى كائن ثنائي الأبعاد (دائرة).

س444/كيف تفلت الشبكة العصبية العميقة/تقاوم لعنة الأبعاد Curse of Dimensionality؟

ج / عادةً ما تحدث لعنة الأبعاد نظرًا لوجود العديد من الميزات غير ذات الصلة (الضوضاء) في البيانات. تستخدم الخلايا العصبية في بنيات التعلم العميق (DL) الكثير من البيانات من أجل نمذجة مشكلة، وبالتالي يقلل نظام DL من تأثير الميزات غير ذات الصلة مع زيادة تأثير الميزات ذات الصلة أثناء التعلم.

س445/ ما هي مكتبات تصوير ورسم البيانات Data Visualization التي تستخدمها؟

ج/ المهم هنا هو تحديد وجهات نظرك حول كيفية تصوير البيانات بشكل صحيح وتفضيلاتك الشخصية عندما يتعلق الأمر بالأدوات. تشمل الأدوات الشائعة R's ggplot seaborn Python's matplotlib وأدوات مثل Plot.ly Tableau.

س446/ ما هو تحويل Box-Cox Transformation؟

ج / تحويل Box-Cox هو طريقة لتسوية البيانات، سميت على اسم اثنين من الإحصائيين الذين قدموها، جورج بوكس وديفيد كوكس. يتم تحويل كل نقطة بيانات، X، باستخدام الصيغة X^a، حيث يمثل a القوة التي يتم رفع كل نقطة بيانات إليها. يناسب تحويل box-cox البيانات للقيم من -5 إلى +5 حتى يتم تحديد القيمة "a" المثلى التي يمكن أن تجعل البيانات طبيعية بشكل أفضل.

س447/ ماذا تفهم بالرسم البياني الحسابي computational graph؟

ج/ سنقوم بتعريف مفهوم الرسم البياني الحسابي، لأن الشبكات العصبية هي شكل خاص منه. الرسم البياني الحسابي هو رسم بياني موجه حيث تتوافق العقد مع العمليات أو المتغيرات. يمكن للمتغيرات أن تغذي قيمتها في العمليات، ويمكن للعمليات أن تغذي ناتجها في عمليات أخرى. بهذه الطريقة، تحدد كل عقدة في الرسم البياني وظيفة المتغيرات.

تسمى القيم التي يتم إدخالها في العقد والخروج من العقد بالموتر tensor، وهي مجرد كلمة رائعة لمصفوفة متعددة الأبعاد. ومن ثم، فإنه يشمل scalars والمتجهات vectors والمصفوفات matrices وكذلك الموترات من رتبة أعلى.

س448/ ما هي الطرق الجماعية Ensemble methods وكيف أنها مفيدة في التعلم العميق؟

ج/ تُستخدم الطرق الجماعية Ensemble methods لزيادة قوة التعميم للنموذج. هذه الأساليب قابلة للتطبيق على كل من خوارزميات التعلم العميق والتعلم الآلي.

بعض الطرق الجماعية التي تم تقديمها في الشبكات العصبية هي Dropout و Dropconnect. يعتمد التحسين في النموذج على نوع البيانات وطبيعة البُنية العصبية.

س449/ ما الفرق بين أخطاء Type1 وType 2؟

ج / يتم تصنيف الخطأ من النوع الأول **Type1** على أنه خطأ إيجابي. بمعنى آخر. يدعي هذا الخطأ أن شيئًا ما قد حدث، لكن الحقيقة هي أن شيئًا لم يحدث. إنه مثل إنذار الحريق الكاذب. يرن جرس الإنذار ولكن لا يوجد حريق.

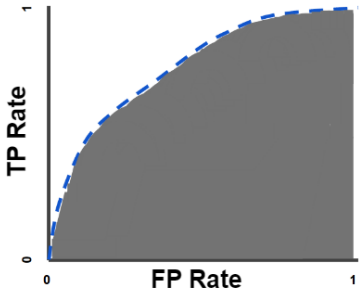
يصنف الخطأ من النوع الثاني **Type 2** على أنه خطأ سلبي. بمعنى آخر. يدعي هذا الخطأ أنه لم يحدث شيء ولكن الحقيقة هي أنه حدث شيء ما في الواقع.

أفضل طريقة للتمييز بين الخطأ من النوع الأول والنوع الثاني هي:

- استدعاء الرجل ليكون حامل- هذا مثال من النوع الأول.
- الاتصال بالنساء الحوامل وإخبارهن بأنها لا تحمل أي طفل - هذا مثال من النوع الثاني.

س450/ ماذا تقصد بمنحنى AUC؟

ج / AUC (المنطقة تحت المنحنى area under curve). كلما زادت المساحة تحت المنحنى، زادت قوة التنبؤ للنموذج.



س451/ لماذا نستخدم دالة التلخيص summary؟

ج/ تُستخدم الدالات التلخيصية **summary** لإعطاء ملخص لجميع القيم الرقمية في إطار البيانات. على سبيل المثال. يمكن استخدام دالة **describe()** لتوفير ملخص لجميع قيم البيانات المعطاة لها.

column_name.describe() سيعطي القيم التالية لجميع البيانات الرقمية في العمود:-

- Count
- Mean
- Std-Standard deviation
- Min-Minimum

- 25%
- 50%
- 75%
- max-Maximum

س452/ كيف يرتبط الاستدعاء Recall بالمعدلات الإيجابية الحقيقية True positive؟

ج / العلاقة هي معدل الإيجابية الحقيقية (TP) = الاستدعاء Recall.

س453/ كيف تختلف AUC عن ROC؟

ج/ منحنى AUC هو قياس الدقة precision مقابل الاسترجاع recall. الدقة = TP / (TP + FP) و TP / (TP + FN). هذا على النقيض من ROC الذي يقيس ويخطط للمعدل الإيجابي الحقيقي TP مقابل المعدل الإيجابي الكاذب FP .

س454/ اشرح كيف يعمل منحنى ROC؟

ج / منحنى ROC (receiver operating characteristic) هو رسم بياني يوضح أداء نموذج التصنيف في جميع عتبات التصنيف. يرسم معلمتين:

- المعدل الإيجابي الحقيقي
- المعدل الإيجابي الكاذب

يتم تعريف المعدل الإيجابي الحقيقي (TPR) على النحو التالي:

$$TPR = TP/(TP+FN)$$

يتم تعريف المعدل الإيجابي الكاذب (FPR) على النحو التالي:

$$FPR = FP/(FP+TN)$$

س455/ عرف الدقة precision والاستدعاء recall؟

ج/

- الدقة precision: يُعرف أيضًا بالقيمة الإيجابية المتوقعة. هذا يعتمد أكثر على التنبؤ. إنه مقياس لعدد الإيجابيات الدقيقة التي يدعيها النموذج عند مقارنتها بعدد الإيجابيات التي يدعيها بالفعل.

❖ الاستدعاء recall: يعرف بأنه المعدل الإيجابي الحقيقي. عدد الإيجابيات التي طالب بها نموذجك مقارنة بالعدد الفعلي المحدد للإيجابيات المتاحة في جميع البيانات.

س456/ ما هي F1 Score؟ كيف يمكنك أن تستفيد منها؟

ج/ F1-score هي مقياس للدقة الشاملة لنموذج التصنيف الثنائي. قبل فهم F1-score، من الضروري فهم مقياسين آخرين للدقة ، أي الدقة precision والاستدعاء recall.

تُعرَّف الدقة بأنها النسبة المئوية للإيجابيات الحقيقية من إجمالي عدد التصنيفات الإيجابية التي تنبأ بها النموذج. بعبارات أخرى،

الدقة = (عدد الإيجابيات الحقيقية / عدد الإيجابيات الحقيقية + عدد الإيجابيات الكاذبة)

يتم تعريف الاستدعاء على أنه النسبة المئوية للإيجابيات الحقيقية إلى العدد الإجمالي للبيانات المعنونة الإيجابية الإيجابية التي تم تمريرها إلى النموذج. بعبارات أخرى،

الاستدعاء = الإيجابيات الحقيقية / (الإيجابيات الحقيقية + السلبيات الكاذبة)

كل من الدقة والاستدعاء هي مقاييس جزئية لدقة النموذج. تجمع F-score بين الدقة والاستدعاء وتوفر درجة إجمالية لقياس دقة النموذج.

F-score = 2 × (الدقة × الاستدعاء) / (الدقة + الاستدعاء)

هذا هو السبب في أن F1-Score هي المقياس الأكثر شيوعًا للدقة في أي نموذج تصنيف ثنائي قائم على التعلم الآلي.

س457/ كيف تجد RMSE و MSE في نموذج الانحدار الخطي؟

ج/

❖ جذر متوسط الخطأ التربيعي (RMSE) هو مقياس يحسب الخطأ في التنبؤ العددي. فيما يلي صيغة RMSE.

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(f_i - o_i)^2}$$

Where:

- ∑ is the summation of all values
- f is the predicted value
- o is observed or actual value
- (fi — oi) 2 are the differences between predicted and observed values and squared
- N is the total sample size

❖ متوسط الخطأ التربيعي (MSE) هو مقياس لدرجة الخطأ الموجودة في النموذج الإحصائي. يمكن العثور عليها بالصيغة التالية:

$$MSE = \frac{1}{n}\Sigma\underbrace{\left(y - \hat{y}\right)}^{2}$$

The square of the difference between actual and predicted

س458/ صِف الحالة الذي ستستخدم فيه MSE كمقياس للجودة؟

ج/ يتم تعريف متوسط مربع الخطأ (MSE) على أنه متوسط مربع الفرق بين القيم الفعلية actual values والقيم المقدرة estimated values.

سنستخدم MSE عند النظر في دقة نموذج الانحدار.

س459/ أعط بعض الأمثلة عن الإيجابية الخاطئة FP، السلبية الخاطئة FN، الإيجابية الحقيقية TP، السلبية الحقيقية TN؟

ج/

❖ الإيجابية الخاطئة FP : يأتي اختبار فحص السرطان إيجابيًا ، لكنك لست مصابًا بالسرطان

❖ السلبية الخاطئة FN : يأتي اختبار فحص السرطان سلبيًا ، لكنك مصاب بالسرطان

❖ الإيجابية الحقيقية TP : يعود اختبار فحص السرطان إيجابيًا ، وأنت مصاب بالسرطان

❖ السلبية الحقيقية TN : يأتي اختبار فحص السرطان سلبيًا ، وليس لديك سرطان

س460/ اشرح TP وTNو FP وFN في مصفوفة الارتباك Confusion Matrix بمثال؟

ج/

- إيجابي حقيقي True Positive (TP)

عندما يتنبأ النموذج بشكل صحيح بالفئة الإيجابية، يقال إنها إيجابية حقيقية.

- سلبي حقيقي True Negative (TN)

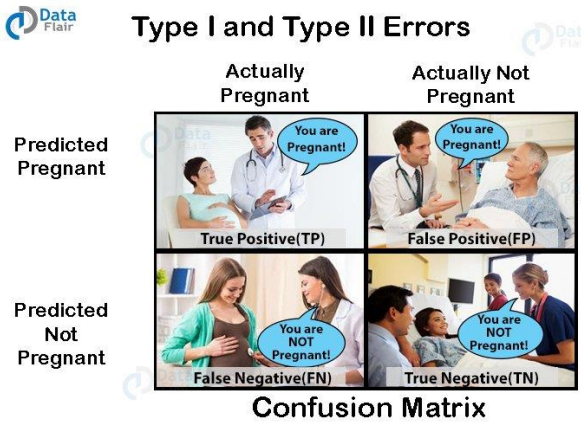
عندما يتنبأ نموذج بالفئة السلبية بشكل صحيح، يقال إنها سلبية حقيقية.

- إيجابي خاطئ False Positive (FP)

عندما يتنبأ النموذج بشكل غير صحيح بالفئة الإيجابية، يقال إنها إيجابية خاطئة. يُعرف أيضًا باسم خطأ "النوع الأول Type I ".

- سلبي خاطئ False Negative (FN)

عندما يتنبأ النموذج بشكل غير صحيح بالفئة السالبة، يقال إنها سلبية خاطئة. يُعرف أيضًا باسم خطأ "النوع الثاني Type II ".



س461/ ما هي مصفوفة الارتباك Confusion Matrix؟

ج/ يتم استخدام مصفوفة الارتباك Confusion Matrix لتحديد فعالية خوارزمية التصنيف. يتم استخدامه لأن خوارزمية التصنيف ليست دقيقة عندما يكون هناك أكثر من فئتين من البيانات، أو عندما لا يكون هناك عدد زوجي من الفئات.

تكون عملية إنشاء مصفوفة الارتباك كما يلي:

1. قم بإنشاء مجموعة بيانات التحقق التي لديك بعض القيم المتوقعة كنتائج لها.
2. توقع نتيجة كل صف موجود في مجموعة البيانات.
3. الآن احسب عدد التنبؤات الصحيحة وغير الصحيحة لكل فئة.
4. قم بتنظيم تلك البيانات في مصفوفة بحيث يمثل كل صف فئة متوقعة وكل عمود يمثل فئة فعلية.
5. املأ العدادات التي تم الحصول عليها من الخطوة الثالثة في الجدول.
6. تُعرف المصفوفة الناتجة عن هذه العملية بمصفوفة الارتباك confusion matrix.

س462/ لماذا نحتاج مجموعات بيانات التحقق من الصحة validation والاختبار test؟

ج / سيؤدي استخدام مجموعات التحقق من الصحة والاختبار إلى زيادة القدرة التعميمية للنموذج على البيانات الجديدة غير المرئية new unseen data. لاحظ أيضًا أن مجموعة التحقق من الصحة ليست ضرورية (زائدة عن الحاجة) إذا كنت لا تريد ضبط النموذج من خلال تجربة مجموعات مختلفة من المعلمات الفائقة.

س463/ ما هو SVM في التعلم الآلي؟

ج/تعد آلة المتجهات الداعمة Support Vector Machine (SVM) خوارزمية تعلم آلي خاضعة للإشراف تستخدم لكل من التصنيف والانحدار. الهدف من خوارزمية SVM هو العثور على hyperplane في مساحة ذات أبعاد N تصنف نقاط البيانات بوضوح.

س464/اشرح الفرق بين KNN وK-mean Clustering؟

ج/

K-Means	KNN algorithms
خوارزمية خاضعة للأشراف	خوارزمية غير خاضعة للأشراف
إنه خوارزمية تصنيف أو الانحدار.	إنها خوارزمية تجميع clustering.
أداؤها بطيء.	أداء أفضل بكثير.

س465/ ما هي النوى (kernels) المختلفة الموجودة في SVM؟

ج/SVM تعني آلة ناقلات الدعم support vector machine. يتم استخدامها لمهام التصنيف والتنبؤ. يتكون SVM من مستوى plane فاصل يميز بين فئتي المتغيرات.

يُعرف هذا المستوى الفاصل بالمستوى الفائق hyperplane. بعض النوى المستخدمة في SVM هي -

- Polynomial Kernel
- Gaussian Kernel
- Laplace RBF Kernel
- Sigmoid Kernel
- Hyperbolic Kernel

س466/ هل يمكنك ذكر بعض مزايا وعيوب أشجار القرار decision trees؟

ج/

المزايا:

- ❖ بالمقارنة مع الخوارزميات الأخرى، فإن أشجار القرار تتطلب جهدًا أقل لإعداد البيانات أثناء المعالجة المسبقة.
- ❖ لا تتطلب شجرة القرار تسوية البيانات.
- ❖ لا تتطلب شجرة القرار قياس البيانات أيضًا.
- ❖ القيم المفقودة في البيانات لا تؤثر أيضًا على عملية بناء شجرة قرار إلى حد كبير.
- ❖ نموذج شجرة القرار بديهي للغاية ويسهل شرحه للفرق الفنية وكذلك أصحاب المصلحة.

العيوب:

- ❖ يمكن أن يتسبب تغيير بسيط في البيانات في حدوث تغيير كبير في بنية شجرة القرار مما يتسبب في عدم الاستقرار.
- ❖ بالنسبة لشجرة القرار في بعض الأحيان، يمكن أن يكون الحساب أكثر تعقيدًا مقارنة بالخوارزميات الأخرى.
- ❖ غالبًا ما تتضمن شجرة القرار وقتًا أطول لتدريب النموذج.
- ❖ يعد تدريب شجرة القرار مكلفًا نسبيًا نظرًا لأن التعقيد والوقت الذي يستغرقه الأمر أكثر.
- ❖ خوارزمية شجرة القرار غير كافية لتطبيق الانحدار والتنبؤ بالقيم المستمرة.

س467/ هل من الممكن استخدام KNN لمعالجة الصور؟

ج/ نعم من الممكن استخدام KNN لمعالجة الصور.

س468/ ما هو نايف بايز Naive Bayes؟

ج / يعد تصنيف Naïve Bayes أحد خوارزميات التصنيف البسيطة والأكثر فاعلية والتي تساعد في بناء نماذج التعلم الآلي السريعة التي يمكنها إجراء تنبؤات سريعة. إنه مصنف احتمالي، مما يعني أنه يتنبأ على أساس احتمال وجود كائن.

س469/ ما هي أهمية جاما Gamma والتنظيم Regularization في SVM؟

ج/ تحدد جاما التأثير. القيم المنخفضة تعني "بعيدة" والقيم العالية تعني "قريبة". إذا كانت جاما كبيرة جدًا، فإن نصف قطر منطقة تأثير نواقل الدعم support vectors يشمل فقط متجه الدعم نفسه ولن يكون أي قدر من التنظيم مع C قادرًا على منع Overfitting. إذا كانت جاما صغيرة جدًا، فإن النموذج مقيد جدًا ولا يمكنه التقاط تعقيد البيانات.

تعمل معلمة التنظيم (lambda) على أنها درجة من الأهمية تُعطى للتصنيفات الخاطئة. يمكن استخدام هذا لرسم الموازنة tradeoff باستخدام Overfitting.

س470/ كيف تحدد عدد المجموعات number of clusters في خوارزمية التجميع clustering algorithm؟

ج / ربما تكون الطريقة الأكثر شهرة هي طريقة الكوع **elbow method**، حيث يتم حساب مجموع المربعات في كل عدد من المجموعات ورسمها، ويبحث المستخدم عن تغيير المنحدر من حاد إلى ضحل (كوع) لتحديد العدد الأمثل لعناقيد المجموعات.

س471/ ما هي الطريقة الافتراضية لتقسيم أشجار القرار decision trees؟

ج / الطريقة الافتراضية لتقسيم أشجار القرار هي مؤشر جيني **Gini Index**. مؤشر جيني هو مقياس شوائب عقدة معينة.

يمكن تغيير هذا عن طريق إجراء تغييرات على معلمات المصنف.

س472/ ما هي اهم المعلمات الفائقة لـ SVM؟

ج/ قيمة جاما وقيمة **c** ونوع النواة **kernel** هي المعلمات الفائقة لنموذج **SVM**.

س473/ ما هي مقاييس المسافة التي يمكن استخدامها في KNN؟

ج/ يمكن استخدام مقاييس المسافة التالية في **KNN**:

- Manhattan
- Minkowski

- Tanimoto
- Jaccard
- Mahalanobis

س474/ ما هي التقنية الجماعية ensemble technique التي تستخدمها Random Forest؟

ج/ الغابة العشوائية Random Forest هي نموذج جماعي يستخدم التعبئة bagging كطريقة للمجموعة وشجرة القرار كنموذج فردي.

س475/ ما هي فوائد التقليم pruning في خوارزمية شجرة القرار؟

ج / يعد تقليم شجرة القرار pruning عملية التخلص من الأشجار الفرعية غير الهامة بحيث لا يتم الضبط الزائد للبيانات قيد الدراسة. في عملية التقليم المسبق، يتم تقليم الشجرة أثناء بنائها، باتباع معايير مثل مؤشر جيني Gini index أو مقاييس اكتساب المعلومات information gain. يستلزم التقليم اللاحق تقليم الشجرة من الأسفل إلى الأعلى بعد بنائها.

س476/ أيهما أفضل، خوارزمية نايف بايز أم أشجار القرار؟

ج/ شجرة القرار هي نموذج تمييزي discriminative model، في حين أن Naive bayes هو نموذج توليدي generative model. أشجار القرار أكثر مرونة وسهولة. قد يؤدي تقليم pruning شجرة القرار إلى إهمال بعض القيم الأساسية في بيانات التدريب، مما قد يؤدي إلى دقة اقل.

س477/ ما هي مزايا استخدام نايف بايز للتصنيف؟

ج/ سوف يتقارب المصنف في Naïve Bayes بشكل أسرع من النماذج التمييزية مثل الانحدار اللوجستي، لذلك تحتاج إلى بيانات تدريب أقل. الميزة الرئيسية هي أنه لا يمكنه تعلم التفاعلات بين الميزات.

س478/ ما هي متجهات الدعم Support Vectors في SVM؟

ج / متجهات الدعم Support Vectors هي نقاط بيانات أقرب إلى المستوي الفائق hyperplane وتؤثر على موضع واتجاه المستوي الفائق. باستخدام متجهات الدعم هذه، نقوم بتعظيم هامش المصنف. سيؤدي حذف ناقلات الدعم إلى تغيير موضع المستوي الفائق. هذه هي النقاط التي تساعدنا في بناء SVM الخاص بنا.

س479/ ما هو الفرق بين Gini Impurity وEntropy في شجرة القرار؟

ج / يعتمد كسب المعلومات information gain على انخفاض الانتروبيا Entropy بعد تقسيم مجموعة البيانات على إحدى السمات. يتعلق إنشاء شجرة قرار بإيجاد السمة التي تُرجع أعلى مكاسب للمعلومات information gain .

Gini	Entropy
لها قيم داخل الفاصل الزمني [0، 0.5]	لها قيم داخل الفاصل الزمني [0، 1]
إنه أكثر تعقيدًا.	انها ليست معقدة.
إنه قياس لحساب نقص المعلومات.	قياسه هو احتمال عينة عشوائية يتم تصنيفها بشكل صحيح.

س480/ كيف يتم تقليم شجرة القرار؟

ج/ في التعلم الآلي، يعني التقليم pruning تبسيط شجرة القرار وتحسينها عن طريق قطع عقد الشجرة التي تسبب فرط التعلم overfitting. يمكن تقسيم عملية التقليم إلى نوعين:

- التقليم من أسفل إلى أعلى Bottom-up pruning : يبدأ الإجراء من العقدة الأخيرة
- التقليم من أعلى إلى أسفل Top-down pruning : يبدأ الإجراء عند العقدة الجذرية

يتم التقليم لزيادة الدقة التنبؤية لنموذج شجرة القرار.

س481/ كيف يمكنك اختيار k لـ k-means؟

ج / الطريقة الأكثر شيوعًا لاختيار k لخوارزمية k-mean هي استخدام طريقة الكوع elbow method. للقيام بذلك، تحتاج إلى حساب مجموع داخل الكتلة للأخطاء التربيعية (WSS) لقيم k مختلفة. يتم وصف WSS على أنها مجموع مربعات المسافة بين كل قيمة بيانات وcentroid الخاص بها.

ستختار بعد ذلك قيمة k التي يبدأ بها خطأ WSS في التلاشي.

س482/ما هي الاختلافات بين تحليلات univariate وbivariate وmultivariate؟

ج/

عندما يتم تحليل متغير واحد فقط من خلال الرسوم البيانية مثل المخططات الدائرية، يسمى التحليل أحادي المتغير univariate.

عندما تتم مقارنة الاتجاهات في متغيرين باستخدام الرسوم البيانية مثل مخططات التبعثر scatter plots، فإنه تحليل النوع ثنائي المتغير bivariate.

عندما يتم النظر في أكثر من متغيرين للتحليل لفهم ارتباطاتهما، يُطلق على التحليل اسم متعدد المتغيرات multivariate.

س483/ ما هي p-value؟

ج / تعبر القيمة P عن احتمال أن تكون الملاحظة التي يتم إجراؤها حول مجموعة البيانات فرصة عشوائية. أي قيمة p أقل من 5٪ هي دليل قوي يدعم الملاحظة وضد فرضية العدم null hypothesis. كلما زادت القيمة الاحتمالية، قل احتمال صحة النتيجة.

س484/ ماذا تعني قيمة P p-value حول البيانات الإحصائية statistical data؟

ج / في الإحصاء statistics، تُستخدم القيمة p لاختبار أهمية الفرضية الصفرية. تشير قيمة p أقل من 0.05 إلى أن هناك فرصة بنسبة 5٪ فقط أن تكون نتائج التجربة عشوائية ويجب رفض فرضية العدم null hypothesis. من ناحية أخرى، تشير القيمة الاحتمالية الأعلى، ولنقل 0.8، إلى أنه لا يمكن رفض الفرضية الصفرية لأن 80٪ من العينة لها نتائج عشوائية.

س485/ ماذا تعني كلمة "Naive" في Naive Bayes؟

ج / يعتمد نموذج خوارزمية Naive Bayes على نظرية بايز. يصف احتمالية وقوع حدث. يعتمد على المعرفة المسبقة بالظروف التي قد تكون مرتبطة بهذا الحدث المحدد.

س486/ اشرح خطوات بناء شجرة القرار decision tree؟

ج /

- خذ مجموعة البيانات بأكملها كمدخلات
- احسب إنتروبيا المتغير المستهدف ، بالإضافة إلى سمات التوقع
- احسب اكتساب المعلومات الخاصة بك لجميع السمات (نحصل على معلومات حول فرز كائنات مختلفة عن بعضها البعض)
- اختر السمة ذات أعلى اكتساب للمعلومات كعقدة جذر
- كرر نفس الإجراء على كل فرع حتى يتم الانتهاء من عقدة القرار لكل فرع

على سبيل المثال، لنفترض أنك تريد بناء شجرة قرار لتقرير ما إذا كان يجب عليك قبول عرض عمل أو رفضه. شجرة القرار لهذه الحالة كما هو موضح:

يتضح من شجرة القرار أنه يتم قبول العرض إذا:

- الراتب أكبر من 50000 دولار
- تستغرق الرحلة أقل من ساعة
- يتم تقديم الحوافز

س487/ كيف تبني نموذج غابة عشوائي random forest model؟

ج / نموذج الغابة العشوائية random forest model عبارة عن خوارزمية تعلم آلي وشكل من أشكال التعلم الخاضع للإشراف. يتم استخدامه بشكل شائع في مشاكل الانحدار والتصنيف. فيما يلي خطوات إنشاء نموذج غابة عشوائي:

1. من مجموعة بيانات ذات سجلات k، حدد n.
2. أنشئ أشجار قرارات فردية لكل من قيم البيانات n قيد الدراسة. يتم الحصول على نتيجة متوقعة من كل منهم.
3. يتم تطبيق خوارزمية التصويت voting algorithm على كل نتيجة.
4. يتم تعيين التنبؤ بأكبر عدد من الأصوات كنتيجة نهائية.

س488/ ما هو skewed Distribution uniform distribution؟

ج / التوزيع المنحرف skewed Distribution هو توزيع لا يتم فيه تسوية القيم الموجودة في مجموعة البيانات ويميل منحنى التوزيع نحو جانب واحد. التوزيع المنتظم uniform distribution من ناحية أخرى هو توزيع متماثل حيث يكون احتمال حدوث كل نقطة هو نفسه بالنسبة لنطاق معين من القيم في مجموعة البيانات.

س489/ كيف يمكنك التعامل مع الأنواع المختلفة من seasonality في نمذجة السلاسل الزمنية؟

ج / تحدث الموسمية **seasonality** في السلاسل الزمنية عندما تعرض السلاسل الزمنية نمطًا متكررًا بمرور الوقت. على سبيل المثال، تنخفض المبيعات الثابتة خلال موسم العطلات، وتزداد مبيعات مكيفات الهواء خلال الصيف وما إلى ذلك، وهي أمثلة قليلة على الموسمية في سلسلة زمنية.

تجعل الموسمية السلاسل الزمنية الخاصة بك غير ثابتة بسبب متوسط قيمة المتغيرات في فترات زمنية مختلفة. يُعرف التفريق بين السلاسل الزمنية **Differentiating a time series** عمومًا بأنه أفضل طريقة لإزالة الموسمية من سلسلة زمنية. يمكن تعريف الفروق الموسمية **Seasonal differencing** على أنها فرق عددي بين قيمة معينة وقيمة بتأخر دوري (أي **12**، في حالة وجود موسمية شهرية)

س490/ في التصميم التجريبي، هل من الضروري عمل التوزيع العشوائي randomization؟؟ إذا كانت الإجابة نعم، فلماذا؟

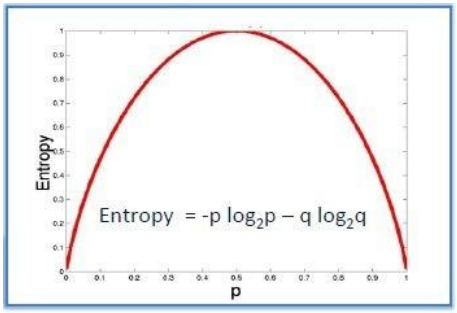
ج / نعم، من الضروري استخدام التوزيع العشوائي **randomization** أثناء تصميم التجارب. عن طريق التوزيع العشوائي، نحاول القضاء على التحيز قدر الإمكان. الغرض الرئيسي من التوزيع العشوائي هو أنه يتحكم تلقائيًا في جميع المتغيرات الكامنة **lurking variables**. تؤسس التجارب مع التوزيع العشوائي علاقة سببية أوضح بين المتغيرات التفسيرية **explanatory variables** ومتغيرات الاستجابة **response variables** من خلال التحكم في المتغيرات التفسيرية.

س491/ ما هو الانتروبيا Entropy واكتساب المعلومات Information Gain في خوارزمية شجرة القرار؟

ج / تسمى الخوارزمية الأساسية لبناء شجرة القرار **ID3**. يستخدم شجرة القرار الإنتروبيا **Enteropy** وكسب المعلومات **Information Gain** لإنشاء شجرة قرار.

- ❖ الإنتروبيا **Enteropy**

يتم إنشاء شجرة القرار من أعلى إلى أسفل من عقدة جذر وتتضمن تقسيم البيانات إلى مجموعات فرعية متجانسة. يستخدم **ID3** الانتروبيا للتحقق من تجانس العينة. إذا كانت العينة متجانسة تمامًا، فإن الإنتروبيا تكون صفرًا وإذا كانت العينة مقسمة بالتساوي فإنها تحتوي على واحد.



Entropy = $-0.5 \log_2 0.5 - 0.5 \log_2 0.5 = 1$

❖ كسب المعلومات **Information Gain**

يعتمد كسب المعلومات على انخفاض الانتروبيا بعد تقسيم مجموعة البيانات على سمة. يدور إنشاء شجرة قرار حول البحث عن السمات التي تُرجع أعلى مكاسب للمعلومات.

		Play Golf	
		Yes	No
Outlook	Sunny	3	2
	Overcast	4	0
	Rainy	2	3
Gain = 0.247			

		Play Golf	
		Yes	No
Temp.	Hot	2	2
	Mild	4	2
	Cool	3	1
Gain = 0.029			

		Play Golf	
		Yes	No
Humidity	High	3	4
	Normal	6	1
Gain = 0.152			

		Play Golf	
		Yes	No
Windy	False	6	2
	True	3	3
Gain = 0.048			

$$Gain(T, X) = Entropy(T) - Entropy(T, X)$$

G(PlayGolf, Outlook) = **E**(PlayGolf) – **E**(PlayGolf, Outlook)

= 0.940 – 0.693 = 0.247

س492/ متى ستستخدم SVM ومتى تستخدم Random Forest؟

ج/

- يمكن استخدام **SVM** إذا كانت البيانات خالية من القيم المتطرفة **outliers** بينما يمكن استخدام **Naïve Bayes** حتى إذا كانت تحتوي على قيم متطرفة (نظرًا لأنها بنيت في حزمة للعناية).
- يناسب **SVM** بشكل أفضل نموذج تصنيف النص والغابة العشوائية تناسب مشكلة التصنيف ذات الحدين / متعدد الحدود.
- تتولى **Random Forest** الاهتمام بمشكلة الضبط الزائد بمساعدة تقليم الأشجار **tree pruning**

س493/ ما هي المعرفة الرياضية الأساسية وراء Naïve Bayes؟

ج/ نظرية بايز **Bayes Theorem**.

س494/ متى تستخدم Random Forest ومتى تستخدم XGBoost؟

ج / إذا كنت تريد استخدام جميع المعالجات الأساسية في نظامك، فانتقل إلى **XGBoost** (نظرًا لأنه يدعم المعالجة المتوازية) وإذا كانت بياناتك صغيرة، فانتقل إلى الغابة العشوائية **Random Forest**.

س495/ ما هو الهامش margin، النواة kernels، التنظيم Regularization في SVM؟

ج/

- الهامش **margin**: يشار إلى المسافة **hyper plane** وأقرب نقاط البيانات باسم "الهامش"
- النواة **Kernel**: هناك ثلاثة أنواع من الانوية تحدد نوع البيانات التي تتعامل معها: **1**) الخطي **Linear**، **2**) الشعاعي **Radial**، **3**) متعدد الحدود **Polynomial**.
- التنظيم **Regularization**: تخبر معلمة التنظيم (التي يطلق عليها غالبًا باسم معلمة **C** في مكتبة **sklearn** في **Python**) تحسين **SVM** إلى أي مدى تريد تجنب سوء التصنيف لكل مثال تدريبي.

س496/ ما هي الطريقة المختلفة لتقسيم الشجرة في شجرة القرار؟

ج/ كسب المعلومات **Information gain** ومؤشر جيني **gini index**.

س497/ ما هو ضعف خوارزمية شجرة القرار؟

ج/

- غير مناسب للمتغير المستمر / المنفصل

- أداء ضعيف على البيانات الصغيرة

س498/ ما هي hyperplane في SVM؟

ج / إنه خط يقسم مساحة متغير الإدخال ويتم تحديده لفصل النقاط في مساحة متغير الإدخال بشكل أفضل عن طريق فئتها (1/0، نعم/لا).

س499/ الغابة العشوائية بها 1000 شجرة، خطأ تدريب: 0.0 وخطأ التحقق 20.00، ما المشكلة هنا؟

ج / إنه المثال الكلاسيكي للضبط الزائد **overfitting**. إنه لا يعمل بشكل جيد على البيانات غير المرئية **unseen data**. قد نضطر إلى ضبط نموذجنا باستخدام التحقق المتقاطع **cross validation** وتقنيات أخرى للتغلب على الضبط الزائد **overfitting**.

س500/ ما هي نظرية بايز Bayes's Theorem وكيف يتم استخدامها في التعلم الآلي؟

ج/ نظرية بايز هي طريقة لحساب الاحتمال الشرطي أي. إيجاد احتمال وقوع حدث بناءً على الاحتمال المعطى لأحداث أخرى وقعت بالفعل. رياضيا، تم ذكره على أنه:

$$P(A|B) = \{P(B|A).\ P(A)\}/P(B)$$

أصبحت نظرية بايز أداة مفيدة للغاية في التعلم الآلي التطبيقي. يوفر طريقة للتفكير في العلاقة المشتركة بين البيانات والنماذج.

نموذج التعلم الآلي هو طريقة محددة للتفكير في العلاقة المهيكلة في البيانات مثل العلاقات المشتركة من خلال الإدخال (x) والمخرج (y).

إذا كان لدينا بعض المعرفة السابقة بالمجال حول الفرضية، فيمكن أن تساعد نظرية بايز في حل مشكلات التعلم الآلي.

س501/ ما هي الحاجة إلى إضافة العشوائية randomness في عملية تهيئة الوزن؟

ج / إذا قمت بضبط الأوزان على الصفر، فإن كل خلية عصبية في كل طبقة ستنتج نفس النتيجة ونفس قيمة التدرج أثناء الانتشار الخلفي. لذلك، لن تكون الشبكة العصبية قادرة على تعلم الدالة حيث لا يوجد عدم تناسق بين الخلايا العصبية. ومن ثم، فإن العشوائية في عملية تهيئة الوزن أمر بالغ الأهمية.

س502/ كيف يؤثر قطع الاتصال العشوائي Randomized Connection Dropping على نموذج التعلم العميق؟

ج / غالبًا ما يؤدي الإسقاط العشوائي للاتصالات بين الطبقات المختلفة في شبكة عصبية متعددة الطبقات إلى نماذج متنوعة تُستخدم فيها مجموعات مختلفة من الميزات لبناء المتغيرات المخفية.

يؤدي قطع الروابط بين الطبقات إلى إنشاء نماذج أقل قوة **less powerful models** بسبب إضافة قيود إلى عملية بناء النموذج. ومع ذلك، نظرًا لإسقاط التوصيلات العشوائية المختلفة من نماذج مختلفة، فإن التنبؤات من النماذج المختلفة متنوعة **diverse** للغاية.

س503/ هل نايف بايز Naïve Bayes سيئة؟ إذا كانت الإجابة بنعم، فما هي الجوانب.

ج / نايف بايز **Naïve Bayes** هي خوارزمية تعلم الآلة تعتمد على نظرية بايز **Bayes Theorem**. يستخدم هذا لحل مشاكل التصنيف. يعتمد على افتراضين، أولاً ، كل ميزة/سمة موجودة في مجموعة البيانات مستقلة عن أخرى ، وثانيًا ، كل ميزة لها أهمية متساوية. ولكن تبين أن افتراض نايف بايز غير مواتٍ. نظرًا لأنه يفترض أن الميزات مستقلة عن بعضها البعض، ولكن في سيناريوهات الحياة الواقعية، لا يمكن أن يكون هذا الافتراض صحيحًا نظرًا لوجود بعض التبعية دائمًا في مجموعة الميزات المحددة. عيب آخر لهذه الخوارزمية هو "مشكلة التردد الصفري **zero-frequency problem**" حيث يخصص النموذج القيمة صفر لتلك الميزات في مجموعة بيانات الاختبار التي لم تكن موجودة في مجموعة بيانات التدريب.

س504/ أيهما أفضل - غابة عشوائية random forest أم أشجار قرارات متعددة multiple decision trees؟

س505/ لماذا أداء XGBoost أفضل من SVM؟

ج / السبب الأول هو أن **XGBoos** هي طريقة تجميعية **ensemble method** تستخدم العديد من الأشجار لاتخاذ قرار حتى تكتسب القوة من خلال تكرار نفسها.

SVM هو فاصل خطي ، عندما لا تكون البيانات قابلة للفصل خطيًا ، يحتاج **SVM** إلى **Kernel** لعرض البيانات في مساحة حيث يمكنه فصلها ، فهناك أكبر قوتها وضعفها ، من خلال القدرة على عرض البيانات في مساحة عالية الأبعاد يمكن لـ **SVM** العثور على فصل خطي لأي بيانات تقريبًا ولكن في نفس الوقت يحتاج إلى استخدام **Kernel** ويمكننا القول بأنه لا توجد نواة **kernel** مثالية لكل مجموعة بيانات.

س506/ ما هي أفكارك حول استخدام GPT 3 لعملك؟

ج/ GPT-3 ، أو الجيل الثالث من المحولات التوليدية المدربة مسبقًا Generative Pre-trained Transformer، هي آلة شبكة عصبية. GPT-3 هو متنبئ بالنص. بالنظر إلى نص أو عبارة، يعرض GPT-3 استجابة شبيهة بالإنسان لإكمال النص بلغة طبيعية. يحتوي GPT-3 على مجموعة واسعة من التطبيقات التي تخدم الصناعة اليوم. إنها أداة قوية يمكنها إنشاء تطبيقات للرد على استفسارات العملاء، ومترجم اللغة (على سبيل المثال، طرح سؤال باللغة الإنجليزية وتوقع إجابة باللغة الإسبانية) وما إلى ذلك.

يمكن لـ GPT3 أيضًا القيام بكل شيء بدءًا من إنشاء جداول البيانات إلى إنشاء CSS المعقدة أو حتى نشر مثيلات Amazon Web Services (AWS). لذا، هل يمكن أن يساعد استخدام GPT-3 في عملك؟ حسنًا، يمكن أن يساعد في نواح كثيرة. كل هذا يتوقف على ما تحتاجه للقيام به، لكنه نموذج تعلم عميق متعدد الاستخدامات يتم تطبيقه على العديد من التطبيقات.

بعض التطبيقات الأخرى لـ GPT-3 التي يمكنك استخدامها على الأرجح في عملك هي:

- ❖ توليد رسائل البريد الإلكتروني من أوصاف قصيرة. تطبيق يمكنه توسيع الوصف المختصر المقدم إلى بريد إلكتروني احترافي منسق وصحيح نحويًا.
- ❖ إنشاء أكواد بيثون من الوصف. قم بإنشاء رمز واجهة برمجة تطبيقات Flask (Python) فقط عن طريق وصف الوظائف باللغة الإنجليزية باستخدام GPT-3.
- ❖ إنشاء نموذج التعلم العميق بناءً على الوصف.

س507/ ما الرؤية الحاسوبية Computer Vision وما علاقته بالذكاء الاصطناعي وماهي تطبيقاته؟

ج/ تُعرّف الرؤية الحاسوبية Computer Vision، التي غالبًا ما يتم اختصارها باسم CV، بأنها مجال دراسة يسعى إلى تطوير تقنيات لمساعدة أجهزة الكمبيوتر على "رؤية" وفهم محتوى الصور الرقمية مثل الصور الفوتوغرافية ومقاطع الفيديو.

ترتبط الرؤية الحاسوبية ارتباطًا وثيقًا بالذكاء الاصطناعي، حيث يجب على الكمبيوتر تفسير ما يراه، ثم إجراء التحليل المناسب أو التصرف وفقًا لذلك.

الهدف من الرؤية الحاسوبية هو محاكاة الرؤية البشرية باستخدام الصور الرقمية من خلال ثلاث مكونات معالجة رئيسية، يتم تنفيذها واحدًا تلو الآخر:

1. اقتناء الصور
2. معالجة الصور
3. تحليل الصورة وفهمها

التقنيات المطورة من أجل رؤية الكمبيوتر لها العديد من التطبيقات في مجالات الروبوتات والتفاعل بين الإنسان والحاسوب والتصور، على سبيل المثال لا الحصر:

1. التعرف على الحركة
2. الواقع المعزز
3. السيارات ذاتية القيادة
4. الروبوتات المحلية/الآلية
5. استعادة الصورة مثل تقليل الضوضاء

س508/ ما هو الفرق بين اكتشاف الوجه Face Detection والتعرف عليه Face Recognition؟

ج/ اكتشاف الوجه هو تقنية كمبيوتر تُستخدم في مجموعة متنوعة من التطبيقات التي تحدد الوجوه البشرية في الصور الرقمية. تركز خوارزميات اكتشاف الوجه على اكتشاف الوجوه البشرية الأمامية. إنه مشابه لاكتشاف الصور حيث يتم مطابقة صورة الشخص شيئًا فشيئًا.

ومع ذلك، فإن أحد أهم تطبيقات اكتشاف الوجه هو التعرف على الوجه. يصف التعرف على الوجوه تقنية المقاييس الحيوية التي تذهب إلى ما هو أبعد من التعرف على وجود وجه بشري. إنها تحاول في الواقع تحديد من هو وجهه. تعمل العملية باستخدام تطبيق كمبيوتر يلتقط صورة رقمية لوجه الفرد (يتم التقاطها أحيانًا من إطار فيديو) ويقارنها بالصور الموجودة في قاعدة بيانات السجلات المخزنة. على الرغم من أن التعرف على الوجه ليس دقيقًا بنسبة 100٪، إلا أنه يمكن أن يحدد بدقة عندما تكون هناك فرصة قوية لأن يتطابق وجه الشخص مع شخص ما في قاعدة البيانات.

هناك الكثير من تطبيقات التعرف على الوجوه. يتم استخدام التعرف على الوجه بالفعل لإلغاء قفل الهواتف وتطبيقات محددة. يستخدم التعرف على الوجه أيضًا للمراقبة الحيوية.

س509/ ما هو YOLOv3؟

ج/ YOLOv3 (You Only Look Once, Version 3) عبارة عن خوارزمية لاكتشاف الكائنات في الوقت الفعلي تحدد كائنات معينة في مقاطع الفيديو أو البث الحي أو الصور. تستخدم خوارزمية التعلم الآلي YOLO الميزات التي تعلمتها شبكة عصبية تلافيفية CNN عميقة لاكتشاف كائن.

س510/ هل الرؤية الحاسوبية تستخدم التعلم العميق؟

ج/ بشكل أساسي، تستخدم الرؤية الحاسوبية شبكات CNN والتعلم العميق لأداء تعلم عالي السرعة وغير خاضع للإشراف على المعلومات المرئية لتدريب أنظمة التعلم الآلي على تفسير البيانات بطريقة تشبه إلى حد ماكيفية عمل العين البشرية.

س511/ هل OpenCV تعلم الآلي أم تعلم عميق؟

ج/ OpenCV هي مكتبة مفتوحة المصدر لمهام الرؤية الحاسوبية ومعالجة الصور في التعلم الآلي. يوفر OpenCV مجموعة ضخمة من الخوارزميات ويهدف إلى الرؤية الحاسوبية في الوقت الفعلي. من ناحية أخرى، يعد Keras إطارًا تعليميًا عميقًا لتمكين التجريب السريع مع التعلم العميق.

س512/ ما هي الآثار الأخلاقية لاستخدام التعلم العميق؟

سؤال

هل تعلّم الآلة فن ام علم؟

المصادر

1. Top 60+ Machine Learning Interview Questions - 2022 https://mindmajix.com/machine-learning-interview-questions#explaim-what-is-precision-and-recall
2. 12 Deep Learning Interview questions you shouldnot be missed : https://medium.com/jp-tech/12-deep-learning-interview-questions-you-should-not-be-missed-part-1-8a61f44cadac
3. 30 Deep Learning Interview Questions (and Sample Answers) : https://www.indeed.com/career-advice/interviewing/deep-learning-interview-questions
4. 30 Questions to test your understanding of Logistic Regression: https://www.analyticsvidhya.com/blog/2017/08/skilltest-logistic-regression/
5. 51 Machine Learning Interview Questions with Answers : https://www.springboard.com/blog/data-science/machine-learning-interview-questions/
6. 60 Advanced Deep Learning Interview Questions (ANSWERED) To Crush Your ML Interview : https://www.mlstack.cafe/blog/deep-learning-interview-questions
7. 87 Commonly Asked Data Science Interview Questions: https://www.springboard.com/blog/data-science/data-science-interview-questions/
8. 100 Deep Learning Interview Questions and Answers for 2022: https://www.projectpro.io/article/100-deep-learning-interview-questions-and-answers-for-2021/419
9. 100+ Data Science Interview Questions and Answers for 2022: https://www.projectpro.io/article/100-data-science-interview-questions-and-answers-for-2021/184
10. 130 Data Science Interview Questions and Answers [Latest] : https://data-flair.training/blogs/data-science-interview-questions/
11. Deep Learning Interview Questions _ Deep Learning Questions: https://www.analyticsvidhya.com/blog/2020/04/comprehensive-popular-deep-learning-interview-questions-answers/
12. Deep Learning Interview Questions and Answers in 2022: https://www.mygreatlearning.com/blog/deep-learning-interview-questions/

13. Interview Query _ Top 50 Machine Learning Interview Questions for 2022: https://www.interviewquery.com/p/machine-learning-interview-questions
14. Machine Learning Engineer Interview Questions (2022 Guide) : https://brainstation.io/career-guides/machine-learning-engineer-interview-questions
15. Popular Machine Learning Interview Questions : https://towardsdatascience.com/popular-machine-learning-interview-questions-91d569afe147
16. Top 50 Data Science Interview Questions and Answers (2022): https://www.guru99.com/data-science-interview-questions.html
17. Top 50 Deep Learning Interview Questions & Answers 2022 [updated]: https://intellipaat.com/blog/interview-question/deep-learning-interview-questions/
18. Top 80 Data Science Interview Questions and Answers 2022 : https://www.simplilearn.com/tutorials/data-science-tutorial/data-science-interview-questions
19. Top 80+ Data Science Interview Questions and Answers for 2022: https://intellipaat.com/blog/interview-question/data-science-interview-questions/
20. Top 100 Machine Learning Interview Questions & Answers: https://www.mygreatlearning.com/blog/machine-learning-interview-questions/
21. Top 100+ Machine learning interview questions and answers: https://www.i2tutorials.com/top-100-machine-learning-interview-questions-and-answers/
22. Top 122 Data Science Interview Questions and Answers in 2022: https://www.edureka.co/blog/interview-questions/data-science-interview-questions/
23. Top 200 Deep Learning interview questions and answers: https://www.i2tutorials.com/top-deep-learning-interview-questions-and-answers/
24. Top Data Science Interview Questions and Answers(2022): https://www.simplilearn.com/tutorials/deep-learning-tutorial/deep-learning-interview-questions

512

Question & Answers

In

ML, DL and DS

By:

Dr. Alaa Taima