

Assignment1_Machine Learning

1. كيف يمكنك تعريف التعلم الآلي؟
➤ التعلم الآلي هو مجال في الذكاء الاصطناعي يتضمن تطوير الخوارزميات والنماذج عن طريق برمجة الآلة لتتعلم من البيانات لحل مشكلة معينة.
2. ما هما المهمتان الخاضعتان للإشراف الأكثر شيوعاً؟
➤ التعلم الإشرافي: Regression, Classification.
3. هل يمكنك تسمية أربع مهام شائعة غير خاضعة للإشراف؟
➤ التعلم غير الإشرافي: Association Rule Learning, Dimensionality Reduction, Clustering, visualization.
4. ما نوع خوارزمية التعلم الآلي التي ستستخدمها للسماح للروبوت بالسير في مختلف التضاريس غير المعروفة؟
➤ خوارزميات التعلم التعزيزي (Reinforcement Learning).
5. ما نوع الخوارزمية التي ستستخدمها لتقسيم عملائك إلى مجموعات متعددة؟ اشرح اجابتك.
➤ إذا كانت لديك بيانات مصنفة حيث تم تعريف labeled بالفعل، فسيتم استخدام خوارزميات التعلم الخاضعة للإشراف مثل Decision Tree، في الحالات التي لا توجد فيها label، يتم استخدام تقنيات التعلم غير الخاضعة للرقابة مثل خوارزميات Clustering.
6. هل يمكنك تأطير مشكلة اكتشاف البريد العشوائي كمشكلة تعليمية خاضعة للإشراف أم مشكلة تعليمية غير خاضعة للإشراف؟ اشرح اجابتك.
➤ تعتبر مشكلة اكتشاف البريد العشوائي (SPAM) كمشكلة تعلم خاضعة للإشراف عن طريق استخدام Classification التي بدورها تصنفه 0 (SPAM) أو 1 (Not SPAM) ... او العكس.
7. هل يمكنك تسمية (مع مثال) أربعة من التحديات الرئيسية في التعلم الآلي؟
➤ الحل:
▪ overfitting the data: مثال توقع أسعار المنازل فإننا نقوم بتدريب Model لتوقع الأسعار وكأنه حافظ البيانات ولم يفهمها فإذا كان معقدا ويستقبل الكثير من Feature الغير ضرورية، وهناك بعض القيم الشاذة في Data فذلك يؤدي الى أداء سيء على البيانات الجديدة.

- **underfitting the data**: مثال توقع أسعار المنازل استخدام Model باستخدام عدد قليل جداً من Feature مما يؤدي إلى توقعات سيئة على مجموعة التدريب والبيانات الجديدة.
- **lacking in data**: مثال تدريب Model يقوم بالتعرف على الكلام وتحويله إلى نص، ولكن لديك مجموعة قليلة من التسجيلات الصوتية، مما يؤثر على قدرة Model في عدم تعلم اللهجات والأصوات المختلفة في الكلام وهذا يؤثر على أدائه عند التعامل مع الناس.
- **nonrepresentative data**: مثال تدريب Model للتعرف على معاملات الاحتمالية في نظام المصارف، مثل إذا كانت البيانات المستخدمة للتدريب تتألف بشكل رئيسي من معاملات قديمة وغير دقيقة فإنها تؤثر على مصدقيه البيانات بشكل عام.

8. ما نوع خوارزمية التعلم التي تعتمد على مقاييس التشابه لإجراء التنبؤات؟
 ➤ خوارزمية التعلم التي تعتمد على مقياس التشابه لإجراء التنبؤات هي : Instance-based algorithm.
9. إذا كان أداء النموذج الخاص بك رائعة فيما يتعلق ببيانات التدريب ولكنه يعم بشكل سيئ على الحالات الجديدة، فماذا يحدث؟ هل يمكنك تسمية ثلاثة حلول ممكنة؟
 ➤ الذي يحصل هو مشكلة ال overfitting the data ولمعالجة المشكلة يمكن التالي:
- التقييم المتقاطع (Cross-Validation): باستخدام أساليب مثل التقييم المتقاطع k-fold يُساعد في تقييم أداء النموذج على مجموعات مختلفة من البيانات.
 - زيادة كمية البيانات: إضافة بيانات متنوعة وتمثيلية إلى مجموعة البيانات التدريبية يمكن أن يساعد النموذج في تعلم أنماط أكثر تعميمًا ويقلل من حدوث التجاوز
 - (هندسة الميزات (feature)): اختيار السمات المهمة بعناية واستبعاد السمات غير المهمة أو المتكررة وإظهار سمات جديدة بدمج بعض السمات ببعض يمكن أن يمنع النموذج من التكيف مع الضوضاء في البيانات.
10. ما هو الفرق بين overfitting و underfitting؟
 ➤ التجاوز والتقليل هما مشكلتان متناقضتان يواجهان النماذج في تعلم الآلة: التجاوز يعني أن النموذج متكيف بشكل زائد مع بيانات التدريب وكأنه حفظ البيانات ولم يفهمها، بينما التقليل يشير إلى أن النموذج لم يتعلم بشكل كافٍ من بيانات التدريب ليقدم تنبؤات دقيقة على البيانات التدريبية والجديدة. الحفاظ على توازن بين هذين الطرفين مهم للنموذج ليتنبأ بشكل جيد بالبيانات الجديدة.

11. ما هي مجموعة الاختبار ولماذا تريد استخدامها؟
 ➤ مجموعة الاختبار (Test set) تشير إلى جزء من مجموعة البيانات يُحفظ به بشكل منفصل عن بيانات التدريب، ويُستخدم لتقييم أداء نموذج التعلم الآلي بعد تدريبه على بيانات التدريب. وهو أمر ضروري لأنه يحدد أو يقيم أداء النموذج (Model) ويمنع التجاوز (overfitting) و مقارنة النماذج واختيار الأفضل من بينهم .