

Groupe 6 : [Partie 1] Examen ML



Brouillon enregistré

***Obligatoire**

Adresse e-mail *

Nom complet *

1. Python est un langage _____? *

- ☐ A. interprété
- ☐ B. machine
- ☐ C. compilé
- ☐ D. binaire

2. Laquelle des fonctions suivantes convertit un « string » en « float » en python? *

- ☐ A. str(x)



- ☐ B. float(x)
- ☐ C. long(x [,base])
- ☐ D. int(x [,base])

3. Quelle est la fonction qui compare les éléments des deux listes? *

- ☐ A. cmp(list1, list2)
- ☐ B. eq(list1, list2)
- ☐ C. len(list1, list2)
- ☐ D. max(list1, list2)

4. En python, quel mot clé est utilisé pour commencer une fonction? *

- ☐ A. function
- ☐ B. fun
- ☐ C. def
- ☐ D. import

5. Qu'est-ce que le machine learning ? *

6. Quel est le lien entre machine learning, deep learning et intelligence artificielle ? *



7. Quelle est la différence entre apprentissage supervisé et apprentissage non supervisé ? *

8. L'apprentissage non supervisé consiste à utiliser un algorithme qui n'a pas besoin d'être supervisé par un scientifique des données ? *

☐ VRAI

☐ Faux

9. L'apprentissage supervisé nécessite des données historiques pour lesquelles on connaît déjà les étiquettes ? *

☐ VRAI

☐ Faux

10. Quels sont les deux catégories d'arbres de décision ? *

☐ Les arbres de transmission

☐ Les arbres de classification

☐ Les arbres de régression

☐ Les arbres de développement



11. Qu'appelle-t-on "Forêt aléatoire" ? *

- ☐ Un arbre de décision basé sur des probabilités
- ☐ Un arbre de décision réalisé de manière aléatoire par un ordinateur
- ☐ Un ensemble de plusieurs arbres de décision

12. Expliquez comment gérer les données manquantes ou corrompues dans un ensemble de données. (les méthodes utilisées et la librairie ...) *

Vous devez identifier les données de recherche et supprimer les lignes / colonnes, ou les remplacer par d'autres valeurs.

13. Quelle est la différence entre la fonction de coût (cost function) et la descente de gradient (Gradient descent) ? *



Groupe 6 : [Partie 1] Examen ML



*Obligatoire

Evaluation d'un modèle ML

Les étapes de création d'un modèle ML

Cela dépendra en grande partie du modèle en question, vous pourrez donc poser des questions de clarification. Mais généralement, le processus est le suivant: 1. Comprendre le modèle commercial et l'objectif final.2. Rassemblez les acquisitions de données. 3. Effectuer le nettoyage des données. 4. Analyse exploratoire de base des données. 5. Utilisez des algorithmes d'apprentissage automatique pour développer un modèle. 6. Utilisez un ensemble de données inconnu pour vérifier l'exactitude

TP / True Positive: le cas était positif, et il a été prédit comme positif. TN / True Negative: le cas était négatif, et il a été prédit comme négatif FN / False Negative: le cas était positif, mais il a été prédit comme négatif FP / False Positive: le cas était négatif, mais il a été prédit comme positif

		Predicted	
		Apples	Oranges
Actual	Apples	TP = 10	FP = 35
	Oranges	FN = 40	TN = 15



14. Calculer l'accuracy, la précision et rappel (recall) *

Votre réponse

15. Pourquoi ne pas utiliser la somme des erreurs au carré (SSE) comme fonction de coût dans la régression logistique ? *

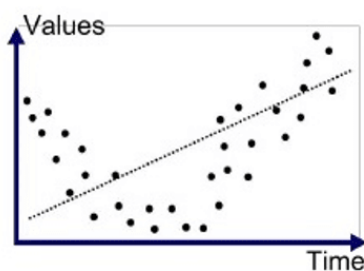
Votre réponse

16. Lors d'une régression simple, si le R^2 vaut 1, les points sont-ils alignés ? *

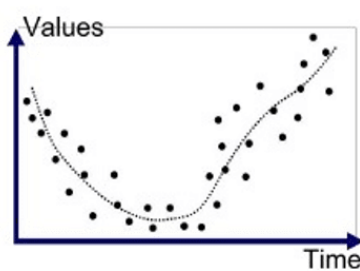
- ☐ Oui
- ☐ Non
- ☐ Pas obligatoirement.

Overfitting

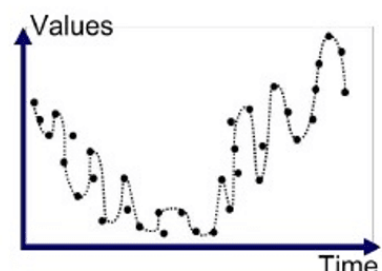
L'overfitting est le plus grand ennemi du data scientist. Il survient lorsque le modèle essaye de trop coller aux données. Si bien qu'il n'est plus généralisable.



Underfitted

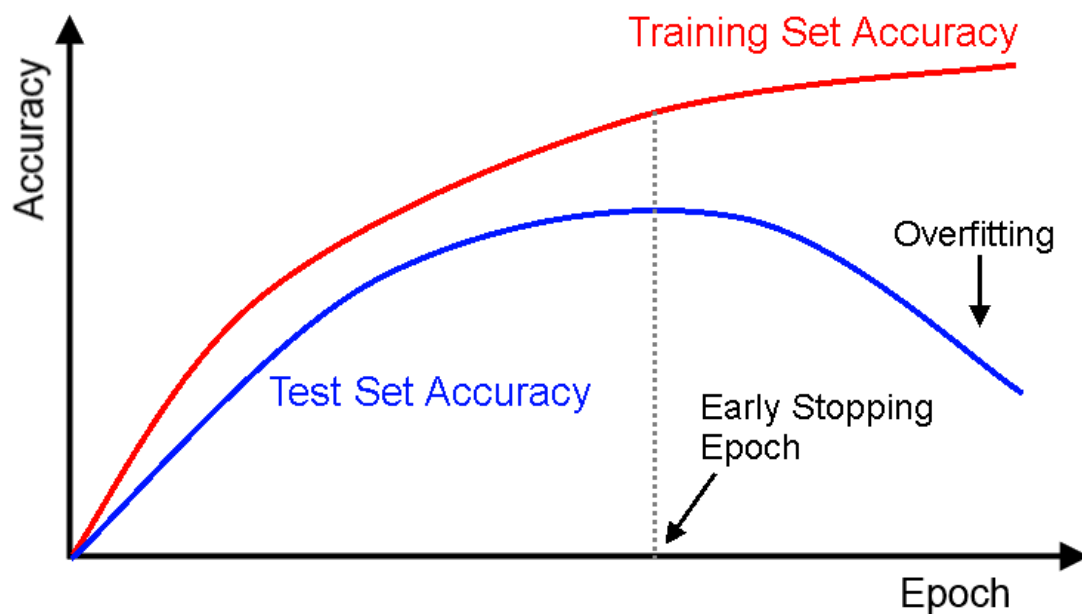


Good Fit/Robust



Overfitted

Avant de savoir comment l'éviter, nous devons apprendre à le détecter. Sur cette courbe (que vous devrez toujours tracer pour vérifier les performances du modèle) on voit qu'à partir d'un certain point, notre précision sur les données de test chutent. Cela veut dire que le modèle commence à être de moins en moins efficace. On fait de l'overfitting.



17. Plusieurs méthodes existent pour éviter l'overfitting, mentionner minimum trois techniques :

Votre réponse

Commentaires et/ou questions

Votre réponse