

Groupe 6 : [Partie 1] Examen ML

sajot.matthieu@gmail.com [Changer de compte](#)



Enregistrement...

***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) *

accuracy = $(TP + TN) / (TP + TN + FN + FP) = 10 + 15 / 10 + 15 + 40 + 35 = 25 / 100 = 1/4$

precision = $TP / (TP + FP) = 10 / 10 + 35 = 10/45 = 0.22$

recall = $TP / (TP + FN) = 10 / 10 + 40 = 1/5$

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

La somme des erreurs au carré est plus efficace sur les fonctions convexes, ce qui n'est pas le cas de la régression logistique. Il existe d'autres méthodes de cost fonction pour la régression logistique comme la matrice de confusion.

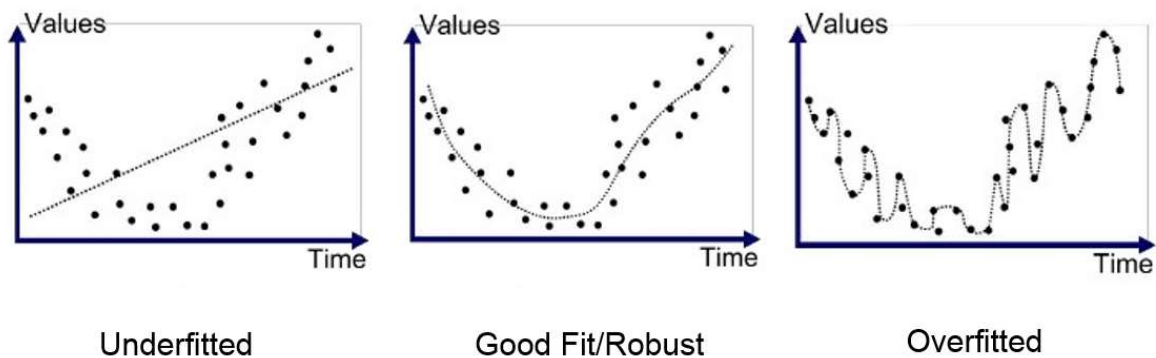
16. Lors d'une régression simple, si le R2 vaut 1, les points sont-ils alignés ? *

- ☒ Oui
- ☐ Non
- ☐ Pas obligatoirement.

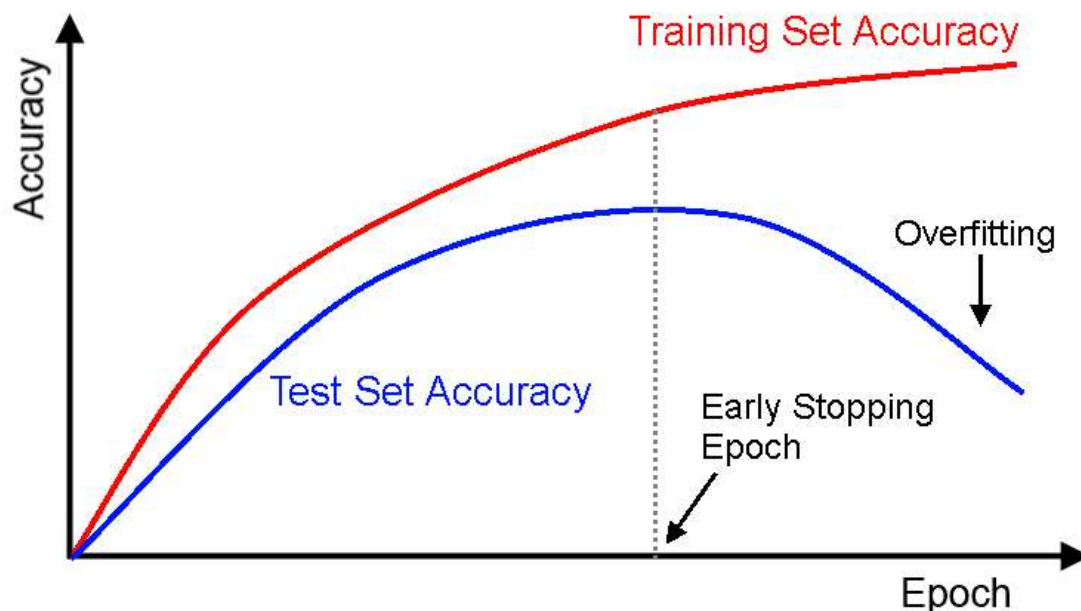
Overfitting



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.



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 :

Réduire le nombre de features, réduire l'importance des features, ajouter des données de train, arrêter le train dès qu'on atteint l'overfitting ou utiliser le cross-validation.

Commentaires et/ou questions

Votre réponse

Retour

Envoyer

Effacer le formulaire

N'envoyez jamais de mots de passe via Google Forms.

Ce contenu n'est ni rédigé, ni cautionné par Google. [Signaler un cas d'utilisation abusive](#) - [Conditions d'utilisation](#) - [Règles de confidentialité](#)

Google Forms

