

Examen - Machine Learning supervisé

Dr. Matthieu Cisel

Décembre 2022

1 Introduction

S'il n'y a pas de limites en termes de nombre de mots, tâchez de donner des explications aussi exhaustives que possible, et d'illustrer quand c'est possible votre propos avec des schémas que vous aurez construits vous-même, ou glanés sur Internet, mais qui doivent être commentés.

2 Arbres de décision : concepts de base

1. Expliquer comme l'indice de Gini ou l'entropie est utilisé (sans entrer dans les formules mathématiques, juste le principe), lors de la création d'un arbre de décision
2. Expliquer le concept de bootstrap
3. Expliquer le concept de bagging
4. Quelle est la différence entre une approche de type bagging et une forêt aléatoire ?

3 Métriques de performance

1. Expliquer la différence entre les métriques RMSE et MAE. Dans quel contexte les emploie-t-on ? Quel est "l'avantage" du RMSE par rapport au MAE ?
2. Expliquer les concepts de faux positif, de faux négatif, de sensibilité et de spécificité
3. Comment une courbe de type ROC est-elle construite ?
4. Que signifie l'AUC ? Qu'est-ce qu'un bon AUC ? L'AUC d'un modèle peu performant ?
5. Expliquer le concept de recall, ou rappel
6. Expliquer le concept de précision

3.1 Tuning d'un modèle

1. A quoi sert une grid search ?
2. Citer plusieurs hyperparamètres que l'on peut faire varier avec un arbre de décision