

SORBONNE UNIVERSITÉ

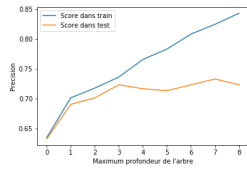
RAPPORT DE ARF

TME 1 - 6

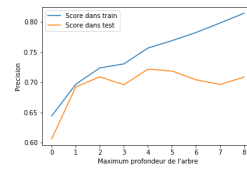
SAMU TAMMINEN

Table des matières

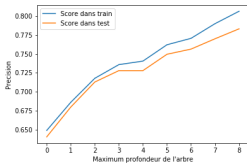
1	TME 1 : Arbres de décision, sélection de modèles	2
1.1	Q1.3 Interpretation de l'entropie	2
1.2	Q1.4 - 6 La profondeur d'arbre	2
1.3	Q1.7 - Sur et sous apprentissage	2
2	TME 2 : Estimation de densité	3
3	TME 3 : Descente de gradient	3
4	TME 4 et 5 : Perceptron	3
5	TME 5 : Support Vector Machine	3
5.1	Pistes de recherche	3



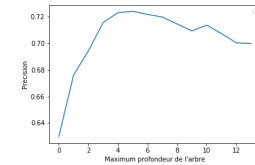
(a) Donnée apprentissage et test : 0.2 / 0.8



(b) Donnée apprentissage et test : 0.5 / 0.5



(c) Donnée apprentissage et test : 0.8 / 0.2



(d) Validation croisee

FIGURE 1 – Score d'arbre de decision en faisant varier le profondeur

1 TME 1 : Arbres de décision, sélection de modèles

1.1 Q1.3 Interpretation de l'entropie

Calculer également la différence entre l'entropie et l'entropie conditionnelle pour chaque attribut. A quoi correspond une valeur de 0 ? une valeur de 1 ? Quel est le meilleur attribut pour la première partition ?

Quand la différence entre l'entropie et l'entropie conditionnelle est 0 quand la partition n'est pas du tout bon et un, quand la partition est le meilleur possible.

1.2 Q1.4 - 6 La profondeur d'arbre

En augmentant la profondeur maximale d'arbre, on obtient meilleur score. Mais ces scores ne sont pas un indicateur fiable, si nous ne faisons pas différence entre les données training et test.

1.3 Q1.7 - Sur et sous apprentissage

Le partition 0.8 / 0.2 dans figure ?? (c) semble d'être le plus robust pour test et training.

Quand il y a que peu d'exemples d'apprentissage, l'erreur est plus élevée dans le test. Par contre quand on a beaucoup des d'exemples d'apprentissage, l'erreur dans le test est moins d'élévé.

Avec validation croisée on obtient un score plus fiable et stable. Dans la figure (d) nous pouvons constater, comment le score augmente au début avec la profondeur maximale, mais à la fin il tombe. Ça c'est à cause de sous et sur apprentissage. Au début, on généralise beaucoup avec un arbre peu profond. Quand la profondeur se monte sur cinq, nous commençons à apprendre les données d'apprentissage à la façon trop détaillée et le score sur le test se tombe.

- 2 TME 2 : Estimation de densité
- 3 TME 3 : Descente de gradient
- 4 TME 4 et 5 : Perceptron
- 5 TME 5 : Support Vector Machine
- 5.1 Pistes de recherche