

Fouille de données et medias sociaux

TP4 : Bagging et Boosting

Olivier Schwander

9 octobre 2017

Exercice 1 *Bagging*

Question 1

Écrivez une fonction (ou une classe) implémentant la méthode Bagging.

Question 2

Appliquez cette méthode sur des arbres de décisions.

Question 3

Évaluez et commentez l'erreur en généralisation par rapport à un arbre unique.

Exercice 2 *Boosting*

Question 1

Montrer que le coefficient de pondération des hypothèses vaut bien ce qu'il vaut.

Question 2

Écrivez une fonction (ou une classe) implémentant la méthode AdaBoost.

Question 3

Comment adapter un classifieur qui ne gère pas naturellement des pondérations pour qu'il puisse en tenir compte ?

Question 4

Appliquez cette méthode sur des *stumps* (arbres de décision à un nœud).

Question 5

Tracez les courbes de l'erreur d'apprentissage et de l'erreur de test en fonction du nombre de classifieurs.

Question 6

Affichez les points associés aux poids les plus élevés (donc les plus difficiles à classer). Sont-ils bien classés désormais ? Étaient-ils bien classés avec un seul arbre de décision ? Avec le Bagging ?

Question 7

Appliquez AdaBoost à des arbres de décision de profondeur plus grande.

Comment se comporte la généralisation quand la profondeur des arbres augmente ?