

Random Forest

Projet

Le but de ce projet est de contruire une forêt aléatoire pour prédire l'occurrence de crimes violents.

1 Présentation du problème

En 1990, le Bureau de Recensement des Etats-Unis lance une enquête sur la situation démographique et sociale des Américains et le nombre de crimes violents par habitant. Les analyses statistiques montraient une corrélation entre ces facteurs. Aujourd'hui, les techniques d'apprentissage automatique devraient nous permettre de compléter cette étude.

2 Travail demandé

Le but du projet est de concevoir et d'entraîner une (ou plusieurs) forêt aléatoire constituée d'arbres de régression pour prédire le nombre de crimes violents par habitant. Celui-ci est supposé dépendre de plusieurs variables démographiques et sociales.

2.1 Réalisation

Les données sont librement accessibles sur le UCI Machine Learning Repository¹. Vous y trouverez plus de détails sur le problème et les données. Le choix du logiciel, de la bibliothèque ou du langage de programmation est libre.

2.2 Rendu

Vous devrez déposer sur la page arche du cours avant le vendredi 20 décembre 2019 à 18h une archive contenant un rapport et les sources de vos programmes.

- Le rapport doit décrire votre démarche en présentant :
 - vos choix (prétraitement des données, paramètres de ou des algorithmes,...) ;
 - les analyses des résultats que vous avez menées ;
 - une conclusion précisant les pointsforts et les points faibles de l'étude et les améliorations possibles.
- Les sources doivent être accompagnées d'un fichier README.txt indiquant comment utiliser vos sources.

¹<https://archive.ics.uci.edu/ml/datasets/Communities+and+Crime>

Références

- [1] U. S. Department of Commerce, Bureau of the Census, Census Of Population And Housing 1990 United States: Summary Tape File 1a & 3a (Computer Files)
U.S. Department Of Commerce, Bureau Of The Census Producer, Washington, DC and Inter-university Consortium for Political and Social Research Ann Arbor, Michigan. (1992)
U.S. Department of Justice, Bureau of Justice Statistics, Law Enforcement Management And Administrative Statistics (Computer File) U.S. Department Of Commerce, Bureau Of The Census Producer, Washington, DC and Inter-university Consortium for Political and Social Research Ann Arbor, Michigan. (1992)
U.S. Department of Justice, Federal Bureau of Investigation, Crime in the United States (Computer File) (1995)
Redmond, M. A. and A. Baveja: A Data-Driven Software Tool for Enabling Cooperative Information Sharing Among Police Departments. European Journal of Operational Research 141 (2002) 660-678.