

Big Data Mining

TP5 : régression en grande dimension

Nous allons travailler dans ce TP sur une base de données relative à la criminologie aux Etats-Unis :

`https://archive.ics.uci.edu/ml/datasets/Communities+and+Crime+Unnormalized`

Votre objectif sera de trouver le meilleur modèle de régression pour prédire le taux de crimes violents aux Etats-Unis :

`ViolentCrimesPerPop`

Les modèles seront comparés sur la base de l'erreur quadratique moyenne évaluée par validation croisée 10-fold.