

Feuille de route

Objectif du travail

L'objectif de ce projet est de travailler sur le jeu de données (modifiées) *The Boston Housing Dataset* afin de prédire le prix d'un appartement. Une description du jeu de données initial est disponible [ici](#). Pour ce projet, nous modifions ce dernier en rajoutant 5 variables indépendantes. Les données, sur lesquels nous travaillerons, sont disponible [ici](#).

Il s'agira en particulier de mettre en œuvre, avec Python, et de comparer les performances de ces quatre méthodes Statistiques :

1. la régression linéaire multiple ;
2. la méthode de sélection de variable pas-à-pas descendante ;
3. la régression Ridge ;
4. la régression LASSO.

Travail écrit

- Date de rendu : 02-03-2021.
- Un fichier au format *.pdf* de moins de 20 pages présentant le travail effectué (L^AT_EX).
- Expliquer la problématique, le fondement théorique de chaque méthode en présentant les forces et faiblesses de celles-ci.
- Un fichier source contenant le code Python commenté convenablement.

Attention : Les projets ne respectant pas l'une de ces règles recevront une pénalité de 20 %

Références

- "The Elements of Statistical Learning : Data Mining, Inference, and Prediction" de T. Hastie, R. Tibshirani et J. Friedman (Springer, 2013) disponible [ici](#).
- PennStat Eberly College of Science : Applied Data Mining and Statistical learning, <https://online.stat.psu.edu/stat857/intro/>