Méthodes pour les modèles de régression

ECTS : 3

Volume horaire : 21

Description du contenu de l'enseignement :

L'objectif de ce cours est de présenter aux étudiants des connaissances fondamentales, sur la régression d'un point de vue théorique ainsi que sur le code lié à ce domaine.

- Rappels théoriques sur le modèle linéaire gaussien multivarié, Anova, Ancova, sélection de modèle, validation du modèle...
- Étude et traitement des outliers en régression.
- Étude des différents critères (AIC, BIC, Cp-Mallows,...) et sélection de modèles.
- Analyse complète de différents modèles linéaires gaussiens multivariées sous R à partir de jeux de données réelles.
- Estimateurs biaisés (Lasso, Ridge, Elastic-Net, PLS,...)
- Performance de généralisation (PRESS sur échantillon tests,...), Validation Croisée,...
- Comparaison des différentes procédures (ML, Lasso,...) sous R à partir de jeux de données réelles.
- Modèles linéaires généralisés (régression poissonnière, régression logistique,...).
- Régression logistique d'un point de vue théorique et sous forme de TP avec des données réelles : déclaration du modèle, validation du modèle, sélection de modèle, odd ratio, matrice de confusion courbe ROC, AUC.

Compétence à acquérir :

A la suite de ce module, les étudiants seront capables de comprendre la régression d'un point de vue théorique et de coder les différentes procédures étudiées. Ils auront le recul nécessaire pour préselectionner des procédures adaptées à la spécifité du jeu de données et sélectionner celles ayant les meilleures performances de généralisation.

Mode de contrôle des connaissances :

Examen