

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 sur les projections dans un espace vectoriel. Projections dans l'espace L^2 des variables aléatoires de carré intégrable. Cas multi-varié.
- Présentation de la régression linéaire avec des hypothèses minimales. Risques empirique et de généralisation. Théorème de Frisch-Waugh-Lovell. Coefficients de détermination centré et non centré.
- Propriétés statistiques des estimations MCO (Moindres Carrés Ordinaires) avec les hypothèses standard de normalité : tests sur les coefficients, théorème de Gauss-Markov. Cas où les hypothèses standard ne sont pas vérifiées : estimateur des moindres carrés généralisé, biais endogène, variables instrumentales.
- Régularisation et sélection de modèles : régressions Ridge et Lasso, régression bayésienne et estimateur de la régression Ridge, validation croisée, critères de validation (AIC, BIC, Cp-Mallows,...).
- Analyse de différents modèles linéaires gaussiens multivariés sous R ou Python à partir de jeux de données réelles.

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 spécificité du jeu de données et sélectionner celles ayant les meilleures performances de généralisation.

Mode de contrôle des connaissances :

Examen

Document susceptible de mise à jour - 02/04/2026

Université Paris Dauphine - PSL - Place du Maréchal de Lattre de Tassigny - 75775 PARIS Cedex 16