Statistique en Grande Dimension et Apprentissage

Niveau d'étude
BAC +5 / master

ECTS
6 crédits

Composante
Faculté des sciences

En bref

- Langue(s) d'enseignement: Français
- Ouvert aux étudiants en échange: Oui

Présentation

Description

Bases mathématiques de l'apprentissage statistique. Bases de la grande dimen- sion (« fléau » de la dimension…). Régression en grande dimension (motivation par l'exemple, théorie du Lasso, seuillage, sparsité, inégalités classiques, régressions ba- sées sur les composantes principales, PLS). Méthodes de pénalisation (généralisation du Lasso (Elastic-Net, Group-Lasso), sélection de variables, Ridge, etc.). Classification et apprentissage en grande dimension (SVM, Logistic-Lasso, forêts aléatoires, k plus proches voisins, ACP à noyaux, Boosting, Gradient Boosting). İntroduction aux réseaux de neurones et au deep learning. Bases de l'optimisation stochastique. Large mise en pratique avec R ou Python.

Objectifs

— Avoir un recul professionnel sur la notion de grande dimension. Comprendre par l'exemple les raisons théoriques mais aussi les raisons techniques qui différencient les méthodes statistiques classiques des méthodes statistiques en grande dimension.

— Comprendre la théorie du Lasso et d'une manière générale la méthodologie consis- tant à pénaliser une estimation par des contraintes sur les coefficients de la régression. Avoir une vision claire des objectifs visés par les méthodes de pénalisation (estimation ou sélection de variables).

— Savoir réduire la dimension des observations par des méthodes classiques d'analyse des données (par exemple l'ACP) et savoir exploiter cette réduction de dimension pour les régressions en grande dimension ou possédant de la multicolinéarité.

İnformations non contractuelles.
Dernière mise à jour le 23 mai 2023

— Savoir ce qui relie et ce qui différencie le Lasso des méthodes de pénalisation ba- sées sur des normes différentes (Elastic-Net, Ridge, etc..). Connaître les points forts et points faibles de chacune des approches pour en permettre une mise en pratique per- tinente sur des jeux de données réelles.

— Connaître les méthodes simples d'apprentissage automatique appliquées à la classi- fication (comme les k plus proches voisins) et les méthodes plus sophistiquées (comme les forêts aléatoires). Avoir une vision d'ensemble de la théorie des SVM appliqués à la classification.

— Connaître les bases du boosting (Adaboost/Gradient Boosting, XGBoost).

— Comprendre le fonctionnement et savoir mettre en oeuvre un algorithme d'apprentissage par deep learning (perceptron et réseaux de neurones convolutionnels).

— Être capable de mobiliser et exploiter les méthodes étudiées dans ce module dans des cas pratiques sous R ou Python, en particulier les méthodes de régression et de sélection de variables en grande dimension.

Heures d'enseignement

CM	Cours magistral	24h
TD	Travaux dirigés	16h
TP	Travaux pratique	16h

Pré-requis obligatoires

Notions et contenus : Algèbre linéaire et analyse (licence mathématiques L3). Statis- tique inférentielle (S1-UE3-DS). Optimisation non linéaire (S1-UE2-DS). Datamining et classification (S2-UE1-DS). Manipulation standard de R et de Python.

Compétences et capacités :

— Maîtriser les compétences enseignées dans les modules de statistique inférentielle et de dataming et classification du M1, particulièrement tout ce qui concerne les modèles de régression linéaire et les méthodes de classification.

— Maîtriser les bases de l'algèbre linéaire, du calcul matriciel, de l'analyse et de l'op- timisation, en particulier : recherche des valeurs propres et singulières, la notion de produit scalaire (sur l'exemple de Rn et de L^2), les propriétés des espaces de Hilbert de fonctions ou encore l'optimisation sous contraintes.

— Maîtriser les notions de loi conditionnelle/espérance conditionnelle.

— Maîtriser les bases de l'optimisation déterministe (descente de gradient et exten- sions).

— Avoir une connaissance minimale des langages R et Python (syntaxe, manipulation élémentaire, calcul matriciel, gestion des graphiques, etc.).

Informations non contractuelles.
Dernière mise à jour le 23 mai 2023

Infos pratiques

Lieu(x)

- Angers

Informations non contractuelles.
Dernière mise à jour le 23 mai 2023