

Apprentissage statistique et classification

Nom : **LAHIANI**

Prénom : **Amine**

Année : **M1**

Semestre : **7**

Nature : **CM**

Volume horaire : **24**

ECTS / Coef : **2**

Prérequis	<ul style="list-style-type: none">- Cours d'analyse de données- Cours d'algèbre linéaire- Cours de Géométrie dans l'espace
Résumé	<p>Ce cours présente les bases des techniques d'apprentissage statistique et de classification les plus répandues. Il détaille tout particulièrement l'analyse discriminante, avec une introduction de concepts bayésiens de classification (via la méthode knn), la validation croisée pour évaluer la robustesse et/ou la précision d'un modèle. Il traite également des arbres de décision (classification and regression trees ou CART). Il s'agit de donner aux étudiants des outils utiles pour analyser des données complexes et éventuellement massives.</p> <p>Afin de parfaire leur compréhension et leur maîtrise, ces techniques d'apprentissage statistique et de classification sont mises en œuvre au moyen des logiciels R</p>
Objectifs	<ul style="list-style-type: none">- Maîtriser les techniques d'apprentissage et de classification- Modélisation pour la prévision- Recherche de modèles optimaux et parcimonieux- Fouille de bases de données volumineuses- Utilisation de logiciels statistiques de fouille des données- Donner une interprétation simple d'une base de données complexe
Bibliographie	<ul style="list-style-type: none">- L. Breiman, J. Friedman, R. Olshen et C. Stone, Classification and regression trees, Wadsworth & Brooks, 1984.- T. M. Cover et P. E. Hart, Nearest neighbor pattern classification, IEEE Trans. Inform. Theory 13 (1967), 21–27- J. H. Friedman, H. Hastie et R. Tibshirani, Additive logistic regression : a statistical view of boosting, The Annals of Statistics 28 (2000), 337–407.- B. Ghattas, Agrégation d'arbres de classification, Revue de Statistique Appliquée 48 (2000), no 2, 85–98.- L. Rokach et O. Maimon ; Data mining with decision trees : theory and applications.

PLAN

1. Introduction

2. k-nearest neighbors

3. Arbres de décision

Section 1 : Introduction aux arbres de décision

Section 2 : Arbre de classification

Section 3 : arbres versus modèles linéaires

Section 4 : Avantages et inconvénients des arbres de décision

Section 5 : Applications : estimation des arbres de classification et de Régression, prévision avec les arbres de classification et de régression

4. Méthodes de classification non linéaires

Section 1 : Régression polynomiale

Section 2 : Fonctions échelons

Section 3 : Régression locale

Section 4 : Modèles généralisés additifs

Section 5 : Applications : Modélisation non linéaire