UE1 - Datamining et Classification

Niveau d'étude
BAC +4

ECTS
5 crédits

Composante
Faculté des sciences

En bref

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

Présentation

Description

Principaux concepts du datamining ; analyse des données (ex : ACM, AFM, MDS) ; classification supervisée (k-plus proches voisins, analyse discriminante, arbres de décision, SVM, etc.) ; courbes ROC et AUC ; classification non-supervisée (k-means, classifications hiérarchiques, DBSCAN, etc.) ; mise en pratique sous R et/ou Python.

Remarque : note plancher de 8/20, sauf appréciation contraire du jury.

Pré-requis obligatoires

Notions et contenus :

Théorie de la mesure et intégration (licence mathématiques L3) ; calcul des probabilités (licence mathématiques L3) ; algèbre linéaire en dimension finie (licence mathématiques L3) ; modules de statistiques et d'analyse numérique matricielle du premier semestre ; langages R et Python.

Compétences et capacités :

Maîtriser les compétences enseignées dans le module de statistiques du semestre 1 : statistique descriptive, modélisation, analyse des données ; maîtriser les bases de l'algèbre linéaire et du calcul matriciel, notamment : changement de base, recherche des valeurs propres, des valeurs singulières, la projection orthogonale de Rn ; avoir une connaissance minimale des langages R et Python.

Informations complémentaires

Section Moodle du M1 DS.

Informations non contractuelles.
Dernière mise à jour le 15 décembre 2023

Compétences visées

Mettre en pratique l'analyse des données vue en statistiques au premier semestre (ACP, AFC) ainsi que les méthodes vues dans le module (ex : ACM, AFM, MDS) : interprétation et démonstration par l'exemple. Être capable d'interpréter graphiquement les sorties fournies par les logiciels de statistique pour en déduire des conclusions pertinentes en analyse des données.

Comprendre le principe de la classification supervisée par l'intermédiaire des méthodes usuelles (k-plus proches voisins, analyse discriminante, arbres de décision, SVM, etc.). Mettre en œuvre ces méthodes sur des exemples concrets.

Connaître le principe des algorithmes simples de classification non supervisée (k-means, classifications hiérarchiques, DBSCAN, etc.). Comprendre par l'exemple la structure d'un modèle de mélange et son estimation par un algorithme EM. Mettre en œuvre ces méthodes sur des exemples concrets.

Savoir utiliser R ou Python pour largement exploiter les méthodes étudiées dans le module, en particulier les méthodes de classification et de réduction de dimension.

Bibliographie

T. Hastie, R. Tibshirani, J. Friedman, *The element of Statistical Learning. Data mining, inference, and prediction.* Second edition. Springer Series in Statistics. Springer, New York, 2009.

R. Garreta, G. Moncecchi, *Learning scikit-learn : machine learning in python.* Packt Publishing, 2013.

Cornillon P.A., Guyader A., Husson F., Jégou N., Josse J., Kloareg M., Matzner-Løber E., Rouvière L., *Statistiques avec R.* PU Rennes, 3e édition revue et augmentée, 2012.

Liste des enseignements

	Nature	CM	TD	TP	Crédits
Datamining et Classification	Matière	20h	12h	16h	

Infos pratiques

Lieu(x)

› Angers

Campus

› Campus Belle-beille

Informations non contractuelles.
Dernière mise à jour le 15 décembre 2023