Analyse des données

ECTS

Volume horaire
58.25h

Présentation

Description

Programme (contenu détaillé) :

- Introduction à l'exploration de données.
- Syntaxe et objets de R et Python, écriture de fonctions, programmation objet et fonctionnelle (Python).
- Méthodes factorielles : rappel de l'analyse en composantes principales (ACP). Variantes de l'ACP pour les données qualitatives (analyse des correspondances), la classification supervisée (analyse linéaire discriminante), les données définies par des distances (positionnement multidimensionnel), le passage au non-linéaire (ACP à noyau).
- Méthodes de clustering : rappel des méthodes de base (k-means, classification hiérarchique). Modèles de mélange et algorithme EM. Découpe en communauté ou clustering de graphes.
- Factorisation non négative de matrices et introduction aux méthodes de recommandation

Objectifs

A la fin de ce module, l'étudiant.e devra avoir compris et pourra expliquer (principaux concepts) :

- Organisation et préparation des données avec R et Python. Syntaxe des langages R et Python, fonctions de leurs principales librairies ;
- Exploration statistique de données multidimensionnelles, réduction de dimension et classification automatique avec R, Python.
- Interprétation statistique des différents types de représentations graphiques en analyse factorielle et classification.

L'étudiant.e devra être capable de :

- Gérer des jeux de données avec R et Python
- Conduire l'analyse exploratoire de données : méthodes uni, bi et multivariées (ACP, AFCM, AFD, NMF, modèles de mélange,...) adaptées à la structure des données.
- Détecter des structures particulières dans des jeux de données complexes et en faire l'interprétation.

Pré-requis nécéssaires

Algèbre linéaire : diagonalisation de matrices, décomposition en valeurs singulières
Analyse en composantes principales
Clustering : méthode k-means et clustering hiérarchique

Évaluation

L'évaluation des acquis d'apprentissage est réalisée en continu tout le long du semestre. En fonction des

enseignements, elle peut prendre différentes formes : examen écrit, oral, compte-rendu, rapport écrit, évaluation par les pairs...

Infos pratiques

Lieu(x)

Toulouse