

Analyse des données



ECTS

3 crédits



Volume horaire

62h

Présentation

Description

Programme (contenu détaillé) :

- Introduction à l'exploration de données.
- Programmation en R et rédaction d'un rapport avec Rmarkdown
- 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)
- Méthodes de clustering : K-means et ses variantes, classification hiérarchique, DBSCAN, modèles de mélange et algorithme EM.

L'étudiant devra être capable de :

- Conduire une analyse exploratoire sur un jeu de données avec le logiciel R et rédiger un rapport avec Rmarkdown
- Savoir expliquer et appliquer les méthodes factorielles PCA, MCA, MFA, MDS, LDA
- Savoir expliquer et appliquer les méthodes de clustering Kmeans et ses variantes, classification hiérarchique, DBSCAN et les modèles de mélange.

Pré-requis nécessaires

Statistiques : statistiques descriptives.

Probabilités : vecteurs aléatoires, loi de probabilité, formule de Bayes, loi normale multidimensionnelle.

Algèbre: espaces vectoriels, espaces euclidiens, calcul matriciel, diagonalisation de matrices.

Géométrie/mécanique : barycentre, inertie, formule de Huygens.

Objectifs

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

- Exploration statistique de données multidimensionnelles, réduction de dimension et classification non supervisée avec R.
- Interprétation statistique des différents types de représentations graphiques en analyse factorielle et clustering.

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