

Analyse de données

Niveau d'étude
Bac +4

ECTS
6 crédits

Composante
**Sciences Fondamentales
et Appliquées**

Période de l'année
Semestre 7

En bref

Méthodes d'enseignement: En présence

Ouvert aux étudiants en échange: Oui

Présentation

Description

Dans le cadre de ce module nous allons étudier différentes méthodes d'exploration de données dans le cadre de la mesure. Etant donné le contexte de mutualisation ainsi que l'origine possible variée des étudiants, le contenu pédagogique débutera par une présentation des concepts élémentaires associés aux probabilités ainsi que les caractéristiques usuelles permettant de décrire des mesures telle que la notion de moyenne, de variance et de corrélation et enfin la description des lois les plus usitées. Il s'agit à travers ces principes issus des statistiques descriptives de fournir à l'étudiant un bagage suffisant pour appréhender la suite de la formation concernant la donnée. Les notions de probabilités seront illustrées à travers les méthodes d'estimation de grandeurs à partir d'une série de mesures.

Nous étudierons ensuite les principales méthodes d'analyse de tableau de mesures comme par exemple l'ACP, à travers à la fois l'analyse des méthodes mais aussi leurs réalisations informatiques. Enfin, nous concluons par la découverte des algorithmes dit de clustering permettant d'extraire des groupes de données à partir de leurs mesures.

Objectifs

L'objectif du module d'analyse des données est de maîtriser les principales techniques d'exploration de données dans le cadre de la mesure.

Comme nous l'avons vu, étant donné le contexte de mutualisation ainsi que l'origine possible variée des étudiants, le premier objectif est l'acquisition des concepts élémentaires associés aux probabilités ainsi que les caractéristiques usuelles permettant de décrire des mesures.

Après ce préambule, le second objectif est d'étudier en détail les principales méthodes d'exploration de tableau de mesures à savoir l'analyse en composante principale ainsi que l'analyse factorielle discriminante (afin de sensibiliser l'étudiant à la problématique de classification automatique). L'objectif est d'extraire des informations, structures, motifs ou indicateurs utiles d'un ensemble de données. Le choix pédagogique sera d'étudier à la fois le concept de la méthode mais aussi sa réalisation pratique.

Le dernier objectif concerne l'étude de la problématique du regroupement automatique des données suivant des critères similitudes au niveau des mesures à travers l'étude de deux méthodes :

- o K-Means
- o Dendogrammes

Heures d'enseignement

CM	CM	14h
TP	TP	24h
TD	TD	8h
P-Proj	Pédagogie par projet	4h

Pré-requis obligatoires

Programmation et algorithmique de niveau licence scientifique.

Mathématiques de niveau licence scientifique.

Compétences visées

Mettre en œuvre informatiquement un algorithme d'analyse de données sur des données réelles.

Maîtriser les outils d'analyse des mesures et la justification associée

Analyser les informations associées à une série de mesures grâce aux outils d'analyse de mesures.

Mettre en œuvre informatiquement un algorithme estimation d'une grandeur à partir de données réelles, en prenant en compte la notion d'intervalle de confiance.

Maîtriser les tests statistiques élémentaires et la justification associée

Mettre en œuvre informatiquement un test statistique sur des données réelles afin de répondre à une question.

Liste des enseignements

	Nature	CM	TD	TP	Crédits
Pratique de l'analyse de données	EC			20h	
Méthodes d'analyse de données	EC	14h	8h	4h	

UE = Unité d'enseignement

EC = Élément Constitutif

Infos pratiques

Lieu(x)

Futuroscope