Analyse de données 1

ECTS : 3

Enseignant responsable : Pierre WOLINSKI (https://dauphine.psl.eu/recherche/cvtheque/wolinski-pierre)

Langue du cours : Français

Volume horaire : 30

Description du contenu de l'enseignement :

Dans ce cours, on s'intéresse à des techniques visant à extraire de l'information d'un jeu de données via l'obtention de modèles linéaires des relations entre les données. Ce choix se justifie par la pertinence et l'efficacité de ces modèles simples dans la pratique. On considèrera que l'on dispose de données pour lesquelles cette extraction d'information est pertinente. À cette fin, les données seront généralement pensées comme des réalisations de variables aléatoires. On présentera ainsi un ensemble de résultats et d'algorithmes mêlant des éléments d'algèbre linéaire, d'optimisation et de statistiques.Durant le cours, on abordera la décomposition en valeurs singulières en tant qu'outil fondamental, ainsi que son utilisation dans les moindres carrés linéaires et l'analyse en composantes principales. On se penchera longuement sur les modèles de régression linéaires, et leur interprétation via les moindres carrés et l'estimation statistique. On illustrera enfin les résultats théoriques via des applications caractéristiques de l'analyse de données.

Compétences à acquérir :

Connaître et manipuler les techniques d'analyse de données liées à des modèles linéaires : méthode des moindres carrés et Analyse en Composantes Principales (ACP).

Pré-requis obligatoires

Algèbre linéaire : calcul matriciel.

Analyse : bases du calcul différentiel.

Pré-requis recommandés

Algèbre linéaire : inversion et diagonalisation de matrices, propriétés des matrices symétriques et orthogonales.

Analyse : calcul différentiel appliqué aux fonctions de plusieurs variables.

Mode de contrôle des connaissances :

Partiel ou CC + Examen.

Bibliographie, lectures recommandées :

S. Boyd and L. Vandenberghe. Introduction to Applied Linear Algebra - Vectors, Matrices and Least Squares. Cambridge University Press, 2018.

G. James, D. Witten, T. Hastie, R. Tibshirani, J. Taylor, An Introduction to Statistical Learning with Applications in Python, 2nd edition, Springer Texts in Statistics, 2023.

G. Strang, Linear algebra and learning from data. Wellesley-Cambridge Press, 2019.

Laurent Rouvière, Cours d'algèbre linéaire, Licence MASS 2e année, Université Rennes 2.