Data and decision track

ECTS credits
8 credits

Semester
Fall

In brief

Course langage: French

Presentation

Prerequisites

- Knowledge in Optimization, Probabillity and Statistics Bachelor level
- Knowledge in Python programming and in algorithmics

Learning objectives

- Know how to acquire, aggregate and manipulate data.
- Know how to model standard regression and classification problems and program their solutions with an adequate programming language.
- Know how to use data to take decisions
- Understand the importance of data governance and data quality
- Understand the basic approach of data engineering in data science projects

Description of the programme

This course unit consists of three courses: Statistical learning, Python for data science and Data-driven decision making, of 24 hours each, and is complemented by the second part of the data science projects (9 hours course and 12 hours project) devoted to data issues.

Statistical learning

1. Introduction
   i. Classical problems: regression, classification
   ii. Supervised, unsupervised and semi-supervised learning

   iii. Curse of dimensionality
2. Regression
   i. Multiple linear regression, OLS method
   ii. Shrinkage-type methods (LASSO, Ridge)
   iii. k-nearest neighbors
3. Classification
   i. Logistic regression
   ii. k-nearest neighbors
   iii. SVM
   iv. Rosenblatt perceptron and neuronal networks

Python for data science

1. Dataframe: data exploration and data description
   i. Spotting patterns using factor
   ii. Principal Component Analysis
   iii. Correspondence analysis
2. Prediction using trend analysis
   i. Linear regression
   ii. Logistic regression
3. Data classification
   i. Classification using partitions
   ii. Hierarchical methods

Data-driven decision making

1. What is data?
2. How do we take decision?
3. Data governance and data quality
4. How to develop data-based decision making?
5. Data platform and data architecture

Data science projects: data issues

1. Starting a data science project
2. The constraints of data science projects
3. Finding data
4. Acquiring information
5. Playing with data

Generic central skills and knowledge targeted in the discipline

- Know how to apply standard supervised and unsupervised classification methods and how to compare several models.

- Know how to apply standard regression methods (OLS) and advanced methods to select variables and cope with the curse of dimensionality (Ridge, LASSO, Elastic Net)
- Know how to apply dimension reduction and data description procedures such as PCA and Correspondence Analysis.
- Be able to build indicators of the performance of a model on a dataset
- Understand and measure the value of data
- Know which data to use to take decisions
- Be able to manipulate data to start data science project

How knowledge is tested

- Tests and projects (Statistical learning): 30%
- Project (Python for data science): 35%
- Group project and presentation (Data-driven decision making): 35%

Bibliography

Statistical Learning

- James G., Witten D., Hastie T. and al. (2013). An introduction to statistical learning: with applications in R. New York: Springer
- Hastie T., Tibshirani R. and Friedman J. (2013). The elements of statistical learning: data mining, inference, and prediction. New York: Springer.
- Cornillon P-A., Matzner-Løber E. et al. (2010). Régression avec R. Paris: Springer.

Python for data science

- Jannach, D., Zanker, M., Felfernig, A. and Friedrich, G. (2010). Recommender Systems: An Introduction. Cambridge.

Data science projects

- Zeng, A and Casari, A. Feature Engineering for Machine Learning. O'Reilly Media.
- Müller, A. and Guido, S. Introduction to Machine Learning with Python. O'Reilly Media.

Teaching team

- Statistical learning: Christophe Pouet (Centrale Marseille)
- Python for data science: François Brucker (Centrale Marseille), Emmanuel Daucé (Centrale Marseille)
- Data-driven decision making: Mickaël Chalamel (Yves Saint-Laurent), Franck Chevalier (EY)
- Data science projects: Maximilien Défourné (Mantiks)

Sustainable Development Goal



Partnerships for the goals

Total des heures		**100h**
CM	Master class	81h
PJ		19h