

Machine Learning

Person in charge: Patricia CONDE-CESPEDES

Prerequisite: II.1102 Algorithmic and programming, IE.1101 Project-based learning (in signal) and IF.1101 Statistics and probability et sciences fondamentales (probabilités et statistiques)

Organization: 14 x 3h Lectures/Tutorials (42 hours+ 80 hours of personal work)

Evaluation: continuous control /project

ECTS: 5 credits

Context

Machine learning or statistical learning consists in the study of methods of mining and learning from data. Data Mining consists in a set of algorithms to analyze and explore data, to understand trends and perform predictions. Machine learning and data mining are becoming widely used in industry with the *big data* phenomenon. Their field of application is widely spread in various areas. For instance, Machine learning is used in medicine for diagnosis assistance, in marketing for the analysis of buyer customer behavior, in finance for credit card fraudulent use detection, in applications for computer vision (autonomous cars, robotics, ...), etc.

Objectives

The main objective of the course is to initiate students into methods and algorithms coming from statistics and learning methods, which are used in data Mining and data analysis as well as in the design of computer-assisted systems. Students will learn the different steps of knowledge discovery from raw data in order to determine trends and perform predictions. We will focus on the following methods: regression, supervised and unsupervised methods, neural networks and an introduction to deep learning.

Knowledge

This module enables students to develop the following concepts and skills.

- **Concepts**
 - Introduction to Machine Learning

- Linear Regression
- Supervised classification: Logistic Regression, Linear Discriminant Analysis, K-nearest Neighbours (KNN), decision trees, Support Vector Machines (SVM)
- Unsupervised Learning: Principal Component Analysis (PCA), k-means, Hierarchical Clustering.
- Neural networks
- **Know-How**
 - Understand and learn the process of knowledge discovery from raw data.
 - Understand the principle of supervised and unsupervised learning as well as their use cases.
 - Acquire knowledge in automatically mining large amount of data.

Pedagogical Approach

14 sessions (one per week) distributed equitably between lectures and practical courses. The practical courses will be taught in R, python or Matlab. By the end, there will be a final project.

References

- James, Gareth; Witten, Daniela; Hastie, Trevor and Tibshirani, Robert (2013). « An Introduction to Statistical Learning with Applications in R ». New York : "Springer texts in statistics". Site web: <http://www-bcf.usc.edu/~gareth/ISL/>
- Hastie, Trevor; Tibshirani, Robert and Friedman, Jerome (2009). « The Elements of Statistical Learning (Data Mining, Inference, and Prediction), 2nd edition». New York: "Springer texts in statistics". Site web : <http://statweb.stanford.edu/~tibs/ElemStatLearn/>
- Leskovec, Jure; Rajaraman, Anand and Ullman, Jeffrey D. (2014). « Mining of Massive Datasets, 2nd edition ». Cambridge University Press. Site web : <http://www.mmds.org>
- Goodfellow, Ian; Bengio, Yoshua and Aaron Courville (2016). « Deep Learning (Adaptive Computation and Machine Learning series). Book in preparation for MIT Press. Site web: <http://www.deeplearningbook.org>
- Mooc Coursera : <https://fr.coursera.org>