IG.2412 – Machine Learning

General information

Title: Machine Learning
Module ID: IG.2412
Person in charge: Jérémie SUBLIME
ECTS: 5
Average amount of work per student: from 100 to 150 hours (42 hours are supervised)
Teamwork: yes
Keywords: learning, classification, clustering, AI

Presentation

Machine learning is the study of algorithms that learn from data to perform various tasks such as classification, regression, predictive models, clustering, and some advanced data visualization methods. This discipline, at the crossroads between statistics and algorithms, is used for a wide range of tasks and professions, ranging from automatic medical diagnosis to bank fraud detection, predictive marketing and DNA sequencing. On the other hand, Machine Learning as a branch of artificial intelligence is also the basis for well-known advanced methods such as deep neural networks.

Educational objectives

The knowledge and skills developed in this module are in the field of Artificial Intelligence and in particular in the entire machine learning part that is not part of deep learning.

Prerequisite
- Python Programming
- Notions of statistics and probability

Subsequent mobilizations at ISEP
- Synergy with the Deep Learning module (A2)

Content/Program

Concepts

- Classification :
    - KNN, naïve bases, decision trees
    - SVMs, dimension de VC
    - Ensemble methods: bagging&boosting, random forests
    - Binary and multiclass indices
    - Cross-validation
- Clustering :
    - Prototype-based models: K-MEANS, GMM and SOM
    - Hierarchical clustering
    - Density-based models: dbscan, optics, meanshift, denclue, CFSFDP
    - Spectral clustering
    - Clustering on Stream: denstream, k-means online
    - Unsupervised indices and the notion of stability
- Visualization:
    - Reminders about the CPA
    - Nonlinear methods: isomap, LLE, MDS, t-SNE

Pedagogical methods

Learning methods

For the 3 components, the theoretical course is followed/accompanied by tutorials and practical work on a machine in Python. In such a way that students can assimilate the theoretical knowledge of the S.A.

This module is based on a problem-based approach, through the systematic use of contextualized problems, particularly in the dimension to improve skills "Ensuring the quality and safety of a system (availability, reliability, maintainability, security, confidentiality – integrity)".

Evaluation methods

Project, graded practical work and final exam.

Language of work

- Although the language of work is French, some bibliographic resources may be in English.
- The data science course materials are in English.

Bibliography, Webography, Other sources

- Christopher M. Bishop, "Pattern Recognition and Machine Learning" (2006)