Machine Learning

ECTS : 6

Enseignant responsable : Fabrice ROSSI (https://www.ceremade.dauphine.fr/en/members/detail-cv/profile/fabrice-rossi.html)

Langue du cours : Français

Volume horaire : 36

Description du contenu de l'enseignement :

The course gives a thorough presentation of the machine learning field and follows this outline:

1. general introduction to machine learning and to its focus on predictive performances (running example: k-nearest neighbours algorithm)
2. machine learning as automated program building from examples (running example: decision trees)
3. machine learning as optimization:
    1. empirical risk minimization
    2. links with maximum likelihood estimation
    3. surrogate losses and extended machine learning settings
    4. regularisation and kernel methods (support vector machines)
4. reliable estimation of performances:
    1. over fitting
    2. split samples
    3. resampling (leave-one-out, cross-validation and bootstrap)
    4. ROC curve, AUC and other advanced measures
5. combining models:
    1. ensemble techniques
    2. bagging and random forests
    3. boosting
6. unsupervised learning:
    1. clustering (hierarchical clustering, k-means and variants, mixture models, density clustering)
    2. outlier and anomaly detection

Compétences à acquérir :

After attending the course the students will

- have a good understanding of the algorithmic and statistical foundations of the main machine learning techniques
- be able to select machine learning techniques adapted to a particular task (exploratory analysis with clustering methods, predictive analysis, etc.)
- be able to design a model selection procedure adapted to a particular task
- report the results of a machine learning project with valid estimation of the performances of their model

Pré-requis obligatoires

- intermediate level in either Python or R. Students are expected to be able to perform standard data management tasks in Python or R, including, but not limited to:
    - loading a data set from a CSV file
    - recoding and cleaning the data set
    - implementing a simple data exploration strategy based on pivot table and on graphical representation
- intermediate level in statistics and probability. Students are expected to be familiar with:
    - descriptive statistics
    - conditional probabilities and conditional expectations
    - core results from statistics: bias and variance concepts, strong law of large numbers, central limit theorem, etc.

Mode de contrôle des connaissances :

- quizzes and tests during the course
- machine learning project