A course on statistical methods for supervised and unsupervised learning.

Content

- Introduction: supervised and unsupervised learning, loss functions, train and test errors, bias-variance tradeoff, model complexity and overfitting, linear regression, k-nearest neighbors.
- Regression: linear regression, model selection, ridge and Lasso.
- Classification: linear discriminant analysis, logistic regression.
- Resampling methods: cross-validation, bootstrap.
- Nonparametric regression: smoothing splines, reproducing kernel Hilbert spaces.
- Support vector machines and kernel logistic regression.
- Tree-based methods: classification and regression trees, bagging, random forests.
- Boosting: AdaBoost, gradient boosting machines.
- Unsupervised learning: principal component analysis, k-means, Gaussian mixtures and the EM algorithm.

Learning Prerequisites

Required courses
Analysis, Linear Algebra, Probability and Statistics, Linear Models

Important concepts to start the course
This is a statistics/mathematics course. Prior to following this course, the student must have very good knowledge of basic probability and statistics (statistical modeling and inference, linear regression).

Learning Outcomes

By the end of the course, the student must be able to:
- Formulate appropriate models for empirical data
- Estimate the parameters of a statistical model
- Interpret the fit of a model to data
- Justify the choice of a model/technique to analyze empirical data
- Implement statistical learning algorithms
• Explain the mathematical/statistical mechanisms of most common machine learning algorithms

Teaching methods
Ex cathedra lectures, exercises and computer practicals in the classroom and at home.

Assessment methods
Written final exam (70%) + Project of implementation or application on real data of a model/algorithm based on a classical research paper describing an important method from the literature. (30%)
Seconde tentative : Dans le cas de l'art. 3 al. 5 du Règlement de section, l'enseignant décide de la forme de l'examen qu'il communique aux étudiants concernés.

Resources

Bibliography
• James, G., Witten, D., Hastie, T. and Tibshirani, R. (2013) An Introduction to Statistical Learning, with Applications in R. Springer.

Ressources en bibliothèque
• Applied Predictive Modeling / Kuhn & Johnson
• Pattern Recognition and Machine Learning / Bishop
• Understanding machine learning
• (electronic version)
• Computer Age Statistical Inference / Efron & Hastie
• (electronic version)
• Introduction to Statistical Learning, with Applications
• (electronic version)
• Elements of Statistical Learning

Notes/Handbook
A polycopié will be available on Moodle.