MATH-412 Statistical learning
Thibaud Emeric Rolland Georges

Summary
An introduction to statistical methods for supervised and unsupervised learning.

Content
• Introduction: supervised and unsupervised learning, motivating examples, train and test errors, bias-variance tradeoff, model complexity and overfitting, k-nearest neighbors;
• Regression: linear regression, model selection, Ridge and Lasso methods, non-linear models;
• Classification: linear discriminant analysis, logistic regression;
• Resampling methods: cross-validation, bootstrap;
• Tree-based methods: classification and regression trees, bagging, random forests;
• Boosting;
• Support vector machines: definition, kernel trick;
• Unsupervised learning: principal component analysis, k-means, Gaussian mixtures and the EM algorithm;
• Other topics as time permits.

Learning Prerequisites
Recommended courses
Analysis, Linear Algebra, Probability, Statistics, Linear Models

Learning Outcomes
By the end of the course, the student must be able to:
• Formulate
• the choice of a model/technique to analyze empirical data
• empirical data using supervised and unsupervised learning methods
• Formulate appropriate models for empirical data
• Estimate the parameters of a statistical model
• Interpret the fit of a model to data

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

Assessment methods
Continuous control, final exam.
Second session: from the rulebook of the Section of Mathematics (art. 3 al. 5), the teacher decides of the form of the exam and communicates it to the concerned students.

Supervision
Assistants Yes

Resources
Virtual desktop infrastructure (VDI)
No

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

Ressources en bibliothèque
• Pattern Recognition and Machine Learning
  (electronic version)
• Understanding machine learning
  (version électronique)
• Introduction to Statistical Learning, with Applications
  (electronic version)
• Elements of Statistical Learning