

Advanced Statistical Learning

Viannet Perchet (ENSAE Paris)

Course: 18 hours - TP: 6 hours

Objectives

The objective of this course is to present the basics of statistical learning by focusing essentially on two problems of supervised learning: binary classification and regression. The most commonly used learning algorithms will be presented and their statistical properties will be discussed. These algorithms will be applied to real data during the TP sessions.

Outline

- The framework and examples
- Minimization of empirical risk
- Nearest Neighbors Algorithm
- Decision trees
- Random Forests
- Support Vector Machines
- Boosting
- Neural networks and deep learning

Bibliography

- Luc Devroye, Laszlo Györfi and Gabor Lugosi. A Probabilistic Theory of Pattern Recognition. Springer 1996.
- Trevor Hastie, Robert Tibshirani and Jerome Friedman. The Elements of Statistical Learning: Data Mining, Inference, and Prediction. Springer 2009.
- Williams, Graham. Data Mining with Rattle and R : The Art of Excavating Data for Knowledge Discovery Springer 2011.
- Luis Torgo. Data Mining with R: Learning with Case Studies. Chapman & Hall/CRC Data Mining and Knowledge Discovery Series 2010.