

Apprentissage Statistique

Arnak Dalalyan - ENSAE - CREST

Cours : 14 heures - TP : 10 heures

Objectif

Ce cours a pour objectif de présenter les bases de l'apprentissage statistique en se focalisant essentiellement sur deux problématiques de l'apprentissage supervisé : classification binaire et régression. Les algorithmes d'apprentissage les plus couramment utilisés seront présentés et leurs propriétés statistiques seront discutées. Ces algorithmes seront appliqués aux données réelles lors des séances de TP.

Plan

- Problématique et exemples
- Minimisation du risque empirique
- Algorithme des plus proches voisins
- Arbres de décision
- Forêts aléatoires
- Machines à vecteurs de support
- Boosting
- Algorithme EM pour l'estimation d'un mélange de gaussiennes

Références

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