UNIVERSITÀ
DEGLI STUDI DI MILANO-BICOCCA

COURSE SYLLABUS

High Dimensional Data Analysis

2122-2-F9101Q016

Learning objectives

This is an advanced course focusing on the analysis of high-dimensional data. The goal is to study modern methods and their underlying theory, drawing together theory, data, computation and recent research.

Contents

This course covers methods for regression and classification which can be applied to high-dimensional data.

Detailed program

1. Linear regression, bias/variance trade-off
2. Regularization, ridge and lasso regression
3. Model selection, cross-validation
4. Nonparametric Regression. k-*nearest neighbors* (k-NN). Kernel smoothing. Regression splines, Smoothing splines, Local regression

Prerequisites

Basic knowledge of statistics and probability, linear algebra and computer programming.

Teaching methods

Theoretical lessons and computer applications in lab with R software.

Assessment methods

Oral individual exam to assess the theoretical knowledge of the student on the topics presented during the course. The grading is based on the correctness, the completeness of the answers and the appropriateness of language.

Textbooks and Reading Materials

- Lecture notes provided by the instructor
- Azzalini, Scarpa (2012) Data analysis and data mining, an introduction . New York: Oxford University Press
- Gareth, Witten, Hastie, Tibshirani (2014) An Introduction to Statistical Learning, with Applications in R . Springer
- Hastie, Tibshirani, Friedman (2009) The Elements of Statistical Learning. Data Mining, Inference and Prediction . Springer
- Hastie, Tibshirani and Wainwright (2015) Statistical Learning with Sparsity: The Lasso and Generalizations . CRC Press

Semester

First semester

Teaching language

Italian