UNIVERSITÀ
DEGLI STUDI DI MILANO-BICOCCA

COURSE SYLLABUS

High Dimensional Data Analysis

2425-2-FDS01Q022

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 statistical inference and probability, statistical modeling, linear algebra, and computer programming with R.

You should take these prerequisites quite seriously. If you don't have a solid intuitive understanding of linear algebra, probability, and inference, as well as substantial programming experience with some attention to data structures, I strongly recommend not taking this course. However, the prerequisites are not formally enforced—rather, they're enforced by the fact that you won't understand the class without them.

Teaching methods

All lessons will integrate theoretical aspects with computational ones through the use of the R software.

12 hours of remote synchronous lecturing sessions;
30 hours of in-person lecturing sessions.

Assessment methods

Attending students: written exam.
Non-attending students: written exam.

During the exam, the correctness and clarity of the answers will be evaluated.
The exam aims to assess the skills described in the learning objectives.
The written exam consists of 2 open-ended questions.

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

Semester

First semester

Teaching language

English

Sustainable Development Goals

QUALITY EDUCATION