UNIVERSITÀ
DEGLI STUDI DI MILANO-BICOCCA

SYLLABUS DEL CORSO

High Dimensional Data Analysis

2021-2-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. Nearest neighbors. 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.

In the **Covid-19 emergency period**, lessons are held remotely asynchronously with synchronous videoconferencing events.

Assessment methods

Presentation of a group project work decided with the lecturer and written individual exam on the theoretical part. Grading is based on the written exam (50%), group project (30%) and project presentation and discussion (20%).

Each project typically comprises a paper or a book chapter on a specific topic from the theme of this class. You are expected to understand the proposed method(s), implement them and evaluate them on data sets. For the work on the project, you are encouraged to form teams consisting of at most three people.

Type of exam:

- Written individual exam with open questions and project work

- Oral individual exam to assess the theoretical knowledge of the student on the topics presented during the course and project work presentation

During the **Covid-19 emergency period**, oral exams will only be online.
They will be carried out using the WebEx platform and on the e-learning page of the course there will be a public link for access to the examination of possible virtual spectators.

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