SYLLABUS DEL CORSO

Statistica Non Parametrica M

2425-2-F8204B041

Learning objectives

The course aims to explore the fundamental principles and methodologies of modern nonparametric statistics. The methods considered will allow addressing a wide range of applied problems, and their implementation will be carried out using the open-source software R. By the end of the course, the student will be able to: i) identify applications where nonparametric methodology is more advantageous than parametric methods, ii) conduct a comprehensive data analysis using nonparametric methods with R, iii) assess the performance of different available procedures using appropriate theoretical tools.

Contents

The course aims to rigorously address the problem of estimating a distribution without making parametric assumptions. The same problem will be formulated in terms of estimating distribution functions and density functions. Various tools and methods will be considered, with attention to their theoretical and computational advantages and disadvantages. Depending on time, other topics that might be covered include nonparametric regression models and goodness-of-fit tests.

The course will be divided into two parts, where the same topics will be covered using both classical and Bayesian approaches. The course will cover both the theoretical properties of nonparametric models and their implementation for data analysis. The labs will require the use of appropriate numerical tools for implementing nonparametric models, including optimization algorithms and Markov chain Monte Carlo (MCMC) simulation algorithms.

Detailed program

Part 1 (Classical approach)

- estimation of a distribution function
- confidence bands for a distribution function
- estimation of statistical functionals
- the bootstrap method
- smoothing techniques
- Histogram
- kernel density estimation
- kernel regression models (time permitting)
- goodness-of-fit tests (time permitting)

Part 2 (Bayesian approach)

- exchangeability and nonparametric priors
- the Dirichlet process
- properties of the Dirichlet process
- predictive distribution of the Dirichlet process
- density estimation using Dirichlet process mixtures
- clustering problems using Dirichlet process mixtures
- regression models using nonparametric mixtures (time permitting)
- Bayesian goodness-of-fit tests (time permitting)

Prerequisites

Knowledge of basic statistics courses at the undergraduate level is required. Familiarity with parametric Bayesian statistics is also recommended, and in this regard, the "Bayesian Statistics" course is **strongly recommended**.

Teaching methods

Lectures and lab sessions. Specifically, the schedule includes:

- 6 lectures of 3 hours each (instructional mode)
- 6 lectures of 2 hours each (instructional mode)
- 6 lab sessions of 2 hours each (interactive mode)

It is planned that 80% of the sessions will be conducted in-person and 20% remotely.

Assessment methods

The exam consists of an oral test divided into two parts:

1. (mandatory) discussion on the topics covered in class;
2. (optional) discussion of additional topics not covered in class.

The mandatory part (Part 1) is individual and aims to assess the student's preparation on all the topics covered

during lectures and labs.

The optional part of the oral exam (Part 2) involves the completion and presentation of a group project on a topic agreed upon with the instructor. This part aims to evaluate the students' ability to study, delve deeper into, and utilize advanced nonparametric statistical tools.

There are no mid-term exams.

Textbooks and Reading Materials

- Wasserman, L., 2006. All of Nonparametric Statistics. Springer Science & Business Media.

Additional materials and references will be provided during the course.

Semester

First semester, first cycle

Teaching language

English

Sustainable Development Goals

QUALITY EDUCATION