COURSE SYLLABUS

Advanced Foundations of Statistics for Ai

2223-1-F9102Q002

Aims

The aim of this course is to provide the theoretical foundations of Statistics and indicate how the theory sheds light on the properties of practical methods used in AI. The topics include estimation, prediction, testing, confidence sets, Bayesian analysis and the general approach to statistical modeling and learning.

Contents

The course consists of a theoretical part with exercises, concerning applications on data. Statistical Inference and statistical modeling are presented, with a view on modern methods for supervised and unsupervised learning widely employed in AI.

Detailed program

- Quick review of basic probability theory: probabilistic conceptions, Bayes theorem, random variables and probability distributions, large sample distributions, LLN and CLT statements
- Statistical inference: Estimators and their properties. Point estimate (average, variance and proportion). Notes on maximum likelihood estimators.
- Interval estimation: confidence intervals, the particular cases for the mean and the proportion, sample size considerations.
- Hypothesis testing: the test statistics, the significance and power of the test. Test on the mean and the proportion, on the difference between averages, independence test.
- Simple linear regression: least squares estimation method, model adequacy measures, sampling distribution of OLS estimators, hypothesis tests and confidence intervals for the regression coefficients, analysis of variance, outliers and influential observations, robust linear regression.

- Linear models for regression and for classification
- Ridge regression, the selection of the model
- Extensions of the linear model
- Unsupervised learning: model-based clustering, latent variables, PCA and factor analysis.

Prerequisites

Statistics background: basis of probability theory and knowledge of the most relevant continuous and discrete random variables.
Mathematics background: linear algebra, matrix theory and advanced calculus.

Teaching form

Lectures and assisted exercises.
All activities will be held face-to-face, unless further COVID-19 related restrictions are imposed.
Attendance to lectures and assisted exercises is highly recommended.

Textbook and teaching resource

- Devore, J. L. (2011). Probability and Statistics for Engineering and the Sciences. Cengage learning.
- James G., Witten D., Hastie T. and Tibshirani R. (2021). An Introduction to Statistical Learning, with applications in R (2nd edition). Springer Verlag.
- Hastie T. , Tibshirani R., Friedman J. (2021). The Elements of Statistical Learning (2nd edition). Springer Verlag.

Semester

First

Assessment method

The exam consists in a written exam with open-ended questions to assess the student's knowledge and understanding of the subject, jointly with the development of an application in R on real data. The exam is closed book. An additional oral discussion can be requested by the lecturer.

Office hours

Office: University of Milan-Bicocca,

Department of Statistics and Quantitative Methods, Via Bicocca degli Arcimboldi, 8
20125 MILANO
Building U7-Civitas, Floor 4, room 4136
Phone: +39-02-6448-3118
just drop me an email (francesca.greselin@unimib.it) and we will agree on how and when to meet

Sustainable Development Goals

QUALITY EDUCATION | SUSTAINABLE CITIES AND COMMUNITIES