UNIVERSITÀ
DEGLI STUDI DI MILANO-BICOCCA

COURSE SYLLABUS

Advanced Data Analysis and Statistics

2526-1-F7603Q032

Aims

The aim of the course is to introduce basic notions of linear algebra, optimization of functions of several variables, multivariate statistical modeling and supervised/unsupervised statistical learning.

Knowledge and understanding
At the end of the course the student will have a fundamental understanding of:
• linear algebra;
• optimization of functions of several variables;
• multivariate statistical modeling;
• supervised/unsupervised statistical learning.

Applying knowledge and understanding
At the end of the course the student will be able to:
• apply linear algebra;
• use functions with several variables;
• perform statistical modeling;
• use and apply statistical learning;
• transfer the concepts and approaches introduced in a certain context to connected fields.

Making judgements
At the end of the course the student will be able to:
• decide where to apply certain mathematical concepts;
• decide whether a statistical learning apporach is fruitful or less;
• where and when to apply statistical modeling.

Communication skills
At the end of the course the student should be able to:
• analyse problems in the ares covered by the course in a clear and concise way.
• explain orally with a suitable language the objectives, the procedures and the results of the elaborations carried

out.

Learning skills
At the end of the course the student should be able to:
• approach problems using multivariable functions;
• approach emerging statistical modeling schemes;
• keep pace with the field of statistcial learning.

Contents

• Linear algebra;
• Differential calculus of several variables;
• Constrained/unconstrained optimization of functions of several variables;
• Linear regression;
• Logistic regression;
• Principal component analysis;
• Linear discriminant analysis;
• K-means clustering;
• Hierarchical Clustering.

Detailed program

Linear Algebra:
• vectors and matrices;
• matrix algebra;
• determinant and rank of a matrix;
• systems of linear equations;
• consistent and inconsistent linear systems;
• eigenvalues and eigenvectors;
• matrix decomposition;
• quadratic forms.

Differential calculus of several variables:
• partial derivatives;
• gradient;
• Jacobian and Hessian matrices;
• implicit function theorem;
• unconstrained optimization: necessary and sufficient conditions;
• constrained optimization: the Lagrange multipliers methodology;
• introduction to linear programming.

Supervised statistical learning:
• simple and multiple linear regression;
• logistic regression;
• discriminant analysis.

Supervised statistical learning:
• principal components analysis:
• K-means clustering;

• hierarchical clustering.

Prerequisites

• Knowledge of calculus for functions of one real variable.
• Elementary functions, limits and continuity, differentiation, optimization.
• Basis of probability theory and knowledge of the most relevant continuous and discrete random variables.
• Basic notions of statistical inference: point estimation, confidence interval hypothesis testing.

Teaching form

6 CFUs of mixed theoretical and interactive lessons in the classroom (48 hours):
• 20 two-hour lectures, in person, Delivered Didactics;
• 4 two-hour lectures, in person, discussing problems/exercises, Mixed Didactics.

Attendance to lectures and interactive sessions is highly recommended.

Textbook and teaching resource

• 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 edi-tion). Springer Verlag.
• Lorenzo Peccati, Sandro Salsa, Annamaria Squellati: Mathematics – Corso di International Economics – Università Milano-Bicocca – EGEA.
• Material provided by lecturers.

Semester

I semester (October - January)

Assessment method

The overall exam of the course is split in two seperate exam for each module at the end of the module, according to the needs for the topics discussed. Students are invited to consult the syllabi of the modules for additional information.

The final score will be the average of the scores obtained for two modules, and will be between 18/30 and 30/30 *cum laude*, based on the overall assessment considering the following criteria:
(1) knowledge and understanding;
(2) ability to connect different concepts;

(3) autonomy of analysis and judgment;
(4) ability to correctly use scientific language.

Office hours

Please refer to the indications provided in the syllabi of the modules.

Sustainable Development Goals

QUALITY EDUCATION