UNIVERSITÀ DI PISA

DATA ANALYSIS

MARCO ROMITO

Anno accademico	2022/23		
CdS	MATHEMATICS		
Codice	699AA		
CFU	6		

Moduli	Settore/i	Tipo	Ore	Docente/i
ANALISI DEI DATI	MAT/06	LEZIONI	42	MARCO ROMITO

Learning outcomes

Knowledge
Students are expected to acquire knowledge of statistical learning, in view of prediction, inference and implementation

Assessment criteria of knowledge
The student will be assessed on his/her demonstrated ability to evaluate the best statistical model to predict based on data, and to provide an algorithmic solution.

Skills
At the end of the course the student

- will be able to formulate a suitable statistical model for the quantitative analysis of data,
- will be able to implement the model through a statistical software,
- will be able to formulate conclusions and predictions backed by the data.

Assessment criteria of skills
Analysis of a statistical model and its implementation through a statistical software will be the subject of the final exam.

Behaviors
After the course, the student will be able to manage the quantitative analysis of datasets through statistical methods.

Assessment criteria of behaviors
During the exams, the student will be assessed over her/his attitude from the formulation of a statistical model to its implementation and prediction.

Prerequisites
The student is required to know and master basic concepts and ideas of probability and statistics, and a basic introduction to data analysis methods (linear regression, principal components analysis, autoregressive methods for time series). The student is also required to have a basic knowledge of R or python.

Teaching methods
The course is delivered face-to-face. The practical part is developed partly during the course, and partly as homework. Homework is done in small working group on problems originated from the content of the course.

Syllabus
Introduction to statistical learning, supervised and unsupervised learning. Learning models, prediction classes, probably approximately correct (PAC) models- Vapnik-Chervonenkis dimension. Analysis of some simple examples (linear regression, non-linear variants, k-nearest-neighbour). Assessment of models (cross-validation, bootstrap, information criteria). Classification problems through logistic regression, discriminant analysis, support vector machines. Tree based methods and forests. Stochastic gradient descent and neural networks. Some problems in unsupervised learning.

Bibliography

S. Ben-David, S. Shalev-Shwartz: Understanding Machine Learning: From Theory To Algorithms
T. Hastie, R. Tibshirani, J. Friedman: The elements of Statistical Learning
J. Gareth, D. Witten, T. Hastie, R. Tibshirani: An introduction to statistical learning

Non-attending students info

Course attendance is highly recommended.

Assessment methods

Students will be assessed through an oral exam. Alternatively, attending students could opt to prepare a personal project on data analysis over a problem provided by the teacher..

Class web page

http://people.dm.unipi.it/romito/Teaching/2023/data

Additional web pages

Course material will be made available on the e-learning page of the course

Updated: 30/08/2022 18:44