Laurea / B.A.
in Global Governance

Academic Year 2017-2018
Syllabus
STATISTICAL LEARNING
CFU 6
Prof. Roberto Rocci

Course Description

The course covers the main statistical models and techniques used to predict a target variable (supervised statistical learning, SSL) or to reduce the dimension of the data (unsupervised statistical learning, USL). In details, the linear and logistic regression models will be discussed in the SSL section, while techniques like principal component analysis and cluster analysis will be discussed in the USL section.

As an example, consider a company and its customer database where on each customer (unit) are recorded a number of characteristics (variables): age, gender, amount spent, online customer (yes/no), product purchased, etc. Statistical learning models/techniques can help us finding answers to questions like:

[Linear regression] can the amount spent be explained/predicted by age and gender?

[Logit regression] can the status of "online customer" be explained/predicted by the type of product purchased?

[Principal component analysis] are there relationships (correlations etc.) among the observed variables? If so, which ones? Can these be summarized in one or more prototype/latent variables able to highlight the different purchasing behaviour of customers?

[Cluster analysis] are there different types of customers? If yes, how many and what are they?

Teaching Method

Emphasis is on principles and specific models/techniques.

Each method is introduced by examples and described in mathematical formulas. Some math is essential but only few derivations are made. Models and techniques are discussed from a theoretical and a practical point of view, describing their definition/properties and their implementation by using a statistical package. Some hours of computer laboratory give to the students the possibility to practice what they learn.

Schedule of Topics

Topic 1	Linear regression
Topic 2	Logistic regression
Topic 3	Principal component analysis
Topic 4	Cluster analysis

Textbook and Materials

The course material will be made available during the course: slides, readings, datasets, supplementary materials (scripts etc).

Suggested reading

Bishop C.M. (2006). Pattern Recognition and Machine Learning. Springer.

Marden J.I. (2015). Multivariate Statistics. http://stat.istics.net/Multivariate/

Duda R.O., Hart P.E., Stork D.G. (2001). Pattern Classification. Wiley, 2nd Edition.

Additional suggestions

Hastie T., Tibshirani R., Friedman J. (2009). The Elements of Statistical Learning: Data Mining, Inference, and Prediction, Second Edition. Springer, Springer Series in Statistics. http://www-stat.stanford.edu/ElemStatLearn/

Witten J.D., Hastie T., Tibshirani R. (2014). An Introduction to Statistical Learning with Applications in R. Springer, Springer Series in Statistics.

Assessment
30% assignment
70% final written exam

Office hours
By appointment.

E-mail
roberto.rocci@uniroma2.it

NOTE: Since the course is provided for a limited number of students, **Erasmus students** who would like to attend this course and take this exam need to contact the Secretariat of Global Governance by e-mail global.governance@uniroma2.it for registration.