Laurea / B.A.
in Global Governance

Academic Year 2023-2024
Syllabus
Statistical Learning
9 CFU
Prof. Alessio Farcomeni - Dott. Marco Stefanucci

Course Description

The course covers some statistical techniques for supervised and unsupervised learning. The R software for statistical computing will be also introduced and used throughout.

Supervised learning techniques are used to predict a target variable (linear and logistic regression) based on predictors, and/or to assess interrelationships among predictors and a target variable (linear and logistic regression). As an example, suppose you want to predict the risk that a family will be materially deprived next year. This can be done by using data that can be measured at baseline (number of family members, disposable income, health status, etc.) and use these to predict material deprivation for a sample of families with known status. Incidentally, you will also understand how health status affects the risk of material deprivation.

Unsupervised learning techniques are used to find groups in data, that is, to predict target categorical variables that are not measured (cluster analysis). Additionally, they are used to summarize data (dimension reduction, done with principal component analysis in this course). As an example, suppose you want to assess an unmeasurable trait, like happiness. Suppose your target units are geographic regions. Happiness can be measured indirectly through a series of variables (questionnaires, indices, etc.). A general score is obtained through dimension reduction by finding the optimal weighted average of all measurements. Cluster analysis will separate regions in few (two, three, four) groups, with respect to levels of happiness. Different policies can then be scheduled for each group.

The last 3 CFU will be dedicated to machine learning methods (classification and regression trees, random forests, shallow and deep neural networks) for supervised learning. Modern applications will be then introduced, where data is extracted from text corpora (natural language processing), images (computer vision), audio tracks.

The main objectives of this course are to provide students with the ability to select the statistical learning technique needed to answer specific questions (based on data), to perform data analysis appropriately, and to interpret the results correctly.

Prerequisites

Prerequisite is an introductory statistics and statistical inference course like "Statistical Tools for Decision Making" of the B. A. in Global Governance. Also some math is essential, but only few derivations are made.

Teaching Method

The course is carried out through lectures and practicums. Techniques will be introduced by examples and described in mathematical formulas. Focus will be on the practical implementation of each technique, and interpretation of results. In the final part of the lesson students will be able to practice the newly introduced topics.

Schedule of Topics

Topic 1	Introduction to R software
Topic 2	Linear regression
Topic 3	Logistic regression

Topic 4	Principal component analysis
Topic 5	Cluster analysis
Topic 6	Machine learning methods for supervised learning
Topic 7	Modern applications: text mining, image processing

Textbook and Materials
Reading material on each course topic (handouts, slides, data sets, R scripts), will be made available to the students by the course instructors during the course.

Suggested books are:
Witten J.D., Hastie T., Tibshirani R. (2014). An Introduction to Statistical Learning with Applications in R. Springer, Springer Series in Statistics
Chatfield, C. and Collins, A. J. (1981) Introduction to Multivariate Analysis, Chapman & Hall/CRC Press
Everitt, B. S. and Hothorn, T. (2006) A Handbook of Statistical Analyses Using R. CRC Press.
Available for free at:http://www.ecostat.unical.it/tarsitano/Didattica/LabStat2/Everitt.pdf
Additional (more technical) reading:
Hastie T., Tibshirani R., Friedman J. (2009). The Elements of Statistical Learning: Data Mining, Inference, and Prediction, Second Edition. Springer, Springer Series in Statistics. Available for free at: https://web.stanford.edu/~hastie/ElemStatLearn/

Assessment
Assessment for attending students will be based on a written exam. This will include closed and open questions. A midterm written exam will be held.
Non attending students will have to take an oral examination in addition to the written exam.

Office hours
upon appointment by e-mail

E-mail
marco.stefanucci@uniroma2.it
alessio.farcomeni@uniroma2.it

NOTE: If you are an Erasmus or a non Global Governance student who would like to attend one or more courses in the Global Governance programme, please be aware that, **before enrolling in the course**, you should have read the code of conduct and the procedural rules characterizing our programme. We assume that, if you enroll in the course, **you have read and accepted all Global Governance values and rules**. Notice that attendance is required from the very first lesson and you need to attend at least 80% of the course to be considered an attending student.

Description of the methods and criteria for testing learning
The examination assesses the student's overall preparation, ability to integrate the knowledge of the different parts of the programme, consequentiality of reasoning, analytical ability and autonomy of judgement. In addition, language property and clarity of presentation are assessed, in accordance with the Dublin descriptors (1. knowledge and understanding; 2. applying knowledge and understanding; 3. making judgements; 4. learning skills; 5. communication skills).
The final grade will be related 70% to the degree of knowledge and 30% to the expressive capacity (written and oral) and autonomous critical judgement demonstrated by the student.
The examination will be graded according to the following criteria:

Unsuitable: important deficiencies and/or inaccuracies in the knowledge and understanding of the topics; limited capacity for analysis and synthesis, frequent generalisations and limited critical and judgement skills; the topics are exposed in an incoherent manner and with inappropriate language.
18-20: barely sufficient knowledge and understanding of the topics, with possible generalisations and imperfections; sufficient capacity for analysis, synthesis and autonomy of judgement; the topics are frequently exposed in an inconsistent manner and with inappropriate/technical language;
21-23: surface knowledge and understanding of the topics; ability to analyse and synthesise correctly with sufficiently coherent logical argumentation and appropriate/technical language.
24-26: fair knowledge and understanding of the topics; good analytical and synthetic skills with rigorously expressed arguments but not always appropriate/technical language.
27-29: complete knowledge and understanding of the topics; considerable capacity for analysis and synthesis. Good autonomy of judgement. Arguments presented in a rigorous manner and with appropriate/technical language.
30-30L: very good level of knowledge and thorough understanding of topics. Excellent analytical and synthetic skills and independent judgement. Arguments expressed in an original manner and in appropriate technical language.