Statistics for Big Data (61209)

Instructors: D. KARLIS

Core Course, 2nd semester, 3.5 ECTS units
Course level: Graduate (MSc)
Language: English

Course Description

The era of big data has led to paradigm shift in statistical methods. New challenges and new problems have occurred due to the abundance of data. The aim of the present course is to discuss the changes on the statistical methods and how the huge volume of data affects the classical statistical methods, while at the same time, how the new problems occurring can be solved via «state of the art» new statistical methodologies. We have selected some methods and problems to exploit showing the new potential and the new dynamics of statistical science towards new problems. Topics that we cover relate to: new challenges in Statistics, Regression for Big Data, Regularization problems, Multiplicity problems and Statistical analysis in networks.

Prerequisites
The course implies a good prior knowledge of statistics. Basic knowledge of programming in R is required.

Target Learning Outcomes
The target of the course is to show the needs and challenges produced by the big data era and how statistical methodologies can handle them . After completing the course, the students will be able to:
- understand the new challenges and problems due to the abundance of data
- apply new techniques defined to handle problems with big data
- implement the new methods using R

Recommended Bibliography
- C. Giraud (2015). Introduction to High-Dimensional Statistics. Philadelphia: Chapman and Hall/CRC.
- T. Tony Cai, Xiaotong Shen, ed. (2011). High-dimensional data analysis. Frontiers of Statistics. Singapore: World Scientific.
- P. Bühlmann and S. van de Geer (2011). Statistics for high-dimensional data: methods, theory and applications. Heidelberg; New York: Springer.
- T. Hastie, R. Tibshirani and R. Friedman (2009) Elements of Statistical Learning, Springer.
- E. D. Kolaczyk (2014) Statistical Analysis of Network Data with R. Springer

Teaching and Learning Activities

Course lasts 6 3-hours lectures (one each week). Every week there will be exercises as homework (some to be submitted).

Assessment and Grading Methods

The final grade is the weighted average of the final examination grade (weight 70%) and the three assignment/projects (weight 30%). There will be two small assignments and a team project.