450SM - STATISTICAL LEARNING FOR DATA SCIENCE

Aims

This course will equip students with both the theory and practical knowledge necessary for starting analyzing complex data generated in modern life. Students will learn how to perform data analysis in an exploratory style, create effective graphs that help them understand the data, build sophisticated predictive models and communicate the extracted information to others. Meanwhile, they will get familiar with R, a very powerful software tool for statistical analysis. In the following, objectives are split.

Knowledge and understanding: know main kinds of problems which can be tackled with SL techniques; know main SL techniques; know design, development, and assessment phases of a SL procedure.

Applying knowledge and understanding: formulate a formal problem statement for simple practical problems in order to tackle them with SL techniques; develop simple end-to-end SL procedures; experimentally assess a simple end-to-end SL procedure.

Making judgements: judge the technical soundness of a SL procedure and of the assessment of a SL procedure.

Communication skills: describe the motivations behind choices in the design, development, and assessment of a SL procedure, possibly exploiting simple plots; being able to communicate the results to experts and to non-experts.

Learning skills: retrieve information from scientific publications about SL techniques (also not explicitly presented in the course) possibly combining them to solve complex problems.

Teaching Format

Frontal lessons with blackboard and slide projection; exercises, under teacher supervision, in dealing with simple problems with SL techniques.

Assessment

Final exam according to one of the following two options (student's choice):
- Written test + project (the final mark is the average of the two marks).
- Written test only with questions on theory and applications with medium- and short-length open answers.

Project (home assignment): the student chooses a problem among a closed, teacher-defined set of problems and proposes a solution based on SL techniques. The expected outcome is a written document (few pages) including the problem statement; a description of the proposed solution; the results and a discussion about the experimental assessment of the solution with, if applicable,

information about used data. Student may form groups for the project. The project will be evaluated according also to clarity.