High Dimensional Statistics

High Dimensional Statistics

Instructor: Panagiotis Papastamoulis

General Course Information
Code: 61227
Type: Elective
Level: Graduate
Year of study: 2nd
Semester: 4th
ECTS: 5
Teaching language: Greek/English

Course Description
Modern statistical applications deal with the analysis of high-dimensional datasets. However,
inference is extremely challenging due to the curse of dimensionality: as the number of variables
increases, the data may become quite noisy and it is difficult to separate the actual underlying
information. The course will present modern statistical techniques particularly suited to problems
with high dimensionality, such as:

• large scale hypothesis testing
• statistical applications in bioinformatics
• applications of (generalized) linear models in big data problems
• Regularization techniques (LASSO, Ridge regression)
• Classification and model-based clustering for multivariate data with dimensionality reduction techniques
• Statistical network analysis

Prerequisites
• Prerequisites
– Hypothesis Testing - Linear Models
– Generalized Linear Models
• Some basic knowledge on
– Computational Statistics
– R programming

• Students will also benefit from the following courses (not required)
– Applied Bayesian Statistics
– Statistical Learning

Key Outcomes
After completing the course, the students will
• know the challenges that big data impose to the statistician
• properly deal with large scale hypothesis testing
• put their hands on many different types of data from cutting edge research areas such as bioinformatics and network analysis
• learn many novel statistical ideas and methods developed in the last 20 years
• know how to apply the relevant methods using R and Bioconductor.

Teaching and Learning Activities
One three-hour lecture per week, study exercises and programming exercises as homework (some to be submitted).

Assessment and Grading Methods
The final grade will be a weighted average of 2 homework assignments and a written examination in the lab.