

Year: 2011-2012 Masters Track Statistical Sciences

Course name: *Statistical Learning Theory*

ECTS: 4

Level: 500

Period: Semester 1

Lecturers: Prof. dr. P. Grünwald (CWI, coordinator), Prof. Dr. J.J. Meulman

This course gives an overview of techniques for automated learning from ill-understood data for which it is hard or impossible to formulate a model that is even approximately correct. Here “learning” means: “finding structure, patterns, regularities” and using these patterns to predict future data. The field is also known as “machine learning”, since many contributions in this field have their origin in computer science areas (pattern recognition, artificial intelligence).

Main topics in the course will be (1) supervised learning (regression and classification, but with a strong focus on the latter); (2) model selection and model averaging, (3) predictive analysis including sequential prediction. The methods discussed will include various classical and state-of-the-art classification methods: naive Bayes, perceptrons (1960s), neural networks, decision trees (1980s), logistic regression, boosting support vector machines, Gaussian processes and other kernel approaches (2000s). We explain interrelations between these methods and analyze their large-sample behaviour. As for model selection and averaging, we again consider both classical and state of the art methods including AIC, BIC, Bayes factor model averaging, Minimum Description Length (MDL), Structural Risk Minimization (SRM), Shrinkage, Lasso and other L1-methods. We explain how all these methods are related to Bayesian and non-Bayesian methods for combining predictors, and again we analyze their large-sample behaviour.

Form:

Lectures and practicals (partly computer practicals, partly exercises)

Each 2 hrs/day; Written exam and practical assignment

Literature: T. Hastie, R. Tibshirani, J. Friedman, *The Elements of Statistical Learning*, 2nd edition, 2009. See: <http://www-stat.stanford.edu/~tibs/ElemStatLearn/>

Handouts of some (very few) papers

Course name: Statistical Genetics

ECTS: 6

Level: 500

Period: Semester 1

Lecturers: Dr Stefan Bohringer (LUMC), Prof Dr Fred van Eeuwijk (WUR), Prof Dr Jeanine Houwing-Duistermaat (coordinator), Prof Dr Aad van der Vaart (VU)

Statistical genetics is an important topic within the Life and Behavioral Sciences. The structure within families (pedigrees) and the dependency structure between genetic markers at the genome require specific statistical tools and software packages for analysis of the data. The topics addressed during this course will be: estimation of heritability, detecting new genetic factors using genome wide linkage and association studies, the analysis of candidate genes, and data bases such as Hapmap. Several software packages will be introduced. Data sets from human genetics and from animal and plant breeding will be analyzed.

Materials: Lecture notes, review papers

Lectures: Three weeks three days and four weeks two days
Each day is from 10.00-15.00 (2 lectures and 2 practical's)

Exam: Two assignments and a written exam. The two assignments will contribute each 25% to the final result.

Course name: High-dimensional data analysis

ECTS: 6

Level: 500

Period: Semester 1

Lecturers: Dr. J. Goeman (LUMC, coordinator), Dr. M.A. van de Wiel (VU University Medical Center), Dr. W.N. van Wieringen (VU University Amsterdam).

This course gives an overview of statistical methods that are used for analyzing high-dimensional data sets in which many variables (often thousands) have been measured for a limited number of subjects. This type of data arises in genomics, where genetic information is measured for many thousands of genes simultaneously, but also in functional MRI imaging of the brain. The first part of the course covers the most important statistical issues in this field, which include multiple testing (empirical Bayes methods, familywise error rate and false discovery rate control), gene set testing, prediction methods in high dimensions (penalized regression methods, principal component regression, cross-validation). The second part explores specific issues in specific types of high dimensional data (gene expression studies, DNA copy number analysis, proteomics and fMRI imaging).

Philosophy:

Teaching students the adjustments to classical statistical methodology, necessary to tackle high-dimensional data

Goals:

Students should be able to perform and understand the most common analysis types: limma, FDR and FWER control methods, clustering, global test, ridge regression, lasso and principal components regression with cross-validation, and be familiar with the specific issues in important types of high dimensional data sets.

The course consists of a series of lectures and practicals (partly computer practicals, partly exercises). Each 2 hrs/week.

Grades will be based on a written exam and a practical assignment.

Course materials:

Review papers and book chapters.

Course name: Psychometrics(including SEM)

ECTS: 6

Level: 500

Period: Semester 1

Lecturers: Prof. dr.H. Kelderman (Free University Amsterdam, coordinator), Prof.dr. W.J. Heiser

This course will provide students with an overview of tools for the analysis of test data. They will learn to understand and apply these tools using statistical software. During the course, students will work on the analysis of empirical data and make exercises about the theory. A more general aim is to enhance the student's psychometric research skills. In psychology and education, attributes of individuals are often measured with tests. A test consists of a number of separate items, questions or problems to be solved. The responses are used to obtain a score that indicates the degree to which a person possesses a certain quality, e.g. compulsiveness or spatial intelligence. Psychologists are interested in various aspects of the scores of such tests. In particular they want to know something about its meaning, reliability, validity, and the best way to obtain such a score. To this end statistical theories for tests and measurements have been developed. In this course you will learn to understand the main test theories and to apply them. Substantive issues are only cursorily discussed; this is primarily an applied statistics course.

The course has two parts: Part I deals with traditional test theory, Part II with modern test theory. The former is most often used, but the latter is much more powerful and elegant and has a usefulness that goes far beyond that of traditional test theory. It also provides great research opportunities for students. Some more advanced applications of modern test theory are discussed at the end of the Part II. All computations and simulations will be performed with R.

Literature:

McDonald R. P. (1999). Test Theory: A Unified Treatment, London: Lawrence Erlbaum.

Henk Kelderman: Psychometrics: Study notes and weekly exercises

Henk Kelderman: Psychometrics: Assignments

Other texts to be obtained from Studion

Course name: Statistical Consulting

ECTS: 6

Level: 500

Period: Semester 1

Lecturers: Prof.dr. W.J. Heiser (coordinator), Prof. dr. Th. Stijnen, Prof. dr. F.A. Van Eeuwijk

Consulting is an integral part of an applied statistician's life, whether working in academia, government or industry. Statistical and methodological consultation may involve brief interactions to respond to very specific technical questions of a single client, or long term associations as a member of research projects. It always involves a form of 'reverse reasoning', that is, figuring out the conclusion you want to come to, and finding a way to get there. Specifically, the consultant's problem is: What statistical procedure(s) could provide an appropriate answer to the client's research question, within all practical constraints?

The goal of this course is to develop some consultancy skills, such as how to effectively communicate with researchers and practitioners. Emphasis will be given to issues of recognizing the client's problem, dealing with assumptions, finding appropriate data analysis options and correct statistics, within a consultation setting. Each student will practice in individual case studies with actual clients.

Literature:

Course materials: