Master´s Programme in Statistics and Data Mining

120 credits

Master´s Programme in Statistics and Data Mining

F7MSG

Valid from:

Determined by
Board of the Faculty of Arts and Sciences

Date determined
2015-06-16

Aim

The rapid growth of databases provides scientists and business people with vast new resources. This programme meets the challenges of turning large or complex data sets into knowledge. Statistical modelling and analysis is integrated with machine learning, data mining and visualization into a solid basis for professional work with the organization and analysis of data, or a career in research.

Learning outcomes

Upon completing the programme the students shall be able to:

- extract and organize large volumes of complexly structured data
- explore, summarize and present large and complex data sets by static, interactive and dynamic graphical facilities
- select a suitable model for a given statistical problem and dataset
- uncover and statistically verify previously unknown patterns and trends in the data
- use advanced statistical and data mining computer software to analyse large data volumes
- implement models suitable for data analysis in some computer language
- combine data information with other sources of prior information to improve inference and prediction performance
- give examples of application areas where analysis of large and complex data sets is needed
- present a written thesis with a theoretical or an applied study of a complex data set

Knowledge and understanding

Upon completing the programme the student shall

- demonstrate knowledge and understanding in statistics, including both broad knowledge of the field and a considerable degree of specialised knowledge in its branch, data mining, as well as insight into current research and development work, and
- demonstrate specialised methodological knowledge in statistics.

Specialized knowledge in data mining shall include modern powerful techniques for prediction, classification, clustering, Bayesian methods and association analysis.

Competence and skills

Upon completing the programme the student shall

- demonstrate the ability to critically and systematically integrate knowledge and analyse, assess and deal with complex phenomena, issues and situations even with limited information
- demonstrate the ability to identify and formulate issues critically, autonomously and creatively as well as to plan and, using appropriate methods, undertake advanced tasks within predetermined time frames and so contribute to the formation of knowledge as well as the ability to evaluate

this work
- demonstrate the ability in speech and writing both nationally and internationally to report clearly and discuss his or her conclusions and the knowledge and arguments on which they are based in dialogue with different audiences, and
- demonstrate the skills required for participation in research and development work or autonomous employment in some other qualified capacity.

Judgement and approach

Upon completing the programme the student shall
- demonstrate the ability to make assessments in statistics informed by relevant disciplinary, social and ethical issues and also to demonstrate awareness of ethical aspects of research and development work
- demonstrate insight into the possibilities and limitations of research, and especially research in statistics and data mining, its role in society and the responsibility of the individual for how it is used, and
- demonstrate the ability to identify the personal need for further knowledge and take responsibility for his or her ongoing learning.

Content

The curriculum joins courses in statistics, computer science and mathematics into a programme in the interface between statistics and computer science. Compulsory courses, introductory courses, and a 30-credit master's thesis ensure progression and depth. Introductory courses are offered to fill in knowledge gaps and ensure that the students are properly prepared for the other courses.

Compulsory Courses

ADVANCED ACADEMIC STUDIES, 3 CREDITS
(given in semester 1)
The aim of the course is to prepare the students for advanced academic studies and also to let the students learn the academic culture in general. A basic ambition is to supply essential tools to the students on the master´s level in Sweden. In addition, practical issues that are specific for the programme are to be discussed.

INTRODUCTION TO MACHINE LEARNING, 9 CREDITS
(given in semester 1)
Basic concepts in machine learning and data mining. Bayesian and frequentist modelling, model selection. Linear regression and regularization. Linear discriminant analysis and logistic regression. Bagging and boosting. Splines, generalized additive models, trees, and random forests. Kernel smoothers and support vector machines. Gaussian process.

DATA MINING – CLUSTERING AND ASSOCIATION ANALYSIS, 15 CREDITS
(given in semester 2)
Principles and tools for dividing objects into groups and discovering relationships hidden in large data sets. Partitional methods and hierarchical clustering. Cluster evaluation. Association analysis using item sets and association rules. Evaluation of association patterns.

PHILOSOPHY OF SCIENCE, 3 CREDITS
(given in semester 2)
Laws of nature and scientific models. Relations between theories and observations. Forces prompting scientific change.

BAYESIAN LEARNING, 6 CREDITS
(given in semester 2)
Bayes' theorem to combine data information with other prior information. Bayesian analysis of conjugate models. Markov Chain Monte Carlo methods for Bayesian computations. Bayesian model comparison.

COMPUTATIONAL STATISTICS, 6 CREDITS
(given in semester 2)
Computer arithmetic. Random number generation and simulation techniques. Markov Chain Monte Carlo methods. Numerical linear algebra. Optimization methods in statistics.

Profile courses

VISUALIZATION, 6 CREDITS
(given in semester 1 for students admitted in an even year and in semester 3 admitted in an odd year)
Advanced visualization techniques for large and complex data sets. Interactive and dynamic statistical graphics. Visualizing spatial information.

ADVANCED MACHINE LEARNING, 6 CREDITS
(given in semester 3)
Bayesian networks and hidden Markov models. State Space models and random fields. Neural networks. Principles of deep learning and its tools: deep neural networks, Boltzman machines.

TIME SERIES ANALYSIS, 6 CREDITS
(given in semester 1 for students admitted in an odd year and in semester 3 admitted in an even year)
Time series decomposition. Autocorrelation and partial autocorrelation. Forecasting using
time series regression, ARIMA models and transfer functions. Intervention analysis. Trend
detection.

MULTIVARIATE STATISTICAL METHODS, 6 CREDITS
(given in semester 1 for students admitted in an odd year and in semester 3 for students admitted in an even year)
Analysis of correlation and covariance structures, including principal components, factor analysis and canonical correlation. Classification and discrimination techniques. Multivariate inference.

PROBABILITY THEORY, 6 CREDITS
(given in semester 3)
Multivariate random variables and conditioning. Order variables. Characteristic functions and other transforms. The multivariate normal
distribution. Probabilistic convergence concepts.

STATISTICAL EVIDENCE EVALUATION, 6 CREDITS
(given in semester 3)
Probabilistic reasoning and likelihood theory. Bayesian hypothesis testing. Bayesian belief networks. Statistical decision theory and influence diagrams.

Elements of forensic theory. Sensitivity analysis.

Complementary courses

WEB PROGRAMMING AND INTERACTIVITY, 4 CREDITS
(given in semester 2)
An overview of general Web architectures, including HTML, DHTML, XML, PHP.

NEURAL NETWORKS AND LEARNING SYSTEMS, 6 CREDITS
(given in semester 2)
Unsupervised learning: principal component analysis, independent component analysis, vector quantization. Supervised learning: neural networks, radial basis functions, support vector machines. Reinforcement learning: Markov processes, Q-learning, genetic algorithms.

DATA MINING PROJECT, 6 CREDITS
(given in semester 3)
Project course in which the student specifies, implements and evaluates a data mining algorithm for a specific data mining problem.

TEXT MINING, 6 CREDITS
(given in semester 3)
Retrieval of textual data from different sources. Text processing by means of computational linguistics. Statistical models for text classification and prediction.

OPTIMIZATION, 6 CREDITS
(given in semester 3)
Fundamental concepts within optimization, such as mathematical modelling, optimality conditions, convexity, and sensitivity analysis, and Lagrangean relaxation. Basic theory and methods for linear and nonlinear optimization, and integer and network optimization.

DATABASE TECHNOLOGY, 6 CREDITS
(given in semester 3)
General database management systems (DBMS). Methods for data modelling and database design. ER-diagrams, relational databases and data structures for databases. Architectures and query languages for the relational model. Relational algebra and query optimization.

BIOINFORMATICS, 6 CREDITS
(given in semester 3)
Basic molecular biology. Basic statistics for biology. Hidden Markov models. QTL-mapping. Sequence alignment. Analysis of Next Generation Sequencing data. Microarray analysis.

INTRODUCTORY COURSES

STATISTICAL METHODS, 6 CREDITS
(given in semester 1)
Concept of probability. Random variable, common statistical distributions and their properties. Point and interval estimation. Hypothesis testing. Simple and multiple linear regression. Resampling. Elements of Bayesian theory.

ADVANCED R PROGRAMMING, 6 CREDITS
(given in semester 1)
R Environment. General programming techniques. Language concepts of R: variables, vectors, matrices, data frames. Language tools: operators, loops, conditions, functions. Importing data from text and spreadsheet files. Using external R packages. Graphics. Object-oriented programming. Performance

enhancement and parallel programming.

MASTER'S THESIS, 30 CREDITS
Theoretical or applied study of a complex data set by using statistical and data mining methods.

Teaching and working methods

Ordinary courses have lectures, seminars, and computer exercises. The lectures are devoted to presentations of theories, concepts, and methods. The seminars comprise presentations and discussions of assignments. The computer exercises provide practical experience of data analysis and other methods taught in the programme. The courses that are named projects have supervision only.

Examination

Ordinary courses yielding a minimum of 4.5 credits have one or more assignments and one written examination. Project courses and the master's thesis are examined through a written report and oral defence of that report.

Grades

As stipulated in the course syllabi.

Entry requirements

A bachelor's degree in one of the following subjects: statistics, mathematics, applied mathematics, computer sicence, enginering or a simlar degree. Courses in culculus and linear algebra, statistics and programming are also required.

Each applicant must enclose a Letter of Intent written in English, explaining why they want to study this programme, and a summary of their bachelor's essay or project. If applicants hold a degree that does not include a bachelor's essay or project, their Letter of Intent should describe previous studies and any academic activities that are related to the master's programme or the programmes applied for.

English B/English 6 or equivalent.

Threshold requirements

The student must have passed at least 6 ECTS credits of the first semester, in order to be admitted to the second semester of the programme.

The student must have passed at least 40 ECTS credits of the first year in order to be admitted to the third semester of the programme.

The student must have passed at least 65 ECTS credits of the programme, including all obligatory courses, in order to be admitted to the fourth semester of the programme.

Degree requirements

The student will be awarded the degree of Master of Science (120 ECTS credits) in Statistics provided all course requirements are completed and that the student fulfils the general and specific eligibility requirements including proof of holding a Bachelor's (kandidat) or a corresponding degree.

To be awarded the degree the students must have passed 90 ECTS credits of courses including 42 ECTS credits of the compulsory courses, a minimum of 6 ECTS credits of the introductory courses, a minimum of 12 ECTS credits of the profile courses, and, possibly, some amount of complementary courses. The students must also have successfully defended a master's thesis of 30 ECTS credits.

Completed courses and other requirements will be listed in the degree certificate.

A degree certificate is issued by the Board of the Faculty of Arts and Sciences on request.

Degree in Swedish

Filosofie masterexamen i huvudområdet statistik

Degree in English

Master of Science (120 Credits) with a major in Statistics

Specific information

Merits

The specific requirements will be assessed as not fulfilled if the average grade is in the lower third of the grading scale used in the country where the degree was awarded, that is grades have to be average/pass or above (the equivalent to the Swedish grade "Godkänd").

Transfer of Credits

The Board of the Faculty of Arts and Sciences or person nominated by the Board decides
whether or not previous education can be transferred into the programme.

Enrolment Procedure

Students are admitted to the programme in its entirety.

Language of instruction

The language of instruction is English.

Curriculum

Semester 1 (Autumn 2016)

Course code	Course name	Credits	Level	Weeks	ECV
732A34	Time Series Analysis	6	A1X	v201634-201641	E
732A93	Statistical Methods	6	A1X	v201634-201641	E
732A94	Advanced Programming in R	6	A1X	v201634-201641	E
732A42	Introduction to Advanced Academic Studies	3	A1X	v201634-201702	C
732A37	Multivariate Statistical Methods	6	A1X	v201643-201651	E
732A95	Introduction to Machine Learning	9	A1X	v201643-201652	C

Semester 2 (Spring 2017)

Course code	Course name	Credits	Level	Weeks	ECV
732A90	Computational Statistics	6	A1X	v201704-201711	C
732A31	Data Mining - Clustering and Association Analysis	15	A1X	v201704-201723	C
720A04	Philosophy of Science	3	A1X	v201718-201723	C

Semester 3 (Autumn 2017)

Course code	Course name	Credits	Level	Weeks	ECV
732A57	Database Technology	6	A1X		E
732A62	Time Series Analysis	6	A1X	v201734-201741	E
732A63	Probability Theory	6	A1X	v201734-201741	E
732A66	Decision Theory	6	A1X	v201734-201751	E
732A96	Advanced Machine Learning	6	A1X	v201735-201741	E
732A65	Data Mining Project	6	A1N	v201735-201751	E
732A92	Text Mining	6	A1X	v201744-201751	E
732A97	Multivariate Statistical Methods	6	A1X	v201744-201751	E

Semester 4 (Spring 2018)

Course code	Course name	Credits	Level	Weeks	ECV
732A64	Master Thesis in Statistics	30	A1X	v201803-201823	C

ECV = Elective / Compulsory /Voluntary
*Kursen läses över flera terminer