

# Masters Programme in Statistics and Machine Learning

120 credits

Masters Programme in Statistics and Machine  
Learning

F7MSL

Valid from: 2018 Autumn semester

**Determined by**  
Board of the Faculty of Arts and Sciences

**Date determined**  
2017-11-24

## Introduction

The rapid IT development has led to the overwhelming of society with enormous volumes of information generated by large or complex systems. Information can be stored in large databases, it can come in a streaming manner or it can be a result of the interaction between the system and the learning environment. This programme meets the challenges of learning from these complex information volumes by means of models and algorithms which enable for efficient prediction, analysis and decision making. Statistical modelling and analysis is integrated with machine learning, data mining and data management into a solid basis for professional work with the information modelling and analysis of data in large or complex systems. The program also provides excellent qualifications for a career in research.

## Aim

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, machine learning, as well as insight into current research and development work, and
- demonstrate specialised methodological knowledge in statistics.

Specialized knowledge in machine learning shall include modern powerful techniques for classification and regression, prediction, methods for statistical simulation and optimization, Bayesian methods and methods for analysis of large databases.

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 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.

Upon completing the programme the students shall be able to:

- model information volumes that are generated by large or complex systems
- select a suitable model in a given context
- 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
- use advanced software to analyse large or complex data volumes
- implement models suitable for data analysis, prediction and decision making 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 it is required to model information volumes that emerge from large or complex systems.
- uncover and statistically verify previously unknown patterns and trends in the data
- present a written thesis with a theoretical or an applied study of large or complex systems or data sets by means of methods from statistics and machine 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. Period

### 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.

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, 6 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.

Big Data Analytics, 6 credits (given in semester 2)

File systems and databases for Big Data. Querying for Big Data. Resource management in a cluster environment. Parallelizing computations for Big Data. Machine Learning for Big Data.

Introduction to Python, 3 credits (given in semester 2)

Python environment. Data structures: numbers, strings, lists, tuples, dictionaries. Basic language elements: loops, conditions, functions. Modules. Input and Output. Debugging. Machine learning and data mining in Python.

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. Gaussian processes. Kalman filtering. Particle methods.

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)

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.

Decision Theory, 6 credits (given in semester 3)

Probabilistic reasoning and likelihood theory. Bayesian hypothesis testing. Decision theoretic elements. Utility and loss functions. Graphical modeling as a tool for decision making. Sequential analysis.

### **Complementary courses**

Web Programming, 6 credits (given in semester 2)

Overview of WWW, HTML, Javascript and other client-server techniques. Programming languages Python, Flask, SQL, Websockets, JSON and other server-side technologies

Bioinformatics, 6 credits (given in semester 1 for students admitted in an even year and in semester 3 admitted in an odd year)

Basics of molecular biology and genetics. Hidden Markov models, genetic sequence analysis. Sequence similarity, sequence alignment. Phylogeny reconstruction. Quantitative trait modelling. Microarray analysis. Network biology.

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.

Research Project, 6 credits (given in semester 3)

Project course in which the student develops, improves or compares machine learning or data mining models and algorithms for a specific research 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.

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.

### **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. Literate programming. Developing R packages.

Master's thesis, 30 credits  
Theoretical or applied study of a complex data set by using statistical, machine learning 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

Bachelor's degree equivalent to a Swedish Kandidatexamen within statistics, mathematics, applied mathematics, computer science, engineering or a similar degree. Courses in calculus and linear algebra, statistics and programming are also required.

English corresponding to the level of English in Swedish upper secondary education (English 6/B).

## 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

### **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 2018)

Course code	Course name	Credits	Level	Weeks	Timetable module	ECV
732A60	Advanced Academic Studies	3	A1X	v201834-201842	-	C
732A76	Research Project	6	A1N	v201834-201843	-	E
732A93	Statistical Methods	6	A1X	v201835-201842	-	E
732A94	Advanced Programming in R	6	A1X	v201835-201842	-	E
732A98	Visualization	6	A1X	v201836-201842	2	E
732A51	Bioinformatics	6	A1X	v201845-201851	-	E
732A99	Machine Learning	9	A1X	v201845-201851	-	C

### Semester 2 (Spring 2019)

Course code	Course name	Credits	Level	Weeks	Timetable module	ECV
732A55	Neural Networks and Learning System	6	A1X	v201904-201911	-	E
732A56	Web programming	6	A1X	v201904-201911	-	E
732A74	Introduction to Python	3	A1X	v201904-201911	-	C
732A75	Advanced Data Mining	6	A1X	v201904-201911	-	C
732A90	Computational Statistics	6	A1X	v201904-201911	-	C
732A54	Big Data Analytics	6	A1X	v201913-201921	-	C
732A91	Bayesian Learning	6	A1X	v201913-201921	-	C
720A04	Philosophy of Science	3	A1X	v201913-201923	-	C

### Semester 3 (Autumn 2019)

Course code	Course name	Credits	Level	Weeks	Timetable module	ECV
732A77	Artificial Intelligence	6	A1N	v201935-201944	-	E
732A62	Time Series Analysis	6	A1X	v201936-201944	2	E
732A63	Probability Theory	6	A1X	v201936-201944	3	E
732A96	Advanced Machine Learning	6	A1X	v201936-201944	1	E
732A66	Decision Theory	6	A1X	v201936-202003	4	E
732A76	Research Project	6	A1N	v201936-202003	4	E
732A57	Database Technology	6	A1X	v201945-202003	1	E
732A92	Text Mining	6	A1X	v201945-202003	2	E
732A97	Multivariate Statistical Methods	6	A1X	v201945-202003	3	E

### Semester 4 (Spring 2020)

Course code	Course name	Credits	Level	Weeks	Timetable module	ECV
732A64	Master Thesis in Statistics	30	A1X	v202004-202023	-	C

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