UNIVERSITY OF GOTHENBURG

DEPARTMENT OF MATHEMATICAL SCIENCES

MSA220 Statistical Learning for Big Data, 7.5 credits

Statistisk slutledning för stora datamängder, 7,5 högskolepoäng

Second Cycle

Confirmation

This course syllabus was confirmed by Department of Mathematical Sciences on 2015-02-27 and was last revised on 2019-05-14 to be valid from 2019-09-02, autumn semester of 2019.

Field of education: Science 100%
Department: Department of Mathematical Sciences

Position in the educational system

The course can be part of the following programmes: 1) Mathematical Sciences, Master's Programme (N2MAT) and 2) Applied Data Science Master's Programme (N2ADS)

Main field of studies	*Specialization*
Mathematical Statistics	A1N, Second cycle, has only first-cycle course/s as entry requirements

Entry requirements

The prerequisites for the course are a basic course in statistical inference and the course *MSG500 Linear Statistical Models*.

Learning outcomes

On successful completion of the course the student will be able to:

- demonstrate understanding of the key concepts and ideas concerning classification, clustering and dimension reduction.

- solve high-dimensional data analysis exercises and interpret the results of such analyses.

Course content

- Overview of high-dimensional data analysis
- Classification:Bayes rule, discriminant analysis methods, nearest neighbour classifier, classification and regression trees.
- Cost functions, greedy searches, gradient descent, cross-validation.
- Logistic regression
- Regularization methods. Sparse logistic regression, sparse discriminant analysis.
- Ensemble methods: bagging, random projections, random forests.
- Clustering: k-means, hierarchical clustering, model-based clustering, spectral methods.
- Dimension reduction: PCA, canonical correlation, multi-dimensional scaling.
- Special topics (subset of the following): networks and graphical models, sparse covariance estimation, network clustering and community detection, neural networks, matrix completion, collaborative filtering.
- Large-scale learning: stochastic searches, batch-methods, online learning.

Form of teaching

The teaching is organized with lectures, discussions, and reading assignments.

Language of instruction: English

Assessment

Oral and/or written examination.

If a student, who has failed the same examined component twice, wishes to change examiner before the next examination, a written application shall be sent to the department responsible for the course and shall be granted unless there are special reasons to the contrary (Chapter 6, Section 22 of Higher Education Ordinance).

In cases where a course has been discontinued or has undergone major changes, the student shall normally be guaranteed at least three examination occasions (including the ordinary examination) during a period of at least one year from the last time the course was given.

Grades

The grading scale comprises: Pass with Distinction (VG), Pass (G) and Fail (U).

Course evaluation

The results of and possible changes to the course will be shared with students who participated in the evaluation and students who are starting the course.