SF1930 Statistical Learning and Data Analysis 6.0 credits

Statistisk inlärning och dataanalys

This is a translation of the Swedish, legally binding, course syllabus.

Establishment

Course syllabus for SF1930 valid from Autumn 2021.

Grading scale

A, B, C, D, E, FX, F

Education cycle

First cycle

Main field of study

Technology

Specific prerequisites

Completed basic course in probability and statistics (SF1918, SF1922 or equivalent).

Language of instruction

The language of instruction is specified in the course offering information in the course catalogue.

Intended learning outcomes

After having passed the course, the student is supposed to be able to:

- formulate and apply concepts in statistical inference and prediction to solve theoretical problems;
- formulate and apply concepts in statistical inference and prediction to solve problems in data analysis;
- design and implement methods in statistical learning for data analysis.

Course contents

The course gives an introduction to the theory of statistical inference and prediction, which constitute the main goals for modern statistical data analysis and machine learning. Particular attention is given to multidimensional probability distributions and exponential families, which are fundamental tools for modeling data analytical problems, and the theory of graphical models is a powerful means for describing conditional dependencies with a bearing on high-dimensional statistical inference problems. Decision theory provides a framework for making optimal decisions under statistical uncertainty, as well as weighting different statistical approaches against each other. In particular, Bayesian decision theory—in which the inference and learning problems are solved through calculation of the posterior and predictive distributions, respectively—plays a central role in today's statistical data analysis and is used to construct Bayesian point estimates, hypothesis tests, and credibility intervals. In parallel with the Bayesian approach, likelihood theory is also discussed, and special attention is given to the asymptotic properties of the maximum likelihood estimate as the amount of data grows towards infinity. The course also introduces basic statistical computation methods, such as stochastic gradient methods and Markov chain Monte Carlo (MCMC) methods, which are of great importance in modern computer-intensive statistics. In the course, these are applied to real data-analytical problems within the framework of a computer-based project.

Examination

- TEN1 - Written exam, 4.0 credits, grading scale: A, B, C, D, E, FX, F
- INL1 - Hand in assignment, 2.0 credits, grading scale: P, F

Based on recommendation from KTH's coordinator for disabilities, the examiner will decide how to adapt an examination for students with documented disability.

The examiner may apply another examination format when re-examining individual students.

If the course is discontinued, students may request to be examined during the following two academic years.

Ethical approach

- All members of a group are responsible for the group's work.

- In any assessment, every student shall honestly disclose any help received and sources used.
- In an oral assessment, every student shall be able to present and answer questions about the entire assignment and solution.