SF2930 Regression Analysis 7.5 credits

Regressionsanalys

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

Establishment

Course syllabus for SF2930 valid from Autumn 2020

Grading scale

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

Education cycle

Second cycle

Main field of study

Mathematics

Specific prerequisites

- Completed basic course in numerical analysis (SF1544, SF1545 or equivalent)
- Completed basic course in probability theory and statistics (SF1922, SF1914 or equivalent)

Language of instruction

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

Intended learning outcomes

To pass the course, the student shallbe able to:

- Formulate and apply statistical regression theory
- Formulate and apply advanced methods in statistical regression modeling
- Design and implement advanced methods in regression analysis for applications

Course contents

The course begins with simple and multiple linear regression models for which fitting, parametric and model inference as well as prediction will be explained. Topics covered are least squares (LS) and generalised LS, the Gauss-Markov theorem, geometry of least squares and orthogonal projections. A special attention is paid to the diagnostic strategies which are key components of good model fitting. Further topics include transformations and weightings to correct model inadequacies, the multicollinearity issue, variable subset selection and model building techniques. Later in the course, some general strategies for regression modelling will be presented with a particular focus on the generalized linear models (GLM) using the examples with binary and count response variables.

As the high-dimensional data, order of magnitude larger than those that the classic regression theory is designed for, are nowadays a rule rather than an exception in computer-age practice (examples include information technology, finance, genetics and astrophysics, to name just a few), regression methodologies which allow to cope with the high dimensionality are presented. The emphasis is placed on methods of controlling the regression fit by regularization (Ridge, Lasso and Elastic-Net), as well as methods using derived input directions (Principal Components regression and Partial Least Squares) that allow to tamp down statistical variability in high-dimensional estimation and prediction problems.

A number of statistical learning procedures with the focus on computer-based algorithms is presented from a regression perspective.

Computer-aided project work with a variety of datasets forms an essential learning activity.

Examination

- TENA - Examination, 4.5 credits, grading scale: A, B, C, D, E, FX, F
- OVN1 - Assignments, 3.0 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.