SF2930 Regression Analysis 7.5 credits

Regressionsanalys

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

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

Establishment

Course syllabus for SF2930 valid from Spring 2018

Grading scale

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

Education cycle

Second cycle

Main field of study

Mathematics

Specific prerequisites

Passed courses in analysis in one and several variables, linear algebra, numerical analysis, differential equations, mathematical statistics

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 should be able to do the following:

- know the sampling properties of point estimators used in linear regression models as well as principles and assumptions behind different estimation techniques applied
- list and understand the assumptions behind standard parametric and model inference in the linear regression models
- assess the fit of a regression model to data and know how to identify and diagnose potential problems with a linear regression model
- design and implement the strategy to correct model inadequacies, and report on the expected accuracy which can be achieved with the suggested model;
- identify and develop regression modelling strategies suitable for large sample as well as for high-dimensional settings
- explain how the multiple linear regression can be generalized to handle a response variable that is categorical or a count variable
- use resampling algorithms, in particular, the bootstrap and cross-validation, for estimation of the model predictive accuracy. Understand the needs for and benefits of resampling methods in regression modelling and assessment
- critically evaluate regression models in a real-world applications, and present the analysis and conclusions in a written report
- read current research papers and understand the issues raised by current research

To receive the highest grade, the student should in addition be able to do the following:

- combine several methods and models in order to gain better results

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.

Disposition

Lectures, presentations, work with computer-aided data analysis.

Course literature

See the course web page

Examination

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

Other requirements for final grade

Passed assignments and final exam.

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.