STAN51, Statistics: Machine Learning from a Regression Perspective, 7.5 credits

Statistik: Maskininlärning ur ett regressionsperspektiv, 7,5 högskolepoäng

Second Cycle / Avancerad nivå

Details of approval

The syllabus was approved by The Board of the Department of Statistics on 2020-12-10 and was last revised on 2023-02-27. The revised syllabus applies from 2023-08-28, autumn semester 2023.

General Information

Second cycle level course in Statistics. The course may be included in a Master's degress in Statistics. The course can also be taken as a single subject course or within other Master's programmes at Lund University.

Language of instruction: English
(Teaching may be in Swedish if all registered students have a good knowledge of Swedish.)

Main field of studies	Depth of study relative to the degree requirements
Statistics	A1N, Second cycle, has only first-cycle course/s as entry requirements

Learning outcomes

Knowledge and understanding
Students shall have an understanding of:

- basic regression and classification methods,
- model training and validation,
- model selection,
- regularization and dimension reduction techniques,
- the basics of non-linear machine learning methods, such as neural networks,
- how to interpret results from the methods covered.

Competence and skills

Students shall have the ability to independently:

- apply the methods covered using a statistical software package,
- appropriately choose between a set of methods for a given empirical problem,
- evaluate the suitability of the chosen method,
- generalise their knowledge to empirical problems that have not been treated during the course,
- draw conclusions from empirical results and provide advice.

Judgement and approach

Students shall have developed the ability to pursue further studies in the subject and should be able to search for and evaluate subject related information with a high degree of independence. Students shall also have developed the ability to write an empirical report where machine learning is an essential element.

Course content

Machine learning refers to statistical model predictions that that improve through experience; as new data arrive, the model learns and adapts. The price that the supermarket can charge for advertisements depends critically on its ability to learn from the data which customers that are likely prospects for a particular supplier's product. Similarly, the price that Google can charge for space for sponsored links is directly tied to their ability to correctly identify people likely to follow the link. That is where machine learning comes in. This course teaches the basics of machine learning and it does so by focusing on those methods that build in one way or another on standard regression analysis. Some of the topics covered are classification based on logistic regression, performance evaluation and model selection, shrinkage methods such as lasso, ridge regression and elastic net, dimension reduction methods such as principal components regression and partial least squares, and neural networks. Theoretical studies are interwoven with empirical applications to problems in business and economics, which are carried out using modern statistical software.

Course design

Teaching: Teaching consists of lectures and exercise classes.

Assessment

1. Examination: The examination consists of a written exam, home assignments and active participation in exercise sessions. The written exam takes place at the end of the course. There will be further opportunities for examination close to this date. Points gained on home assignments and active participation in exercise sessions count at exams during the current term. Other forms of examination may be used to a limited extent.
2. Limitations on the number of examination opportunities: –

The University views plagiarism and other academic dishonesty very seriously, and will take disciplinary action against students for any kind of attempted malpractice in connection with examinations and assessments. Plagiarism is considered to be a very serious academic offence. The penalty that may be imposed for this, and other unfair

practices in examinations or assessments, includes suspension from the University for a specified period.

The examiner, in consultation with Disability Support Services, may deviate from the regular form of examination in order to provide a permanently disabled student with a form of examination equivalent to that of a student without a disability.

Subcourses that are part of this course can be found in an appendix at the end of this document.

Grades

Marking scale: Fail, E, D, C, B, A.

A (Excellent) 85-100 points/percent. A distinguished result that is excellent with regard to theoretical depth, practical relevance, analytical ability and independent thought.

B (Very good) 75-84 points/percent. A very good result with regard to theoretical depth, practical relevance, analytical ability and independent thought.

C (Good) 65-74 points/percent. The result is of a good standard with regard to theoretical depth, practical relevance, analytical ability and independent thought and lives up to expectations.

D (Satisfactory) 55-64 points/percent. The result is of a satisfactory standard with regard to theoretical depth, practical relevance, analytical ability and independent thought.

E (Sufficient) 50-54 points/percent. The result satisfies the minimum requirements with regard to theoretical depth, practical relevance, analytical ability and independent thought, but not more.

F (Fail) 0-49 points/percent. The result does not meet the minimum requirements with regard to theoretical depth, practical relevance, analytical ability and independent thought.

To pass the course, the students must have been awarded the grade of E or higher.

Entry requirements

90 credits in Statistics with at least 7.5 credits in Regression Analysis or Econometrics, or the equivalent.

Further information

This course replaces together with STAN52 the course STAN45. The course can not be combined with STAN45 in a degree.

Subcourses in STAN51, Statistics: Machine Learning from a Regression Perspective

Applies from H21

2101   Machine Learning from a Regression Perspective, 7,5 hp
        Grading scale: Fail, E, D, C, B, A