



**LUND**  
UNIVERSITY

School of Economics and Management

## STAN52, Statistics: Advanced Machine Learning, 7.5 credits

*Statistik: Avancerad maskininlärning, 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 to be valid from 2020-12-10, autumn semester 2020.

### General Information

Second cycle level course in Statistics. The course may be included in a Master's degree 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*

Statistics

*Depth of study relative to the degree requirements*

A1F, Second cycle, has second-cycle course/s as entry requirements

### Learning outcomes

#### Knowledge and understanding

Students shall have an understanding of:

- bootstrap methods,
- classification and regression trees,
- ensemble methods,
- unsupervised learning methods,
- applications of machine learning to problems that are relevant in business and economics.

## Competence and skills

Students shall have the ability to independently:

- apply the methods covered using the R 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

This course covers advanced machine learning methods that are relevant for applications in business and economics, and is intended as a continuation of Machine Learning from a Regression Perspective. Some of the topics covered include bootstrapping, ensemble methods such as boosting and random forests, unsupervised machine learning methods such as principal components analysis and clustering algorithms as well as applications of machine learning methods to problems that are relevant for business and economics, such as causal inference and text analysis. Theoretical studies are interwoven with empirical applications to problems in business and economics, which are carried out using the R software.

## Course design

Teaching: Teaching consists of lectures and exercises.

## Assessment

1. Examination: The examination consists of a written exam and home assignments. 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 the home assignments 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 and also STAN51 Statistics: Machine Learning from a Regression Perspective, or the equivalent.

## **Further information**

This course replaces together with STAN51 the course STAN45. The course can not be combined with STAN45 in a degree.

## Subcourses in STAN52, Statistics: Advanced Machine Learning

Applies from H21

2101 Advanced Machine Learning, 7,5 hp  
Grading scale: Fail, E, D, C, B, A