INSTITUTE OF ENVIRONMENTAL MEDICINE

C6F5650, Introduction to Machine Learning, 1.5 credits (hec)

Introduktion till maskininlärning, 1,5 högskolepoäng

Third-cycle level / Forskarnivå

Approval

This syllabus was approved by the The Committee for Doctoral Education on 2023-12-21, and was last revised on 2025-03-05. The revised course syllabus is valid from autumn semester 2025.

Responsible department

Institute of Environmental Medicine, Faculty of Medicine

Prerequisite courses, or equivalent

Biostatistics I and Biostatistics II or corresponding courses.

Purpose & Intended learning outcomes

Purpose

The purpose of this course is to give an introduction to machine learning without heavy-mathematics. A main focus is on machine learning algorithms for regression analyses using large datasets, both in terms of the number of variables observed and/or the number of units (sample size).

Intended learning outcomes

After successfully completing this course, the student is expected to be able to:

- Recognize and formulate well defined questions that can be solved using machine learning algorithms: both prediction and inference questions.
- Explain key concepts in machine learning, including curse of dimensionality, out of sample validation, generalization, uncertainty.
- Choose relevant machine learning algorithms for prediction and inference.
- Conduct and interpret simple analyses using machine learning algorithms on real data.

Course content

This course focuses on machine learning algorithms for regression analyses using large datasets, both in terms of the number of variables observed and/or the number of units (sample size). Register data studies are a typical example where such large datasets are analysed. The course will start by going through key concepts of statistical learning, including regression and prediction problems, curse of dimensionality, out of sample validation, generalization, uncertainty. These are problems and concepts that students need to be able to recognize, formulate and explain. The course will then present some central machine learning algorithms, including Lasso, trees, random forest, bagging, together with methods to validate the algorithms and to draw inference. The students will in particular learn how to choose relevant algorithms for specific situations.

Forms of teaching and learning

Lectures, quizzes, group sessions and computer labs.

Language of instruction

The course is given in English

Grading scale

Pass (G) /Fail (U)

Compulsory components & forms of assessment

Compulsory components

Individual examination (summative assessment).

Forms of assessment

To pass the course, the student must show that the learning outcomes have been achieved. Assessments methods used are group tasks (formative assessments) along with a written individual task (summative assessment). The examination is viewed as a contribution to the development of knowledge, rather than as a test of knowledge. Students who do not obtain a passing grade in the first examination will be offered a second examination within two months of the final day of the course. Students who do not obtain a passing grade at the first two examinations will be given top priority for admission the next time the course is offered.

Course literature

Suggested reading:

An introduction to statistical learning by James, Witten, Hastie and Tibshirani.
https://www.statlearning.com