



## DEPARTMENT OF MEDICAL EPIDEMIOLOGY AND BIostatISTICS

### **C8F5694, Multivariate Prediction Models, Machine Learning and AI with Applications in Precision Medicine, 1.5 credits (hec)**

Multivariata prediktionsmodeller, maskininlärning och AI med tillämpningar inom precisionsmedicin, 1,5 högskolepoäng

*Third-cycle level / Forskarnivå*

---

#### **Approval**

This syllabus was approved by the The Committee for Doctoral Education on 2023-12-27, and was last revised on 2025-03-11. The revised course syllabus is valid from autumn semester 2025.

#### ***Responsible department***

Department of Medical Epidemiology and Biostatistics, Faculty of Medicine

#### **Prerequisite courses, or equivalent**

Epidemiology I, Introduction to epidemiology; Epidemiology II, Design of epidemiological studies; Biostatistics I, Introduction for epidemiologists; Biostatistics II, Logistic regression for epidemiologists; and Biostatistics III: Survival analysis for epidemiologists, or equivalent courses

#### **Purpose & Intended learning outcomes**

##### **Purpose**

This course aims to provide an introduction to both supervised and unsupervised methodologies for prediction modelling with a focus on biomedical applications, molecular epidemiology and personalised medicine. The main objective of the course is to provide basic theory and to facilitate for the course participants to acquire practical knowledge that will enable to apply covered methodologies in their own research.

##### **Intended learning outcomes**

After successfully completing this course you as a student are expected to be able to:

- Perform and assess basic quality control and outlier detection

- Apply unsupervised and supervised statistical learning methods to detect patterns in data
- Devise cross-validation strategies for parameter estimation, model selection and prediction performance evaluation
- Make informed judgement of how to apply basic principles for variable selection
- Critically evaluate prediction models and artificial intelligence in real-world applications
- Conceptually design and devise applications of machine learning and deep learning in real-world applications

## Course content

Personalised medicine is a cornerstone of tomorrow's health care, and is based on the idea of stratifying patients into groups based on e.g. disease risk, prognosis or probability of treatment response and administer the most suitable therapy for each individual. The capabilities to generate vast amounts of quantitative clinical, imaging, and molecular data from DNA- and RNA-sequencing and other molecular profiling methods are providing unprecedented opportunity for implementation of personalized precision medicine approaches in the health care system. Molecular profiling typically generates data with tens of thousands of variables of which only a subset is relevant for treatment decisions. Similarly, imaging data from e.g. radiology and digital pathology provides information rich data to inform patient management. The promise of personalised medicine relies on our ability to turn the vast molecular datasets into clinically actionable predictive models of individualised diagnostics, prognostication, and therapy response. Development and application of statistical learning methods, prediction modelling, artificial intelligence, and deep learning are central components in developing these models, and in developing the biomarker panels that can be used for subtyping, risk stratification and prediction of treatment response. This course provides an introduction to statistical learning methods, prediction models, and deep learning that are relevant for personalised medicine with a focus on real-world applications.

The course covers basic theory and introduction to modern statistical and machine learning methods for prediction modelling and deep learning in high-dimensional data, together with applied data analysis through computer-based exercises. Lectures and exercises will cover the full process going from the initial data set and through data normalisation, quality control, outlier detection, application of unsupervised learning methods, application of supervised learning methods, variable selection, cross-validation, model evaluation, and recently developed methods (e.g. deep learning, conformal prediction).

## Forms of teaching and learning

The course is based on a combination of lectures, which cover methods and theory, together with computer-based exercises in R, where real-world data are analysed and interpreted. Previous experience from practical experience applying statistical models in a computer-based environment (e.g R, SAS, Stata, Matlab, Python) is strongly recommended.

### *Language of instruction*

The course is given in English

## Grading scale

Pass (G) /Fail (U)

## Compulsory components & forms of assessment

### Compulsory components

The individually written examination.

### Forms of assessment

The individual examination will be performed as a take-home examination. It consists of an individually written lab-report where results from an applied data analysis mini-project should be summarised and critically evaluated. 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.

## Course literature

Suggested course literature:

Elements of Statistical Learning, Hastie, Tibshirani and Friedman (2009). Springer-Verlag,  
Freely available at <https://statweb.stanford.edu/~tibs/ElemStatLearn/>