

Course Syllabus

Statistical and Machine Learning 10 Credits, Third Cycle

Learning Outcomes

Upon completion of the course, the PhD-student shall be able to:

- Select a suitable statistical models, and methods for a data analysis problem in the real world based on reasoned argument, especially when the underlying data generating mechanism is unknown.
- Apply various supervised and unsupervised statistical learning algorithms in a range of real world problems.
- Evaluate and optimise the performances of the learning models and algorithms, and communicate the expected accuracy of the model/algorithm.
- Combine several models to achieve higher predictive accuracy.
- Apply Neural Networks to real world problem solving.
- Conduct comparative analysis, both theoretical and empirical, in order to decide which Neural Network is most suitable for a particular task.
- Design different kinds of Neural Network, evaluate their performance, and use them to solve complex problems.
- Apply deep learning to real world problems.

Course Content

The course consists of two parts, statistical learning (Part 1, corresponding to Learning Outcomes 1-4) and Machine Learning (Part 2, Learning outcomes 5-8). Both parts of the course focuses mainly on the applied aspects of statistical learning and machine learning with Part 2 emphasizing on neural networks and deep learning.

In Part 1, the most important basic properties of, and relations between different statistical learning models and algorithms are also included, however. This part covers supervised learning algorithms, with special emphasis on classification methods such as logistic regression, classification trees, linear discriminant analysis, quadratic discriminant analysis, K nearest neighbour, support vector machine, and regression methods such as linear regression, smoothing splines, generalised additive model, and regression trees. Part 1 also covers unsupervised learning methods such as principal component analysis, k mean clustering, and hierarchical clustering. Model validation through cross validation, and bootstrap methods are covered. Regularisation for model selection, high dimensional data analysis, and improving prediction performance through model averaging, bagging, and boosting techniques are also covered.

Part 2 of the course gives an introduction to machine learning and an overview of neural networks. The perceptron as the basic element for linear separability and its limitations in classification is discussed. Then, different activation functions and the sigmoid perceptron is studied to solve non-linear classification problems.

The different types of learning paradigms such as supervised, unsupervised, and reinforcement is covered in a machine learning context. Feed-forward neural networks and the back-propagation algorithm will be presented. The course will also cover recurrent neural networks. Finally, deep learning is discussed with emphasis on the basic principles and different types of deep learning neural networks.

Assessment

Part 1 and Part 2 are assessed independently.

Part 1 on statistical learning is assessed by assignment and written examination giving a total of 5 credits.

Part 2 on machine learning is assessed by a reported project work, written reflection and seminar giving another 5 credits.

Forms of Study

The course comprises lectures, labs, project work and seminars.

Grades

The Swedish grades U–G.

Prerequisites

Degree of Master 60 credits in Microdata Analysis

Subject:

Microdata Analysis

Approved:

Approved 29 October 2019

Valid from 29 October 2019