



Course information 2020-21

ST3189 Machine learning

General information

COURSE LEVEL: 6

CREDIT: 30

NOTIONAL STUDY TIME: 300 hours

Summary

In the last decade there has been a remarkable growth in machine learning. Following recent advances in gathering, storing and managing vast amounts of observations, the ability to process high dimensional data and deal with uncertainty becomes increasingly important. Despite the increase of available information, inference may still lead to false conclusions in the absence of a suitable methodology. This course covers a wider range of such model based and algorithmic machine learning methods, illustrated in various real-world applications and datasets. At the same time, the theoretical foundation of the methodology is presented in some cases.

Conditions

Prerequisites: If taken as part of a BSc degree, the following courses must be passed before this course may be attempted:

- ST104a Statistics 1
- **AND** ST104b Statistics 2
- **AND (EITHER** MT1174 Calculus **OR** MT1186 Mathematical methods **OR (BOTH** MT105a Mathematics 1 **AND** MT105b Mathematics 2)

Aims and objectives

- To provide an in-depth introduction to supervised and unsupervised learning
- To present some of the main models and algorithms for regression, classification and clustering
- Other topics include Bayesian inference, Monte Carlo methods and dimension reduction

Learning outcomes

At the end of the course and having completed the essential reading and activities students should be able to:

- develop an understanding of the process to learn from data
- be familiar with a wide variety of algorithmic and model based methods to extract information from data
- apply and evaluate suitable methods to various datasets by model selection and predictive performance evaluation

Please consult the current EMFSS Programme Regulations for further information on the availability of a course, where it can be placed on your programme's structure, and other important details.

Essential reading

James G., Witten D., Hastie T. and Tibshirani R. *An introduction to Statistical Learning: with Applications in R*, Springer (2013) [ISBN 9781461471387]

Assessment

This course is assessed by an individual case study piece of coursework (30%) and a two hour unseen written examination (70%).

Syllabus

A list of topics to be covered is given below:

1. Linear regression and regularisation (via least squares and maximum likelihood)
2. Bayesian Inference
3. Classification
4. Resampling methods
5. Clustering
6. Non-linear models
7. Tree-based methods
8. Support Vector Machines
9. Random forests
10. Gaussian Processes

The coursework will involve several computer exercises in R (no prior knowledge is required).

Please consult the current EMFSS Programme Regulations for further information on the availability of a course, where it can be placed on your programme's structure, and other important details.