Course information 2018–19
ST3189 Machine learning

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 is some cases.

Prerequisite
If taken as part of a BSc degree, courses which must be passed before this course may be attempted:

ST104a Statistics 1 and ST104b Statistics 2 and either MT105a Mathematics 1 with MT105b Mathematics 2 or MT1174 Calculus.

Exclusion

None

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

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

Further reading
Rogers S. and Girolami M. *A First Course in Machine Learning*, Chapman & Hall/CRC Press, second edition (2011) [ISBN 9781498738484].

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

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

Syllabus
This is a description of the material to be examined, as published in the *Regulations*. On registration, students will receive a detailed subject guide which provides a framework for covering the topics in the syllabus and directions to the essential reading

A list of topics to be covered is given below

1. Week 1: Essential concepts of Statistical inference, Matrix Algebra and Calculus, and computing in R.

2. Weeks 2 to 4: Linear regression via least squares and maximum likelihood.

3. Weeks 5 and 6: Sampling and Resampling Methods

4. Weeks 7 to 9: Bayesian Inference

5. Weeks 10 to 12: Linear Regression in High Dimensions and Regularisation

6. Weeks 13 and 14: Classification

7. Weeks 15 and 16: Unsupervised Learning: Principal Component Analysis and Clustering

8. Weeks 17 and 18: Non-linear models and Gaussian Processes

9. Week 19: Tree-Based Methods and Random Forests

10. Week 20: Support Vector Machines

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

Course convenor: Dr Kostas Kalogeropoulos.

Course will be added to selection group N.