Statistical Learning in Practice (L26)

Dr. W. G. Underwood

Statistical learning aims to build statistical models for learning relationships and structures from data, mainly with the goal of predicting future outcomes. It blends classical statistics with techniques from modern machine learning with the aim of providing concrete algorithms, for example for regression or classification tasks. This course consists of 12 lectures and 12 practical classes. We will get to know some of the most successful and widely used statistical methodologies in modern applications, with a focus on practical aspects and concrete problems. The practical classes will cover an introduction to the programming language R, exploratory data analysis and the implementation of the statistical methods discussed in the lectures. We aim to cover a selection of the following topics:

- Generalised linear models for regression and classification
- Model selection and regularisation
- Mixed effects models
- Nearest neighbours classifiers
- Support vector machines
- Random forests
- Introduction to deep learning

Prerequisites

Elementary probability theory, maximum likelihood estimation, hypothesis tests and confidence intervals, linear models. Previous experience with R is not essential and can be obtained during the course (only a good working knowledge of R will be necessary, as provided, for instance, in the lab sessions of the first reference below).

Literature

1. G. James, D. Witten, T. Hastie and R. Tibshirani. *An Introduction to Statistical Learning (with Applications in R).* Springer, 2013. Available at https://www.statlearning.com.
2. T. Hastie, R. Tibshirani and J. Friedman. *The Elements of Statistical Learning.* Springer, Second Edition.
3. A. J. Dobson and A. Barnett. *An Introduction to Generalized Linear Models.* Third edition. Chapman & Hall/CRC, 2008.

Additional support

This course includes practical classes, in which statistical methods are introduced in a practical context and students carry out analysis of data sets using R. In these classes, the students have the opportunity to discuss statistical questions with the lecturers. Four example sheets will be provided and there will be four associated example classes (the last one probably to be held in Easter term). There will be a revision class in the Easter Term.