

Statistical Learning in Practice (L24)

Randolf Altmeyer

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 and tries to provide 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 and quasi-likelihood methods
- Linear discriminant analysis and Support Vector Machines
- Introduction to Deep learning and Random forests
- Introduction to time series

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. Also available at <https://www.statlearning.com>.
2. T. Hastie, R. Tibshirani and J. Friedman. *The Elements of Statistical Learning*. Springer, Second Edition.
3. C. Robert and G. Casella. *Introducing Monte Carlo Methods with R*. Springer, 2009.
4. P.J. Brockwell and R.A. Davis. *Introduction to time series and forecasting*. Springer, 1996.
5. A.J. Dobson and A. Barnett. *An Introduction to Generalized Linear Models*. Third edition. Chapman & Hall/CRC, 2008.
6. R.H. Shumway D.S. Stoffer. *Time Series Analysis and Its Applications: with R Examples*. Springer Science & Business Media, 2010.
7. L. Wasserman. *All of Statistics*. Springer, 2003.

Additional support

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