

Complex Data Analysis
Summer term 2024
PD Dr. Ralf Stecking

Lecture:

Wednesday 10.15-11.45 o'clock in room A05 0-056

Exercise:

Tuesday 12.15-13.45 o'clock in room A04 1-139

Starts: 3rd April 2024, ends: 3rd July 2024

Content of the module

Students of this course shall be able to analyse complex empirical data sets, like aggregated data, privacy constrained data, distance information, distributions, tables, symbolic or granular data. Students will also learn to handle issues of big data challenges: large number of cases or variables, unknown dependencies, redundancy, missing values, small or no variance.

The module consists of two parts: the lecture imparts knowledge about the basic theoretical principles of complex data analysis. The exercise includes case studies using real data sets with appropriate software packages.

Structure

The following parts are the content of the course *Complex Data Analysis*:

- Cluster analysis,
- Linear discriminant analysis,
- Principal component analysis,
- Correspondence analysis,
- Classification and regression trees,
- Decision trees,
- Symbolic data analysis.

Literature

Hastie, T., Tibshirani, R. and Friedman, J. (2001): *The Elements of Statistical Learning*, New York

Tuffery, S. (2011): *Data Mining and Statistics for Decision Making*, West Sussex

Hennig, C., Meila, M., Murtagh, F. and Rocci, R. (2016): *Handbook of Cluster Analysis*, Boca Raton

Jolliffe, I.T. and Cadima, J. (2016): Principal component analysis: a review and recent developments. *Phil. Trans. R. Soc. A* 374:20150202

Greenacre, M.J. (1984): *Theory and Applications of Correspondence Analysis*, London

Breiman, L., Friedman, J.H., Olshen, R.A. and Stone, C.J. (1984): *Classification and Regression Trees*, Pacific Grove, California

Billard, L. and Diday, E. (2006): *Symbolic Data Analysis*, West Sussex

Pedrycz, W. (2017): *Granular Computing*, Boca Raton