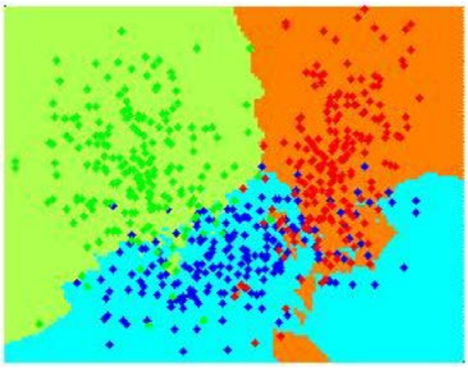
Computational Statistics and Data Analysis (MVComp2)

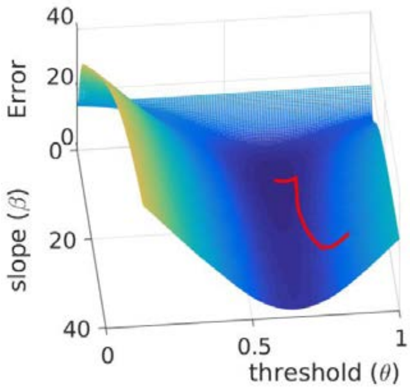
Wednesdays from 22/4 – 29/7 2020
Format: Virtual (Explain-EDU + GoToMeeting/ Zoom + Slack)
Lecture (2 hrs): **Wed 11.00ct-13.00**
Exercises (2 hrs): **Wed 14.00ct-16.00**

Credit Points: 6

Lecturers:
Main lecture: Prof. Daniel Durstewitz, daniel.durstewitz@zi-mannheim.de
Exercises: Dr. Georgia Koppe, georgia.koppe@zi-mannheim.de
Leonard Bereska, leonard.bereska@zi-mannheim.de

Summary
This lecture introduces students to basic methods and techniques in computational statistics and data analysis, as widely applicable to empirical problems in the natural sciences. It provides an overview from relevant concepts and theorems in probability theory and mathematical statistics to modern deep learning techniques. The lecture is accompanied by computational exercises in Python or Matlab. It will enable students to analyze small and large data sets and interpret the results from a solid, theoretically grounded statistical perspective, to devise statistical & machine learning models of experimental situations, to infer the parameters of these models from empirical observations, and to test hypotheses about them.





(source: from ref. 2 below)

Prerequisites
- Linear (Matrix) Algebra
- Basic calculus (derivatives & integrals)
- Basic programming skills in Python or Matlab

Literature
1) Wackerly, D.D., Mendenhall III, W., Scheaffer, R.L. (2008) Mathematical Statistics with Applications. Brooks/Cole.
2) Durstewitz, D. (2017) Advanced Data Analysis in Neuroscience: Integrating Statistical and Computational Models. Springer.
3) Hastie, T., Tibshirani, R., and Friedman, J. (2009). The Elements of Statistical Learning: Data Mining, Inference, and Prediction (New York: Springer).
4) Bishop, C.M. (2007). Pattern Recognition and Machine Learning. Springer.
5) Goodfellow, I., Bengio, Y., Courville, A. (2016). Deep Learning. MIT Press.

(Approximate) Table of Contents

All concepts will be introduced along particular, motivating examples and data sets from the natural sciences.

1) Basic concepts & axioms in probability theory

2) Discrete & continuous univariate probability distributions

3) Moment-generating functions and multivariate distributions

4) Statistical models & inference, properties of statistical estimators

5) Statistical models & inference cont., principles of parameter estimation (Maximum Likelihood, Bayesian inference) and numerical solution methods

6) Hypothesis tests; test construction, central limit theorem, asymptotic tests, log-likelihood-ratio principle

7) Hypothesis tests cont.; exact tests, bootstrap & permutation methods

8) Linear regression & General Linear Model

9) Nonlinear regression models; Locally Linear Regression, basis expansions & splines, (deep) neural networks for regression

10) Classification models: Linear & Quadratic Discriminant Analysis, k-nearest neighbors

11) Classification models cont.: Support Vector Machines, logistic regression, (deep) neural networks for classification

12) Model selection & bias-variance tradeoff

13) Regularization techniques: Ridge & LASSO regression;

14) Dimensionality reduction: PCA, Auto-Encoders

15) Generative models: Factor Analysis, Variational Auto-Encoders & re-parameterization trick for stochastic gradient descent