QBio404: Data Science and Machine Learning

Module Responsible: Prof. Dr. Achim Tresch	**Version:** 02/01/2021
Module Organizer: Prof. Dr. Achim Tresch,	**Type:** Compulsory

Lecturer:
Prof. Dr. Achim Tresch, Prof. Dr. Andreas Beyer, Dr. Hajira Jabeen

Total Working Time	**Credit Points**	**Contact Time**	**Self Study**	**Duration**
180 h	6 CP	75 h	105 h	1 Semester

Course Components		**Group Size**	**Frequency**
Lecture:	3 SWS	P: 40	Every Winter Semester
Exercise:	2 SWS	P: 20	

Learning Competencies:
Students can analyse high-dimensional, biological data by methods of machine and statistical learning using the statistical software R. They acquire a repertoire of computer-based methods for dimension reduction, for supervised and unsupervised learning tasks and estimation problems as they occur in omics data. Students are aware of the peculiarities of high-dimensional statistics and large data sets (e.g. curse of dimensionality and multiple testing) and can critically assess third-party analyses.

Content:
Principles & basic techniques for Machine Learning:

- Risk minimization
- Optimization (convexity, gradient descent, backpropagation, Nelder-Mead)
- Sampling (rejection sampling, Markov Chain Monte Carlo)
- Cross validation, Bootstrap
- Boosting
- Statistical testing

Supervised learning:

- Classification (logistic regression, Support Vector Machines, decision trees)
- Regression ((generalized) linear models, regularization, interactions)
- Random forests
- Neural networks

Unsupervised learning:

- Clustering (k-means, Gaussian mixtures, hierarchical clustering)
- Hidden Markov models
- Autoencoders

Dimension reduction:

- Principal Components Analysis
- Stochastic neighborhood embedding (SNE/t-SNE)

Practical skills: Analysis of high-dimensional data, statistical packages in R / Bioconductor
Conditions of Participation: Passed Module QBio203 and QBio103
Examination: Learning portfolio consisting of • Written Exam about the content of the lectures (50% of the final grade) • Exercises (20% of the final grade)
Prerequisites for Awarding Credits for this Module: • Passing Exercises (50 % of Exercise Sheets) • Passing Written Exam
Factor for the Overall Grade: The grade is weighted according to the credit points (CP) in the overall grade.
Language: English
Literature:
Further Information: