

BIOENG-210

Biological data science I: statistical learning

La Manno Gioele

Cursus	Sem.	Type
Life Sciences Engineering	BA4	Obl.

Language of teaching	English
Credits	4
Session	Summer
Semester	Spring
Exam	Written
Workload	120h
Weeks	14
Hours	4 weekly
Lecture	2 weekly
Exercises	2 weekly
Number of positions	

Summary

Processing, analyzing, and interpreting large biological datasets is an essential skill for modern biologists. This course aims to provide the theoretical foundations, analytical techniques, and software tools necessary to effectively manage and derive insights from complex biological data.

Content

Biological data types
 Probability Distributions in Biology data
 Maximum Likelihood estimators for Univariate and Bivariate Distributions
 Statistical tests
 Multivariate data analysis
 Multivariate Linear Regression
 Principal Component Analysis (PCA)
 Clustering
 Priors, Bayes, and Maximum a Posteriori Estimation
 Advanced Linear Regression
 Logistic Regression and Classification
 Model Selection
 Resampling and Simulations
 Time series and 1D Signal Processing
 ND-image Processing
 Generative Models and MCMC

Keywords

Biological data, statistical learning, probability distributions, maximum likelihood estimation, multivariate analysis, multivariate normal, PCA, SVD, multivariate regression, classification, Bayesian inference, time series, image processing, resampling methods, MCMC.

Learning Prerequisites**Required courses**

Analysis, Linear Algebra, Probability and Statistics

Learning Outcomes

By the end of the course, the student must be able to:

- Analyze multidimensional biological data
- Apply regression and classification models
- Perform model selection
- Use PCA and interpret it
- Visualize multivariate data
- Explore different types of biological data
- Implement basic routines of ML and MAP estimation
- Choose the most appropriate model to specific situation
- Plan an analysis end-to-end
- Interpret statistical tests and posterior distributions

Teaching methods

Lectures and exercises

Assessment methods

Written examination at the exam session (70%) and graded exercises (30%).

Supervision

Office hours	No
Assistants	Yes
Forum	Yes

Resources

Virtual desktop infrastructure (VDI)

No

Bibliography

Main:

"Elements of Statistical Learning" by Trevor Hastie, Robert Tibshirani, and Jerome Friedman

"Methods of Multivariate Analysis" By Alvin C. Rencher

Other resources:

"Computer Age Statistical Inference" by Bradley Efron and Trevor Hastie

"Data-Driven Science and Engineering" by Steven L. Brunton and J. Nathan Kutz

Ressources en bibliothèque

- ["Elements of Statistical Learning" by Trevor Hastie, Robert Tibshirani, and Jerome Friedman](#)
- ["Methods of Multivariate Analysis" By Alvin C. Rencher](#)
- ["Computer Age Statistical Inference" by Bradley Efron and Trevor Hastie](#)
- ["Data-Driven Science and Engineering" by Steven L. Brunton and J. Nathan Kutz](#)

Notes/Handbook

Course notes in pdf format

Moodle Link

- <https://go.epfl.ch/BIOENG-210>