Faculty of Science

BERN02, Computational Science: Reproducible Data Science and Statistical Learning, 7.5 credits

Beräkningsvetenskap: Reproducerbar dataanalys och statistisk inlärning, 7,5 högskolepoäng

Second Cycle / Avancerad nivå

Details of approval

The syllabus was approved by Study programmes board, Faculty of Science on 2022-12-14 to be valid from 2022-12-14, autumn semester 2023.

General Information

The course is a compulsory course in the second cycle for a degree of Master of Science in Applied Computational Science and an alternative-compulsory course in the second cycle for a degree of Master of Science in Computational Science. The course is also an elective course in the second cycle for a degree of Master of Science in Mathematical Statistics.

Language of instruction: English

Main field of studies	*Depth of study relative to the degree requirements*
Applied Computational Science	A1N, Second cycle, has only first-cycle course/s as entry requirements
Mathematical Statistics	A1N, Second cycle, has only first-cycle course/s as entry requirements
Computational Science	A1N, Second cycle, has only first-cycle course/s as entry requirements

Learning outcomes

The overall learning outcome for the course is to let the students work with and combine two fields for data analysis in computational science: reproducible work flows and statistical learning. This includes to be able to create reports where programming code, results and text are combined in the same document, applied on a selection of common methods in statistical parametric modelling and machine learning.

Knowledge and understanding
On completion of the course, the students shall be able to:

- describe principles of reproducible and interoperable work flows for data analysis and reporting of results
- account for a selection of methods for statistical modelling and algorithms for machine learning
- identify appropriate models for classification, estimation of parameters, and prediction for a given research question.

Competence and skills
On completion of the course, the students shall be able to:

- create, divide and use tools for "literate programming" e.g. Jupiter Notebooks, analytical work flows, and version management for data analysis and reporting
- handle, present and visualise data to emphasise important properties in a data set
- apply a selection of common methods for statistical parametric modelling and machine learning
- use a programming language e.g. Python or R, for statistical analysis and machine learning
- present method, results and conclusions from a data analysis in a reproducible and interoperable written report.

Judgement and approach
On completion of the course, the students shall be able to:

- discuss the benefits and limitations for models and methods chosen for a practical problem
- critically review reproducibility and interoperability in work flows to handle and analyse scientific data

Course content

The course introduces basic principles of reproducible and interoperable work flows with a clear focus on application. The students will obtain an overview in import, transformation and visualisation of data, where realistic data are prepared for analysis in electronic "notebooks". These electronic "notebooks" use tools for "literate programming", analytical work flows and version management.

Students are introduced to a selection of methods for statistical learning. These include generalised linear regression with maximum likelihood and Bayesian inference to estimate parameters, machine learning methods for regression and classification, and methods for dimension reduction and clustering. General methods for model evaluation (e.g. cross validation) and model selection are also discussed. The assessment is based on projects where the students should choose suitable methods for analysis of given data. The analyses should be perfomed with a reproducible and interoperable work flow and then summarised in a report.

Course design

Teaching consists of lectures, computer laboratory sessions and projects. Participation in computer laboratory sessions and project work are compulsory.

Assessment

Assessment takes the form of project work during the course, which includes written project reports, an oral presentation and peer review of another project report, as well as laboratory sessions and an oral assessment at the end of the course.

Students who do not pass a regular assessment will be offered another opportunity for assessement soon thereafter.

The examiner, in consultation with Disability Support Services, may deviate from the regular form of examination in order to provide a permanently disabled student with a form of examination equivalent to that of a student without a disability.

Subcourses that are part of this course can be found in an appendix at the end of this document.

Grades

Marking scale: Fail, Pass, Pass with distinction.
For a Pass grade on the whole course, the student must have Pass grades on the project reports, the oral presentation, the peer review, the oral assessment and the laboratory sessions with associated compulsory components.

The grading scale for the project reports, the oral presentation, the peer review and the laboratory sessions is Fail, Pass, whereas the oral assessment is graded according to the scale Fail, Pass, and Pass with Distinction.

The final grade is determined by the grade on the exam.

Entry requirements

To be admitted to the course, students must have passed 90 credits in natural science or technical studies, including 43.5 credits in mathematics, where of 7.5 credits in statistics and 6 credits in programming, and English 6/B.

or

a bachelor's degree in physics and English 6/B.

Further information

Knowledge in Python or R is recommended, but is not a requirement.

Subcourses in BERN02, Computational Science: Reproducible Data Science and Statistical Learning

Applies from H23

2301 Oral exam, 3,0 hp
Grading scale: Fail, Pass, Pass with distinction
2302 Project - reports, oral presentation and peer review, 4,0 hp
Grading scale: Fail, Pass
2303 Laboratory work, 0,5 hp
Grading scale: Fail, Pass