Course syllabus

Dataanalys: statistisk inlärning och visualisering med projekt
Data Analysis: Statistical Learning and Visualization with project

FMSF90, 7,5 credits, G2 (First Cycle)

Valid for: 2023/24
Faculty: Faculty of Engineering, LTH
Decided by: PLED I
Date of Decision: 2023-04-14

General Information

Main field: Technology.
Elective for: F4, Pi4, R4
Language of instruction: The course will be given in English

Aim

The course begins with an overview of basic data wrangling and visualisation. With a focus on the student's ability to identify and illustrate important features of the data.

Then important methods in statistical learning are introduced. Emphasis is given to dimension reduction, supervised and unsupervised learning. Issues arising from fitting multiple models (i.e. multiple testing) as well as the methods relationship to regression are discussed. Computer based labs and projects form an imporant part of the learning activities. The course concludes with a project where the students will select suitable methods to analyze a given data material.

Learning outcomes

Knowledge and understanding
For a passing grade the student must

- Describe different ways of aggregating, summarising and visualising data.
- Explain the principles of dimension reduction.

- Explain the principles of supervised and unsupervised learning.

Competences and skills
For a passing grade the student must

- be able to wrangle, present and visualise data to highlight important features in a complex data material.
- be able to perform dimension reduction and imputation of missing data.
- be able to use common methods for classification, supervised and unsupervised learning.
- use methods for classification and statistical learning to draw conclussion regarding a data material.
- present the analysis and conclusions of a practical problem in a written report.

Judgement and approach
For a passing grade the student must

- Reflect over the limitations of the chosen model and method, as well as alternative solutions.
- Reflect over the possible issues with fitting multiple models to the same data material.

Contents

- Basic methods for data handling and common visualisation methods for data
- Methods for data reduction such as Principal Component Analysis (PCA) and their use for imputation of missing data.
- Methods for unsupervised and supervised learning/classification such as: Support Vector Machines (SVM), clustering (K-means), hierarchical clustering, simpler regression methods, and methods for decision trees (bagging, boosting, and random forests).
- Multiple testing and common solutions such as Benjamini-Hochberg and Bonferroni.

Examination details

Grading scale: TH - (U,3,4,5) - (Fail, Three, Four, Five)
Assessment: The final grade is determined by the final project.

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.

Parts
Code: 0123. **Name:** Computer Lab 1.
Credits: 2. **Grading scale:** UG. **Assessment:** Written report. **Contents:** Data handling and visualisation.
Code: 0223. **Name:** Computer Lab 2.
Credits: 2. **Grading scale:** UG. **Assessment:** Written report. **Contents:** Supervised learning.
Code: 0323. **Name:** Computer Lab 3.

Credits: 2. **Grading scale:** UG. **Assessment:** Written report. **Contents:** Unsupervised learning.
Code: 0423. **Name:** Project.
Credits: 1,5. **Grading scale:** TH. **Assessment:** Written project report **Contents:** Final project

Admission

Admission requirements:

- FMAB30 Calculus in Several Variables or FMAB35 Calculus in Several Variables or FMSF20 Mathematical Statistics, Basic Course or FMSF25 Mathematical Statistics - Complementary Project or FMSF32 Mathematical Statistics or FMSF45 Mathematical Statistics, Basic Course or FMSF50 Mathematical Statistics, Basic Course or FMSF55 Mathematical Statistics, Basic Course or FMSF70 Mathematical Statistics or FMSF75 Mathematical Statistics, Basic Course or FMSF80 Mathematical Statistics, Basic Course
- FMAA20 Linear Algebra with Introduction to Computer Tools or FMAA21 Linear Algebra with Numerical Applications or FMAB20 Linear Algebra or FMSF20 Mathematical Statistics, Basic Course or FMSF25 Mathematical Statistics - Complementary Project or FMSF32 Mathematical Statistics or FMSF45 Mathematical Statistics, Basic Course or FMSF50 Mathematical Statistics, Basic Course or FMSF55 Mathematical Statistics, Basic Course or FMSF70 Mathematical Statistics or FMSF75 Mathematical Statistics, Basic Course or FMSF80 Mathematical Statistics, Basic Course

Assumed prior knowledge: A basic course in mathematical statistics and knowledge in linear algebra.
The number of participants is limited to: 50
Selection: Completed university credits within the program. (Note that only credits which according to Ladok have been included in the program before the selection process count. For students taking master's programmes 180 credits corresponding to a bachelor's degree are added.) Priority is given to students enrolled on programmes that include the course in their curriculum. Among these students place is guaranteed to those in the specialisation on Riskmodellering at Risk, säkerhet och krishantering education.
The course overlaps following course/s: FMSF86, FMAN45, EDAN96

Reading list

- Gareth James, Daniela Witten, Trevor Hastie, Robert Tibshirani: An Introduction to Statistical Learning, with Applications in R (2ed, 2021 Edition). Springer, 2021, ISBN: 978-1071614174. Available as e-book: https://web.stanford.edu/~hastie/ISLRv2_website.pdf.
- Jake VanderPlas: Python Data Science Handbook, Essential Tools for Working with Data. O'Reilly, 2016, ISBN: 978-1491912058. Available as e-book.

Contact and other information

Director of studies: Johan Lindström, studierektor@matstat.lu.se
Course coordinator: Linda Hartman, linda.hartman@matstat.lu.se
Course homepage:
https://www.maths.lu.se/utbildning/civilingenjoersutbildning/matematisk-statistik-paa-civilingenjoersprogram/
Further information: Given in parallell with FMSF86. Only one of the courses FMSF86 and FMSF90 may be included in a degree. The course overlaps with EDAN96.