FID3214 Programming for Data Science 7.5 credits

Programmering för data science

This is a translation of the Swedish, legally binding, course syllabus.

If the course is discontinued, students may request to be examined during the following two academic years

Establishment

Course syllabus for FID3214 valid from Autumn 2020

Grading scale

P, F

Education cycle

Third cycle

Specific prerequisites

Language of instruction

The language of instruction is specified in the course offering information in the course catalogue.

Intended learning outcomes

Having passed the course, the student should be able to:

1. account for and discuss the application of

- techniques to convert data to appropriate format for data analysis
- algorithms to analyse data through supervised and unsupervised machine learning, as well as
- techniques and performance measurements for evaluation of data analysis results

2. implement and apply

- techniques to convert data to an appropriate format for data analysis
- algorithms for supervised and unsupervised machine learning, as well as
- techniques and performance measurements for evaluation of data analysis results

Course contents

Syntax and semantics for programming languages that are particularly suited for data science, e.g., Python. Routines to import, combine, convert and make selection of data. Algorithms for handling of missing values, discretisation and dimensionality reduction. Algorithms for supervised machine learning, e g naive Bayes, decision trees, and random forests. Algorithms for unsupervised machine learning e g k-means clustering. Libraries for data analysis. Evaluation methods and performance metrics. Visualisation and analysis of results of data analysis.

Course structure

Ten lectures (non-mandatory)

One mandatory seminar

Four assignments, of which one is to be presented at the seminar

Course literature

I. Witten, E. Frank, M. Hall and C. Pal, Data Mining: Practical Machine Learning Tools and Techniques (4th ed.), Morgan Kaufmann, 2016 ISBN: 9780128042915. J. VanderPlas, Python Data Science Handbook: Essential tools for working with data (1st ed.), O'Reilly Media Inc., 2016 ISBN: 9781491912058.

Required equipment

Own computer

Examination

- EXA1 - Examination, 7.5 credits, grading scale: P, F

Based on recommendation from KTH's coordinator for disabilities, the examiner will decide how to adapt an examination for students with documented disability.

The examiner may apply another examination format when re-examining individual students.

Three of the four assignments are the same as for the corresponding course on the advanced level (ID2214). The fourth assignment consists of a small-scale research project, in which a research question has to be identified and results have to be presented in a short scientific article.

Other requirements for final grade

Pass on all assignments, including presentation at the final seminar.

Ethical approach

- All members of a group are responsible for the group's work.
- In any assessment, every student shall honestly disclose any help received and sources used.
- In an oral assessment, every student shall be able to present and answer questions about the entire assignment and solution.