UNIVERSITY OF GOTHENBURG

DEPARTMENT OF PHILOSOPHY, LINGUISTICS AND THEORY OF SCIENCE

LT2222 Machine learning for statistical NLP: introduction, 7.5 credits

Maskininlärning för statistisk datalingvistik: inledning, 7,5 högskolepoäng

Second Cycle

Confirmation

This course syllabus was confirmed by Department of Philosophy, Linguistics and Theory of Science on 2020-05-25 to be valid from 2020-08-31, autumn semester of 2020.

Field of education: Science 100%
Department: Department of Philosophy, Linguistics and Theory of Science

Position in the educational system

This course is part of the 'Master in Language Technology' program (H2MLT). It can also be offered as a freestanding course.

The course can be part of the following programme: 1) Master in Language Technology (One year or Two years) (H2MLT)

Main field of studies	*Specialization*
Language Technology	A1F, Second cycle, has second-cycle course/s as entry requirements

Entry requirements

Admission to the course requires either a pass in both the courses:

LT2001 Introduction to programming, 7.5 credits
LT2002 Introduction to formal linguistics, 7.5 credits

or equivalent language technology skills.

Learning outcomes

On successful completion of the course the student will be able to:

Knowledge and understanding

- account for the use of linguistic data in common NLP applications,
- show awareness of basic practices in, and commonalities between, NLP and machine learning,
- show evidence of understanding the mathematical basis for quantitative approaches to NLP at a conceptual level,

Competence and skills

- apply standard evaluation techniques and interpret their results,
- apply simple statistical NLP algorithms and approaches,
- use data analysis and machine learning toolkits and collaborative tools for developing well-documented software,

Judgement and approach

- compare and select appropriate machine learning and statistical approaches.

Course content

This course will cover the following areas:

- Data representation for machine learning over human language tasks.
- Data pipeline design and scientific practice in machine learning for NLP with consideration of ethical and professional matters.
- Foundations of learning in terms of information theory, probability and statistics, and vector spaces.
- "Classic" machine learning techniques such as support vector machines (SVM) and logistic regression.
- Basics of perceptrons, neural networks, and stochastic gradient descent.

Course content corresponding to learning outcomes of knowledge and understanding will enable students to:

- enumerate a sample of simple algorithms and approaches for NLP tasks,
- describe simple statistical NLP algorithms and approaches in terms of their inputs and outputs,
- explain at a conceptual level, how statistical NLP algorithms function, in terms of their mathematical foundations in probability, statistics, linear algebra, and calculus,
- account for the use of linguistic data in common NLP applications,
- identify basic practices in NLP applications research and development,
- identify commonalities between NLP tasks and other common machine learning tasks.

Course content corresponding to learning outcomes of competence and skills will enable students to:

- develop software in current well-known toolkits for data analysis and machine learning,
- design a data pipeline from linguistic input data to output data,
- apply established experimental paradigms from NLP and computational linguistics research,
- use machine learning techniques as building blocks in well-known basic NLP tasks,
- interpret NLP system output using standard evaluation approaches,
- document code, system design, and results,
- use collaborative software development tools and code repositories.

Course content corresponding to learning outcomes of judgement and approach will enable students to:

- select machine learning and statistical tools appropriate to common NLP and data processing tasks,
- evaluate problem-solving approaches in statistical NLP using standard measures,
- assess the success or failure of an approach to solving NLP application problems using quantitative and qualitative methods,
- compare multiple machine learning approaches to NLP application problem-solving.

Form of teaching

The principal modes of teaching will be: lecture, demonstration, and practical take-home assignments, which will also be part of the assessment.

Language of instruction: English
This course will be taught in English.

Assessment

This course will be assessed in terms of graded individual take-home assignments, assessing practical mathematical and programming skills and capacity for independent work. Students may also be asked to write in-class written tests or quizzes as part of the assessment.

A student who has failed a test twice has the right to change examiner, unless weighty argument can be adduced. The application shall be sent to the board of the department and has to be in writing. The total number of exam sessions is five, when feasible. Completion of examined student achievement is admitted.

Grades

The grading scale comprises: Pass with Distinction (VG), Pass (G) and Fail (U).

Course evaluation

Students participating in, or having completed the course, are given the chance to anonymously submit their opinions of and suggestions for the course in a course evaluation. A short version of the course evaluation, together with the reflections of the course coordinator, is published and made available to the students within a reasonable time after the course has finished. The next time the course will be given, a short version of the course evaluation will be presented together with any measures implemented.