University of St.Gallen

Course and Examination Fact Sheet: Spring Semester 2019

10,376: Statistical Learning and Applications

ECTS credits: 4

Overview examination/s

(binding regulations see below)
Decentral - examination paper written at home (individual) (100%)

Attached courses

Timetable -- Language -- Lecturer
10,376,1.00 Statistical Learning and Applications -- Englisch -- De Mol Christine

Course information

Course prerequisites

Prerequisites (knowledge of topic):

The course will focus on the statistical and mathematical foundations of machine learning theory. The aim is to provide the students with a thorough understanding of the basic principles so as to prepare them to develop innovative methods and algorithms in their own field of applications. The course will be reasonably self-contained and does not require any specific prior knowledge in learning theory. It is nevertheless targeted at students with a sufficient quantitative background and it will rely on a basic knowledge of statistics and mathematics (probability, regression methods, linear algebra, elements of optimization theory, etc.) such as provided by standard undergraduate courses. Although students will be encouraged to perform some numerical experiments on their own, this will by no means be compulsory and could be made by the software/hardware of their choice.

Hardware:none mandatory

Software:none mandatory

Course content

This is a first course on statistical/machine learning and high-dimensional data analysis,aiming at providing a mathematical toolkit to deal with large datasets.

A tentative list of topics which will be covered is as follows (NB. some might be optional and reserved to students with a more mathematical background):

INTRODUCTION

- What is 'Statistical Learning' aka 'Learning Theory' or 'Machine Learning'?
- The goal: extract meaningful information/infer good models from huge datasets and understand how a human brain/machine can 'learn from examples'
- Examples from various scientific disciplines
- Similarities and differences with related fields: multivariate statistics, high-dimensional analysis, computational statistics, Bayesian statistics, data mining, computer vision, artificial intelligence, etc.

MATHEMATICAL FRAMEWORK

- Learning from examples
- Supervised versus unsupervised setting
- Classification and regression problems

- The ideal case: the Bayes risk and the Minimum Mean Square Error Predictor
- Empirical Risk Minimization
- Probabilistic concentration inequalities: Markov, Chebyshev, Hoeffding and McDiarmid
- The PAC ('Probably Approximately Correct') paradigm

CLASSIFICATION PROBLEMS AND SUPPORT VECTOR MACHINES (SVM)

- The Perceptron, ancestor of neural networks
- Convex optimization in a nutshell
- Linear SVM: the separable ('hard margin') and nonseparable ('soft margin') cases
- Nonlinear SVM: the 'kernel trick'

KERNEL METHODS

- Kernels
- Reproducing Kernel Hilbert Spaces (RKHS)
- The Representer Theorem
- Examples of popular kernels

CONSISTENCY AND COMPLEXITY ISSUES

- Uniform Convergence of Empirical Means
- 'Overfitting' and capacity of a class of functions (VC dimension, Rademacher complexity)
- Risk bounds from Rademacher averages
- Rademacher complexity of a ball in a RKHS

LINEAR REGRESSION PROBLEMS

- The trouble with Ordinary Least Squares (OLS) for high-dimensional datasets
- Linear regularization/shrinkage methods: Truncated SVD ('Singular Value Decomposition') aka Principal Component Regression and Ridge Regression
- Nonlinear regularization/shrinkage methods: Lasso and Elastic Net Regression; variable selection; other (structured-) sparsity enforcing methods
- Computational issues

APPLICATIONS

- High-dimensional time series prediction in Economics and other disciplines
- Portfolio optimization/selection in Finance
- Optimal combination of forecasts
- Bioinformatics: gene selection, etc.
- Computer Vision: face detection and authentication

RISK BOUNDS

- Risk bounds via surrogate loss functions
- The contraction principle
- Convex risk minimization
- Least-squares regression in a RKHS
- The hinge loss: revisiting SVM

Course structure

There will be 5 whole-day lessons (tentatively).

Course literature

Mandatory: none

Supplementary / voluntary:

Besides the lecture notes/slides which will be distributed to the students, the main

supplementary reference book for the course will be

M. Mohri, A. Rostamizadeh, and A. Talwalkar. Foundations of Machine Learning. MIT Press, 2012.

Also recommended is

T. Hastie, R. Tibshirani, and M. Wainwright. Statistical Learning with Sparsity:
The Lasso and Generalizations. Chapman & Hall/CRC Monographs on Statistics & Applied Probability. CRC Press, 2015.

as well as the more encyclopedic references

T. Hastie, R. Tibshirani, and J. Friedman. The Elements of Statistical Learning: Data Mining, Inference, and Prediction. Springer Series in Statistics. Springer New York, 2013.

K. Murphy. Machine learning. A Probabilistic Perspective. MIT Press 2012.

The following textbooks

H. Kobayashi, B.L. Mark, and W. Turin. Probability, Random Processes, and Statistical Analysis: Applications to Communications, Signal Processing, Queueing Theory and Mathematical Finance. Cambridge University Press, 2011.

B. Efron and T; Hastie. Computer Age Statistical Inference. Algorithms, Evidence, and Data Science. Cambridge UP 2016.

can be useful as refreshers or to learn more about basic probability and statistics.

Additional course information

Only for PhD students of the University of St.Gallen

PEF students may register via regular bidding for the courses offered together by PEF and Global School in Empirical Research Methods (GSERM). Enrolment in a course is binding: students have to attend the course and take the exam. The credits will be shown on the scorecard.

All other PhD students should register for the courses offered by Global School in Empirical Research Methods (GSERM), **both via bidding and via GSERM** for:

- courses for the curriculum and

- optional courses **with** an examination. These will be listed on the scorecard under optional work (**only possible if all required elective courses have already been completed**).

Please register **only via GSERM** for:

- optional courses **without** an examination and

- optional courses **if not all required elective courses have been completed** (not shown on the scorecard)

The registration via GSERM can only be made starting **March 1st 2019**. Earlier registrations have to be kept pending and will not be confirmed.

Examination information

Examination sub part/s

1. Examination sub part (1/1)

Examination time and form

Decentral - examination paper written at home (individual) (100%)

University of St.Gallen

Remark
--

Examination-aid rule
Term papers

- Term papers must be written without anyone else's help and in accordance with the known quotation standards, and they must contain a declaration of authorship.
- The documentation of sources (quotations, bibliography) has to be done throughout and consistently in accordance with the APA or MLA standards. The indications of the sources of information taken over verbatim or in paraphrase (quotations) must be integrated into the text in accordance with the precepts of the applicable quotation standard, while informative and bibliographical notes must be added as footnotes (recommendations and standards can be found, for example, in METZGER, C. (2017), Lern- und Arbeitsstrategien (12th ed., Cornelsen Schweiz).
- For any work written at the HSG, the indication of the page numbers both according to the MLA and the APA standard is never optional.
- Where there are no page numbers in sources, precise references must be provided in a different way: titles of chapters or sections, section numbers, acts, scenes, verses, etc.
- For papers in law, the legal standard is recommended (by way of example, cf. FORSTMOSER, P., OGOREK R. et SCHINDLER B. (2018, Juristisches Arbeiten: Eine Anleitung für Studierende (6. Auflage), Zürich: Schulthess, or the recommendations of the Law School).

Supplementary aids
--

Examination languages
Question language: English
Answer language: English

Examination content
Content of the lectures.

Examination relevant literature
To be defined during the lectures.

Please note
We would like to point out to you that this fact sheet has absolute priority over other information such as StudyNet, faculty members' personal databases, information provided in lectures, etc.When will the fact sheets become binding?

- Information about courses and examination time (central/decentral and grading form): from the start of the bidding process on 24 January 2019
- Information about decentral examinations (examination-aid rule, examination content, examination relevant literature): after the 4th semester week on 18 March 2019
- Information about central examinations (examination-aid rule, examination content, examination relevant literature): from the start of the enrolment period for the examinations on 08 April 2019

Please look at the fact sheet once more after these deadlines have expired.