University of St.Gallen

Course and Examination Fact Sheet: Spring Semester 2020

10,365: Computational Statistics

ECTS credits: 4

Overview examination/s

(binding regulations see below)
Decentral - Group examination paper (all given the same grades) (100%)

Attached courses

Timetable -- Language -- Lecturer
10,365,1.00 (GSERM) Computational Statistics -- Englisch -- Audrino Francesco

Course information

Course prerequisites

Advanced knowledge in statistics and econometrics.

Learning objectives

Students will gain an advanced knowledge on the statistical aspects related to the use of machine learning techniques needed to analyze large or high-dimensional datasets.

Students will learn how to apply machine learning tools in a responsible way and will properly apply the methods on a concrete dataset of their choice using the statistical R software and prepare a research paper summarizing their results.

Course content

Computational Statistics is the area of specialization within statistics that includes statistical visualization and other computationally-intensive methods of statistics for mining large, nonhomogeneous, multi-dimensional datasets so as to discover knowledge in the data. As in all areas of statistics, probability models are important, and results are qualified by statements of confidence or of probability. An important activity in computational statistics is model building and evaluation.

First, the basic multiple linear regression is reviewed. Then, some nonparametric procedures for regression and classification are introduced and explained. In particular, Kernel estimators, smoothing splines, classification and regression trees, additive models, projection pursuit and eventually neural nets will be considered, where some of them have a straightforward interpretation, other are useful for obtaining good predictions.

The main problems arising in computational statistics like the curse of dimensionality will be discussed. Moreover, the goodness of a given (complex) model for estimation and prediction is analyzed using resampling, bootstrap and cross-validation techniques.

Course structure

Outline:

1. Overview of supervised learning

Introductory examples, two simple approaches to prediction, statistical decision theory, local methods in high dimensions, structured regression models, bias-variance tradeoff, multiple testing and use of p-values.

2. Linear methods for regression

Multiple regression, analysis of residuals, subset selection and coefficient shrinkage.

3. Methods for classification

Bayes classifier, linear regression of an indicator matrix, discriminant analysis, logistic regression.

4. Nonparametric density estimation and regression

Histogram, kernel density estimation, kernel regression estimator, local polynomial nonparametric regression estimator, smoothing splines and penalized regression.

5. Model assessment and selection

Bias, variance and model complexity, bias-variance decomposition, optimism of the training error rate, AIC and BIC, cross-validation, boostrap methods.

6. Flexible regression and classification methods

Additive models; multivariate adaptive regression splines (MARS); neural networks; projection pursuit regression; classification and regression trees (CART).

7. Bagging and Boosting

The bagging algorithm, bagging for trees, subagging, the AdaBoost procedure, steepest descent and gradient boosting.

8. ***Introduction to the idea of a Superlearner***

Course literature

Main references:

- F. Audrino, Lecture Notes.
- Hastie T., Tibshirani, R. and Friedman, J. (2001). *The elements of statistical learning: data mining, inference and prediction*, Springer Series in Statistics, Springer, Canada.
- Bühlmann, P. and van de Geer, S. (2011).*Statistics for High-Dimensional Data: Methods, Theory and Applications.* Springer.
- van der Laan, M.J. and Rose, S. (2011). *Targeted Learning: Causal Inference for Observational and Experimental Data.* Springer.

References to related published papers / chapters of books will be given during the course.

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)

University of St.Gallen

Examination information

Examination sub part/s

1. Examination sub part (1/1)

Examination time and form
Decentral - Group examination paper (all given the same grades) (100%)

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

Outline:

1. Overview of supervised learning

Introductory examples, two simple approaches to prediction, statistical decision theory, local methods in high dimensions, structured regression models, bias-variance tradeoff, multiple testing and use of p-values.

2. Linear methods for regression

Multiple regression, analysis of residuals, subset selection and coefficient shrinkage.

3. Methods for classification

Bayes classifier, linear regression of an indicator matrix, discriminant analysis, logistic regression.

4. Nonparametric density estimation and regression

Histogram, kernel density estimation, kernel regression estimator, local polynomial nonparametric regression estimator,

smoothing splines and penalized regression.

5. Model assessment and selection

Bias, variance and model complexity, bias-variance decomposition, optimism of the training error rate, AIC and BIC, cross-validation, boostrap methods.

6. Flexible regression and classification methods

Additive models; multivariate adaptive regression splines (MARS); neural networks; projection pursuit regression; classification and regression trees (CART).

7. Bagging and Boosting

The bagging algorithm, bagging for trees, subagging, the AdaBoost procedure, steepest descent and gradient boosting.

8. Introduction to the idea of a superlearner

Examination relevant literature

F. Audrino, Lecture Notes, available on Canvas.

Please note

Please note that this fact sheet alone is binding and has priority over any other information such as StudyNet (Canvas), personal databases or faculty members' websites and information provided in their lectures, etc.

Any possible references and links within the fact sheet to information provided by third parties are merely supplementary and informative in nature and are outside the University of St.Gallen's scope of responsibility and guarantee.

Documents and materials that have been submitted no later than the end of term time (CW21) are relevant to central examinations.

Binding nature of the fact sheet:

- Information about courses and examination time (central/decentral) and examination typestarting from the beginning of the bidding on 23 January 2020
- Information about examinations (examination aid regulations, examination content, examination-relevant literature) for decentral examinations after the 4th semester week on 16 March 2020
- Information about examinations (examination aid regulations, examination content, examination-relevant literature) for central examinations as from the starting date for examination registration on 6 April 2020

Please consult the fact sheet again after these deadlines have expired.