University of St.Gallen

Course and Examination Fact Sheet: Spring Semester 2014

10,365: Computational Statistics

ECTS credits: 4

Overview examination/s

(binding regulations see below)
Decentral - Group examination paper (100%)

Attached courses

Timetable -- Language -- Lecturer
10,365,1.00 Computational Statistics -- English -- Audrino Francesco

Course information

Course prerequisites

Advanced knowledge in statistics and econometrics.

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.

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.

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.

References to related published papers will be given during the course.

Additional course information

--

Examination information

Examination part/s

1. Examination part (1/1)

Examination time and form

Decentral - Group examination paper (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. (2010), Lern- und Arbeitsstrategien (11. Aufl.). Aarau: Sauerländer).
- 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. und VOGT H. (2008, Juristisches Arbeiten: Eine Anleitung für Studierende (4. 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.

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.

University of St.Gallen

Examination relevant literature

F. Audrino, Lecture Notes.

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 23 January 2014
- Information about decentral examinations (examination-aid rule, examination content, examination relevant literature): after the 4th semester week on 17 March 2014
- Information about central examinations (examination-aid rule, examination content, examination relevant literature): from the start of the enrolment period for the examinations on 7 April 2014

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