



Analysis of High Dimensional Data (I002004)

Cursusomvang (nominale waarden; effectieve waarden kunnen verschillen per opleiding)

Studiepunten 3.0 **Studietijd** 75 u **Contacturen** 30.0 u

Aanbodsessies en werkvormen in academiejaar 2017-2018

A (semester 2)	werkcollege: PC-klasoefeningen	15.0 u
	hoorcollege	15.0 u

Lesgevers in academiejaar 2017-2018

Thas, Olivier	LA10	Verantwoordelijk lesgever
---------------	------	---------------------------

Aangeboden in onderstaande opleidingen in 2017-2018

	stptn	aanbodssessie
Master of Science in Bioinformatics (afstudeerrichting Bioscience Engineering)	3	A
Master of Science in Chemical Engineering	3	A
Master of Science in de ingenieurswetenschappen: chemische technologie	3	A
Master of Science in de bio-ingenieurswetenschappen: land- en waterbeheer	3	A
Master of Science in de bio-ingenieurswetenschappen: milieutechnologie	3	A
Uitwisselingsprogramma Bioinformatics (niveau master)	3	A

Onderwijstalen

Engels

Trefwoorden

Statistics, multivariate data analysis, data mining, data science

Situering

Multivariate and high-dimensional data analysis is basically a collection of many statistical methods that are applicable to large and/or high dimensional data sets. All methods that are covered in this course, are often applied in industry and research institutions. Good knowledge of basic statistical methods and linear regression models are required, as well as notions of matrix algebra (matrix multiplications, matrix inverse, inner product).

Applications are important in this course, and they are related to several different subject fields (e.g. analytical chemistry, ecology, environmental sciences, biotechnology, ...).

Reporting of the results of a statistical analysis is considered to be very important.

Inhoud

1. Association versus prediction
2. The singular value decomposition (SVD), multidimensional scaling (MDS), principal component analysis (PCA) and the biplot
3. Sparse SVD, sparse PCA
4. High-dimensional linear prediction models (including classification): principal component regression, ridge, lasso and elastic net
5. Model building / feature selection for linear models in high-dimensional settings
6. (Sparse) Fisher's Discriminant analysis
7. Large scale hypothesis testing, the false discovery rate and empirical Bayes
8. One of the following topics (depending on the interest of the students): functional data analysis, canonical correlation analysis, correspondence analysis, biclustering, factor analysis, cluster analysis, ...

Begincompetenties

A basic course in probability theory and statistics (linear models, i.e., regression analysis and analysis of variance), and a good basic knowledge of matrix algebra.

Eindcompetenties

- 1 The student has knowledge of methods for analysing and exploring high-dimensional data sets.
- 2 The student can detect and quantify structures in large high dimensional/multivariate datasets, using the software R.
- 3 The student can value and interpret the statistical data analyses of high-dimensional data correctly.
- 4 The student can correctly report the results of the data analyses according to scientific standards.
- 5 The student can take responsibility and initiative in a group effort.

Creditcontractvoorwaarde

Toelating tot dit opleidingsonderdeel via creditcontract is mogelijk mits gunstige beoordeling van de competenties

Examencontractvoorwaarde

Dit opleidingsonderdeel kan niet via examencontract gevolgd worden

Didactische werkvormen

Hoorcollege, werkcollege: PC-klasoefeningen

Leermateriaal

A syllabus is available at ca. 10 €. All learning materials are made available through Minerva.

Referenties

- Efron, B. and Hastie, T. (2016). Computer Age Statistical Inference. Cambridge University Press.
Efron, B. (2010). Large Scale Inference. IMS Monographs.
Johnson, R. and Wichern, D. (1998). Applied Multivariate Statistical Analysis. Prentice Hall, 816pp.
Hastie, T., Tibshirani, R. and Friedman, J. (2009). The Elements of Statistical Learning (2nd edition). Springer.
Ramsay, J. and Silverman, B. (2002). Applied Functional Data-analysis. Springer-Verlag

Vakinhoudelijke studiebegeleiding

In the practical sessions in the PC classes the students are coached by an assistant. Students can make an appointment to ask questions to the lecturer. Questions and answers can be exchanged in Minerva.

Evaluatiemomenten

periodegebonden en niet-periodegebonden evaluatie

Evaluatievormen bij periodegebonden evaluatie in de eerste examenperiode

Schriftelijk examen met open vragen, openboekexamen

Evaluatievormen bij periodegebonden evaluatie in de tweede examenperiode

Schriftelijk examen met open vragen, openboekexamen

Evaluatievormen bij niet-periodegebonden evaluatie

Werkstuk, peer-evaluatie

Tweede examenkans in geval van niet-periodegebonden evaluatie

Examen in de tweede examenperiode is mogelijk

Toelichtingen bij de evaluatievormen

The final examination consists of a written examination with open questions (part of it is an open book examination). Most of the questions aim to assess the insight of the student in the statistical analysis and how they result in the conclusions. This happens at the borderline between theory and application. The assignment relates to a real life problem. Students may work together in groups of 3 to 4 persons. The result of the project work is a written report that should satisfy conventional scientific and professional standards.

Eindscoreberekening

The total mark is a weighted average of:
- Final exam (12/20)

