Universitat Autònoma de Barcelona

2018/2019

Multivariate Analysis

Code: 100122
ECTS Credits: 6

Degree	Type	Year	Semester
2500149 Mathematics	OT	4	0
2501919 Applied Statistics	OB	3	1

Contact

Name: Mercè Farré Cervelló
Email: Merce.Farre@uab.cat

Use of languages

Principal working language: catalan (cat)
Some groups entirely in English: No
Some groups entirely in Catalan: Yes
Some groups entirely in Spanish: No

Prerequisites

A previous course of Linear Algebra is essential, as well as courses in Probability and Statistical Inference. Also, a good knowledge of the R software is assumed.

Objectives and Contextualisation

Most of collected data sets are multivariate, that is, for the same experimental unit, perhaps a complex nature object, we observe simultaneously the values of several variables. Multivariate Analysis deals with the methods that are most appropriate for describing, exploring and modelling vectorial data, as well as for applying statistical inference.

The interest in processing large amounts of observations in many variables of a diverse nature, together with the aim of reducing the information that is not relevant or discovering patterns of association between variables or between cases, they have recently promoted the development of a series of multivariate techniques.

This subject is intended as a first contact of the student with the statistical learning theory. Students must understand the power and applicability as well as the limitations of the multivariate tools, some of which are based on very simple heuristic ideas. The subject focuses in the applications, mostly in the practical computer sessions using the R free software resources. Theoretical and problems sessions are devoted to formalize the models, derive their properties, and study some models validation techniques.

Content

Multivariate data: **Statistical learning and dimension reduction.**

- Supervised and unsupervised learning. Multivariate methods. Examples.
- Random vectors. Expectation vector and covariance-correlation matrices. Properties.
- Multivariate data. Sample expectation and covariance-correlation matrices. Maximum likelihood estimation in the Gaussian case.
- Spectral decomposition (SD) and singular value decomposition (SVD).
- Maximizing quadratic forms: The fundamental theorem.

Factorial methods I: Principal components analysis (PCA).

- Introduction to PCA. Definition of components. The fundamental result.
- Criteria for deciding on the number of components: The principal components.
- Variables and individuals plots. Standardizations.
- Row and column analysis of the eigenvectors matrix and other related matrices.
- A geometric point of view of the principal components.

Factorial methods II: Factorial analysis (FA).

- The factorial model. Communalities and specificities.
- The covariance matrix decomposition theorem.
- Discussing the existence and uniqueness of the factorial model.
- Rotations.
- Parameters estimation methods.
- Factorial scores estimation or prediction.
- Interpreting the results. Comparing PCA and FA.

Factorial methods III: Multidimensional scaling (MDS) and correspondence analysis (CA).

- Objectives and methods.
- Classic and metric multidimensional scaling.
- Non-metric multidimensional scaling.
- Distances, proximities and dissimilarities.
- Categorical data: *Chi-square* distance and other dissimilarities.
- Correspondence analysis as a MDS: Profiles and inertias.
- Decomposing inertia.
- Graphical representationsand interpretation of results in CA.

Cluster analysis (CLA).

- Comparing different approaches: Examples.
- Results' analysis and validation.
- Hierarchical clustering: Link functions.
- Centroid based methods: The -means algorithm.
- Model based methods: Expectation and maximization (EM).

Multivariate inference basics.

- The likelihood ratio test.
- Tests for expectation vectors.
- Tests for covariance matrices.
- ANOVA and MANOVA.

Discriminant analysis (DA) and other supervised methods.

- Objectives and criteria of discriminant analysis.
- Discriminant analysis in Gaussian models.
- Fisher's linear discriminant.
- Partial least squares regression (PLS).
- Other methods.