STATISTICAL MODELLING II

Francesca Greselin

August 2023

MIXTURE MODELS (6h)

Theoretical Aspects:

- Introduction to Cluster Analysis and Classification
- General definition of a mixture model, examples and properties
- Focus on Gaussian Mixtures: parsimonious models and their geometric interpretation.
- Inferential methods: the Expectation-Maximization algorithm, its initialization, convergence criteria.
- Open issues: degeneracies and spurious solutions.
- Model selection: number of components and patterned covariance matrix
- Mixtures of student t: Expectation-Maximization algorithm
- Mixtures of factor analyzers: Expectation-Maximization algorithm
- Brief review of mixtures with other non Gaussian components

Lab with R: implementation of EM algorithm from scratch and use of dedicated software (mclust)

ROBUST MIXTURE MODELS (6h)

Theoretical Aspects:

- Theoretical foundation of robust statistics: outliers in multivariate data, Minimum Covariance Determinant, Forward Search
- Robust classification and clustering
  - Robust Gaussian mixtures with impartial trimming and constrained estimation
  - Cluster weighted model
  - Mixtures of factor analyzers
  - Mixtures of skew normal
- Robust semi-supervised classification and clustering
  - Gaussian parsimonious models
  - Adaptive mixtures to detect novel classes
  - Variable selection methods

Lab with R: implementation of trimmed k-means from scratch, use of dedicated software packages (tclust, fsda, Raedda, Varsel)

The lectures will be supported by extensive examples on data, with listings of code mobilizing software packages, that can be run by the reader. The chosen language for codes is the R software, which is one of the most popular languages for data science. It is an excellent tool for data science since the most recent statistical learning techniques are provided on the R platform (named CRAN). Using R is probably the best way to be directly connected to current research in statistics and data science through the packages provided by researchers.

Lecturer: Francesca Greselin

Department of Statistics and Quantitative Methods
University of Milano Bicocca
20126 Milan, Italy

Tel: +39 2 6448 3118 [office]
Fax: +39 2 6448 3105 [department]
Email: francesca.greselin@unimib.it
ORCID ID orcid.org/0000-0003-2929-1748
Scopus Author ID: 25936155400
ResearcherID: A-8770-2015