MIXMOD

a Software for Model-Based Clustering and Classification with Continuous and Categorical Data

Christophe Biernacki[a], Gilles Celeux[b], Gérard Govaert[c], Florent Langrognet[a]

[a]**Laboratoire de Mathématiques de Besançon** (Université de Franche-comté - CNRS)
[b]**INRIA Futurs, Université Paris-Sud**
[c]**Laboratoire Heudiasyc** (Université de Compiègne - CNRS)

Overview

- A software for modelling heterogeneous data
- Density estimation, cluster and discriminant analysis
- Written in C++ and interfaced with Matlab and Scilab
- Different user levels from beginner to expert
- Publicly available under the GPL licence at **http://www-math.univ-fcomte.fr/mixmod/index.php**
- Distributed for Linux, Unix and Windows
- Numerous graphical displays available
- Documentation: user guide, statistical presentation, ...

Models

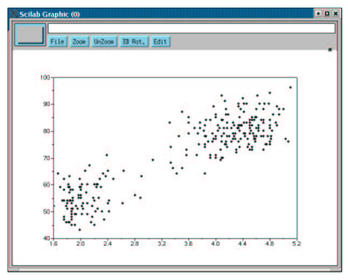
- Data are assumed to arise from a **mixture of distributions**
- Clustering context: Labels of the mixture components are unknown
- Classification context: Labels of the mixture components are known
- **Gaussian** mixture are modelling **continuous** data: **14** different models available
- **Latent** class model are modelling **discrete** data: **5** different multivariate multinomial models available

Estimation and Model Selection

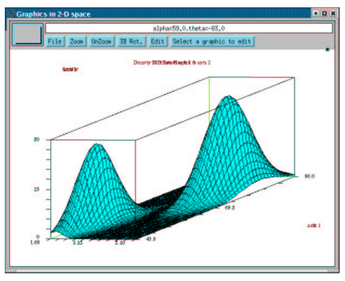
- Estimation of the mixture parameters is considered through **Maximum Likelihood**
- The **EM** algorithm and Stochastic and Classification variants are proposed in the unsupervised or semi supervised contexts
- Many **initialization** strategies combining those algorithms are possible
- Choosing the number of **mixture components** or selecting a **parsimonious** model can be achieved with various criteria
- Penalized likelihood criteria as **BIC**
- Entropy criteria as **ICL**
- **Cross validated** error rates

Density Estimation

Data in the 1rst PCA 2D space

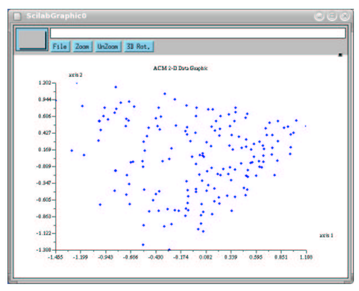


Mixture density in the 1rst PCA 2D space

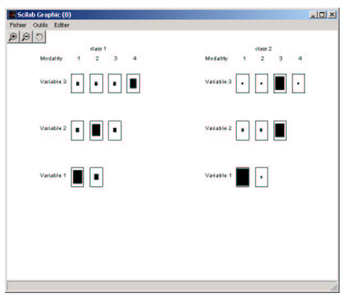


Continuous data

Data in the 1rst MCA 2D space



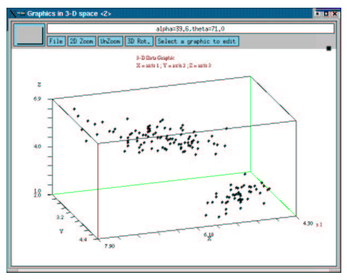
Component parameters



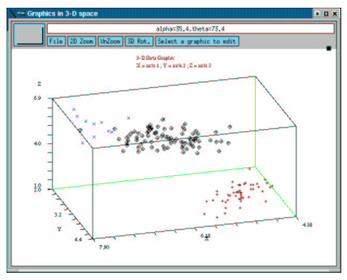
Discrete data

Clustering

Data in the 1rst PCA 3D space

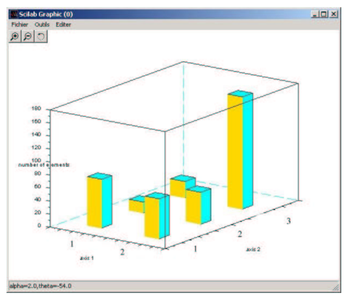


Data & labels in the 1rst PCA 3D space

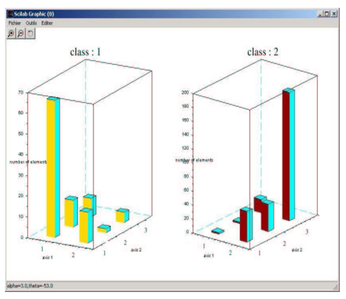


Continuous data

Data in the 1rst 2D-space



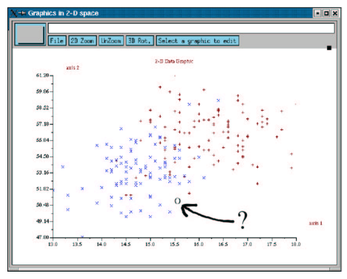
Data in the 1rst 2D-space



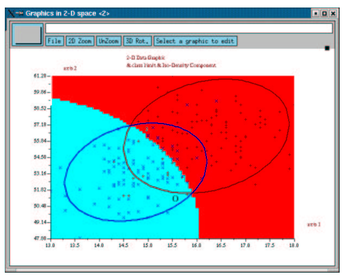
Discrete data

Discriminant Analysis

Training data & labels and new individual

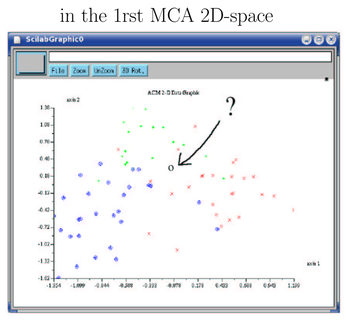


Class limit, and iso-density component

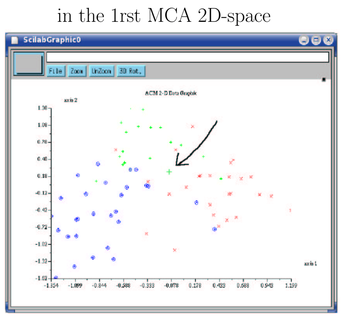


Continuous data

Training data & labels and new individual in the 1rst MCA 2D-space



Label of new individual in the 1rst MCA 2D-space



Discrete data