

Model-based clustering

Presentation of the MIXMOD software

Gilles Celeux

INRIA Saclay-Île-de-France

In this mini course, I will present model-based clustering methods:

- Gaussian mixtures models for continuous data sets and the multivariate multinomial distributions for categorical data sets;
- EM-like algorithms to estimate the parameters of the numerous possible mixture models through maximum likelihood methodology. The important problems of spurious local maximizers and the dependence of the algorithms over their starting positions will be discussed;
- Model selection issue and the choice of the number of cluster according to various model selection criteria will be considered;
- Extensions as clustering of mixed data, variables selection, and alternative mixture models will be presented;
- If enough time, Bayesian inference for mixtures as the block clustering model will be introduced.

Along the lecture, the main features of the software MIXMOD and its extension RMIXMOD in R will be presented and used to illustrate the models and methods introduced.

References

BIERNACKI, C., CELEUX, G., GOVAERT, G. & LANGROGNET, F. (2006). Model-based cluster and discriminant analysis with the mixmod software. *Computational Statistics and Data Analysis*, 51, 587-600.

FRALEY, C., RAFTERY, A. E. (2002). Model-Based Clustering, Discriminant Analysis, and Density Estimation. *J. Am. Statistical Association*, 97, 611-631.

LEBRET, R., IOVLEFF, S., LANGROGNET, F., BIERNACKI, C., CELEUX, G. & GOVAERT, G. (2015) Rmixmod: The R package of the Model-Based Unsupervised, Supervised and Semi-supervised Classification Mixmod Libray. *Journal of Statistical Software*. (In Press).

MCLACHLAN, G. J., Peel, D. (2000). *Finite Mixture Models*. Wiley, New York.