

Mixture-based Clustering for Ordinal Data

Daniel Fernández

University at Albany, State University of New York (SUNY)
dfdez23@hotmail.com

Abstract: Many of the methods which deal with the reduction of dimensionality in matrices of data are based on mathematical techniques such as distance-based algorithms or matrix decomposition and eigenvalues. In general, it is not possible to use statistical inferences or select the appropriateness of a model via information criteria with these techniques because there is no underlying probability model. Additionally, the use of ordinal data is very common (e.g. Likert or pain scale). Recent research has developed a set of likelihood-based finite mixture models for a data matrix of ordinal data (Fernández *et al.*, 2016). This approach applies fuzzy clustering via finite mixtures to the stereotype model. Fuzzy allocation of rows, columns (called one-dimensional clustering) and rows, and columns simultaneously (called two-dimensional clustering, biclustering, or block clustering) to corresponding clusters is obtained by performing the expectation-maximization (EM) algorithm and, also by Bayesian approaches (Reversible-Jump MCMC sampler, Fernández and Arnold, 2016). Examples with ordinal data sets will be shown to illustrate the application of this approach. Finally, different visualization tools for depicting the fuzziness of the clustering results for ordinal data will be demonstrated.

Keywords: Classification; Dimension reduction; Fuzzy clustering; Likert scale; Mixture models; Ordinal data; RJMCMC; Stereotype model

References

Fernández, D., Arnold, R. and Pledger, S. (2016). Mixture-based clustering for the ordered stereotype model. *Computational Statistics & Data Analysis*, 93, 46-75.

Fernández, D. and Arnold, R. (2016). Model selection for mixture-based clustering for ordinal data. *Australian and New Zealand Journal of Statistics*, In Press.