

Mixture-based clustering with covariates for ordinal responses

Kemmawadee Preedalikit^a, Daniel Fernández^b, Ivy Liu^c, Louise McMillan^c, Marta Nai Ruscone^d,
and Roy Costilla^e

^aUniversity of Phayao; kemmawadee@gmail.com

^bUniversitat Politècnica de Catalunya - BarcelonaTech; daniel.fernandez.martinez@upc.edu

^cVictoria University Wellington; ivy.liu@vuw.ac.nz, mcmillo@ecs.vuw.ac.nz

^dUniversity of Genoa; marta.nairuscone@unige.it

^eAgResearch NZ; roy.costilla@agresearch.co.nz

Existing methods can perform likelihood-based clustering on a multivariate data matrix of ordinal responses, using finite mixtures to cluster the rows and columns of the matrix [1, 2]. Those models can incorporate the main effects of individual rows and columns and the cluster effects to model the matrix of responses [3]. However, many real-world applications also include available covariates. In this study, we have extended mixture-based models to include covariates and test what effect this has on the resulting clustering structures. We focus on clustering the rows of the data matrix, using the proportional odds cumulative logit model for ordinal data. We fit the models using the Expectation-Maximization (EM) algorithm and assess their performance. Finally, we also illustrate an application of the models to the well-known arthritis clinical trial data set.

Keywords: cluster analysis, mixture models, EM algorithm, ordinal responses, proportional odds model

- [1] D. Fernández, R. Arnold, and S. Pledger, “Mixture-based clustering for the ordered stereotype model,” *Computational Statistics & Data Analysis*, vol. 93, pp. 46–75, 2016.
- [2] D. Fernández, R. Arnold, S. Pledger, I. Liu, and R. Costilla, “Finite mixture biclustering of discrete type multivariate data,” *Advances in Data Analysis and Classification*, vol. 13, pp. 117–143, 2019.
- [3] R. Rocci and M. Vichi, “Two-mode multi-partitioning,” *Computational Statistics & Data Analysis*, vol. 52, no. 4, pp. 1984–2003, 2008.