

Mixture model selection criteria focusing on a clustering purpose

Gilles Celeux

INRIA Saclay-Île-de-France

In the framework of model-based clustering, I will review model selection criteria for choosing the number of clusters. I will pay special attention to the Integrated Completed Likelihood (ICL) of a mixture model. Different ways to approximate ICL with a full Bayesian view or with a maximum likelihood view will be presented and discussed. The position of ICL with respect to BIC criterion will be discussed too and the differences between these two criteria will be highlighted. Finally, some extensions of ICL, useful in applied contexts will be presented.

References

BAUDRY, J.-P. (2015). Estimation and model selection for model-based clustering with the conditional classification likelihood. (Submitted).

BAUDRY, J.-P., CARDOSO, M., CELEUX, G., AMORIM, M. J. & SOUSA FERREIRA, A. (2014). Enhancing the selection of a model-based clustering with external categorical variables. *Advances in Data Analysis and Classification*, on line.

BAUDRY, J.-P., RAFTERY, A., CELEUX, G., LO, K. & GOTTARDO, R. (2010). Combining mixture components for clustering. *Journal of Computational and Graphical Statistics*, 19, 332-353.

BERTOLETTI, M., FRIEL, N. & RASTELLI, R. (2014). Choosing the number of components in a finite mixture model using an exact Integrated Completed Likelihood criteria. *arXiv preprint arXiv:1411.4257* [stat.CO].

BIERNACKI, C., CELEUX, G. & GOVAERT, G. (2000). Assessing a mixture model for clustering with the integrated completed likelihood. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 22, 719-725.

BIERNACKI, C., CELEUX, G. & GOVAERT, G. (2010) Exact and Monte Carlo calculations of integrated likelihoods for the latent class model. *Journal of Statistical Planning Inference*, 140, 2991-3002.

CELEUX, G., SOROMENHO, G. (1996). An entropy criterion for assessing the number of clusters in a mixture model. *Journal of Classification*, 13, 195-212.