

A Hierarchical Bayes Model for Two-Way Clustering

This article proposed a hierarchical Bayes procedure for two-way clustering. We begin with embedding two-way cluster analysis into a framework of plaid models with random effects. The likelihood is then regularised by assigning hierarchical priors to the expression levels of genes for all samples in each layer. The theoretical analysis demonstrates that the resulting posterior can attenuate the effect of high dimensionality on cluster predictions. We provide an empirical Bayes algorithm for sampling posteriors, which is based on an explicit posterior for the layer memberships of all genes and samples. The new algorithm makes the estimation of hierarchical Bayes plaid models computationally feasible and efficient. The performance of our procedure is tested on both simulated and real microarray gene expression datasets [1]. The numerical results show that our proposal substantially out-performs the original plaid model [2] in terms of error rates across a range of scenarios.

References

- [1] Chu, S., DeRisi, J., Eisen, M., Mulholland, J., Botstein, D., Brown, PO, Herskowitz, I (1998). The transcriptional program of sporulation in budding yeast. *Science* 282:699-705.
- [2] Lazzeroni, L., Owen, A. (2002) Plaid models for gene expression data. *Statist. Sinica* 12:61-86.