

Variable selection in penalized model-based clustering via regularization on grouped parameters

Benhuai Xie¹, Wei Pan^{*1}, Xiaotong Shen²

¹Division of Biostatistics, School of Public Health, University of Minnesota

²School of Statistics, University of Minnesota

E-Mail: weip@biostat.umn.edu

Abstract: Penalized model-based clustering has been proposed for high-dimensional but small sample-sized data, such as arising from genomic studies; in particular, it can be used for variable selection. A new regularization scheme is proposed to group together multiple parameters of the same variable across clusters, which is shown both analytically and numerically to be more effective than the conventional L_1 penalty for variable selection. In addition, we develop a strategy to combine this grouping scheme with grouping structured variables. Simulation studies and applications to microarray gene expression data for cancer subtype discovery demonstrate the advantage of the new proposal over several existing approaches.