

Dimension Reduction in Survival Analysis: with Applications in Microarrays

Steven (Shuangge) Ma
Yale University

An important goal of microarray studies is to discover genes that are associated with clinical outcomes such as disease status and patient survival. While a typical experiment surveys gene expressions on a global scale, there may be only a small number of genes that have significant influence on a clinical outcome. Moreover, expression data have cluster structures and the genes within a cluster have correlated expressions and coordinated functions, but the effects of individual genes in the same cluster may be different. For gene expression data without pathway information, we divide genes into clusters using commonly used methods such as K-means or hierarchical approaches. The optimal number of clusters is determined using the Gap statistic. We propose a Clustering Threshold Gradient Descent Regularization (CTGDR) method, for simultaneous cluster selection and within cluster gene selection. Compared to the standard TGDR and other regularization methods, the CTGDR takes into account the cluster structure and carries out feature selection at both the cluster level and within-cluster gene level. We demonstrate the CTGDR on two studies correlating survival of lymphoma patients with microarray expressions. The study is joint with Dr. Jian Huang, University of Iowa.

[Steven Ma, 60 College ST, LEPH 209. New Haven, CT 06520; shuangge.ma@yale.edu]

<http://www.stat.uiowa.edu/techrep/tr348.pdf>