Variable Selection in the Accelerated Failure Time Model via the Bridge Method

by
Shuangge Ma
Yale University
60 College ST, LEPH 209, New Haven CT 06510
shuangge.ma@yale.edu

Abstract

In high throughput genomic studies, an important goal is to identify a small number of genomic markers that are associated with development and progression of diseases. A representative example is microarray prognostic studies, where the goal is to identify genes whose expressions are associated with disease free or overall survival. Because of the high dimensionality of gene expression data, standard survival analysis techniques cannot be directly applied. In addition, among the thousands of genes surveyed, only a subset are disease-associated. Gene selection is usually needed along with estimation. In this study, we model the association between gene expressions and survival using accelerated failure time (AFT) models. We propose using the bridge penalization for regularized estimation and gene selection. An efficient iterative computational algorithm is proposed. Tuning parameters are selected using V-fold cross validation. We use a resampling-based approach to evaluate prediction performance of the proposed approach and the relative stability of identified genes. Analysis of two lymphoma prognostic studies suggests that the proposed approach can identify a small number of genes. We also present some theoretical results concerning the variable selection properties of the proposed method in "small n, large p" settings.

This is joint work with Jian Huang (University of Iowa).