

Variable Selection in Semi-parametric Regression Modeling

by

Hua Liang

University of Rochester Medical Center

601 Elmwood Avenue, Box 630

Rochester, NY 14642, USA

`hliang@bst.rochester.edu`

Abstract

We are concerned with how to select significant variables in semi-parametric modeling. Variable selection for semi-parametric regression models consists of two components: model selection for nonparametric components and selection of significant variables for parametric portion. Thus, it is much more challenging than that for parametric models such as linear models and generalized linear models because traditional variable selection procedures including stepwise regression and the best subset selection require model selection to nonparametric components for each sub-model. This leads to very heavy computational burden. In this paper, we propose a class of variable selection procedures for semi-parametric regression models using non-concave penalized likelihood. The newly proposed procedures are distinguished from the traditional ones in that they delete insignificant variables and estimate the coefficients of significant variables simultaneously. This allows us to establish the sampling properties of the resulting estimate. We first establish the rate of convergence of the resulting estimate. With proper choices of penalty functions and regularization parameters, we then establish the asymptotic normality of the resulting estimate, and further demonstrate that the proposed procedures perform as well as an oracle procedure. Semi-parametric generalized likelihood ratio test is proposed to select significant variables in the nonparametric component. We investigate the asymptotic behavior of the proposed test and demonstrate its limiting null distribution follows a chi-squared distribution, which is independent of the nuisance parameters. Extensive Monte Carlo simulation studies are conducted to examine the finite sample performance of the proposed variable selection procedures.

This is joint work with Dr. Runze Li.