

Variable screening with multiple studies and its application in survival analysis

Tianzhou Ma

*University of Maryland School of Public Health
E-mail: tma0929@umd.edu*

Abstract: Advancement in technology has generated abundant high-dimensional data that allows integration of multiple relevant studies. Due to huge computational advantage, variable screening methods based on marginal correlation have become promising alternatives to the popular regularization methods for variable selection. However, all screening methods are limited to single study so far. We consider a general framework for variable screening with multiple related studies, and further propose a novel two-step screening procedure using a self-normalized estimator for high-dimensional regression analysis in this framework. Compared to the one-step and rank-based procedures, our procedure greatly reduces false negative errors while keeping a low false positive rate. Theoretically, we show that our procedure possesses the sure screening property with weaker assumptions on signal strengths and allows the number of features to grow at an exponential rate of the sample size. Simulations and a real transcriptomic application illustrate the advantage of our method. Other than a linear model setting, our proposed framework is readily extensible to Cox model or threshold regression model in survival analysis for high-dimensional variable selection.