

Dimension reduction based linear surrogate variable approach for model free variable selection

Pengjie Dai*

School of Business, Renmin University of China, Beijing, China,
daipengjie@rbs.org.cn

Xiaobo Ding

Academy of Mathematics and Systems Science, Chinese Academy of
Sciences, Beijing, China

Qihua Wang

Academy of Mathematics and Systems Science, Chinese Academy of
Sciences, Beijing, China,
Yunnan University, Kunming, China

variable selection methods are popular to be applied in the problem of high dimensional data sets in the past years. Most of those methods depend on the model assumptions, while sufficient dimension reduction is a nonparametric method to deal with high dimensional data. In this topic, We aim at integrating sufficient dimension reduction into variable selection. A two stage procedure is proposed. First, we obtain dimension reduction directions and integrate them to construct a variable which is linearly dependent on predictors. Then by treating this constructed variable as a new response, we use the traditional variable selection methods such as adaptive LASSO to conduct variable selection. We call such a procedure as dimension reduction based linear surrogate variable (LSV) method. To illustrate that it has wide application, we also apply it to variable selection for the problem of missing responses. Extensive simulation studies show that it is more robust than the variable selection methods depending on model assumptions, and more efficient than the other model-free variable selection methods. Another advantage of the LSV is that it can be easily implemented. We also consider about the situation of missing data, within which LSV still has a good performance. A real example is given to illustrate the proposed method.

Key Words: adaptive LASSO, central subspace, linear surrogate variable, sufficient dimension reduction, variable selection.