Journal of Multivariate Analysis 125 (2014) 89–99

Contents lists available at ScienceDirect

Journal of Multivariate Analysis

journal homepage: www.elsevier.com/locate/jmva

A new sparse variable selection via random-effect model

Youngjo Lee, Hee-Seok Oh *

Department of Statistics, Seoul National University, Seoul 151-747, Republic of Korea

ARTICLE INFO

Article history:
Received 24 December 2010
Available online 17 December 2013

AMS subject classifications:
62J07
62F30

Keywords:
Maximum likelihood estimator
Prediction
Random-effect models
Sparsity
Variable selection

ABSTRACT

We study a new approach to simultaneous variable selection and estimation via random-effect models. Introducing random effects as the solution of a regularization problem is a flexible paradigm and accommodates likelihood interpretation for variable selection. This approach leads to a new type of penalty, unbounded at the origin and provides an oracle estimator without requiring a stringent condition. The unbounded penalty greatly enhances the performance of variable selections, enabling highly accurate estimations, especially in sparse cases. Maximum likelihood estimation is effective in enabling sparse variable selection. We also study an adaptive penalty selection method to maintain a good prediction performance in cases where the variable selection is ineffective.

1. Introduction

Consider the regression model

$$y_i = \boldsymbol{x}_i^T \boldsymbol{\beta} + \varepsilon_i, \quad i = 1, \ldots, n, \tag{1}$$

where $\boldsymbol{\beta}$ is a $d \times 1$ vector of fixed unknown parameters and the ε's are white noises with mean 0 and finite variance ϕ. This study aims to effectively select significant variables, while maintaining good estimation and prediction accuracy.

Many variable selection procedures can be described as penalized least squares (PLS) estimation methods that minimize

$$Q_\lambda(\boldsymbol{\beta}) = \frac{1}{2}\sum_{i=1}^{n}(y_i - \boldsymbol{x}_i^T\boldsymbol{\beta})^2 + \sum_{j=1}^{d} p_\lambda(|\beta_j|), \tag{2}$$

where $p_\lambda(\cdot)$ is a penalty function controlling model complexity. With the entropy or L_0-penalty, namely, $p_\lambda(|\beta_j|) = \lambda I(|\beta_j| \neq 0)$, the PLS becomes

$$\frac{1}{2}\sum_{i=1}^{n}(y_i - \boldsymbol{x}_i^T\boldsymbol{\beta})^2 + \lambda|M|,$$

where $M = \sum_{j=1}^{d} I(|\beta_j| \neq 0)$ denotes the size of the candidate model. This leads to traditional variable selection procedures, which have two fundamental limitations. First, when the number of predictors d is large, it is computationally infeasible to perform subset selection. Second, subset selection is extremely variable because of its inherent discreteness [1,6]. To overcome these difficulties, several other penalties have been proposed. With the L_1-penalty, specifically, the PLS estimator becomes the least absolute shrinkage and selection operator (LASSO), which sets thresholds for predictors with small estimated coefficients [14]. LASSO is a popular technique for simultaneous estimation and variable selection, ensuring high prediction accuracy, and enabling the discovery of relevant predictive variables. Prediction accuracy is often improved by

* Corresponding author.
E-mail addresses: heeseok@stats.snu.ac.kr, heeseok.oh@gmail.com (H.-S. Oh).

0047-259X/$ – see front matter
http://dx.doi.org/10.1016/j.jmva.2013.11.016

shrinking [4] or setting some coefficients to zero by thresholding [2]. Tibshirani [14] gave a comprehensive overview of LASSO as a method of PLS.

LASSO has been criticized on the grounds that it typically ends up selecting a model with too many variables to prevent over shrinkage of the regression coefficients [13]; otherwise, regression coefficients of selected variables are often over shrunk. To improve LASSO, various other penalties have been proposed. Fan and Li [6] proposed the smoothly clipped absolute deviation (SCAD) penalty for oracle estimators. More recently, Zou [16] proposed the adaptive LASSO.

Ridge regression often achieves good prediction performance through a bias–variance trade-off, whereas it cannot produce a parsimonious model. Variable selection is particularly important in the interpretation of the model, especially when the true underlying model has a sparse representation. Identifying null predictors enhances the estimation accuracies of the fitted model. We show that the use of the new penalty greatly improves variable selection to enhance estimation performance, especially in sparse cases. Zou and Hastie [17] noted that the prediction performances of the LASSO can be poor in cases where variable selection is ineffective. To overcome this difficulty, they proposed the elastic net which improves the prediction of LASSO. We propose an adaptive penalty selection for better prediction without hampering the variable selection performance of our method. Through numerical analysis, we show that the proposed adaptive method outperforms LASSO uniformly in variable selection, estimation, and prediction.

Until now, finite penalties, leading to unimodal penalized likelihoods (PL), have been studied. Singularities (unbounded likelihood) have been believed to occur when the description of the process generating the observation is not adequate. This may be considered to be the product of unacceptable probability models [3]. In this paper, we show that the use of singular likelihood (unbounded penalty) at the origin greatly enhances the performance of variable selection. To be specific, by using the unbounded penalty, we define a local minimizer $\hat{\boldsymbol{\beta}}$ that satisfies

$$\frac{\partial Q_\lambda(\boldsymbol{\beta})}{\partial \beta_j}\Big|_{\boldsymbol{\beta}=\hat{\boldsymbol{\beta}}} = 0, \quad \text{for } \hat{\beta}_j \neq 0.$$

Moreover, we investigate the property of $\hat{\boldsymbol{\beta}}$, discuss the practical algorithm for obtaining $\hat{\boldsymbol{\beta}}$, and show its empirical performance through the numerical study.

To achieve these goals, we employ a new random-effect model that generates a family of penalties; the normal-type (bell-shaped), LASSO-type (cusped), and a new (singular) unbounded penalty at the origin. The new unbounded penalty gives an oracle estimators without requiring a stringent condition of Fan and Li [6]. An enhancement of LASSO by Zou [16] could be viewed as the use of an asymptotically unbounded penalty.

There are some analogies between the PL and random-effect model approaches. For example, the iterative weighted least squares (IWLS) estimation for random-effect models can be used for the estimation of $\boldsymbol{\beta}$ [12]. Random-effect models provide new insights and interpretations of IWLS, which explain how the algorithm overcomes the difficulty in nonconvex optimization problem. However, there are differences between the two approaches. In the PL approach, the penalty need not stem from a statistical model; hence, the tuning parameters cannot be estimated by model likelihood. However, in the random-effect model approach, likelihood methods can be used. In this paper, we follow the PL approach for the tuning parameter to compare various methods under a uniform condition, and highlight that the new unbounded penalty is better than the existing penalties in variable selection.

In Section 2, we present a new unbounded penalty. Numerical studies are presented in Section 3. In Section 4, we show that the resulting PLS estimators satisfy the oracle property of Fan and Li [6], but under mild requirements. Conclusion remarks are given in Section 5. In the Appendix, a new random-effect model is introduced, which can be fitted by an IWLS procedure.

2. Unbounded penalty for variable selection

For simplicity of notation, we omit the subscript when deemed unnecessary. Suppose that β is a random variable such that

$$\beta | u \sim N(0, u\theta), \tag{3}$$

where θ is a dispersion parameter, and u follows the gamma distribution with a parameter w such that

$$f_w(u) = (1/w)^{1/w} \frac{1}{\Gamma(1/w)} u^{1/w-1} e^{-u/w}$$

with $\mathrm{E}(u) = 1$ and $\mathrm{Var}(u) = w$. Throughout the paper, $f_\theta(\cdot)$ denote the density function with a parameter θ. The h-likelihood of the above random-effect model leads to a new type of unbounded penalty function $p_\lambda(\beta)$ that is indexed by w and is parameterized by ϕ of (1) and θ of (3). By the result (8) in the Appendix and Stirling's approximation, the resultant unbounded penalty for a given w can be defined as

$$p_\lambda(|\beta|) = \frac{\phi}{2\theta}\frac{\beta^2}{u} + \frac{\phi(w-2)}{w}\log u + \frac{\phi}{w}u,$$

where λ is a function of ϕ and θ, and $u = u(\beta)$ is given by (7) in the Appendix. In this study, we set $\lambda = \phi/\theta$. The detailed procedure for a derivation of $p_\lambda(|\beta|)$ from the random-effect model is given in the Appendix.