COMSTA: 6065 Model 3Gsc pp. 1–11 (col. fig: NIL)

Computational Statistics and Data Analysis xx (xxxx) xxx–xxx

Contents lists available at ScienceDirect

Computational Statistics and Data Analysis

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

A random-effect model approach for group variable selection

Q1 Sangin Lee [a], Yudi Pawitan [b], Youngjo Lee [c,*]

[a] University of Iowa, United States
[b] Karolinska Institutet, Sweden
[c] Seoul National University, Republic of Korea

ARTICLE INFO

Article history:
Received 3 January 2014
Received in revised form 22 October 2014
Accepted 4 February 2015
Available online xxxx

Keywords:
Hierarchical likelihood
LASSO
Model selection
Sparse regression

ABSTRACT

We consider regression models with a group structure in explanatory variables. This structure is commonly seen in practice, but it is only recently realized that taking the information into account in the modeling process may improve both the interpretability and accuracy of the model. In this paper, we study a new approach to group variable selection using random-effect models. Specific distributional assumptions on random effects pertaining to a given structure lead to a new class of penalties that include some existing penalties. We also develop an efficient computational algorithm. Numerical studies are provided to demonstrate better sensitivity and specificity properties without sacrificing the prediction accuracy. Finally, we present some real-data applications of the proposed approach.

1. Introduction

Q2

Variable selection is an important issue in statistical modeling, and many penalized methods have been proposed as tools for variable selection and estimation. Tibshirani (1996) introduced the least-absolute shrinkage and selection operator (LASSO), which performs shrinkage and variable selection simultaneously. It is well known that LASSO tends to select a model with more variables than the true underlying model. To overcome the deficiency of variable selection via the LASSO, various other penalized approaches have been proposed and demonstrated to achieve the oracle property in Fan and Li (2001). These include the smoothly clipped absolute deviation (SCAD) (Fan and Li, 2001), the adaptive LASSO (Zou, 2006), the Bridge (Huang et al., 2008) and the minimax concave penalty (MCP) (Zhang, 2010). Recently, Lee and Oh (2014) introduced an approach to variable selection using random-effect models. The model provides a wider and more flexible class of penalty functions, including the LASSO as a special case and a new unbounded penalty at the origin, and it achieves oracle variable selection without losing prediction accuracy.

In many regression problems, the explanatory variables often possess a natural group structure. For example, (i) categorical factors are often represented by a group of indicator variables, and (ii) to capture flexible functional shapes, continuous factors can be represented by a linear combination of basis functions such as splines or polynomials. In these situations, the problem of selecting relevant variables is that of selecting groups rather than selecting individual variables. Depending on the situation, the individual variables in a group may or may not be meaningful scientifically. If they are not, we are typically not interested in selecting individual variables and the interest is limited to *group selection*. In fact, most recent papers considered only this selection problem; see, e.g. Yuan and Lin (2006), Wang et al. (2007), Wang and Leng (2008) and Huang et al. (2012). However, if the individual variables are meaningful, then we would be interested in selecting individual variables within each selected group; we refer to this as *bi-level selection* (Breheny and Huang, 2009; Huang et al., 2012).

* Corresponding author.
E-mail address: youngjo@snu.ac.kr (Y. Lee).

http://dx.doi.org/10.1016/j.csda.2015.02.020

In this paper, we introduce how the group and bi-level selections can be achieved using the random-effect model approach. Specific distributional assumptions reflecting a given structure on the random effects produce a flexible class of penalties that includes, for example, the group LASSO and group Bridge as special cases. First, we study random-effect models for group-only selection. Second, we propose random-effect models for bi-level selection, which enable simultaneous selection at both the group level and the variable level within each selected group. Unlike the existing penalized approaches, the use of random effects makes it obvious how sparseness at group and variable levels is achieved. In addition, we show some applications of the proposed random-effect model approaches.

The rest of the paper is organized as follows. We review the penalty approach and describe the random-effect model approach for structured variable selection in Section 2. Section 3 describes the implied penalties derived from the random-effect models and the computational algorithms to evaluate the proposed estimate. In Sections 4 and 5, we show the results of the simulated and real data analysis, respectively. Concluding remarks are given in Section 6.

2. Methods for group variable selection

Suppose that the explanatory variables can be divided into K groups, and the outcome $\boldsymbol{y} = (y_1, \ldots, y_n)^T$ has mean $\boldsymbol{\mu} = (\mu_1, \ldots, \mu_n)^T$, which follows a generalized linear model (GLM) with link function $\eta_i \equiv h(\mu_i)$, such that we have linear predictor

$$\boldsymbol{\eta} = \mathbf{X}\boldsymbol{\beta} = \mathbf{X}_1\boldsymbol{\beta}_1 + \cdots + \mathbf{X}_K\boldsymbol{\beta}_K, \tag{1}$$

where $\boldsymbol{\eta} = (\eta_1, \ldots, \eta_n)^T \in \mathbb{R}^n$ is the vector of linear predictors; $\mathbf{X} = (\mathbf{X}_1, \ldots, \mathbf{X}_K) \in \mathbb{R}^{n\times p}$ is the design matrix of the predictors; $\boldsymbol{\beta} = (\boldsymbol{\beta}_1, \ldots, \boldsymbol{\beta}_K)^T \in \mathbb{R}^p$ is the vector of regression coefficients. Here, $\mathbf{X}_k \in \mathbb{R}^{n\times p_k}$ and $\boldsymbol{\beta}_k \in \mathbb{R}^{p_k}$ are the corresponding design matrix and vector of coefficients for the kth group, respectively. Note that with this GLM specification our method immediately works with various types of data, including continuous, binary and count data.

2.1. Penalty approach

We first review penalized maximum likelihood approaches to group variable selection, which can be achieved by maximizing

$$Q_\lambda(\boldsymbol{\beta}) = \ell(\boldsymbol{\beta}) - \sum_{k=1}^{K} J_{\lambda_k}(\|\boldsymbol{\beta}_k\|_2), \tag{2}$$

where $\ell(\boldsymbol{\beta}) = \sum_{i=1}^n \log f_\phi(y_i|\boldsymbol{\beta})$ is the corresponding log-likelihood, with $f_\phi(y_i|\boldsymbol{\beta})$ being the density function with the dispersion parameter ϕ, $\lambda_k > 0$ is the regularization parameter of the kth group and $\|\cdot\|_2$ stands for the ℓ_2-norm. Yuan and Lin (2006) proposed the group LASSO, using penalty function $J_{\lambda_k}(t) = \lambda_k t,\ t > 0$. To adjust for different group sizes they chose $\lambda_k \equiv \lambda\sqrt{p_k}$, with $\lambda > 0$. The group LASSO can be constructed by applying the LASSO penalty to the ℓ_2-norm of sub-coefficients within each group. For other methods, Wang et al. (2007) and Huang et al. (2012) proposed the group SCAD and group MCP with penalty function $J_{\lambda_k}(\cdot)$ associated with the SCAD and MCP penalty functions, respectively. These penalties enforce sparsity at group level only, but are incapable of selecting important variables within the selected groups.

To achieve bi-level selection, Huang et al. (2012) introduced the following ℓ_1-norm criterion

$$Q_\lambda(\boldsymbol{\beta}) = \ell(\boldsymbol{\beta}) - \sum_{k=1}^{K} J_{\lambda_k}(\|\boldsymbol{\beta}_k\|_1), \tag{3}$$

where $\|\cdot\|_1$ stands for the ℓ_1-norm. The so-called group Bridge of Huang et al. (2009) is obtained by using $J_{\lambda_k}(t) = \lambda_k t^\nu,\ t > 0$ for $\lambda_k = \lambda p_k^{1-\nu}$. For later comparisons with other methods we use a commonly suggested value $\nu = 0.5$. Another examples are the ℓ_1-norm group SCAD and MCP proposed by Huang et al. (2012), where $J_{\lambda_k}(\cdot)$ are SCAD and MCP functions, respectively.

Another approach for bi-level selection is to combine both the group penalty and individual variable penalty. As a special case, Friedman et al. (2010) proposed the sparse group LASSO, which is defined as a maximizer of

$$Q_\lambda(\boldsymbol{\beta}) = \ell(\boldsymbol{\beta}) - \lambda_1 \sum_{k=1}^{K} \sqrt{p_k}\|\boldsymbol{\beta}_k\|_2 - \lambda_2 \sum_{k=1}^{K}\sum_{j=1}^{p_k} |\beta_{kj}|, \tag{4}$$

where $\lambda_1 \geq 0$ and $\lambda_2 \geq 0$ are regularization parameters. The group LASSO penalty can be replaced with any group penalties in (2), and similarly the LASSO penalty for variable selection can be replaced with any penalties such as SCAD, MCP and Bridge.

Please cite this article in press as: Lee, S., et al., A random-effect model approach for group variable selection. Computational Statistics and Data Analysis (2015), http://dx.doi.org/10.1016/j.csda.2015.02.020