Journal of Statistical Planning and Inference 139 (2009) 2665 – 2674

Contents lists available at ScienceDirect

Journal of Statistical Planning and Inference

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

Bayesian variable selection using an adaptive powered correlation prior

Arun Krishna, Howard D. Bondell*, Sujit K. Ghosh

Department of Statistics, North Carolina State University, Raleigh, NC 27695-8203, USA

ARTICLE INFO

Article history:
Received 10 March 2008
Received in revised form
11 December 2008
Accepted 11 December 2008
Available online 24 December 2008

Keywords:
Bayesian variable selection
Collinearity
Powered correlation prior
Zellner's *g*-prior

ABSTRACT

The problem of selecting the correct subset of predictors within a linear model has received much attention in recent literature. Within the Bayesian framework, a popular choice of prior has been Zellner's *g*-prior which is based on the inverse of empirical covariance matrix of the predictors. An extension of the Zellner's prior is proposed in this article which allow for a power parameter on the empirical covariance of the predictors. The power parameter helps control the degree to which correlated predictors are smoothed towards or away from one another. In addition, the empirical covariance of the predictors is used to obtain suitable priors over model space. In this manner, the power parameter also helps to determine whether models containing highly collinear predictors are preferred or avoided. The proposed power parameter can be chosen via an empirical Bayes method which leads to a data adaptive choice of prior. Simulation studies and a real data example are presented to show how the power parameter is well determined from the degree of cross-correlation within predictors. The proposed modification compares favorably to the standard use of Zellner's prior and an intrinsic prior in these examples.

1. Introduction

Consider the linear regression model with n independent observations and let $\boldsymbol{y} = (y_1, \ldots, y_n)'$ be the vector of response variables. The canonical linear model can be written as

$$\boldsymbol{y} = \boldsymbol{X}\boldsymbol{\beta} + \epsilon, \tag{1.1}$$

where $\boldsymbol{X} = (\boldsymbol{x}_1, \ldots, \boldsymbol{x}_p)$ is an $n \times p$ matrix of explanatory variables with $\boldsymbol{x}_j = (x_{1j}, \ldots, x_{nj})'$ for $j = 1, \ldots, p$. Let $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_p)'$ be the corresponding vector of unknown regression parameters, and $\epsilon \sim \mathrm{N}(0, \sigma^2 \boldsymbol{I})$. Throughout the paper, we assume $\boldsymbol{y}$ to be empirically centered to have mean zero, while the columns of $\boldsymbol{X}$ have been standardized to have mean zero and norm one, so $\boldsymbol{X}'\boldsymbol{X}$ will be the empirical correlation matrix.

Under the above regression model, it is assumed that only an unknown subset of the coefficients are non-zero, so that the variable selection problem is to identify this unknown subset. Bayesian approaches to the problem of selecting variables/predictors within a linear regression framework has received considerable attention over the years, for example see, Mitchell and Beauchamp (1988), Geweke (1996), George and McCulloch (1993, 1997), Brown et al. (1998), George (2000), Chipman et al. (2001) and Casella and Moreno (2006).

* Corresponding author. Tel.: +1 919 515 1914; fax: +1 919 515 1169.
E-mail address: bondell@stat.ncsu.edu (H.D. Bondell).

0378-3758/$ - see front matter © 2009 Elsevier B.V. All rights reserved.
doi:10.1016/j.jspi.2008.12.004

For the linear model, Zellner (1986) suggested a particular form of a conjugate normal-Gamma family called the g-prior which can be expressed as

$$\boldsymbol{\beta}|\sigma^2, \boldsymbol{X} \sim \mathrm{N}\left(0, \frac{\sigma^2}{g}(\boldsymbol{X}'\boldsymbol{X})^{-1}\right),$$
$$\sigma^2 \sim IG(a_0, b_0), \tag{1.2}$$

where $g > 0$ is a known scaling factor and $a_0 > 0$, $b_0 > 0$ are known parameters of the inverse Gamma distribution with mean $a_0/(b_0 - 1)$. The prior covariance matrix of $\boldsymbol{\beta}$ is the scalar multiple σ^2/g of the inverse Fisher information matrix, which concurrently depends on the observed data through the design matrix $\boldsymbol{X}$.

This particular prior has been widely adopted in the context of Bayesian variable selection due to its closed form calculations of all marginal likelihoods which is suitable for rapid computations over a large number of submodels, and its simple interpretation that it can be derived from the idea of a likelihood for a pseudo-dataset with the same design matrix $\boldsymbol{X}$ as the observed sample (see, Zellner, 1986; George and Foster, 2000; Smith and Kohn, 1996; Fernandez et al., 2001).

In this paper, we point out a drawback of using Zellner's prior on $\boldsymbol{\beta}$ particularly when the predictors ($\boldsymbol{x}_j$) are highly correlated. The conditional variance of $\boldsymbol{\beta}$ given σ^2 and $\boldsymbol{X}$ is based on the inverse of the empirical correlation of predictors and puts most of its prior mass in the direction that causes the regression coefficients of correlated predictors to be smoothed away from each other. So when coupled with model selection, Zellner's prior discourages highly collinear predictors to enter the models simultaneously by inducing a negative correlation between the coefficients.

We propose a modification of Zellner's g-prior by replacing $(\boldsymbol{X}'\boldsymbol{X})^{-1}$ by $(\boldsymbol{X}'\boldsymbol{X})^{\lambda}$ where the power $\lambda \in \mathbb{R}$, controls the amount of smoothing of collinear predictors towards or away from each other accordingly as $\lambda > 0$ or $\lambda < 0$, respectively. For $\lambda > 0$, the new conditional prior variance of $\boldsymbol{\beta}$ puts more prior mass in the direction that corresponds to a strong prior smoothing of regression coefficients of highly collinear predictors towards each other. Therefore, by choosing $\lambda > 0$ our proposed modification in contrast, forces highly collinear predictors entering or exiting the model simultaneously (see Section 2). Hence, the use of the power hyperparameter λ to the empirical correlation matrix helps us to determine whether models with high collinear predictors are preferred or not.

The hyperparameter λ is further incorporated into the prior probabilities over model space with the same intentions of encouraging or discouraging the inclusion of groups of correlated predictors. The choice of hyperparameter is obtained via an empirical Bayes approach and the inference regarding model selection is then made based on the posterior probabilities. By allowing the power parameter λ to be chosen by the data, we let the data decide whether to include collinear predictors or not.

The remainder of the paper is structured as follows. In Section 2, we describe in detail the powered correlation prior and provide a simple motivating example, when $p = 2$. Section 3, describes the choice of new prior specifications for model selection. The Bayesian hierarchical model and the calculation of posterior probabilities are presented in Section 4. The superior performance of using the powered correlation prior over Zellner's g-priors is illustrated with the help of simulation studies and real data examples in Section 5. Finally, in Section 6 we conclude with a discussion.

2. The adaptive powered correlated prior

Consider again a normal regression model as in (1.1), where $\boldsymbol{X}'\boldsymbol{X}$ represents the correlation matrix. Let $\boldsymbol{X}'\boldsymbol{X} = \boldsymbol{\Gamma}\boldsymbol{D}\boldsymbol{\Gamma}'$ be the spectral decomposition, where the columns of $\boldsymbol{\Gamma}$ are the p orthonormal eigenvectors and $\boldsymbol{D}$ is the diagonal matrix with eigenvalues $d_1 \geqslant \cdots \geqslant d_p \geqslant 0$ as the diagonal entries. The powered correlation prior for $\boldsymbol{\beta}$ conditioned on σ^2 and $\boldsymbol{X}$ is defined as

$$\boldsymbol{\beta}|\sigma^2, \boldsymbol{X} \sim \mathrm{N}\left(0, \frac{\sigma^2}{g}(\boldsymbol{X}'\boldsymbol{X})^{\lambda}\right), \tag{2.1}$$

where $(\boldsymbol{X}'\boldsymbol{X})^{\lambda} = \boldsymbol{\Gamma}\boldsymbol{D}^{\lambda}\boldsymbol{\Gamma}'$, with $g > 0$ and $\lambda \in \mathbb{R}$ controlling the strength and the shape, respectively, of the prior covariance matrix, for a given $\sigma^2 > 0$.

There are several priors which are special cases of the powered correlation prior. For instance, $\lambda = -1$ produces the Zellner's g-prior (1.2). By setting $\lambda = 0$ we have $(\boldsymbol{X}'\boldsymbol{X})^0 = I$ which gives us the ridge regression model of Hoerl and Kennard (1970), under this model β_j are given independent $\mathrm{N}(0, \sigma^2/g)$ priors. Next we illustrate how λ controls the model's response to collinearity which is the main motivation for using the powered correlation prior.

Let $\boldsymbol{T} = \boldsymbol{X}\boldsymbol{\Gamma}$ and $\boldsymbol{\theta} = \boldsymbol{\Gamma}'\boldsymbol{\beta}$. The linear model can be written in terms of the principal components as

$$\boldsymbol{y} \sim \mathrm{N}(\boldsymbol{T}\boldsymbol{\theta}, \sigma^2) \quad \text{with } \boldsymbol{\theta} \sim \mathrm{N}\left(0, \frac{\sigma^2}{g}\boldsymbol{D}^{\lambda}\right). \tag{2.2}$$

The columns of the new design matrix $\boldsymbol{T}$ are the principal components, and so the original prior on $\boldsymbol{\beta}$ can be viewed as independent mean zero normal priors on the principal component regression coefficients, with prior variance proportional to the power of the corresponding eigenvalues, $d_1^{\lambda}, \ldots, d_p^{\lambda}$. Principal components with d_i near zero indicate a presence of a near-linear relationship