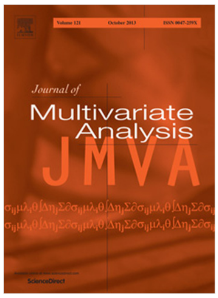
Accepted Manuscript



High-dimensional multivariate posterior consistency under global-local shrinkage priors

Ray Bai, Malay Ghosh

PII: S0047-259X(17)30690-5
DOI: https://doi.org/10.1016/j.jmva.2018.04.010
Reference: YJMVA 4357

To appear in: *Journal of Multivariate Analysis*

Received date : 21 November 2017

High-dimensional multivariate posterior consistency under global-local shrinkage priors

Ray Bai[a,*], Malay Ghosh[a]

[a]*Department of Statistics, University of Florida, Gainesville, FL 32611, USA*

Abstract

We consider sparse Bayesian estimation in the classical multivariate linear regression model with p regressors and q response variables. In univariate Bayesian linear regression with a single response y, shrinkage priors which can be expressed as scale mixtures of normalnormal densities are popular for obtaining sparse estimates of the coefficients. In this paper, we extend the use of these priors to the multivariate case to estimate a $p \times q$ coefficients matrix $\mathbf{B}$. We derive sufficient conditions for posterior consistency under the Bayesian multivariate linear regression framework and prove that our method achieves posterior consistency even when $p > n$ and even when p grows at nearly exponential rate with the sample size. We derive an efficient Gibbs sampling algorithm and provide the implementation in a comprehensive R package called `MBSP`. Finally, we demonstrate through simulations and data analysis that our model has excellent finite sample performance.

Keywords: Heavy tail, High-dimensional data, Posterior consistency, Shrinkage estimation, Sparsity, Variable selection

1. Introduction

1.1. Background

We consider the classical multivariate normal linear regression model, viz.

$$\mathbf{Y} = \mathbf{X}\mathbf{B} + \mathbf{E}, \tag{1}$$

where $\mathbf{Y} = (y_1, \ldots, y_q)$ is an $n \times q$ response matrix of n samples and q continuous response variables, $\mathbf{X}$ is an $n \times p$ matrix of n samples and p covariates, $\mathbf{B} \in \mathbb{R}^{p \times q}$ is the coefficient matrix, and $\mathbf{E} = (\varepsilon_1, \ldots, \varepsilon_n)^\top$ is an $n \times q$ noise matrix. Under normality, we assume that $\varepsilon_1, \ldots, \varepsilon_n$ are iid $\mathcal{N}_q(\mathbf{0}, \mathbf{\Sigma})$. In other words, each row of $\mathbf{E}$ is identically distributed with mean $\mathbf{0}$ and covariance $\mathbf{\Sigma}$. Throughout this paper, we also assume that $\mathbf{Y}$ and $\mathbf{X}$ are centered so there is no intercept term in $\mathbf{B}$.

Our focus is on sparse Bayesian estimation and variable selection on the coefficients matrix $\mathbf{B}$ in (1). In practical settings, particularly in high-dimensional settings when $p > n$, it is important not only to provide robust estimates of $\mathbf{B}$, but to choose a subset of regressor variables from the p rows of $\mathbf{B}$ which are good for prediction on the q responses. Although p may be large, the number of predictors that are actually associated with the responses is generally quite small. A parsimonious model also tends to give far better estimation and prediction performance than a dense model, which further motivates the need for sparse estimates of $\mathbf{B}$.

In frequentist approaches to univariate regression, the most commonly used method for inducing sparsity is through imposing regularization penalties on the coefficients of interest. Popular choices of penalty functions include the lasso [40] and its many variants, e.g., [38, 47, 50, 51]. Many of these penalized regression methods include

*Corresponding author.
Email address: raybai07@ufl.edu (Ray Bai)