

Multivariate Bayesian Variable Selection and Prediction

Caleb Jin

2-10-2019

Contents

1	Foreword	5
2	Multivariate Bayesian Variable Selection and Prediction	7

Chapter 1

Foreword

I am Caleb Jin. After I read this paper, **Multivariate Bayesian variable selection and prediction** (Brown and Vannucci, 1998), I write down the nodes of the key idea and R code to realize it.

Chapter 2

Multivariate Bayesian Variable Selection and Prediction

Consider a multivariate variable regression model,

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

where $\mathbf{Y} = (\mathbf{y}_1, \mathbf{y}_2, \dots, \mathbf{y}_n)^\top$ is the $n \times q$ response matrix, $\mathbf{X} = (\mathbf{x}_1, \dots, \mathbf{x}_p)$ is the $n \times p$ design matrix, $\mathbf{B} = (\boldsymbol{\beta}_1, \dots, \boldsymbol{\beta}_p)^\top$ is a $p \times q$ unknown coefficient matrix, and $\mathbf{E} = (\mathbf{e}_1, \mathbf{e}_2, \dots, \mathbf{e}_n)^\top$ is an $n \times q$ error matrix with $\mathbf{e}_i \stackrel{i.i.d}{\sim} \mathcal{N}_q(\mathbf{0}, \boldsymbol{\Sigma})$. To identify the variables selected among p regressors, (Brown and Vannucci, 1998) introduced a latent binary vector $\boldsymbol{\xi} = (\xi_1, \dots, \xi_p)^\top$ with

$$\xi_j = \begin{cases} 1 & \text{if } \boldsymbol{\beta}_j \text{ is active} \\ 0 & \text{if } \boldsymbol{\beta}_j \text{ is not active} \end{cases}.$$

The conjugate priors are considered as follows:

$$\begin{aligned} \pi(\mathbf{B}_{\boldsymbol{\xi}} | \boldsymbol{\Sigma}, \boldsymbol{\xi}) &\sim \mathcal{MN}(\mathbf{0}, \mathbf{H}_{\boldsymbol{\xi}}, \boldsymbol{\Sigma}) \\ \pi(\boldsymbol{\Sigma}) &\sim \mathcal{IW}(\mathbf{Q}, \delta) \\ \pi(\boldsymbol{\xi}) &\stackrel{ind}{\sim} \prod_{j=1}^p \text{Ber}(\omega_j), \end{aligned}$$

where \mathbf{Q} , \mathbf{H}_ξ and δ are hyperparameters. Then, by Bayesian theorem, the marginal posterior distribution can be obtained by

$$\begin{aligned}
& m(\xi|\mathbf{Y}) \propto m(\mathbf{Y}|\xi)\pi(\xi) \\
&= \iint f(\mathbf{Y}|\mathbf{B}_\xi, \Sigma)\pi(\mathbf{B}_\xi|\Sigma, \xi)\pi(\Sigma)d\mathbf{B}_\xi d\Sigma\pi(\xi) \\
&= \iint (2\pi)^{-\frac{nq}{2}}|\Sigma|^{-\frac{n}{2}}\exp\left[-\frac{1}{2}\text{tr}\{(\mathbf{Y} - \mathbf{X}_\xi\mathbf{B}_\xi)^\top(\mathbf{Y} - \mathbf{X}_\xi\mathbf{B}_\xi)\}\Sigma^{-1}\right] \times \\
&\quad (2\pi)^{-\frac{|\xi|q}{2}}|\Sigma|^{-\frac{|\xi|}{2}}|\mathbf{H}_\xi|^{-\frac{q}{2}}\exp\left[\frac{1}{2}\text{tr}\{\mathbf{B}_\xi^\top\mathbf{H}_\xi^{-1}\mathbf{B}_\xi\Sigma^{-1}\}\right] \times \\
&\quad |\mathbf{Q}|^{\frac{\delta}{2}}2^{-\frac{1}{2}\delta q}\xi_q^{-1}\left(\frac{\delta}{2}\right)|\Sigma|^{-\frac{\delta+q+1}{2}}\exp\{-\frac{1}{2}\text{tr}(\mathbf{Q}\Sigma^{-1})\}d\mathbf{B}_\xi d\Sigma\pi(\xi) \\
&\propto (2\pi)^{-\frac{|\xi|q}{2}}|\mathbf{H}_\xi|^{-\frac{q}{2}}\iint |\Sigma|^{-\frac{1}{2}(n+|\xi|+\delta+q+1)}\exp\{-\frac{1}{2}\text{tr}(\mathbf{Q}\Sigma^{-1})\} \\
&\quad \exp\left[-\frac{1}{2}\text{tr}\left\{\left(\mathbf{B}_\xi^\top(\mathbf{X}_\xi^\top\mathbf{X}_\xi + \mathbf{H}_\xi^{-1})\mathbf{B}_\xi - 2\mathbf{B}_\xi^\top\mathbf{X}_\xi^\top\mathbf{Y}\right)\Sigma^{-1}\right\}\right] \times \\
&\quad \exp\left[-\frac{1}{2}\text{tr}\left\{\mathbf{Y}^\top\mathbf{Y}\Sigma^{-1}\right\}\right]d\mathbf{B}_\xi d\Sigma\pi(\xi) \\
&= (2\pi)^{-\frac{|\xi|q}{2}}|\mathbf{H}_\xi|^{-\frac{q}{2}}\iint |\Sigma|^{-\frac{1}{2}(n+|\xi|+\delta+q+1)}\exp\{-\frac{1}{2}\text{tr}[(\mathbf{Y}^\top\mathbf{Y} + \mathbf{Q})\Sigma^{-1}]\} \times \\
&\quad \exp\left[-\frac{1}{2}\text{tr}\left\{\left(\mathbf{B}_\xi^\top\mathbf{K}_\xi\mathbf{B}_\xi - 2\mathbf{B}_\xi^\top\mathbf{K}_\xi\mathbf{K}_\xi^{-1}\mathbf{X}_\xi^\top\mathbf{Y}\right)\Sigma^{-1}\right\}\right]d\mathbf{B}_\xi d\Sigma\pi(\xi) \\
&= (2\pi)^{-\frac{|\xi|q}{2}}|\mathbf{H}_\xi|^{-\frac{q}{2}}\iint |\Sigma|^{-\frac{1}{2}(n+|\xi|+\delta+q+1)} \times \\
&\quad \exp\{-\frac{1}{2}\text{tr}[(\mathbf{Y}^\top\mathbf{Y} - \tilde{\mathbf{B}}_\xi^\top\mathbf{K}_\xi\tilde{\mathbf{B}}_\xi + \mathbf{Q})\Sigma^{-1}]\} \times \\
&\quad \exp\left[-\frac{1}{2}\text{tr}\left\{((\mathbf{B}_\xi - \tilde{\mathbf{B}}_\xi)^\top\mathbf{K}_\xi(\mathbf{B}_\xi - \tilde{\mathbf{B}}_\xi))\Sigma^{-1}\right\}\right]d\mathbf{B}_\xi d\Sigma\pi(\xi) \\
&= (2\pi)^{-\frac{|\xi|q}{2}}|\mathbf{H}_\xi|^{-\frac{q}{2}}(2\pi)^{\frac{|\xi|q}{2}}|\mathbf{K}_\xi|^{-\frac{q}{2}} \times \\
&\quad \int |\Sigma|^{\frac{|\xi|}{2}}|\Sigma|^{-\frac{1}{2}(n+|\xi|+\delta+q+1)}\exp\{-\frac{1}{2}\text{tr}[(\mathbf{Y}^\top\mathbf{Y} - \tilde{\mathbf{B}}_\xi^\top\mathbf{K}_\xi\tilde{\mathbf{B}}_\xi + \mathbf{Q})\Sigma^{-1}]\}d\Sigma\pi(\xi) \\
&= |\mathbf{H}_\xi|^{-\frac{q}{2}}|\mathbf{K}_\xi|^{-\frac{q}{2}}\int |\Sigma|^{-\frac{1}{2}(n+\delta+q+1)} \times \\
&\quad \exp\{-\frac{1}{2}\text{tr}[(\mathbf{Y}^\top\mathbf{Y} - \mathbf{Y}^\top\mathbf{X}_\xi\mathbf{K}_\xi^{-1}\mathbf{X}_\xi^\top\mathbf{Y} + \mathbf{Q})\Sigma^{-1}]\}d\Sigma\pi(\xi) \\
&= |\mathbf{H}_\xi\mathbf{K}_\xi|^{-\frac{q}{2}}\int |\Sigma|^{-\frac{1}{2}(n+\delta+q+1)} \times \\
&\quad \exp\{-\frac{1}{2}\text{tr}[(\mathbf{Y}^\top(\mathbf{I}_n - \mathbf{X}_\xi\mathbf{K}_\xi^{-1}\mathbf{X}_\xi^\top)\mathbf{Y} + \mathbf{Q})\Sigma^{-1}]\}d\Sigma\pi(\xi) \\
&\propto |\mathbf{H}_\xi\mathbf{K}_\xi|^{-\frac{q}{2}}|\mathbf{Y}^\top(\mathbf{I}_n - \mathbf{X}_\xi\mathbf{K}_\xi^{-1}\mathbf{X}_\xi^\top)\mathbf{Y} + \mathbf{Q}|^{-\frac{n+\delta}{2}}\pi(\xi) \\
&\equiv g(\xi|\mathbf{Y}),
\end{aligned}$$

where $\mathbf{K}_\xi = \mathbf{X}_\xi^\top \mathbf{X}_\xi + \mathbf{H}_\xi^{-1}$ and $g(\xi|\mathbf{Y})$ is the proportional form of $\pi(\xi|\mathbf{Y})$. (Brown and Vannucci, 1998) set $\mathbf{Q} = k\mathbf{I}_q$ and $\mathbf{H}_\xi = c(\mathbf{X}_\xi^\top \mathbf{X}_\xi)^{-1}$ and $\omega_j = \omega$, then $\pi(\xi) = \prod_{j=1}^p \omega_j^{\xi_j} (1 - \omega_j)^{1-\xi_j} = \omega^{|\xi|} (1 - \omega)^{p-|\xi|}$. Apply $\mathbf{Q} = k\mathbf{I}_q$ and $\mathbf{H}_\xi = c(\mathbf{X}_\xi^\top \mathbf{X}_\xi)^{-1}$ into (2.1), hence we have

$$\begin{aligned} \log g(\xi|\mathbf{Y}) &= -\frac{n+\delta}{2} \log |k\mathbf{I}_q + \mathbf{Y}^\top \mathbf{Y}| - \frac{c}{c+1} \mathbf{Y}^\top \mathbf{X}_\xi (\mathbf{X}_\xi^\top \mathbf{X}_\xi^{-1}) \mathbf{X}_\xi^\top \mathbf{Y} - \\ &\quad \frac{q|\xi|}{2} \log(c+1) + |\xi| \log \omega + (p - |\xi|) \log(1 - \omega). \end{aligned}$$

Exhaustive computation of $g(\xi|\mathbf{Y})$ for 2^p values of ξ becomes prohibitive even in a super computer when p is larger than 40. In this circumstances, using MCMC sampling to explore marginal posterior is possible. (Brown and Vannucci, 1998) use Gibbs sampler to generate each ξ value component-wise from full conditional distributions $\pi(\xi_j|\xi_{-j}, \mathbf{Y})$, where $\xi_{-j} = \xi \setminus \xi_j$. It is easy to show that

$$\pi(\xi_j = 1|\xi_{-j}, \mathbf{Y}) = \frac{\theta_j}{\theta_j + 1},$$

where

$$\theta_j = \frac{g(\xi_j = 1, \xi_{-j}|\mathbf{Y})}{g(\xi_j = 0, \xi_{-j}|\mathbf{Y})}.$$

The complete algorithm of (Brown and Vannucci, 1998) is described in Algorithm 3:

Algorithm 3 Brown: MCMC sampling for ξ

1. Initialize $\xi = (\xi_1, \xi_2, \dots, \xi_p)$.
 2. **Repeat** for $t = 1, 2, \dots$ until convergence
 - 1). $a \leftarrow g(\xi_1^{(t)} = 1, \xi_{-1}^{(t)}|\mathbf{Y})$ and $b \leftarrow g(\xi_1^{(t)} = 0, \xi_{-1}^{(t)}|\mathbf{Y})$.
 - 2). $\theta_1 \leftarrow \frac{a}{b}$; $p_1 \leftarrow \frac{\theta_1}{\theta_1 + 1}$.
 - 3). $\xi^* \sim \text{Ber}(p_1)$; $\xi_1^{(t)} \leftarrow \xi^*$.
 - 4). **Repeat** for $j = 2, \dots, p$:
 - i) $\kappa = a\xi^* + b(1 - \xi^*)$; $\tilde{\xi} \leftarrow \xi^{(t)}$.
 - ii) **If** $\xi_j^{(t)} = 1$,

$$\tilde{\xi}_j \leftarrow 0; a \leftarrow \kappa; b \leftarrow g(\tilde{\xi}|\mathbf{Y}).$$
 - else**

$$\tilde{\xi}_j \leftarrow 1; a \leftarrow g(\tilde{\xi}|\mathbf{Y}); b \leftarrow \kappa.$$
 - iii) $\theta_j \leftarrow \frac{a}{b}$; $p_j \leftarrow \frac{\theta_j}{\theta_j + 1}$.
 - iv) $\xi^* \sim \text{Ber}(p_j)$; $\xi_j^{(t)} \leftarrow \xi^*$.
 - 5). Get $\xi^{(t)}$
-

For the hyperparameter, we follow the same setting with (Brown and Vannucci, 1998), which are $\delta = 2 + q$, $\omega = 20/p$, $k = 2$ and $c = 10$. Note that since our definition of inverse-wishart distribution is different from that of (Brown and Vannucci, 1998), the values of q and k we set here are after adjustment to make them consistent with (Brown and Vannucci, 1998).

Bibliography

Brown, P. J. and Vannucci, M. (1998). Multivariate bayesian variable selection and prediction. *Journal of the Royal Statistical Society: Series B (Methodological)*, pages 627–641.