Supplementary material to Bayesian variable selection for disease classification using gene expression data

Yang Aijun and Song Xinyuan

ajyang81@gmail.com; xysong@sta.cuhk.edu.hk

Department of Statistics, The Chinese University of Hong Kong

1 Method

(i) Proof of equation (8).

Since the prior distributions for α, β_γ and γ are

$$\alpha \sim N(0,h), \quad \beta_\gamma|\gamma \sim N(0, c(\mathbf{X}_\gamma^{'}\mathbf{X}_\gamma)^{+}), \quad \gamma_i \sim {\pi_i}^{\gamma_i}(1-\pi_i)^{1-\gamma_i}, \tag{A 1}$$

and conditional on parameters α, β_γ, and γ,

$$Z_i = \alpha + X_{i,\gamma}\beta_\gamma + \varepsilon_i, \qquad i = 1, \cdots, n, \tag{A 2}$$

we have

$$Z_i|Y, \mathbf{X}, \alpha, \beta_\gamma, \gamma \sim N(\alpha + X_{i,\gamma}\beta_\gamma, 1)\mathrm{I}(A_i), \tag{A 3}$$

where A_i is either equal to $\{Z_i : Z_i > 0\}$ or $\{Z_i : Z_i \leq 0\}$ corresponding to $Y_i = 1$ or $Y_i = 0$, respectively; and $\mathrm{I}(\cdot)$ is an indicator function which truncates the univariate normal distribution of Z_i to the appropriate region.

The joint posterior distribution of $(Z, \alpha, \beta_\gamma, \gamma)$ given $(Y, \mathbf{X})$ is

$$\begin{aligned} p(Z, \alpha, \beta_\gamma, \gamma|Y, \mathbf{X}) &\propto \prod_{i=1}^{n} p(Z_i|Y, \mathbf{X}, \alpha, \beta_\gamma, \gamma) p(\alpha) p(\beta_\gamma|\mathbf{X}, \gamma) \prod_{i=1}^{p} p(\gamma_i) \\ &\propto \left[\exp\{-\frac{\sum_{i=1}^{n}(Z_i - \alpha - X_{i,\gamma}\beta_\gamma)^2}{2}\} \prod_{i=1}^{n} \mathrm{I}(A_i) \right] \times \exp(-\frac{\alpha^2}{2h}) \\ &\times \left[\exp(-\frac{\beta_\gamma^{'}\mathbf{X}_\gamma^{'}\mathbf{X}_\gamma\beta_\gamma}{2c}) \prod_{i=1}^{m_\gamma} \lambda_i^{-\frac{1}{2}} \right] \times \prod_{i=1}^{p} \pi_i^{\gamma_i}(1-\pi_i)^{1-\gamma_i}, \end{aligned} \tag{A 4}$$

where $\lambda_1, \cdots, \lambda_{m_\gamma}$ $(m_\gamma \leq p_\gamma)$ are the nonzero eigenvalues of $(\mathbf{X}_\gamma^{'}\mathbf{X}_\gamma)^{+}$. We first integrate out α given Z, β_γ, γ. The exponentiated terms that are associated with α in above equation can be rewritten as follows:

$$\begin{aligned} -\frac{\sum_{i=1}^{n}(Z_i - \alpha - X_{i,\gamma}\beta_\gamma)^2}{2} - \frac{\alpha^2}{2h} &= -\frac{(Z - 1\alpha - \mathbf{X}_\gamma\beta_\gamma)'(Z - 1\alpha - \mathbf{X}_\gamma\beta_\gamma)}{2} - \frac{\alpha^2}{2h} \\ = -\frac{(h^{-1}+n)\{\alpha - (h^{-1}+n)^{-1}1'(Z - \mathbf{X}_\gamma\beta_\gamma)\}^2}{2} &- \frac{(1+nh)^{-1}(Z - \mathbf{X}_\gamma\beta_\gamma)'(Z - \mathbf{X}_\gamma\beta_\gamma)}{2}. \end{aligned} \tag{A 5}$$

The exponential of the first term in expression (A 5) forms the kernel of a Gaussian probability density of α and can be integrated out. Thus, the integration of α is done.

Using a special case of binomial inverse theorem (see Woodbury 1950; Plackeett, 1950), the second term of expression (A 5) can be expressed as

$$-\frac{(Z-\mathbf{X}_\gamma\beta_\gamma)'(\mathbf{I}_n+h11')^{-1}(Z-\mathbf{X}_\gamma\beta_\gamma)}{2}. \tag{A 6}$$

Turning to the integration of β_γ, the expression (A 6) plus the third term of expression (A 4) can be rewritten as

$$\begin{aligned} -\frac{\beta_\gamma'\mathbf{X}_\gamma'\{(\mathbf{I}_n+h11')^{-1}+c^{-1}\mathbf{I}_n\}\mathbf{X}_\gamma\beta_\gamma-2\beta_\gamma'\mathbf{X}_\gamma(\mathbf{I}_n+h11')^{-1}Z}{2}-\frac{Z'(\mathbf{I}_n+h11')^{-1}Z}{2} \\ =-\frac{(\beta_\gamma-A^{-1}B)'A(\beta_\gamma-A^{-1}B)}{2}-\frac{Z'(\mathbf{I}_n+h11')^{-1}Z-B'A^{-1}B}{2}, \end{aligned} \tag{A 7}$$

where $A=\mathbf{X}_\gamma'\{(\mathbf{I}_n+h11')^{-1}+c^{-1}\mathbf{I}_n\}\mathbf{X}_\gamma$, $B=\mathbf{X}_\gamma'(\mathbf{I}_n+h11')^{-1}Z$.

The first term of expression (A 7) is a completed quadratic form in β_γ, which forms a Gaussian probability density and can be integrated out. The second term forms the kernel of a posterior probability density of $Z|\mathbf{X},\gamma$ as

$$-\frac{Z'\{(\mathbf{I}_n+h11')^{-1}-(\mathbf{I}_n+h11')^{-1}\mathbf{X}_\gamma[\mathbf{X}_\gamma'\{(\mathbf{I}_n+h11')^{-1}+c^{-1}\mathbf{I}_n\}\mathbf{X}_\gamma]^{-1}\mathbf{X}_\gamma'(\mathbf{I}_n+h11')^{-1}\}Z}{2}. \tag{A 8}$$

From expression (A 8), we obtain that $p(Z|\mathbf{X},\gamma)\sim N(0,\mathbf{\Sigma}_\gamma)$, with

$$\mathbf{\Sigma}_\gamma{}^{-1}=(\mathbf{I}_n+h11')^{-1}-(\mathbf{I}_n+h11')^{-1}\mathbf{X}_\gamma[\mathbf{X}_\gamma'\{(\mathbf{I}_n+h11')^{-1}+c^{-1}\mathbf{I}_n\}\mathbf{X}_\gamma]^{-1}\mathbf{X}_\gamma'(\mathbf{I}_n+h11')^{-1}.$$

Denote $\mathbf{\Sigma}_\gamma{}^*=\mathbf{I_n}+h11'+c\mathbf{X}_\gamma(\mathbf{X}_\gamma'\mathbf{X}_\gamma)^+\mathbf{X}_\gamma'$. Then

$$\begin{aligned} \mathbf{\Sigma}_\gamma{}^{-1}\mathbf{\Sigma}_\gamma{}^*=\{&(\mathbf{I}_n+h11')^{-1}-(\mathbf{I}_n+h11')^{-1}\mathbf{X}_\gamma[\mathbf{X}_\gamma'\{(\mathbf{I}_n+h11')^{-1}+c^{-1}\mathbf{I}_n\}\mathbf{X}_\gamma]^{-1}\mathbf{X}_\gamma'(\mathbf{I}_n+h11')^{-1}\} \\ &\times[(\mathbf{I_n}+h11')+c\mathbf{X}_\gamma(\mathbf{X}_\gamma'\mathbf{X}_\gamma)^+\mathbf{X}_\gamma'] \end{aligned}$$

$$= \mathbf{I}_n + c(\mathbf{I_n} + h11^{'})^{-1}\mathbf{X}_\gamma(\mathbf{X}_\gamma^{'}\mathbf{X}_\gamma)^{+}\mathbf{X}_\gamma^{'} - (\frac{1}{1+nh} + \frac{1}{c})^{-1}(\mathbf{I_n} + h11^{'})^{-1}\mathbf{X}_\gamma(\mathbf{X}_\gamma^{'}\mathbf{X}_\gamma)^{+}\mathbf{X}_\gamma^{'}$$

$$- (\frac{1}{1+nh} + \frac{1}{c})^{-1}(\mathbf{I_n} + h11^{'})^{-1}\mathbf{X}_\gamma(\mathbf{X}_\gamma^{'}\mathbf{X}_\gamma)^{+}\mathbf{X}_\gamma^{'}(\mathbf{I_n} + h11^{'})^{-1}c\mathbf{X}_\gamma(\mathbf{X}_\gamma^{'}\mathbf{X}_\gamma)^{+}\mathbf{X}_\gamma^{'}$$

$$= \mathbf{I}_n + c(\mathbf{I_n} + h11^{'})^{-1}\mathbf{X}_\gamma(\mathbf{X}_\gamma^{'}\mathbf{X}_\gamma)^{+}\mathbf{X}_\gamma^{'} - (\frac{1}{1+nh} + \frac{1}{c})^{-1}(\mathbf{I_n} + h11^{'})^{-1}\mathbf{X}_\gamma(\mathbf{X}_\gamma^{'}\mathbf{X}_\gamma)^{+}\mathbf{X}_\gamma^{'}$$

$$- (\frac{1}{1+nh} + \frac{1}{c})^{-1}\frac{c}{1+nh}(\mathbf{I_n} + h11^{'})^{-1}\mathbf{X}_\gamma(\mathbf{X}_\gamma^{'}\mathbf{X}_\gamma)^{+}\mathbf{X}_\gamma^{'}$$

$$= \mathbf{I}_n + c(\mathbf{I_n} + h11^{'})^{-1}\mathbf{X}_\gamma(\mathbf{X}_\gamma^{'}\mathbf{X}_\gamma)^{+}\mathbf{X}_\gamma^{'} - c(\mathbf{I_n} + h11^{'})^{-1}\mathbf{X}_\gamma(\mathbf{X}_\gamma^{'}\mathbf{X}_\gamma)^{+}\mathbf{X}_\gamma^{'} = \mathbf{I}_n.$$

Therefore, $\mathbf{\Sigma}_\gamma = \mathbf{\Sigma}_\gamma{}^*$ and

$$p(Z|\mathbf{X}, \gamma) \sim N(0, \mathbf{\Sigma}_\gamma). \tag{A 9}$$

Hence, the joint posterior distribution of $(Z, \gamma|Y, \mathbf{X})$ is

$$\begin{aligned} p(Z, \gamma|Y, \mathbf{X}) &\propto p(Z|Y, \mathbf{X}, \gamma)p(\gamma) \\ &\propto \frac{1}{|\mathbf{\Sigma}_\gamma|^{\frac{1}{2}}}\exp(-\frac{Z^{'}\mathbf{\Sigma}_\gamma^{-1}Z}{2})\prod_{i=1}^{n} I(A_i) \times \prod_{i=1}^{p}\pi_i^{\gamma_i}(1-\pi_i)^{1-\gamma_i}. \end{aligned} \tag{A 10}$$

(ii) Proof of equation (10).

From equations (A 1) and (A 10), we have

$$\begin{aligned} p(\gamma_i|\gamma_{(-i)}, Y, \mathbf{X}, Z) &\propto p(Z|\mathbf{X}, \gamma)p(\gamma_i) \\ &\propto \frac{1}{|\mathbf{\Sigma}_\gamma|^{\frac{1}{2}}}\exp(-\frac{Z^{'}\mathbf{\Sigma}_\gamma^{-1}Z}{2}) \times \pi_i^{\gamma_i}(1-\pi_i)^{1-\gamma_i}, \end{aligned} \tag{A 11}$$

and

$$p(\gamma_i = 1|\gamma_{(-i)}, Y, \mathbf{X}, Z) \propto \frac{1}{|\mathbf{\Sigma_{\gamma^1}}|^{\frac{1}{2}}}\exp(-\frac{Z^{'}\mathbf{\Sigma}_{\gamma^1}^{-1}Z}{2}) \times \pi_i, \tag{A 12}$$

$$p(\gamma_i = 0|\gamma_{(-i)}, Y, \mathbf{X}, Z) \propto \frac{1}{|\mathbf{\Sigma_{\gamma^0}}|^{\frac{1}{2}}}\exp(-\frac{Z^{'}\mathbf{\Sigma}_{\gamma^0}^{-1}Z}{2}) \times (1-\pi_i), \tag{A 13}$$

where $\gamma^1 = (\gamma_1, \cdots, \gamma_{i-1}, \gamma_i = 1, \gamma_{i+1}, \cdots, \gamma_p)$ and $\gamma^0 = (\gamma_1, \cdots, \gamma_{i-1}, \gamma_i = 0, \gamma_{i+1}, \cdots, \gamma_p)$.

As γ_i is binary, we have

$$p(\gamma_i = 1|\gamma_{(-i)}, Y, \mathbf{X}, Z) + p(\gamma_i = 0|\gamma_{(-i)}, Y, \mathbf{X}, Z) = 1. \quad \text{(A 14)}$$

From equations (A 12)-(A 14), we get

$$\begin{aligned} p(\gamma_i = 1|\gamma_{(-i)}, Y, \mathbf{X}, Z) &= \frac{p(\gamma_i = 1|\gamma_{(-i)}, Y, \mathbf{X}, Z)}{p(\gamma_i = 1|\gamma_{(-i)}, Y, \mathbf{X}, Z) + p(\gamma_i = 0|\gamma_{(-i)}, Y, \mathbf{X}, Z)} \\ &= \frac{|\mathbf{\Sigma_{\gamma^1}}|^{-\frac{1}{2}} \exp(-\frac{Z'\mathbf{\Sigma}_{\gamma^1}^{-1}Z}{2}) \times \pi_i}{|\mathbf{\Sigma_{\gamma^1}}|^{-\frac{1}{2}} \exp(-\frac{Z'\mathbf{\Sigma}_{\gamma^1}^{-1}Z}{2}) \times \pi_i + |\mathbf{\Sigma_{\gamma^0}}|^{-\mathbf{\frac{1}{2}}} \exp(-\frac{Z'\mathbf{\Sigma}_{\gamma^0}^{-1}Z}{2}) \times (1-\pi_i)} \\ &= (1 + \frac{1-\pi_i}{\pi_i}\rho)^{-1}, \end{aligned}$$

where

$$\rho = |\mathbf{\Sigma}_{\gamma^1}\mathbf{\Sigma}_{\gamma^0}^{-1}|^{\frac{1}{2}} \exp\{\frac{Z'(\mathbf{\Sigma}_{\gamma^1}^{-1} - \mathbf{\Sigma}_{\gamma^0}^{-1})Z}{2}\}. \quad \text{(A 15)}$$

(iii) Proof of equation (13).

Let $Y_{(-i)}$ be the vector of Y without the i-th element. Since $\mathbf{X}$ is a matrix of fixed covariates, it is treated as constant and is not included in the following derivation for notation simplicity. A cross validation predictive probability can be calculated as

$$\begin{aligned} p(Y_i|Y_{(-i)}) &= \frac{p(Y)}{p(Y_{(-i)})} = \left(\frac{p(Y_{(-i)})}{p(Y)}\right)^{-1} = \left(\frac{\iint p(Y_{(-i)}, Z, \gamma) dZ d\gamma}{p(Y)}\right)^{-1} \\ &= \left(\iint \frac{p(Y_{(-i)}, Z, \gamma)}{p(Y)} dZ d\gamma\right)^{-1} = \left(\iint \frac{p(Y_{(-i)}, Z, \gamma)}{p(Y, Z, \gamma)} \frac{p(Y, Z, \gamma)}{p(Y)} dZ d\gamma\right)^{-1} \\ &= \left(\iint \frac{p(Y_{(-i)}|Z, \gamma)p(Z, \gamma)}{p(Y|Z, \gamma)p(Z, \gamma)} p(Z, \gamma|Y) dZ d\gamma\right)^{-1} \\ &= \left(\iint \frac{p(Y_{(-i)}|Z, \gamma)}{p(Y_i|Y_{(-i)}, Z, \gamma)p(Y_{(-i)}|Z, \gamma)} p(Z, \gamma|Y) dZ d\gamma\right)^{-1} \\ &= (\iint p(Y_i|Y_{(-i)}, Z, \gamma)^{-1} p(Z, \gamma|Y) dZ d\gamma)^{-1}. \end{aligned} \quad \text{(A 16)}$$

2 General algorithm

The gsg-SSVS algorithm is as follows:

Input: data matrix ($\mathbf{X}$) of size $n \times p$ which is preprocessed and the class values vector (Y) of size n;

Parameters: nstart, the number of burn-in iterations; nend, total number of iterations; c, h and π_i, the hyperparameters;

Initialization: $\gamma^{(0)}$ can be obtained by randomly selecting a small number of genes and assigning 1 to the corresponding entries of $\gamma^{(0)}$ and 0 otherwise;

For k=1:nend

step (a): For $i = 1, 2, \cdots, n$, draw $Z_i^{(k+1)}$ from $p(Z_i^{(k)}|Z_{(-i)}^{(k)}, Y, \mathbf{X}, \gamma^{(k)})$ which is a multivariate truncated normal distribution. Direct sampling from this distribution is known to be difficult. In this paper, we follow the method given in Devroye (1986) to simulate samples from the univariate truncated normal distribution;

step (b): For $i = 1, 2, \cdots, p$, generate a random number u_i from a uniform distribution $U[0, 1]$, calculate the probability $p_i^{(k+1)} = p(\gamma_i^{(k+1)} = 1|\gamma_{(-i)}^{(k)}, Y, \mathbf{X}, Z^{(k+1)})$ via (10) and (11), and update γ_i as follows:

$$\gamma_i^{(k+1)} = \begin{cases} 1 & \text{if } \ p_i^{(k+1)} < u_i, \\ 0 & \text{otherwise.} \end{cases}$$

step (c): If $i >$nstart, update $\gamma_j^{(k+1)} = \gamma_j^{(k)}, j = 1, \cdots, p$.

end

3 Results

Table A: Initial (randomly selected) gene indices, selected gene indices by gsg-SSVS, and the LOOCV error rates for three different chains in the analysis of Colon Cancer Data.

Chain No.	Initial (randomly selected) gene index	selected gene index by gsg-SSVS	LOOCV error rate
Chain 1	203 1879 385 1743 100 1621 513 1021 1642 531 213 909 1172 933 1254 346 1583 1207 949 1329 493 227 16 1306 846	377 493 1843 1772 576 792 1423 1346 1635 353 1042 **822** 249 1924 **1210** 14 1400 1549	0.1290 (6 genes) 0.1129 (10 genes)
Chain 2	1023 980 726 1946 657 1702 1224 1522 807 1773 1144 240 1573 825 1272 728 1962 1790 771 889 1219 1504 941 1724 1684	377 493 1843 1772 576 792 1423 1346 1635 353 1549 1042 **1210** 249 1924 1400 14 **625**	0.1290 (6 genes) 0.1129 (10 genes)
Chain 3	388 1459 683 1572 882 718 1723 1950 275 1024 1569 1755 342 29 778 633 1080 734 1365 1286 116 82 557 1372 1207 406 95 1769 985 1398	377 493 1843 1772 1423 792 576 1635 353 1346 249 1549 1924 1400 1042 **625** 14 **739**	0.1290 (6 genes) 0.1129 (10 genes)

*The gene indices in boldface indicate non-overlapping genes in the three sets of the 18 most significant genes. Note that the ten top-ranked selected genes are the same, only minor differences appeared in relation to genes with lower ranks.

References

Le Cao K-A. and Chabrier,P. (2008) ofw: an R package to selection continuous variables for multiclass classification with a stochastic wrapper method. *Journal of Statistical Software*, **28**, 1-16.

Plackett, R. L. (1950). Some theorems in least squares. *Biometrika*, **37**, 149-157.

Woodbury, M. (1950). Inverting modified matrices. Technical Report, **42**. Statistical Research Group, Princeton University.