線形回帰モデルにおけるベイズ型変数選択規準

東京大学・空間情報科学研究センター　丸山　祐造

Suppose the normal linear regression model is used to relate Y to the potential predictors $X_1, \ldots, X_p$,

$$Y \sim N_n(\alpha 1_n + X\beta, \sigma^2 I) \tag{1}$$

where α is an intercept parameter, 1_n is a $n \times 1$ vector each component of which is one, $X = (X_1, \ldots, X_p)$ is an $n \times p$ design matrix, β is a $p \times 1$ vector of unknown regression coefficients, and σ^2 is an unknown positive scalar. The variable selection problem arises when there is some unknown subset of the predictors with regression coefficients so small or useless that it would be preferable to ignore them. It would be convenient throughout to index each of these 2^p possible subset choices by the vector $\gamma = (\gamma_1, \ldots, \gamma_p)$, where $\gamma_i = 0$ or 1 according to where β_i is small or large, respectively. We use $q_\gamma = \gamma' 1_p$ to denote the size of the γth subset.

The goal is to ignore those X_i for which $\beta_i = 0$ in (1). In effect, the problem then becomes that of selecting a submodel of (1) of the form

$$p(Y|\alpha, \beta_\gamma, \sigma^2, \gamma) = N_n(\alpha 1_n + X_\gamma \beta_\gamma, \sigma^2 I) \tag{2}$$

where X_γ is the $n \times q_\gamma$ matrix whose columns corresponds to the γth subset of $X_1, \ldots, X_p$, β_γ is a $q_\gamma \times 1$ vector of unknown regression coefficients. Let denote $\mathcal{M}_\gamma$ the submodel given by (2).

In the linear model (2), α, β_γ and σ^2 are unknown parameters. Eventually we will give prior distributions for all of them, which means full Bayes method. First of all, we provide prior distributions for α and β_γ. The prior measure of α is the Lebesgue measure,

$$p_\alpha(\alpha) = 1 \tag{3}$$

by convention, which leads that the level of predictive values $\hat{y}$ is not shrunk averagely.

The most tractable prior distribution of β_γ is normal conjugate. In the traditional situation $p < n - 1$, which means $q_\gamma < n - 1$ for any $\mathcal{M}_\gamma$, so-called Zellner's g-prior

$$p_{\beta_\gamma}(\beta_\gamma|\sigma^2, g) = N_{q_\gamma}(0, g\sigma^2(Z_\gamma' Z_\gamma)^{-1}), \tag{4}$$

where Z_γ is a centered matrix of X_γ by subtracting the corresponding mean from each of them, is often used.

Additionally treating (very) many regressors case $p > n - 1$ becomes more and more important in modern statistics. Since the residual sum of squares is zero in the case Even worse, where $q_\gamma \geq n - 1$, naive AIC and BIC methods do not work. When $q_\gamma > n - 1$, where the inverse matrix $(Z_\gamma' Z_\gamma)^{-1}$ which is covariance matrix in Zellner's g-prior, does not exist.

In this paper, we consider full Bayes method which is applicable for any case ($p \geq n-1$ and $p < n-1$). Needless to say, full marginal density is expressed by multiple integration and so has been always calculated by numerical method like MCMC. Such numerical full Bayes methods cannot clearly show us what is happening in terms of data y and X_γ whereas the AIC and BIC can do that. This is the motivation of this paper and for this purpose we will give a special variant of Zellner's g-prior which enables us to not only calculate the marginal density analytically but also treat many regressors case. In order to give the variant of Zellner's which is applicable for any case ($q_\gamma > n-1$ and $q_\gamma \leq n-1$), we will make the use of the singular value decomposition of Z_γ,

$$Z_\gamma = U_\gamma D_\gamma W_\gamma' = \sum_{i=1}^{r} d_i[\gamma] u_i[\gamma] w_i'[\gamma] \tag{5}$$

where r is assumed to be equal to $\min(q_\gamma, n-1)$. Notice that the $n-1$ is from the fact that Z is the centered matrix. Here U_γ and W_γ are $n \times r$ and $r \times q_\gamma$ orthogonal matrices, with the columns of U_γ, $(u_1[\gamma], \ldots, u_r[\gamma])$, spanning the column space of Z_γ, and the columns of W_γ, $(w_1[\gamma], \ldots, w_r[\gamma])$, spanning the row space. D_γ is an $r \times r$ diagonal matrix whose diagonal components satisfy $d_1[\gamma] \geq \cdots \geq d_r[\gamma] > 0$.

Under our special prior using the singular value decomposition, we will propose the default Bayesian criterion as follows:

$$\mathrm{BC}[\mathcal{M}_\gamma] = \begin{cases} \prod_{i=1}^{n-1} \nu_i[\gamma]^{-1/2} \left(\mathrm{GESS}_\gamma\right)^{-(n-1)/2} & \text{if } q_\gamma \geq n-1 \\ \prod_{i=1}^{q_\gamma} \nu_i[\gamma]^{-1/2} (\mathrm{RSS}_\gamma + \mathrm{GESS}_\gamma)^{-1/4-q_\gamma/2} & \\ \quad \times (\mathrm{RSS}_\gamma)^{-(n-q_\gamma)/2+3/4} \frac{B(q_\gamma/2+1/4,(n-q_\gamma)/2-3/4)}{B(1/4,(n-q_\gamma)/2-3/4)} & \text{if } q_\gamma \leq n-2. \end{cases}$$

Here $\nu_i[\gamma]$ is $d_i^2[\gamma]/d_r^2[\gamma]$ for $i = 1, \ldots, r$, RSS_γ is the residual sum of squares under $\mathcal{M}_\gamma$ and GESS_γ is the generalized explained sum of squares

$$\mathrm{GESS}_\gamma = \sum_{i=1}^{r} \frac{(u_i[\gamma]' y)^2}{\nu_i[\gamma]}. \tag{6}$$

The model, which maximizes $\mathrm{BC}[\mathcal{M}_\gamma]$ among the class of candidates, is chosen. In particular, when $q_\gamma \geq n-1$, $\mathrm{BC}[\mathcal{M}_\gamma]$ is also expressed as

$$\mathrm{BC}[\mathcal{M}_\gamma] = \left\{ \prod_{i=1}^{n-1} d_i[\gamma] \ \|\hat{\beta}_{LSE}[\gamma]\|^{n-1} \right\}^{-1}$$

where $|\cdot|$ denotes the determinant of matrix, $\hat{\beta}_{LSE}[\gamma]$ is the least squares estimator

$$\hat{\beta}_{LSE}[\gamma] = \sum_{i=1}^{n-1} \frac{w_i[\gamma] u_i'[\gamma] v}{d_i[\gamma]} = (Z_\gamma' Z_\gamma)^- Z_\gamma' v$$

and A^- denotes the Moore-Pennrose inverse matrix of A. The most advantage of our criterion $\mathrm{BC}[\mathcal{M}_\gamma]$ is that it is from exact analytical calculation of the marginal density.