STA 601 - Homework 12

Kedar Prabhudesai

October 25, 2013

Linear Regression:

Model Specification:

This is the general form of Linear Regression Model:

$$
\begin{aligned}
Y_i &= X\beta + \epsilon_i, i = 1, 2, \ldots, n \\
X &= [X_1, X_2, \ldots, X_p], \\
\beta &= [\beta_1, \beta_2, \ldots, \beta_p] \\
\epsilon_i &\sim \mathcal{N}(0, \sigma^2)
\end{aligned}
$$

X is $(n \times p)$ matrix of predictor variables and β is a $(p \times 1)$ vector of regression co-efficients.

Now if we want to incorporate variable selection in this model. We wish to select an unknown subset of predictors. We can thus set an indicator variable which denotes whether or not a predictor has been selected.

$$
\gamma_k = [\gamma_1, \gamma_2, \ldots, \gamma_p].
$$

$\gamma_j = 0$ or 1, based on whether the pth predictor is selected. Hence, the total number of models we can have is 2^p, which corresponds to the total number of combinations. Hence for each model, γ_k, we have the following,

$$
\begin{aligned}
Y_i &= X_{\gamma_k}\beta_{\gamma_k} + \epsilon_i, i = 1, 2, \ldots, n \\
X_{\gamma_k} &= \{X_{ij}, \gamma_j = 1\}, \\
\beta_{\gamma_k} &= \{\beta_j, \gamma_j = 1\}, \\
\epsilon_i &\sim \mathcal{N}(0, \sigma^2), \\
p_{\gamma_k} &= \sum_{j=1}^{p} \gamma_j.
\end{aligned}
$$

Where, p_{γ_k} is the number of predictors selected for the γ_kth model. Hence, X_{γ_k} is a $n \times p_{\gamma_k}$ matrix and β_{γ_k} is a $p_{\gamma_k} \times 1$ vector corresponding to regression co-efficients of the kth model.

To clarify the notations, k is the model index and j is the predictor index within the kth model.

Prior Specification:

Since the appropriate model is unknown, we can model this using a mixture prior.

$$
\begin{aligned}
\pi(\beta_{\gamma_k}, \sigma^2, \gamma_k) &= \pi(\beta_{\gamma_k} \mid \sigma^2, \gamma_k) \times \pi(\sigma^2 \mid \gamma_k) \times \pi(\gamma_k) \\
\pi(\beta_{\gamma_k} \mid \sigma^2, \gamma_k) &\sim \mathcal{N}_{p_{\gamma_k}}(\beta_0, \Sigma_{p_{\gamma_k}} \sigma^2) \\
\pi(\sigma^2 \mid \gamma_k) &\sim Inv - Ga(\alpha, \beta) \\
\pi_{\gamma_k} &= \prod_{j=1}^{p} p_0^{\gamma_j} (1 - p_0)^{(1-\gamma_j)} \\
\pi_{\gamma_k} &= p_0^{\sum_{j=1}^{p} \gamma_j} (1 - p_0)^{\sum_{j=1}^{p}(1-\gamma_j)} \\
\therefore \pi_{\gamma_k} &= p_0^{p_{\gamma_k}} (1 - p_0)^{p - p_{\gamma_k}}
\end{aligned}
$$

Where, p_0 is the probability of selecting the pth predictor for any model γ_k.

Posterior Computation:

For each model the posterior probability of selecting γ_kth model given data y is given as follows.

$$
P(\gamma_k \mid Y) = \frac{\pi_{\gamma_k} L_{\gamma_k}(Y \mid \gamma_k)}{L(Y)}
$$

Where, $L_{\gamma_k}(Y \mid \gamma_k)$ is the marginal likelihood of data Y under the model γ_k. This can be further expanded as follows,

$$
P(\gamma \mid Y) = \frac{\pi_{\gamma_k} \int L_{\gamma}(Y \mid \beta_{\gamma_k}, \sigma^2, \gamma_k) \pi(\beta_{\gamma_k} \mid \sigma^2, \gamma_k) \pi(\sigma^2 \mid \gamma_k) d\beta_{\gamma_k} d\sigma^2}{\sum_{k=1}^{2^p} \pi_{\gamma_k} L_{\gamma_k}(Y \mid \gamma_k)}
$$

Since we have a conjugate priors we can evaluate $L_{\gamma_k}(Y \mid \gamma_k)$. It will be in Normal-Gamma form. For the denominator we need to find the marginal likelihood for each model, then weight it with the probability of each model and sum over them. As p increases, 2^p increases even more. This becomes computationally very demanding. Also for a fixed p_0, we do not get a very accurate model selection because the posterior will always be high for pp_0 value of π_{γ_k}, which is the expected value of the Binomial distribution. This can be improved if we put a Beta prior on p_0.