Big Data Bayesian Linear Regression and Variable Selection by Normal-Inverse-Gamma Summation

Mickey Warner

12 March 2018

Review of the paper by Hang Qian (2017)

Linear Regression with Big Data

With n independent observations and k covariates, fitting the typical linear regressional model

$$y|X, \beta, \sigma^2 \sim N_n\left(X\beta, \sigma^2 I\right) \quad (1)$$

can be problematic when n is so large that we cannot load all the data into memory to perform standard computations.

Need a way to break up the data and perform computations on separate processors.

Normal-Inverse-Gamma (NIG) prior

If β and σ^2 are defined in the following way

$$\begin{aligned} \beta|\sigma^2 &\sim N_k(\mu, \sigma^2\Lambda^{-1}) \\ \sigma^2 &\sim IG(a, b) \end{aligned} \tag{2}$$

then the joint density function is given by

$$p(\beta, \sigma^2) \propto (\sigma^2)^{-(a+k/2+1)} e^{-\frac{1}{\sigma^2}\left[b+\frac{1}{2}(\beta-\mu)^\top\Lambda^{-1}(\beta-\mu)\right]} \tag{3}$$

and we write $(\beta, \sigma^2) \sim NIG(\mu, \Lambda, a, b)$. The NIG distribution is a conjugate prior to the linear model.

NIG posterior

The posterior is given by

$$\beta, \sigma^2 | X, y \sim NIG(\overline{\mu}, \overline{\Lambda}, \overline{a}, \bar{b}) \tag{4}$$

where

$$\begin{aligned} \overline{\mu} &= (\Lambda + X^\top X)^{-1}(\Lambda\mu + X^) \\ \overline{\Lambda} &= \Lambda + X^\top X \\ \overline{a} &= a + \frac{n}{2} \\ \bar{b} &= b + \frac{1}{2}y^\top y + \frac{1}{2}\mu^\top \Lambda \mu - \frac{1}{2}\overline{\mu}^\top \overline{\Lambda}\,\overline{\mu} \end{aligned} \tag{5}$$

NIG summation

Consider the k-dimensional distributions $NIG(\mu_1, \Lambda_1, a_1, b_1)$ and $NIG(\mu_2, \Lambda_2, a_2, b_2)$. If a distribution $NIG(\mu, \Lambda, a, b)$ satisfies

$$\begin{aligned}
\mu &= (\Lambda_1 + \Lambda_2)^{-1}(\Lambda_1\mu_1 + \Lambda_2\mu_2) \\
\Lambda &= \Lambda_1 + \Lambda_2 \\
a &= a_1 + a_2 + \frac{k}{2} \\
b &= b_1 + b_2 + \frac{1}{2}(\mu_1 - \mu_2)^\top(\Lambda_1^{-1} + \Lambda_2^{-1})^{-1}(\mu_1 - \mu_2)
\end{aligned}$$

then it is said to be the sum of two NIG distributions

$$NIG(\mu, \Lambda, a, b) = NIG(\mu_1, \Lambda_1, a_1, b_1) + NIG(\mu_2, \Lambda_2, a_2, b_2)$$

NIG summation, continued

Commutative, associative, identity element $\mu = 0_k$, $\Lambda = 0_{k\times k}$, $a = -k/2$, $b = 0$.

$$NIG(\mu, \Lambda, a, b) + NIG(0_k, 0_{k\times k}, -k/2, 0) = NIG(\mu, \Lambda, a, b)$$