HW6 Part II

Friday, 8 March 2024 23:10

Show that for $\delta = 1$ and hyperparameters $b_r = 0$ the log posterior distribution for β using either normal or Laplace prior distributions have analogous forms to the above expressions

Recall:

$$\beta_i \sim \text{Normal}(b_i, s_i) \quad \text{and} \quad y_i \sim \text{Normal}(x_i^T \beta, \delta)$$

Normal: $$f(\beta_i \mid b_i, s_i) = \frac{1}{\delta\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{\beta_i - b_i}{s_i}\right)^2}$$

Laplace: $$f(\beta_i \mid b_i, s_i) = \frac{1}{2s} \exp\left(-\frac{|\beta_i - b_i|}{s_i}\right)$$

With $\delta = 1$ and $b_i = 0$:

$$f(\beta_i \mid s_i) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{\beta_i}{s_i}\right)^2}$$

and $$f(\beta_i \mid s_i) = \frac{1}{2b}\left(-\frac{|\beta_i|}{s_i}\right)$$

We know the posterior $\propto$ prior $\cdot$ likelihood

so log posterior $\propto$ log prior + log likelihood

i.e. $\log p(\beta \mid x) \propto \log p(x \mid \beta) + \log p(\beta)$

We know $$\log p(x \mid \beta) = -\frac{1}{2\delta^2} \sum_{i=1}^{n} (y_i - x_i^T \beta)^2$$

and for the normal distribution:

$$\log p(\beta) = -\frac{1}{2s^2} \sum_{j=1}^{p} (\beta_j - b_j)^2$$

$$= -\frac{1}{2s^2} \sum_{j=1}^{p} \beta_j^2$$

whereas for the Laplace:

$$\log p(\beta) = -\frac{1}{s} \sum_{j=1}^{p} |\beta_j - b_j| = -\frac{1}{s} \sum_{j=1}^{p} |\beta_j|$$

So for the normal dis:

$$\log p(\beta \mid x) \propto -\frac{1}{2\delta^2} \sum_{i=1}^{n} (y_i - x_i^T \beta)^2 - \frac{1}{2s^2} \sum_{j=1}^{p} \beta_j^2$$

$$= -\frac{1}{2\delta^2}(y - X\beta)^T(y - X\beta) - \frac{1}{2s^2}\sum_{j=1}^{n}\beta_j^2$$

$$= -\frac{1}{\delta^2}\left(\frac{1}{2}(y - X\beta)^T(y - X\beta) + \frac{\delta^2}{2s^2}\sum_{j=1}^{n}\beta_j^2\right)$$

$$= -\frac{1}{2\delta^2}\left(\|y - X\beta\|_2^2 + \frac{\delta^2}{2\delta^2}\|\beta\|_1\right)$$

and for the Laplace:

$$\log p(\beta|x) \propto -\frac{1}{2\delta^2}\sum_{i=1}^{n}(y_i - x_i^T\beta)^2 - \frac{1}{2s^2}\sum_{j=1}^{n}|\beta_j|$$

$$= -\frac{1}{\delta^2}\left(\frac{1}{2}(y - X\beta)^T(y - X\beta) + \frac{\delta^2}{2s^2}\sum_{j=1}^{n}|\beta_j|\right)$$

$$= -\frac{1}{\delta^2}\left(\|y - X\beta\|_2^2 + \frac{\delta^2}{2s^2}\|\beta\|_1\right)$$

Hence these are analogous to the lasso and ridge regularizations respectively

Bayesians do not optimize posterior distributions, they sample from them; but, the posterior distributions are nonetheless 'regularizations' of the likelihood through the prior