High-Dimensional Bayesian

Bayesian Lasso

Lasso minimizes the penalized least squares to achieve *sparsity*

$$\min_{\beta} \sum_{i=1}^{n} \frac{1}{2} \left(y_i - \mathbf{x}_i^T \boldsymbol{\beta}\right)^2 + \lambda \sum_{j=1}^{p} |\beta_j|$$

The objective function is equivalent to maximizing

$$\prod_{i=1}^{n} \underbrace{\exp\left(-\left(y_i - \mathbf{x}_i^T \boldsymbol{\beta}\right)^2/2\right)}_{likelihood} \times \underbrace{\exp\left(-\lambda \sum_{j=1}^{p} |\beta_j|\right)}_{prior}$$

It's not easy to directly sample from the posterior. But it can be written as a mixture of normals, mixed with exponential distribution,

$$\frac{a}{2} e^{-a|z|} = \int_0^\infty \frac{1}{\sqrt{2\pi s}} e^{-z^2/(2s)} \times \frac{a^2}{2} e^{-a^2 s/2} ds$$

Hence, the full model is

$$\begin{aligned} \mathbf{y} \mid \mu, \mathbf{X}, \boldsymbol{\beta}, \sigma^2 &\sim N\left(\mu \mathbf{1} + \mathbf{X}\boldsymbol{\beta}, \sigma^2 \mathbf{I}\right) \\ \beta_j \mid s_j^2, \sigma^2 &\sim N\left(\mathbf{0}, \sigma^2 s_j^2\right) \\ s_j^2 \mid \tau &\sim \frac{\tau^2}{2} e^{-\tau^2 s_j^2/2} \\ \sigma^2 &\sim \pi\left(\sigma^2\right) \end{aligned}$$

The Horseshoe Prior

To be completed.

Bayesian Robust Regression

We want to simulate the uncertainty with t distribution(heavy tail) in this case.

$$f(t) = \frac{\Gamma((v+1)/2)}{\sqrt{\pi v}\Gamma(v/2)} \left(1 + x^2/v\right)^{-\frac{v+1}{2}}$$

Analogous to the Bayesian Lasso, it's not a easy task to sample from the posterior directly.

Here, we introduce an integral (mixture of $x \sim \mathcal{N}(0, \sigma^2)$ and $1/\sigma^2 \sim \Gamma(v/2, v/2)$):

$$\int_0^\infty \frac{\sqrt{\tau}}{\sqrt{2\pi}} \exp\left(-\tau x^2/2\right) \times \frac{(v/2)^{v/2}}{\Gamma(v/2)} \tau^{v/2-1} e^{-v\tau/2} d\tau$$

Thus, for robust regression

$$y_i = \mathbf{x}_i^T \boldsymbol{\beta} + \epsilon_i, v \sim t_v(0)$$

The full model is:

$$\begin{gathered} y_i \sim N\left(\mathbf{x}_i^T \boldsymbol{\beta}, \sigma_i^2\right) \\ 1/\sigma_i^2 \sim \text{Gamma}(v/2, v/2) \\ \boldsymbol{\beta} \sim N\left(\mathbf{0}, \delta^{-2}\mathbf{I}\right) \end{gathered}$$

Then, the full conditional density for each variables is:

$$\begin{aligned} &\tau_i \mid \boldsymbol{\beta}, \mathbf{y}, \mathbf{X} \sim \text{Gamma}\left(\frac{v+1}{2}, \frac{v + \left(y_i - \mathbf{x}_i^T \boldsymbol{\beta}\right)^2}{2}\right) \\ &\boldsymbol{\beta} \mid \tau_1, \cdots, \tau_n, \mathbf{y}, \mathbf{X} \sim N\left(\boldsymbol{\mu}_n, \Sigma_n^2\right) \end{aligned}$$

where

$$\boldsymbol{\mu}_n = \mathrm{A}^{-1}\mathbf{b},\ \Sigma_n = \mathrm{A}^{-1},\ \mathrm{A} = \sum_{i=1}^n \tau_i \mathbf{x}_i \mathbf{x}_i^T + \delta^2 \mathbf{I},\ \mathbf{b} = \sum_{i=1}^n \tau_i y_i \mathbf{x}_i$$

<u>Note</u>: MH algorithm is also realizable in this case.