# SDS 387 HW 4

Maurice Diesendruck, Sukyung Park, Bowei Yan, Michael Zhang

March 24, 2015

## 3 Comparison

This is an analysis of the LASSO, BLASSO, and Horseshoe priors on diabetes data.

1. LASSO
   The LASSO is an L1-penalized least squares estimate of $\beta$, with $\tilde{y}$ as the mean-centered outcome variable, where the following is optimized:

$$\min_{\beta}(\tilde{y} - XB)'(\tilde{y} - XB) + \lambda \sum_{j=1}^{p} |\beta_j|$$

   (a) Strengths
      i. Easy to implement.
   (b) Weaknesses
      i. Proposed standard error estimators are not considered satisfactory.

2. BLASSO
   The BLASSO interprets the L1-penalty as a prior on $\beta$ and $\sigma^2$. In this case, $y$ is normal, the prior $\pi(\beta|\sigma^2)$ is Laplace (i.e. double exponential), and the prior $\pi(\sigma^2)$ is non-informative and defined as such to ensure a unimodal posterior:

$$\begin{aligned} y|\mu, X, \beta, \sigma^2 &\sim \ N(\mu 1_n + X\beta, \sigma^2 I_n) \\ \pi(\beta|\sigma^2) &= \prod_{j=1}^{p} \frac{\lambda}{2\sqrt{\sigma^2}} e^{-\lambda|\beta_j|/\sqrt{\sigma^2}} \qquad \text{(Laplace distribution)} \\ \pi(\sigma^2) &= \frac{1}{\sigma^2} \end{aligned}$$

   (a) Strengths
      i. Easy to implement
      ii. Automatically provides interval estimates for all parameters, including the error variance.
   (b) Weaknesses
      i. More computationally intensive than LASSO.

3. Horseshoe
   The Horseshoe (HS) prior assumes $y|\theta \sim \ N(\theta, \sigma^2 I)$, and aims to (1) estimate $\theta$, and (2) predict future realizations of $y$. It is especially useful in cases where most covariates are nearly zero (i.e. sparse $\theta$). The model is as follows:

$$\begin{aligned} \theta_i|\lambda_i &\sim \ N(0, \lambda_i^2) \\ \lambda_i|\tau &\sim \ C^+(0, \tau) \\ \tau &\sim \ C^+(0, \sigma) \\ E(\theta_i|y) &= \int_0^1 (1 - \kappa_i) y_i p(\kappa_i|y) \mathrm{d}\kappa_i = (1 - E(\kappa_i|y)) y_i \end{aligned}$$

   Where $\kappa_i = 1/(1 + \lambda_i^2)$, assuming fixed values $\sigma^2 = \tau^2 = 1$. Given $\lambda_i \geq 0$, this implies that $\lambda_i$'s are indirectly related to amount of shrinkage (e.g. high $\lambda_i$ means low $\kappa_i$, and less shrinkage).

(a) Strengths

  i. Easy to implement.
  ii. Has fewer hyperparameters.
  iii. Robust and flexible to high or low sparsity situations.
  iv. Converges efficiently.

(b) Weaknesses

  i.
  ii.

# 4 Perform Gibbs Sampling for Bayesian Lasso

Reference: Park and Casella (2008) (Section 2).

1. Model and Prior for BLASSO
   See 3.2, above.

2. Complete Conditional for each (set of) Parameters
   To get complete conditionals, the prior $\pi(\beta|\sigma^2)$ is represented differently, as a scaled mixture of normals (with exponential mixing density). Again, the Laplace can be re-written as the integrated product of a Normal and Exponential distribution:

$$
\begin{aligned}
\pi(\beta|\sigma^2) &= \prod_{j=1}^{p} \frac{\lambda}{2\sqrt{\sigma^2}} e^{-\lambda|\beta_j|/\sqrt{\sigma^2}} && \text{(Laplace distribution)} \\
&= \prod_{j=1}^{p} \int_0^{\infty} \text{ Normal * Exponential}
\end{aligned}
$$

   This way, full conditionals become:

$$
\begin{aligned}
y|\mu, X, \beta, \sigma^2 \sim&\ N(\mu 1_n + X\beta, \sigma^2 I_n), \\
\beta|\sigma^2, \tau_1^2, \ldots, \tau_p^2 \sim&\ N_p(0_p, \sigma^2 D_\tau), && \text{(Normal part of Laplace)} \\
&\text{where } D_\tau = diag(\tau_1^2, \ldots, \tau_p^2), \\
\sigma^2, \tau_1^2, \ldots, \tau_p^2 \sim&\ \pi(\sigma^2) d\sigma^2 \prod_{j=1}^{p} \frac{\lambda^2}{2} e^{\lambda^2 \tau_j^2/2} d\tau_j^2, && \text{(Exponential part of Laplace)} \\
\sigma^2, \tau_1^2, \ldots, \tau_p^2 > 0.
\end{aligned}
$$

## Results and Discussion

Our model is run with the following command:

```
> # R code here!
```