Statistics 700 Homework 4

SMC and MCMC

Due date: 6:00 pm (EST) Oct. 31, 2017

Bayesian Lasso. The Lasso method estimates linear regression coefficients through L_1 constrained least squares, also known as the penalized regression. It provides a way of obtaining a sparse solution for high-dimensional regression problems, which are prevalent in modern applications such as genetics. Here we investigate on the Bayesian interpretation and computation of the LASSO method.

Consider a regression model

$$\boldsymbol{y} = \mu \mathbf{1}_n + X\boldsymbol{\beta} + \boldsymbol{\epsilon},$$

where $\boldsymbol{y}$ is the $n \times 1$ vector of responses, $\mathbf{1}_n$ is the $n \times 1$ vector of 1s, μ is the overall mean, $\boldsymbol{\beta}$ is $p \times 1$ vector of regression coefficients, X is the $n \times p$ matrix of standardized regressors, and $\boldsymbol{\epsilon}$ is the $n \times 1$ vector of independent and identically distributed normal errors with mean 0 and unknown variance σ^2. Lasso estimates achieve

$$\min_{\beta} \left\{ (\tilde{\boldsymbol{y}} - X\boldsymbol{\beta})^T (\tilde{\boldsymbol{y}} - X\boldsymbol{\beta}) + \lambda \sum_{j=1}^{p} |\beta_j| \right\}$$

for some $\lambda \geq 0$, where $\tilde{\boldsymbol{y}} = \boldsymbol{y} - \bar{y}\mathbf{1}_n$, $\bar{y}$ is sample average of $\boldsymbol{y}$.

We use the Diabetes Data from Efron et al. (2004). It contains $p = 10$ variables and $n = 442$ measurements. The data is standardized such that the means of all variables are zero, and all variances are equal to one. Refer to `https://artax.karlin.mff.cuni.cz/r-help/library/care/html/efron2004.html` for detailed information about loading the data. References: Efron, B., et al. 2004. Least angle regression (with discussion). Ann. Statist. 32:407–499.

1. Implement a linear regression with the Diabetes data.

2. Implement a Lasso method with the Diabetes data.

3. Now we consider a full Bayesian analysis of the regression model

$$\tilde{y}_i \sim N(X_i\boldsymbol{\beta}, \sigma^2).$$

   Assume that the priors for $\boldsymbol{\beta}$ and σ^2 are as follows.

$$\pi(\boldsymbol{\beta}|\sigma^2) = \prod_{j=1}^{p} \frac{\lambda}{2\sqrt{\sigma^2}} \exp\left[-\lambda \frac{|\beta_j|}{\sqrt{\sigma^2}}\right], \ \pi(\sigma^2) \propto \sigma^{-2}.$$

   Implement a Metropolis-Hastings algorithm to sample from the posterior of $(\boldsymbol{\beta}, \sigma^2|X, \boldsymbol{y})$ with the Diabetes data ($\lambda = 0.237$).

   - Optional (≤ 5 bonus points). Try a reparametrization, e.g. $\log(\sigma^2)$, and discuss whether or not the reparametrization improves the sampling.

4. In fact, we can represent the Bayesian Lasso model as

$$\begin{aligned}
y_i|\mu, X_i, \boldsymbol{\beta}, \sigma^2 &\overset{iid}{\sim} N(\mu + X_i\boldsymbol{\beta}, \sigma^2), 1 \leq i \leq n;\\
\beta_j|\sigma^2, \tau_j^2 &\overset{iid}{\sim} N(0, \sigma^2\tau_j^2), 1 \leq j \leq p;\\
\tau_j^2 &\overset{iid}{\sim} \frac{\lambda^2}{2} \exp\left\{-\frac{\lambda^2\tau_j^2}{2}\right\}, 1 \leq j \leq p;\\
\sigma^2 &\sim \pi(\sigma^2).
\end{aligned}$$

   Assume a flat prior on μ. Describe and implement a Gibbs sampler of the posterior using the Diabetes data ($\lambda = 0.237$).

   - Hint: you can integrate out the grand mean μ first from the joint posterior to reduce the number of parameters in the Gibbs sampler.

5. Compare the four methods above. What do you find?

Remark: This homework problem is based on "The Bayesian Lasso" by Trevor PARK and George CASELLA, Journal of the American Statistical Association, June 2008, Vol. 103, No. 482, pp. 681-686. **Please work out your own solutions before referring to the original paper.**