Bayesian LASSO

STA 4241

Shenyu Zhou

Introduction

Linear Regression Model:

$$y_i = \beta_0 + \sum_{j=1}^{p} \beta_j x_{ij} + \epsilon, \epsilon_i \sim N(0, \sigma^2)$$

OLS Estimation:

$$RSS(\beta) = \sum_{i=1}^{n} \left(Y_i - \beta_0 - \sum_{j=1}^{p} \beta_j X_{ij} \right)^2$$

LASSO

$$\hat{\beta}_{lasso} = \arg\min_{\beta} \left\{ RSS(\beta) + \lambda \sum_{j=1}^{p} \left| \beta_j \right| \right\}$$

Bayesian LASSO

Bayesian Theorem:

$$p(\beta|X,Y) \propto p(Y|X,\beta)p(\beta)$$

Laplace Prior:

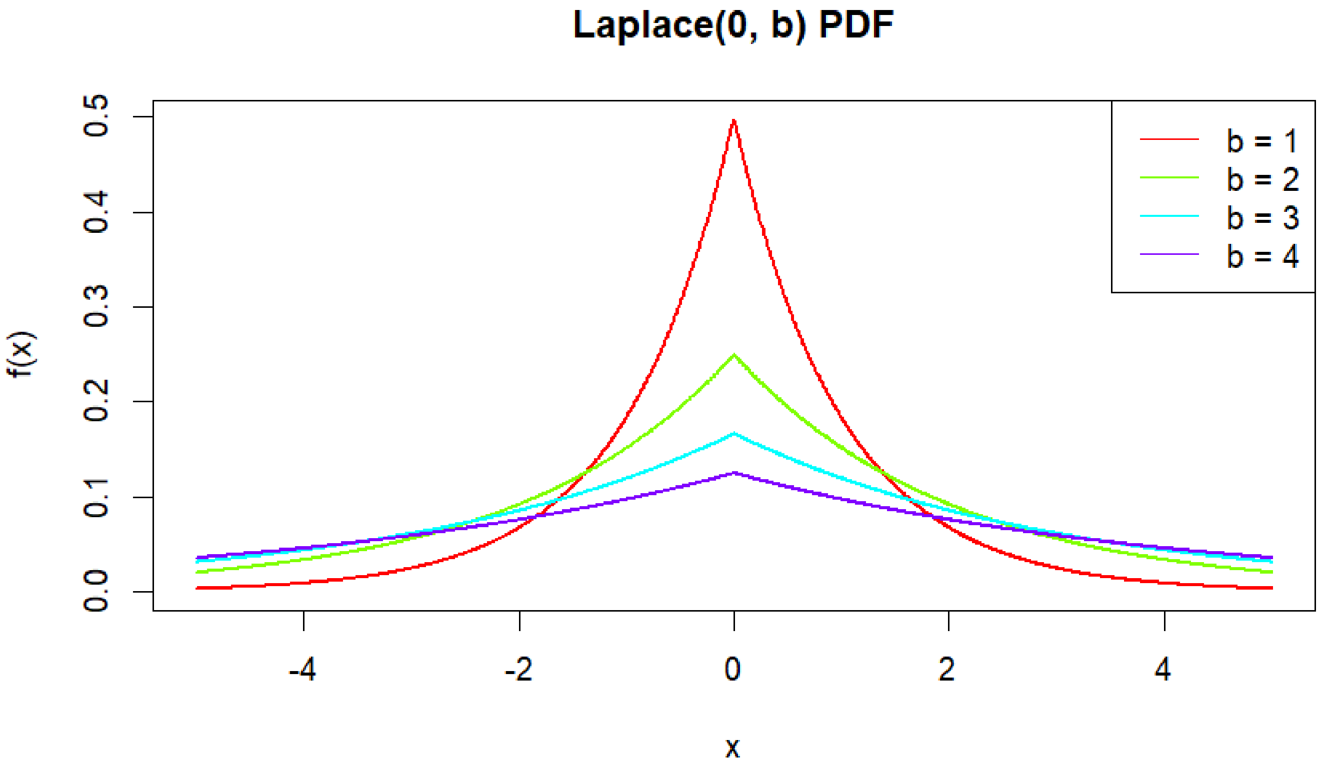
$$\beta_j \sim Laplace(0,b)$$

$$p(\beta_j) = \frac{1}{2b}\exp\left(-\frac{|\beta_j|}{b}\right)$$

Equivalence to frequentist LASSO:

$$\hat{\beta}_{MAP} = \arg\min\left\{RSS(\beta) + \frac{b}{2\sigma^2}\sum_{j=1}^{p}|\beta_j|\right\}$$

with $\lambda = \frac{2\sigma^2}{b}$



Hierarchical Bayesian Model

$$p(\beta, \lambda, \sigma^2 | y) \propto p(y|\beta, \sigma^2) \prod_{j=1}^{Y} p\left(\beta_j \middle| \lambda_j, \sigma^2\right) p\left(\lambda_j\right) p(\sigma^2)$$

Likelihood:

$$y|\beta, \sigma^2 \sim N(X\beta, \sigma^2 I_n)$$

Prior for β:

$$\beta_j | \lambda_j, \sigma^2$$

Prior for λ_j:

$$\lambda_j \sim Exp\left(\frac{\lambda^2}{2}\right)$$

Prior for σ^2:

$$\sigma^2 \sim IG(a_0, b_0)$$

Gibbs Sampling Algorithm

1. Update β:
   1. Full conditional is Gaussian:

$$p(\beta|\lambda,\sigma^2,y) \sim N\left(\left(X^TX + D_\lambda^{-1}\right)^{-1}X^Ty, \sigma^2\left(X^TX + D_\lambda^{-1}\right)^{-1}\right)$$

2. Update λ:
   1. Full conditional is Inverse-Gaussian:

$$\lambda|\beta,\sigma^2 \sim InverseGaussian\left(\sqrt{\frac{\sigma^2}{\lambda|\beta|}}, \lambda^2\right)$$

3. Update σ^2:
   1. Full conditional is Inverse-Gamma:

$$\sigma^2|\beta,\lambda,y \sim IG\left(a_0 + \frac{n+p}{2}, b_0 + \frac{\|y - X\beta\|^2}{2} + \frac{1}{2}\sum_{j=1}^{p}\frac{\beta_j^2}{\lambda_j}\right)$$

Simulation Example

Set-up:

- $n = 500, p = 20$
- $\beta^* = (1.5, 2.5, 3.5, 0, \dots, 0)$
- $X_{ij} \sim N(0, 1), Y = X\beta^* + \epsilon \sim N(0, 1)$

Methods Compared:

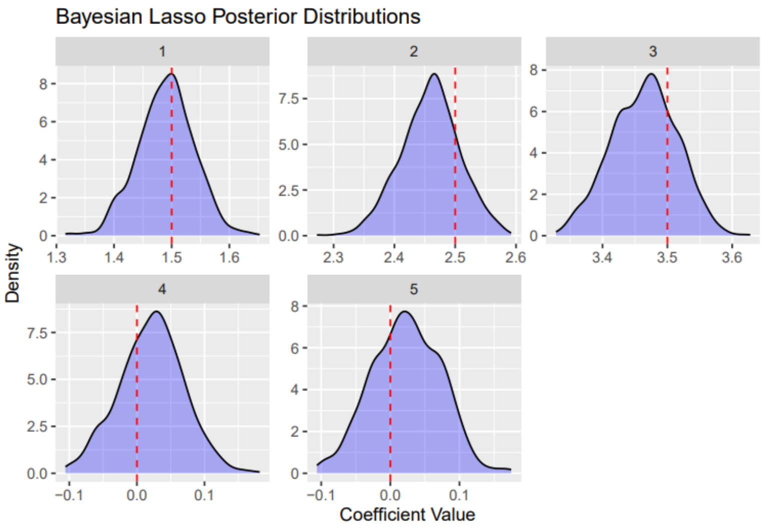
1. Frequentists LASSO
    1. Bootstrap (empirical intervals)
2. Bayesian LASSO

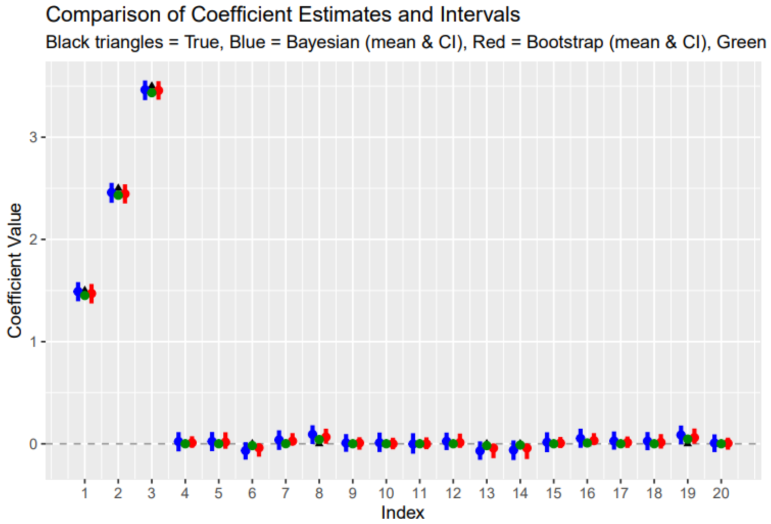
Simulated Example: results

```
## Frequentist Lasso Estimates:

##  [1]  1.4513956897  2.4324151397  3.4355699619  0.0000000000  0.0000000000
##  [6] -0.0181063290  0.0008068596  0.0413408229  0.0000000000  0.0000000000
## [11]  0.0000000000  0.0000000000 -0.0179382157 -0.0125672986  0.0000000000
## [16]  0.0080060524  0.0000000000  0.0000000000  0.0458186691  0.0000000000

##
## Bootstrapped Lasso Means and 95% Intervals:

##                [,1]     [,2]     [,3]         [,4]        [,5]        [,6]
## mean     1.471133 2.446005 3.457501  0.009607898  0.01613904 -0.03935249
## ci_lower 1.373514 2.352196 3.367407 -0.039184316 -0.04886426 -0.12620823
## ci_upper 1.563989 2.539126 3.548582  0.073839278  0.11288932  0.00000000
##                 [,7]       [,8]         [,9]        [,10]        [,11]
## mean      0.02589403 0.06547572  0.007932695  2.642478e-05 -0.001745804
## ci_lower -0.01290876 0.00000000 -0.058153373 -5.425888e-02 -0.053470641
## ci_upper  0.10468195 0.14804731  0.064584540  5.947082e-02  0.062938355
##                [,12]       [,13]         [,14]       [,15]      [,16]
## mean      0.01161351 -0.04186976 -0.0421790768  0.00667215 0.03133238
## ci_lower -0.03995028 -0.14040612 -0.1492642160 -0.04159837 0.00000000
## ci_upper  0.10004003  0.00000000  0.0008926861  0.06717149 0.10452442
##                [,17]       [,18]      [,19]        [,20]
## mean      0.01217462  0.01348161 0.06011742  0.004675554
## ci_lower -0.04063421 -0.04715406 0.00000000 -0.057687255
## ci_upper  0.07150023  0.09498035 0.14959496  0.059895024
```

```
##
## Bayesian Lasso Posterior Means and 95% Credible Intervals:

##              [,1]     [,2]     [,3]        [,4]        [,5]        [,6]
## mean     1.489961 2.458935 3.464444  0.02139147  0.02305214 -0.06632575
## ci_lower 1.395016 2.358254 3.363268 -0.07458665 -0.07218371 -0.15487936
## ci_upper 1.582228 2.553421 3.555510  0.11428027  0.11560833  0.01576948
##                 [,7]         [,8]         [,9]       [,10]        [,11]
## mean      0.03774316  0.091826876  0.008084577  0.01041076 -0.002268714
## ci_lower -0.06027940 -0.002965315 -0.078762204 -0.08087050 -0.097045694
## ci_upper  0.13211690  0.181392104  0.095022368  0.11024738  0.103488751
##                [,12]       [,13]       [,14]       [,15]       [,16]
## mean      0.02418686 -0.06898059 -0.06154494  0.01520955  0.05280079
## ci_lower -0.06159162 -0.15712832 -0.15902972 -0.08227585 -0.03957666
## ci_upper  0.10982833  0.02381479  0.03293029  0.11291128  0.14802858

##                [,17]       [,18]        [,19]        [,20]
## mean      0.02704109  0.02560291  0.086079654  0.007602938
## ci_lower -0.05760811 -0.06307989 -0.004802462 -0.081495506
## ci_upper  0.12069999  0.11518782  0.178723196  0.092350040
```

Simulated Example: results






```
##
## MSE of Estimates:

## Lasso: 0.000788102

## Lasso (Bootstrap mean): 0.001056956

## Bayesian Lasso (Posterior mean): 0.001973984

##
## Variable Selection (qualitative):

## Lasso sets coefficients exactly to zero or not, Bayesian gives a probability.

## Number of exact zeros (Lasso): 10

## Bayesian CI containing zero (count): 17 out of 20 coefficients
```

Conclusion

- Advantage:
  - Interpretability (probabilistic, standard error)
  - Flexibility
- Limitation:
  - Computationally intensive
  - Does not automatically zero out coefficients.

Reference

Chen, Y. (2021). STA521: Predictive Modelling and Statistical Learning, Lecture 9: Bayesian Regression I. Duke University. Retrieved from https://www2.stat.duke.edu/courses/Fall21/sta521.001/post/week05- 1/main.pdf

Kyung, Minjung & Gill, Jeff & Ghosh, Malay & Casella, George. (2010). Penalized Regression, Standard Errors, and Bayesian Lassos. Bayesian Analysis. 5. 369-412. 10.1214/10-BA607.

Park, T., & Casella, G. (2008). The Bayesian Lasso. Journal of the American Statistical Association, 103(482), 681–686.

Tibshirani, R. (1996). Regression shrinkage and selection via the Lasso. Journal of the Royal Statistical Society. Series B (Methodological), 58(1), 267–288.