BIOSTAT 615 - Final Project
Re-implementing Spike-and-Slab Group Lasso Algorithms with Rcpp

Group 10: Yuan Zhong, Yilei Zhang

1 Introduction

1.1 Background

Group structures of predictors arise in many statistical applications. For example, in a regression model, a multi-level categorical predictor is usually presented by a group of dummy variables. In genomics, genes within the same pathway may form a group at the pathway or gene set level and act in tandem to regulate a biological system. In these scenarios, the response $Y_{n\times 1}$ can be modeled as a linear regression problem with G groups:

$$Y = \sum_{g=1}^{G} X_g \beta_g + \epsilon, \tag{1}$$

where $\epsilon \sim N_n\left(0, \sigma^2 I_n\right)$, β_g is a coefficient vector of length m_g, and X_g is an $n \times m_g$ covariate matrix corresponding to group $g = 1, \ldots, G$. Let p be the total number of predictors, then $p = \sum_{g=1}^{G} m_g$ We are often interested in selecting groups of variables that are most significantly associated with the response.

Penalized regression methods have been very popular for the power to select relevant variables and estimate regression coefficients simultaneously. Among them the lasso proposed by Tibshirani (1996) draws much attention for its ability to both select and estimate. Yuan and Lin (2006) proposed the group lasso which generalizes the lasso in order to select grouped variables for accurate prediction in regression. The group lasso estimator is obtained by solving:

$$\min_{\beta} \|Y - \sum_{g=1}^{G} X_g \beta_g\|_2^2 + \lambda \sum_{g=1}^{G} \|\beta_g\|_2.$$

In the Bayesian framework, lasso estimator is equivalent to the posterior mode with independent double exponential prior for each regression coefficient. For selecting relevant groups under model (1), Xu and Ghosh (2015) proposed the following hierarchical Bayesian group lasso model with an independent spike-and-slab type prior for each factor β_g:

$$Y|X, \beta, \sigma^2 \ \sim \ N_n\left(X\beta, \sigma^2 I_n\right), \tag{2}$$

$$\beta_g|\sigma^2, \tau_g^2 \ \overset{ind}{\sim} \ (1-\pi_0) N_{m_g}\left(0, \sigma^2\tau_g^2 I_{m_g}\right) + \pi_0 \delta_0(\beta_g), \ g = 1, \ldots, \ G, \tag{3}$$

$$\tau_g^2 \ \overset{ind}{\sim} \ Gamma\left(\frac{m_g+1}{2}, \ \frac{\lambda^2}{2}\right), \ g = 1, \ldots, \ G, \tag{4}$$

$$\sigma^2 \ \sim \ Inverse\ Gamma(\alpha, \ \gamma), \ \sigma^2 > 0, \tag{5}$$

where $\delta_0(\beta_g)$ denotes a point mass at $0 \in R^{m_g}$ (the "spike"), and $N_{m_g}\left(0, \sigma^2\tau_g^2 I_{m_g}\right)$ is the "slab density", and $\pi_0 \sim Beta(a, b)$. For Bayesian variable selection, point mass spike-and-slab priors (3) are interpretable, but they are computationally intractable in high dimensions, due in large part to the discrete nature of the spike-and-slab density. As an alternative, fully continuous variants of spike-and-slab models have been developed. For continuous spike-and-slab models, the point mass spike δ_0

is replaced by a continuous density heavily concentrated around 0_{m_g}. Ročková (2018) introduced the univariate spike-and-slab lasso (SSL), which places a mixture prior of two Laplace densities on the individual coordinates β_j:

$$\pi(\beta|\theta) = \prod_{j=1}^{p}[(1-\theta)\psi(\beta_j|\lambda_0) + \theta\psi(\beta_j|\lambda_1)], \tag{6}$$

where $\psi(\beta|\lambda) = \frac{\lambda}{2}e^{-\lambda|\beta|}$, and $\theta \in (0,1)$ is the mixing proportion. Typically we set $\lambda_0 \gg \lambda_1$ so that the spike is heavily concentrated around zero. Though the SSL prior model (6) does not place any mass on exactly sparse vectors, the global posterior mode under the SSL prior may be exactly sparse.

While the SSL (6) induces sparsity on individual coefficients, it does not account for group structure of covariates. Bai et al. (2020) introduced the spike-and-slab group lasso (SSGL) for Bayesian grouped regression and variable selection. Under the SSGL prior, the global posterior mode is exactly sparse, thereby allowing the mode to automatically threshold out insignificant groups of coefficients. Bai et al. (2020) introduced a highly efficient block coordinate ascent algorithm for the global maximum a posterior (MAP) estimation and variable selection, which allows us to rapidly identify significant groups of coefficients, and they also adopted a non-separable penalty which allows us to automatically adapt to ensemble information about sparsity.

1.2 Our Contributions

Bai et al. (2020) provided an R package for implementing their algorithms: `https://github.com/jantonelli111/SSGL`. Their functions are totally implemented in R, and include multiple layers of "for loops", which may highly reduce the time efficiency. In this project, we made the following contributions:

1. Re-implement their algorithms with Rcpp.
2. Conduct some simulation studies to compare our algorithm efficiency with the original functions in their package.
3. Compare our SSGL's performance with the group lasso algorithm implemented in the R package gglasso.

2 Algorithm Description and Analysis

The coordinate ascent algorithm in R package SSGL is implemented in the following way:

Algorithm 1 Spike-and-Slab Group Lasso (SSGL)

Input: a fixed small value λ_1, a grid of increasing λ_0 values $I = \{\lambda_0^1, \ldots, \lambda_0^L\}$, updating frequency M, response vector $Y_{n\times 1}$, and covariate matrix X, number of groups G, a, b, error, forceGroup, a bool variable updateSigma
Initialize: $\beta^* = 0_p$, $\theta^* = 0.5$, $\sigma^{*2} = \sqrt{var(Y) * qchisq(0.1,\ df = 3)/(df+2)}$, $\sigma^2_{min} = var(Y)/n$
for l in 1:L **do**
    for g in 1:G **do**
        Apply SVD on X_g: $X_g = U_{X_g} D_{X_g} V_{X_g}^T$, and assign $\tilde{X}_g = U_{X_g}$
    end for
    intercept $= mean\left(Y - \tilde{X}\beta^*\right)$
    if $l \geq 2$ **then**
        Z=number of groups having non-zero coefficients
        $\theta^* = \frac{a+Z}{a+b+G}$
    end if
    counter=0, diff=10*error
    $\lambda_{0base} = \lambda_0$

Algorithm 2 SSGL - Part 2

while diff>error & counter<300 **do**
  $\beta^*_{old} = \beta^*$
  for g in 1:G **do**
    update intercept = $mean\left(Y - \tilde{X}\beta^*\right)$
    m = number of predictors in group g, $\lambda_0 = \sqrt{m}\lambda_{0_{base}}$
    if g is in forceGroup **then**
      $yResid \ = \ Y - intercept - \tilde{X}_{\setminus g}\beta^*_{\setminus g}$
      $\beta^*_g = \left(\tilde{X}_g^T \tilde{X}_g\right)^{-1} \tilde{X}_g^T yResid$
    else
      $p^* = \frac{\theta^* \psi(0_m|\lambda_1)}{\theta^* \psi(0_m|\lambda_1) + (1-\theta^*)\psi(0_m|\lambda_0)}$ ▷ $\psi(\beta_g|\lambda) = \frac{2^{-m}\pi^{-(m-1)/2}}{\Gamma((m+1)/2)} \cdot \lambda^m exp(-\lambda\|\beta_g\|_2)$
      $l^* = \lambda_1 p^* + \lambda_0 \left(1 - p^*\right)$
      $g = \left(l^* - \lambda_1\right)^2 + \left(2n/\sigma^{*2}\right) \cdot log\left(p^*\right)$
      if $g > 0$ **then**
        $\Delta = \sqrt{2n\sigma^2 log\left(1/p^*\right)} + \sigma^{*2}\lambda_1$
      else
        $\Delta = \sigma^{*2} l^*$
      end if
      $z_g = X_g^T \left[Y - intercept - \tilde{X}_{\setminus g}\beta^*_{\setminus g}\right]$
      $p^*_g = \frac{\theta^* \psi\left(\beta^*_g|\lambda_1\right)}{\theta^* \psi\left(\beta^*_g|\lambda_1\right) + (1-\theta^*)\psi\left(\beta^*_g|\lambda_0\right)}$
      $l^*_g = \lambda_1 p^*_g + \lambda_0 \left(1 - p^*_g\right)$
      $\beta^*_g \leftarrow \frac{1}{n}\left(1 - \frac{\sigma^{2*} l^*_g}{\|z_g\|_2}\right)_+ z_g I(\|z_g\|_2 > \Delta)$
    end if
    update Z
    diff = $\sqrt{\left(\beta^* - \beta^*_{old}\right)^T \left(\beta^* - \beta^*_{old}\right)}$
    if g %% M == 0 **then**
      $\theta^* = \frac{a + Z - length(forceGroup)}{a + b + G - length(forceGroup)}$
    end if
    $\sigma^{*2}_{tempt} = sum\left(\left(Y - \tilde{X}\beta^* - intercept\right)^2\right)/(n+2)$
    if updateSigma=F & $(\sigma^{*2}_{tempt} < \sigma^2_{min} \mid \sigma^{*2}_{tempt} > 100var(Y))$ **then**
      set σ^{*2} back to the initial value
      $\beta^* = 0_p$
      converged = F
      diff = 0
    end if ▷ Check to make sure algorithm doesn't explode for too small λ_0
    if updateSigma=T & $(\sigma^{*2}_{tempt} < \sigma^2_{min} \mid \sigma^{*2}_{tempt} > 100var(Y))$ **then**
      set σ^{*2} and β^* back to initial values
      updateSigma=F
      diff = 10*error
      counter=0
    end if
    if number of non-zero coefficient in $\beta^* \ >= \ min(n, p)$ **then**
      set σ^{*2} back to the initial value
      $\beta^* = 0_p$
      converged = F
      break
    end if
  end for
  counter = counter + 1
end while

Algorithm 3 SSGL - Part 3

63: **for** g in 1:G **do**
64: re-standardize β^*
65: **end for**
66: re-standardize the intercept
67: **if** converged == T **then**
68: $\beta_{start} = \beta^*$
69: **if** updateSigma **then**
70: $\sigma^2_{start} = \sigma^{*2}$
71: **end if**
72: **end if**
73: **end for**
74: estimate σ^2

Complexity Analysis of this algorithm:

In this algorithm, there are several places we can improve:

1. For each l in the loop on line 3, we apply the SVD on X once. This operation repeated L times, which is not necessary, and we can remove the loop from line 4 to line 6 before the loop on line 3.
2. There are multiple layers of loops, and re-implementing these loops with Rcpp can highly improve the efficiency.

3 Simulation Studies

4 Conclusion

References

Bai, R., G. E. Moran, J. L. Antonelli, Y. Chen, and M. R. Boland (2020, Jun). Spike-and-slab group lassos for grouped regression and sparse generalized additive models. *Journal of the American Statistical Association*, 1–14.

Ročková, V. (2018, 02). Bayesian estimation of sparse signals with a continuous spike-and-slab prior. *Ann. Statist. 46*(1), 401–437.

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

Xu, X. and M. Ghosh (2015, Dec). Bayesian variable selection and estimation for group lasso. *Bayesian Analysis 10*(4), 909–936.

Yuan, M. and Y. Lin (2006). Model selection and estimation in regression with grouped variables. *Journal of the Royal Statistical Society: Series B (Statistical Methodology) 68*(1), 49–67.