Bayesian Variable Selection and Estimation for Group Lasso

Authors: Xiaofan Xu and Malay Ghosh

Presented by: SANNA TOURAY

Date: 11/21/2024

Objective of the Study

- **Aim:** Enhance group lasso via Bayesian modeling with spike and slab priors.
- **Goals:** Accurate group and bi-level (group and within-group) variable selection.

Introduction to Group Lasso

- **Group lasso** is used in regression for grouped variable selection.
- In a regression model, a multi-level categorical predictor is usually represented by a group of dummy variables.
- Commonly applied in genetics, imaging, etc., where predictors form natural groups.
- **Lasso** :mean Least absolute shrinkage and selection operator

Motivation for Bayesian Approaches

- In many applications(e.g. genomics, finance), predictors often form grouped naturally(e.g. gene pathway in biological data, economic sector)..
- Bayesian methods provide credible intervals, better probabilistic interpretation.
- **Challenges with traditional lasso**: lacks standard errors, variable selection issues

Model Formulation of BGL-SS

- Mathematical model includes group structure and priors.
- **Regression Model** is defined $Y|X,\beta,\sigma^2 \sim N_n(X\beta,\sigma^2 I_n)$
- Where, **Y** is the response vector
- **X** is the design matrix, $\boldsymbol{\beta}$ *is the coefficient vector* ,$\boldsymbol{\sigma^2}$ is the variance
- **Prior Distribution:** a hierarchical model with spike and slab prior for the coefficients:

$$\boldsymbol{\beta}_g|\sigma^2,\tau_g^2 \overset{ind}{\sim} (1-\pi_0)\boldsymbol{N}_{m_g}(\mathbf{0},\sigma^2\tau_g^2\boldsymbol{I}_{m_g}) + \pi_0\delta_0(\boldsymbol{\beta}_g),$$

Prior distribution con’t

- π_0 is the probability of a coefficient being exactly zero(**the spike**). The **slab** is represented by a normal distribution allowing for non-zero coefficient.

Advantages of Posterior Median in BGL-SS

- Consistency in selecting true variables.
- Achieves lower false positive rates than other methods
- Superior variable selection and predictive performance

Bi-Level Selection Approach

- Extends BGL-SS to select variables at both group and within-group levels.

a. **Group levels**: Select relevant groups of predictors.
b. **Within-group level:** Select individual predictors within the chosen groups.

- Important for complex data ,such as genetic studies.

Bayesian Sparse Group Selection with Spike and Slab (BSGS-SS)

- Model formulation for bi-level selection.
- Uses independent priors at both group and within-group levels.
- To enable shrinkage both at the group level and within a group , we propose the following Bayesian hierarchical model which we refer to as Bayesian sparse group lasso(BSGL).

$$\boldsymbol{Y} | \boldsymbol{\beta}, \sigma^2 \sim N\left(\boldsymbol{X}\boldsymbol{\beta}, \sigma^2 \boldsymbol{I}_n\right),$$
$$\boldsymbol{\beta}_g | \boldsymbol{\tau}_g, \gamma_g, \sigma^2 \sim N\left(\boldsymbol{0}, \sigma^2 \boldsymbol{V}_g\right), \quad g = 1, \ldots, G,$$

Gibbs Sampling for Posterior Inference

- Gibbs sampling used for posterior inference in BGL-SS and BSGS-SS.
- The Gibbs sampler we used to generate from the posterior distribution is given below
- Let $\boldsymbol{\mu}_g = \boldsymbol{\Sigma}_g \boldsymbol{X}_g^T(\boldsymbol{Y} - \boldsymbol{X}_{(g)}\boldsymbol{\beta}_{(g)}), \boldsymbol{\Sigma}_g = (\boldsymbol{X}_g^T\boldsymbol{X}_g + \frac{1}{\tau_g^2}\boldsymbol{I}_{m_g})^{-1}$, then the conditional posterior distribution of $\boldsymbol{\beta}_g$

Is a spike and slab distribution,

$$\boldsymbol{\beta}_g|\text{rest} \sim (1 - l_g)\boldsymbol{N}(\boldsymbol{\mu}_g, \sigma^2\boldsymbol{\Sigma}_g) + l_g\delta_0(\boldsymbol{\beta}_g), \quad g = 1, \ldots, G,$$

where

$$\begin{aligned} l_g &= p(\boldsymbol{\beta}_g = 0|\text{rest}) \\ &= \frac{\pi_0}{\pi_0 + (1 - \pi_0)(\tau_g^2)^{-\frac{m_g}{2}}|\boldsymbol{\Sigma}_g|^{\frac{1}{2}}\exp\left\{\frac{1}{2\sigma^2}\|\boldsymbol{\Sigma}_g^{\frac{1}{2}}X_g^T(\boldsymbol{Y} - \boldsymbol{X}_{(g)}\boldsymbol{\beta}_{(g)})\|_2^2\right\}} \end{aligned}$$

Simulation

We simulate data from the following true model

$$\boldsymbol{Y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}, \text{ where } \epsilon_i \overset{iid}{\sim} N(0, \sigma^2), \; i = 1, 2, \ldots, n.$$

For the following examples, we compare the variable selection accuracy and prediction performance of BGL-SS, BSGL, BSGS-SS with 4 other models: linear regression, the Group Lasso (GL), the Sparse Group Lasso (SGL) and the Bayesian Group Lasso (BGL), when applicable. Five examples are considered in our simulations

Simulation Studies and Experimental Setup

To compare BGL-SS, BSGS-SS with frequentist methods (GL, SGL).

	BGL-SS		BSGS-SS		GL	SGL
	MTM	HPPM	MTM	HPPM		
Example 1						
TPR	0.96	0.98	0.79	0.89	0.97	0.90
FPR	0.23	0.48	0.09	0.19	0.65	0.53
Example 2						
TPR	0.90	0.91	0.82	0.92	0.98	0.87
FPR	0.06	0.12	0.02	0.02	0.39	0.16
Example 3						
TPR	1.00	1.00	1.00	1.00	1.00	1.00
FPR	0.00	0.00	0.02	0.03	0.44	0.26
Example 4						
TPR	1.00	1.00	1.00	1.00	1.00	1.00
FPR	0.34	0.34	0.22	0.34	0.79	0.32
Example 5						
TPR	0.97	0.99	0.91	0.94	0.99	0.94
FPR	0.14	0.54	0.02	0.02	0.40	0.30

Table 1: Mean True/False Positive Rate for six methods in five simulation examples, based on 50 simulations.

Results: Variable Selection Accuracy

- Comparison of true and false positive rates for BGL-SS, BSGS-SS, GL, and SGL.
- **Lower False Positive Rates**: The BGL-SS method significantly reduces the number of false positive compared to the group lasso.
- **Improved Variable Selection Accuracy:**The BGL-SS method demonstrates better accuracy in selecting relevant variables.

Results: Prediction Performance

	Example 1	Example 2	Example 3	Example 4	Example 5
BGL-SS with mean	9.69(0.35)	6.79(0.39)	6.45(0.29)	6.41(0.34)	5.24(0.17)
BGL-SS with median	9.76(0.40)	6.60(0.43)	6.46(0.25)	6.40(0.32)	5.08(0.18)
BSGS-SS with mean	10.07(0.38)	5.51(0.21)	6.83(0.42)	5.37(0.15)	4.83(0.16)
BSGS-SS with median	10.37(0.34)	5.59(0.32)	6.51(0.38)	5.38(0.12)	4.92(0.15)
Group Lasso	9.82(0.51)	5.99(0.33)	5.91(0.38)	6.98(0.46)	5.30(0.16)
Sparse Group Lasso	10.48(0.55)	5.75(0.45)	6.88(0.34)	5.90(0.28)	5.22(0.23)
Bayesian Group Lasso	10.53(0.34)	8.24(0.51)	7.89(0.24)	7.48(0.41)	6.46(0.23)
Bayesian Sparse Group lasso	10.08(0.47)	10.55(0.56)	10.21(0.37)	8.65(0.41)	6.03(0.16)
Linear Regression	11.19(0.42)	–	12.71(0.96)	12.68(1.03)	8.71(0.54)

Table 2: Median mean squared error for nine methods in five simulation examples, based on 50 replications.

Con’t of Prediction performance

- The Bayesian Group Lasso with spike and Slab (BGL-SS) and Bayesian Sparse Group Selection with Spike and Slab(BSGS-SS) generally perform comparably or better than traditional frequentist group lasso and sparse group lasso (GL and SGL) models in prediction accuracy.
- Posterior mean estimator and posterior median estimator have close prediction error.
- BGL and BSGL does not predict as well as their frequentist counterpart, GL and SGL

Conclusion

The Bayesian methods, especially **BGL-SS and BSGS-SS**, provide effective prediction performance with lower MSE than **traditional GL and SGL** in Scenarios with high noise, high dimensionally, and complex grouping. The use of spike and slab priors and adaptive shrinkage makes them particularly suitable for datasets with varying group structures and sparsity requirements.

Floor Open For Questions

- Thank you for your Kind Attendance !!!