**THE BAYESIAN LASSO**

**SUMMARY**

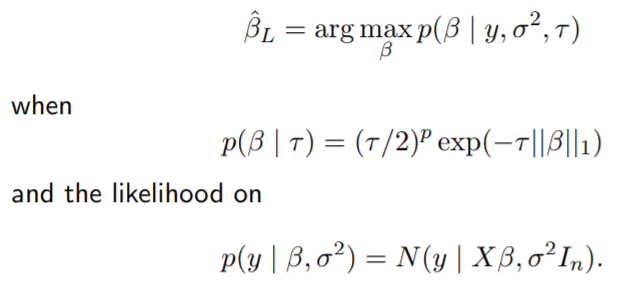
"The Bayesian Lasso" by Trevor Park and George Casella implements the Bayesian Lasso considering the linear models by assigning the parameters on the scale mixture of normal (SMN) priors and variances by their exponential priors which are independent. These parameters can be explained as a Bayesian posterior mode estimate, which can also be obtained by using conjugate normal priors on regression parameters and hyperpriors that are exponentially independent in a Gibbs Sampler. This paper gives formulation of comprehensive hierarchical model and an overview of how Gibbs sampler is implemented. It also summarizes the ways to select the Lasso parameter by providing both Bayesian and likelihood methodologies. Also, the methods characterized can be implemented to other methods relating to estimation of Lasso, such as bridge regression, Huberized Lasso and robust variants. The study presents a brief explanation of the ordinary Lasso, ridge regression and Bayesian lasso regression in consideration of their model formulation and deriving insightful results using these models. It also highlights the practical implementation of Gibbs Sampler for Bayesian lasso and provides a theoretical approach of methods to choose the lambda values. The paper then talks about the ways of choosing the Bayesian Lasso parameter, namely the empirical Bayes through marginal maximum likelihood and by using an appropriate hyperprior.

In this literature review project, we wanted to implement the Bayesian Lasso regression model and its methods in comparison to the ordinary least square regression, ridge regression and ordinary lasso regression. As a result, we implemented the diabetes example given in this paper onto a real-world dataset to gain a proper understanding of how the models work and then comparatively estimated the results via graphs and plots.

**Related Work**

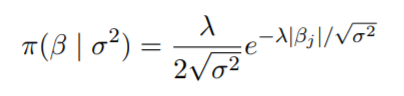
**The Bayesian Lasso**

Specifically, the lasso estimate can be viewed as the mode of the posterior distribution of β,



For any values of σ^2 > 0 and τ > 0, the posterior mode of β is the lasso estimate with penalty λ=2τ(σ^2).

The Bayesian Lasso was inspired by a conditional Laplace prior, where



This formula derived the conclusion that lack of unimodality slows the convergence of Gibbs sampler and makes point estimates less important.

**Implementation of diabetes example from the paper**

This paper puts into effects the methodologies on the lars diabetes dataset. It consisted data of 442 patients which were measured on 10 baseline variables like age, sex, blood pressure etc. The response variable is a measure of disease progression one year after the baseline.

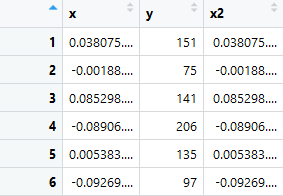
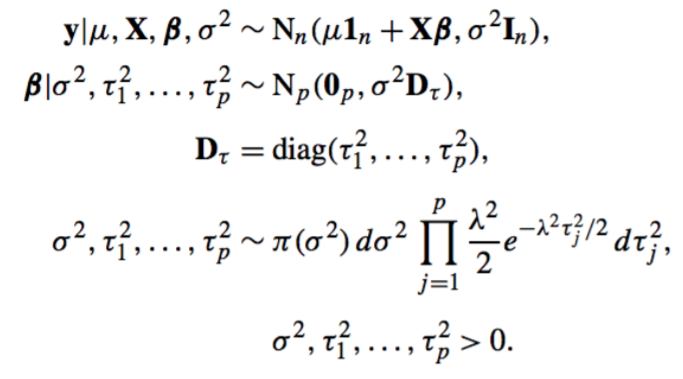


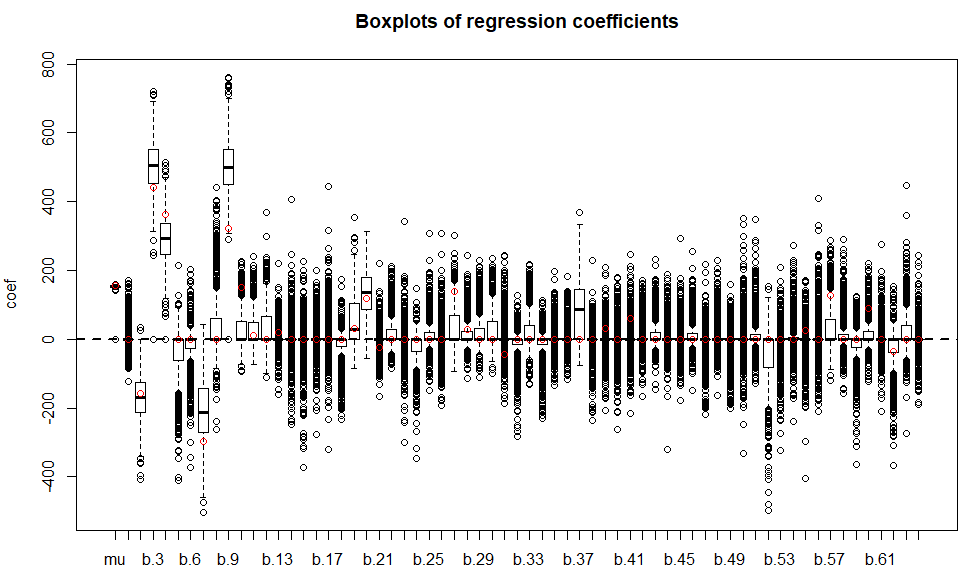
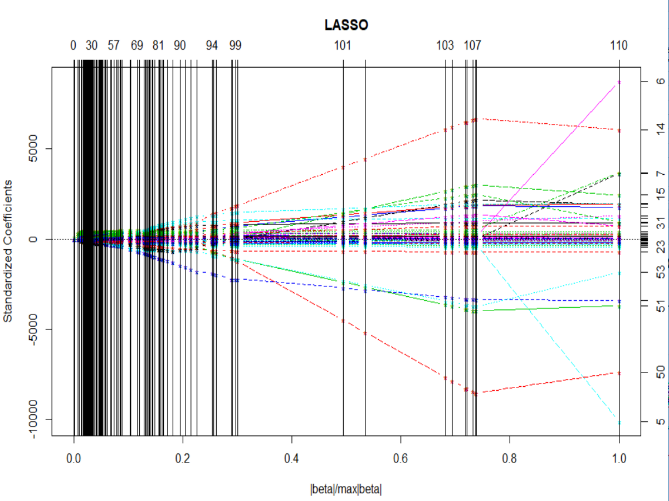
Fig. Screenshot of diabetes dataset.

In order to implement the regression techniques on this dataset, the packages that were used are 'monomvn' (Estimation for Multivariate Normal and Student-t Data with Monotone Missingness) and 'lars' (Least Angle Regression, Lasso and Forward Stagewise). The monomvn package provides the evaluation of multivariate normal data and student-t data of randomly chosen dimensions where there exists a monotonous pattern of missing data. The present version of this provides a full Bayesian approach that applies scale-mixtures for Gibbs Sampling. In short, it is completely a functional standalone interface to the Bayesian Lasso regression technique. Whereas, the lars package provides effective methods to fit an entire sequence of lasso which requires single least squares fit. Basically, Lasso is related with the least angle regression and infinitesimal forward stagewise regression. The full model implementation is as shown below.

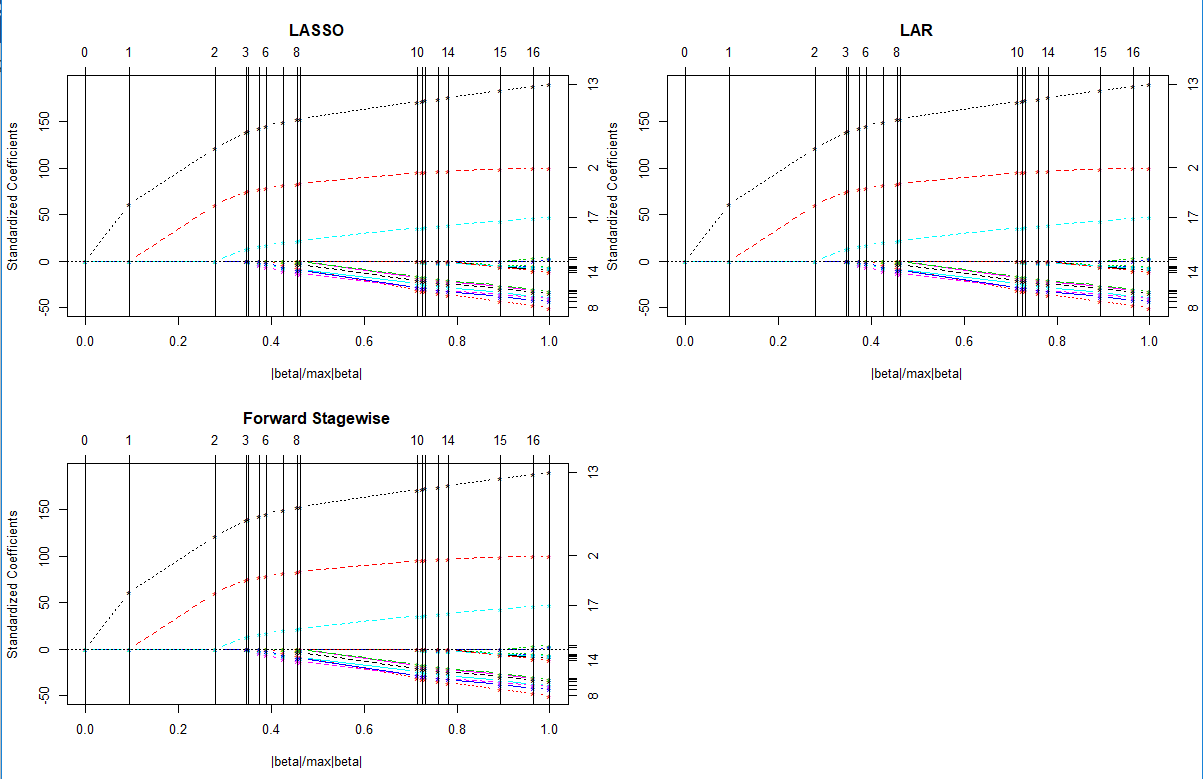


This study made use of the improper prior density π (σ^2) = 1/(σ^2), but any inverse-gamma prior for σ^2 also would maintain conjugacy.

The lars diabetes example application and the plots demonstrate a minute difference between Lasso and Bayesian Lasso by comparing and contrasting both the regression techniques on measure of their interval estimates, error in variance and posterior estimates. The first plot shows an example of how the standardized coefficients are plotted along the relative L1 norm and the second plot shows the regression coefficients plotted using the Bayesian lasso function, over their respective means.



The following plots gives a detailed explanation of the lasso estimates by using the lars package, which is implementing the least angle regression and forward stagewise mechanism. It shows how the coefficients are distributed on the relative L1 norm scale.



The below figure represents the trace plots for estimates of the diabetes data regression parameters versus the relative L1 norm, with vertical lines for the Lasso and Bayesian Lasso indicating the estimates chosen by n-fold cross-validation and marginal maximum likelihood. The Bayesian Lasso estimates were posterior medians computed over a grid of λ values, using 10,000 consecutive iterations of the Gibbs sampler of Section 2 (after 1,000 burn-in iterations) for each λ.

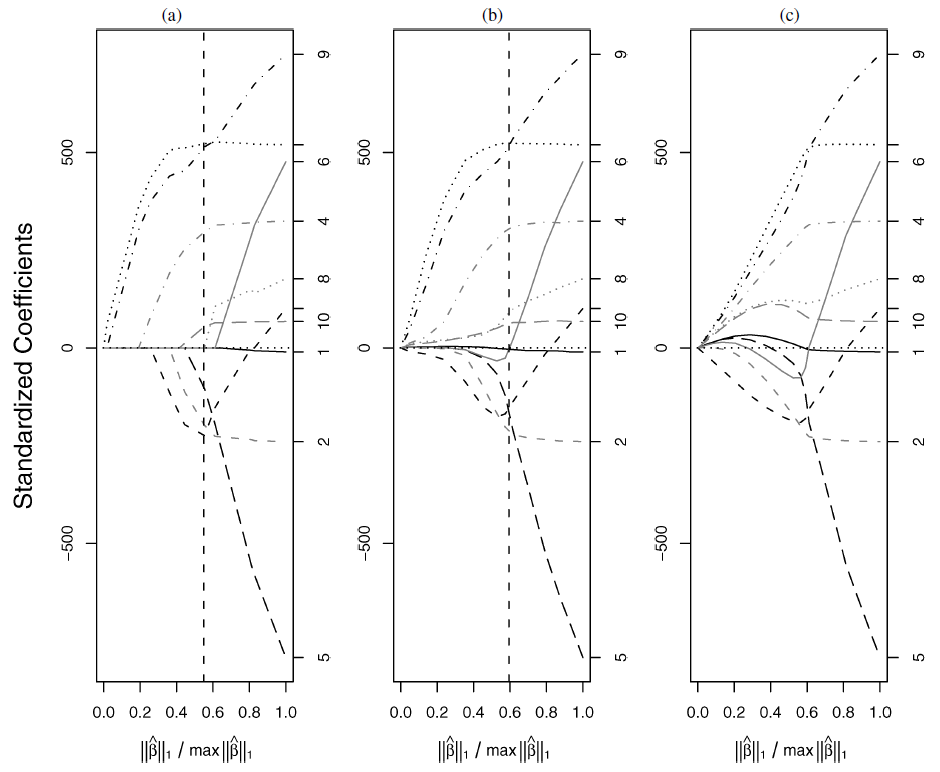
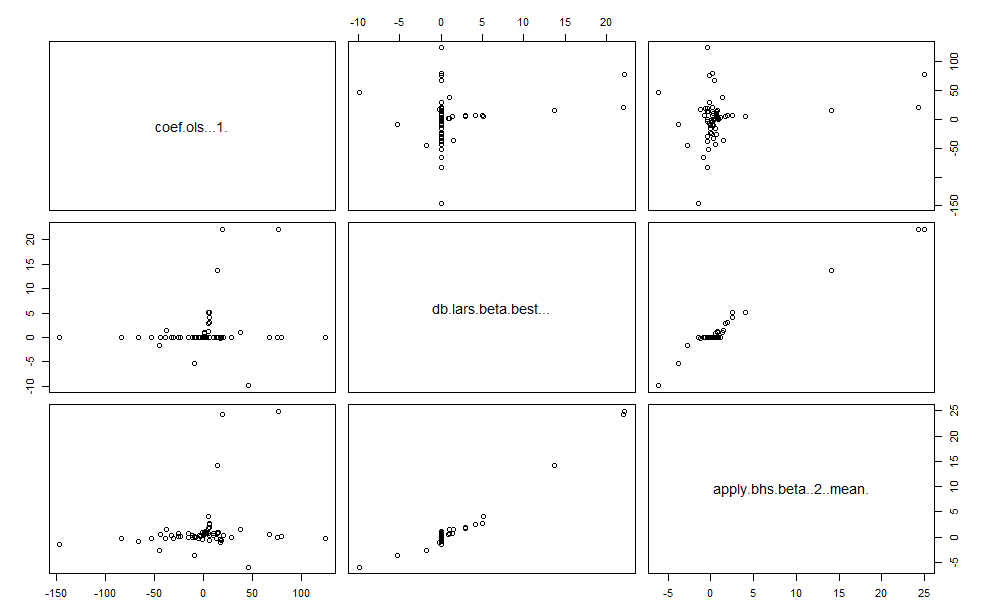
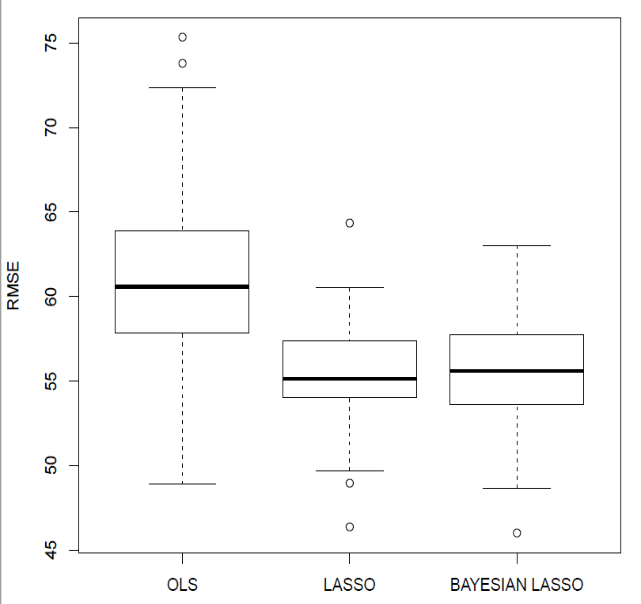


Fig (a) Lasso (b) Bayesian Lasso and (c) Ridge regression

The above figure compares the posterior median estimates of Bayesian Lasso with the ordinary lasso regression and ridge regression estimates of the diabetes dataset that has n = 442 and p=10. The plots above explain that the paths of all these estimates are varied as per their respective shrinkage parameters. As a matter of comparison, each model is plotted as a function of its relative L1 norm.



The above plots portray the boxplot of root mean square errors of the three models and also compares the coefficient estimates of all the aforementioned models.

**Example Results**

Results from the Bayesian Lasso are strikingly similar to those from the ordinary Lasso. Although more computationally intensive, the Bayesian Lasso is easy to implement and automatically provides interval estimates for all parameters, including the error variance.

**Implementation of the above diabetes example on a real-world dataset**

The dataset we chose to work on is from Kaggle. We chose a dataset from health analytics dataset in order to correlate with the diabetes dataset. This file consisted of 18 columns, which were general health status, blood cholesterol, age, sex, bmi, strength etc. and n = 7594.

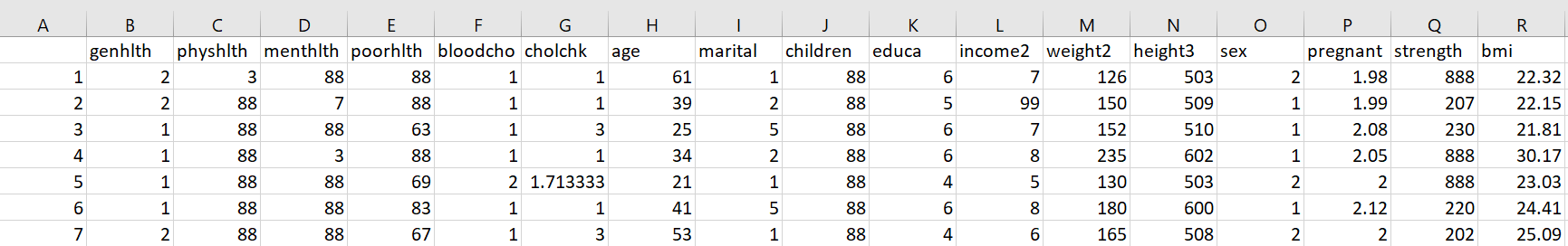
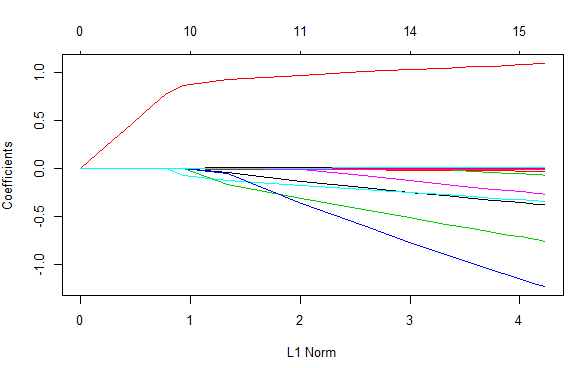
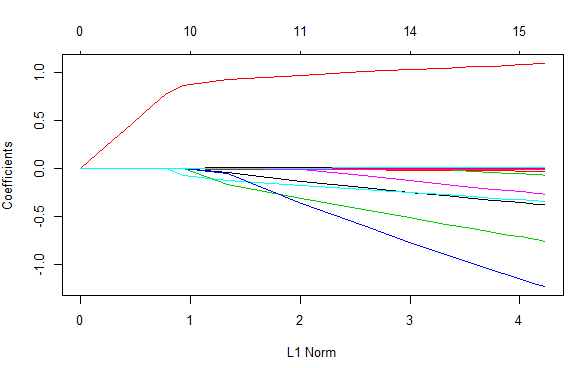
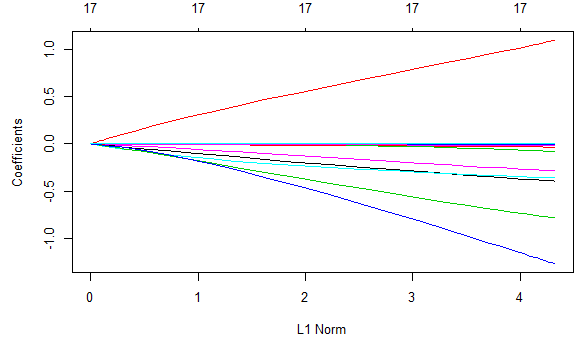


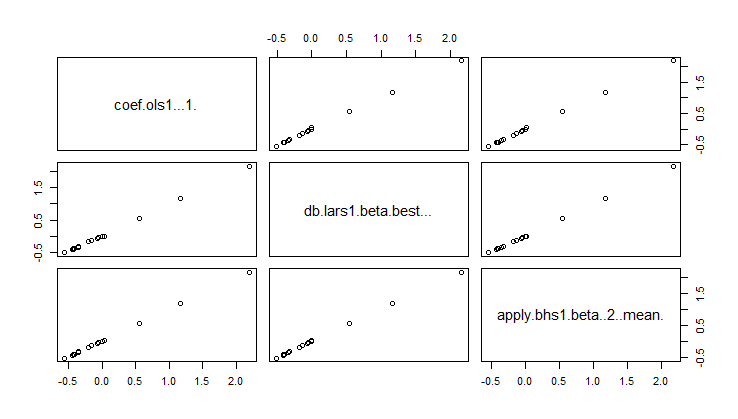
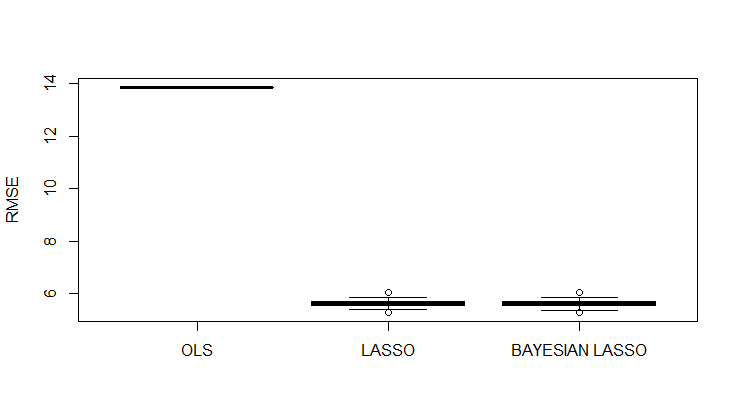
Fig: Screenshot of the new dataset used

The above dataset is a cleaned dataset. Originally, there were 208 columns. So, we applied 'missforest' method to clean the dataset and removed all the missing values. After cleaning, we performed correlation to find out the highly correlated columns.

As in the diabetes dataset, there was only 1 predictor and response variables. But, in this dataset we had to form a matrix of the predictor variables and 1 continuous response variable. Here, we chose the variable 'bmi' as the response variable and the remaining columns as the predictor variables. After choosing our variables, we then fitted our models.

   Fig (1) Lasso (2) Bayesian Lasso (3) Ridge regression

The above figure compares the posterior median estimates of Bayesian Lasso with the ordinary lasso regression and ridge regression estimates of the new health dataset that has n = 7594 and 18 columns. The plots above explain that the paths of all these estimates are varied as per their respective shrinkage parameters. As a matter of comparison, each model is plotted as a function of its relative L1 norm.



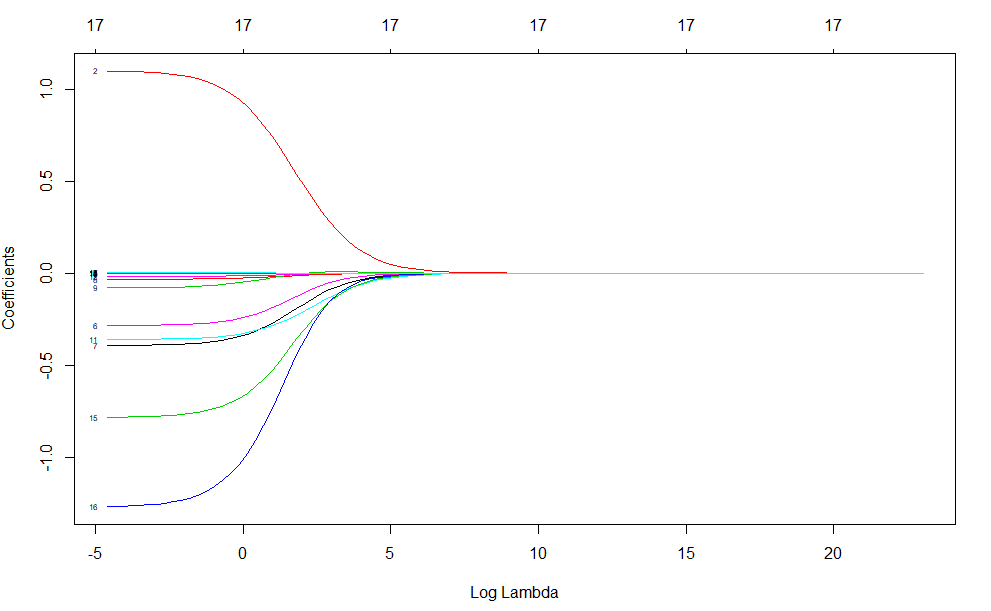
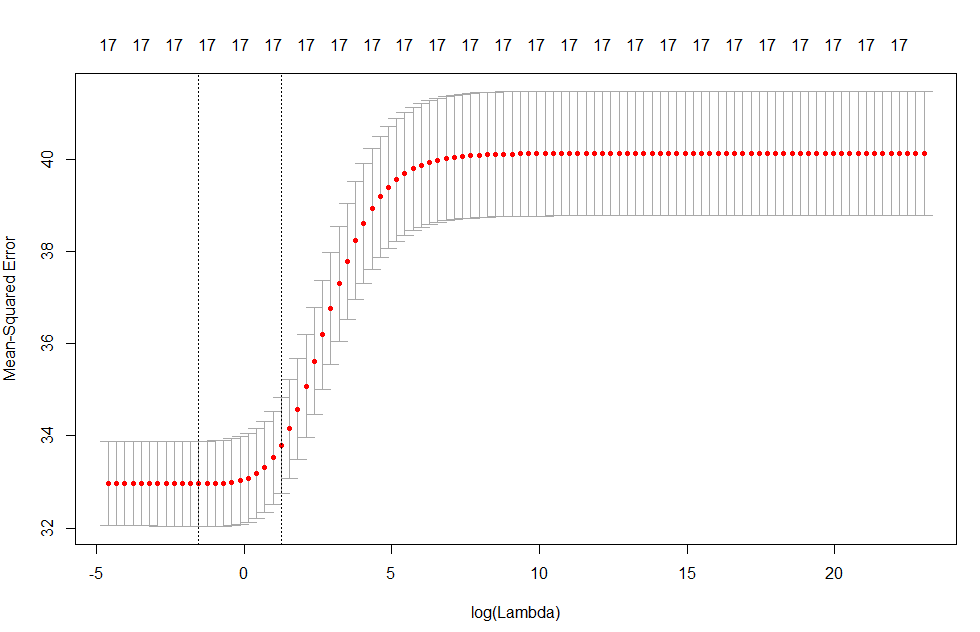
The above plots show the boxplot of root mean square errors of the three models and also compares the coefficient estimates of all the aforementioned models on our real dataset.

**Results derived**

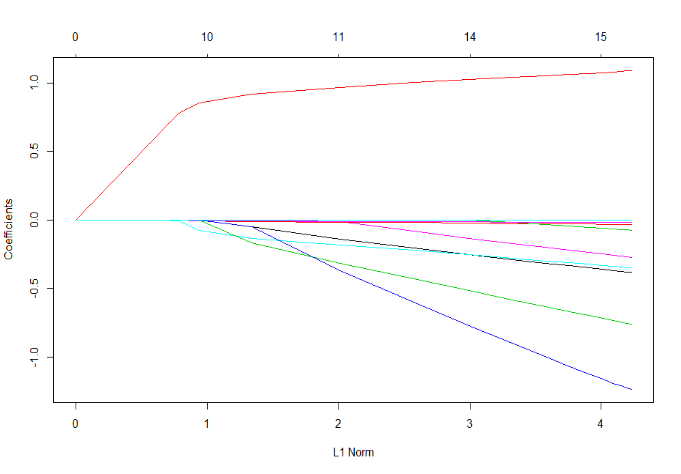
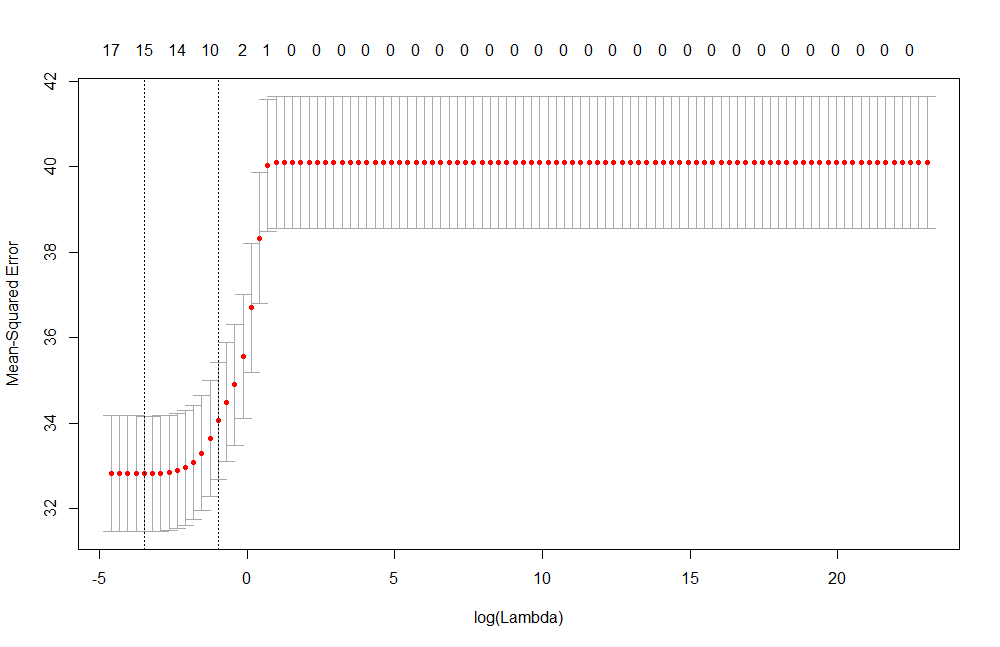
The posterior estimates of the lasso and ridge regression were compared to the Bayesian Lasso regression and interestingly the Bayesian Lasso posterior estimates were found to be quite similar to the ordinary Lasso Regression estimates, leaving a small amount of change. Not only the Lasso, but also the Bayesian Lasso seems to compromise over ridge regression too. It is found that the paths are as smooth as in ridge, specifically when there is relatively a smaller L1 norm and it was also similar in shape to the paths of Lasso. We can see that the Bayesian Lasso appears to pull the more weakly parameters to 0, rather than ridge regression. The vertical line in the model represents the estimates chosen by the marginal maximum likelihood function. We see that the least square estimates and the lasso estimates are compared and the Bayesian posterior medians are similar in value to the Lasso regression estimates.

Additionally, for our further understanding we divided our dataset into training and testing data and then performed cross-validation on the three models. Ridge regression after cross-validation gave the following plots. We use cross validation to pick the best value for lambda, the resulting plot indicates that the unregularized full model does pretty well in this case and also Ridge keeps all the variables and shrinks the coefficients to 0. In this case, the best lambda was equal to 0.2222325. After applying the model on the training dataset, we then predicted the values using our testing dataset.

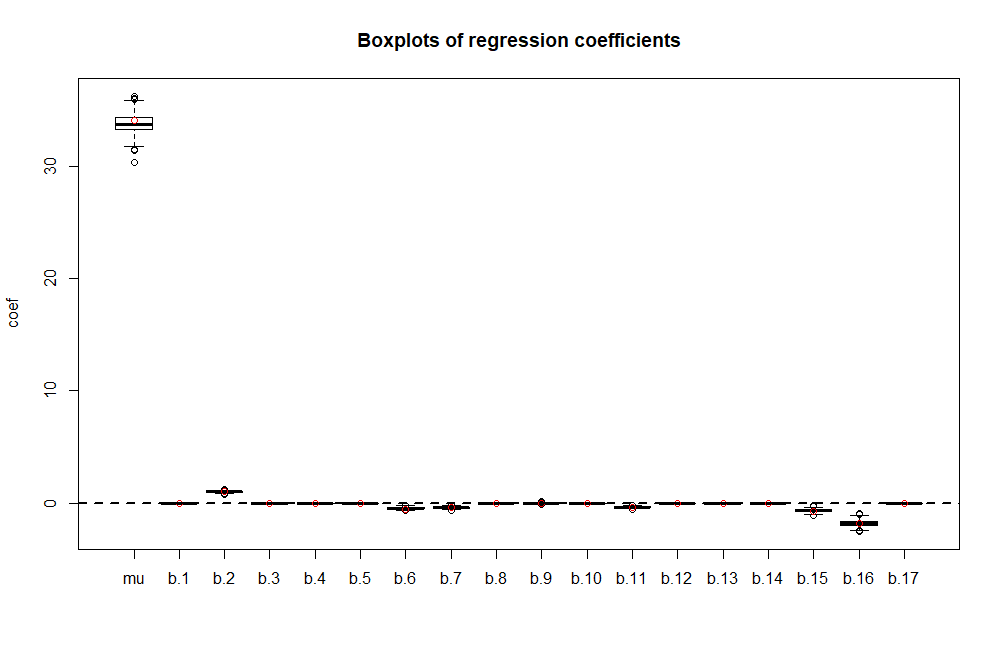
Then we compared the mean square errors of the ridge regression to the ordinary least squares. We found that Ridge performs better. According to our dataset, we found that bmi variable was highly correlated to strength, weight, cholesterol check, blood cholesterol and age.



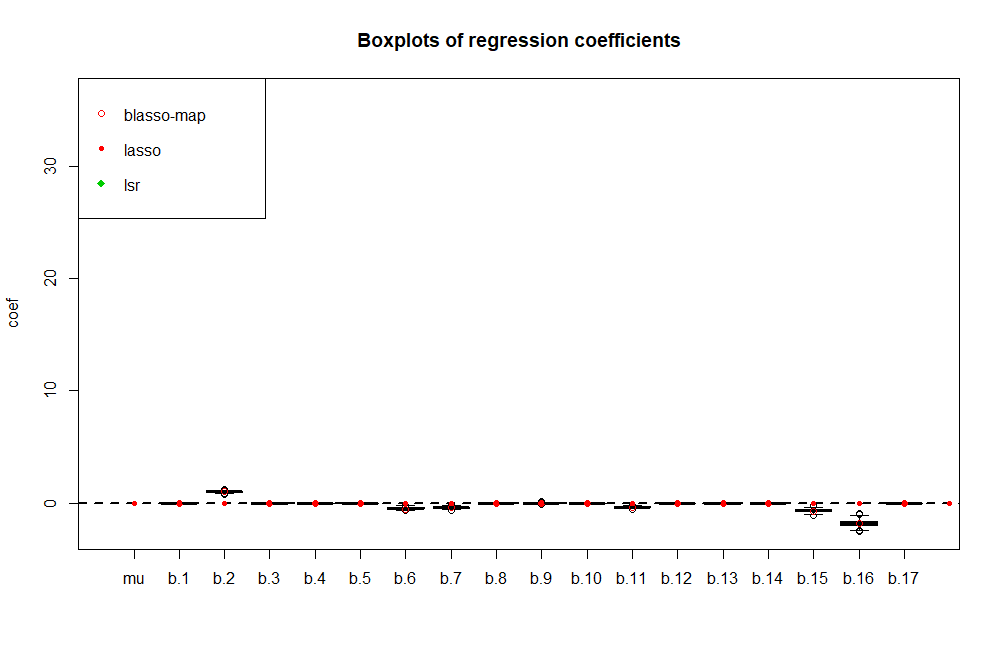
Similarly, Lasso was implemented. Here the cross-validation plot will indicate which variables to include and picks the coefficients from the best model. After predicting the values using this model, we found that the mean square error is bit higher for the lasso estimate than the ordinary least squares and ridge regression MSE's. From this, we found out that cholesterol check, age, strength and weight might be of high importance to bmi variable.



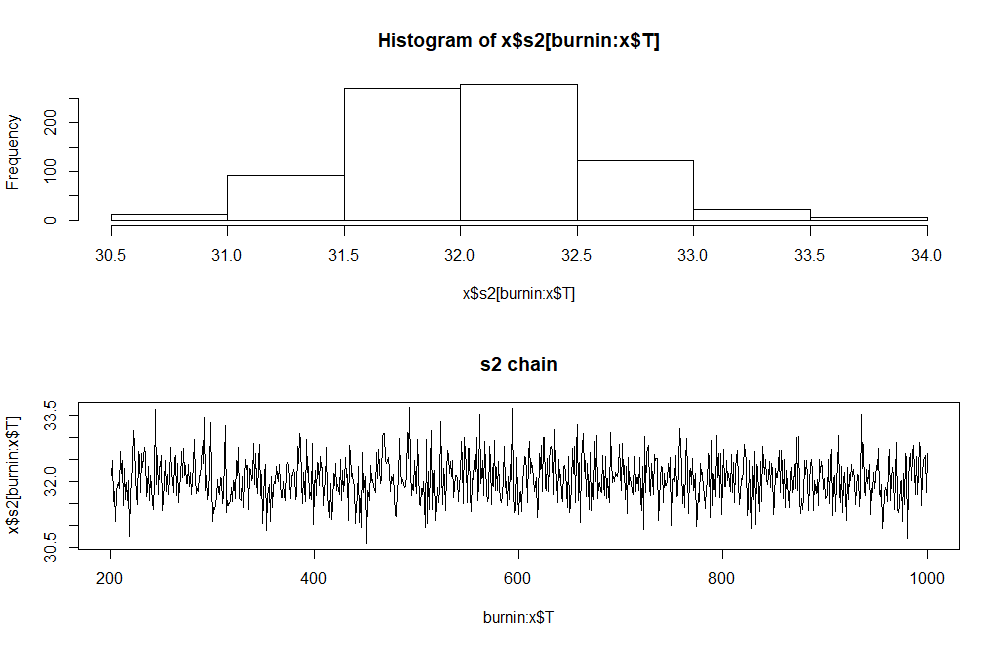
In order to apply Bayesian Lasso to our dataset, we used blasso() function. The first plot shows the regression coefficients of the Bayesian lasso model.



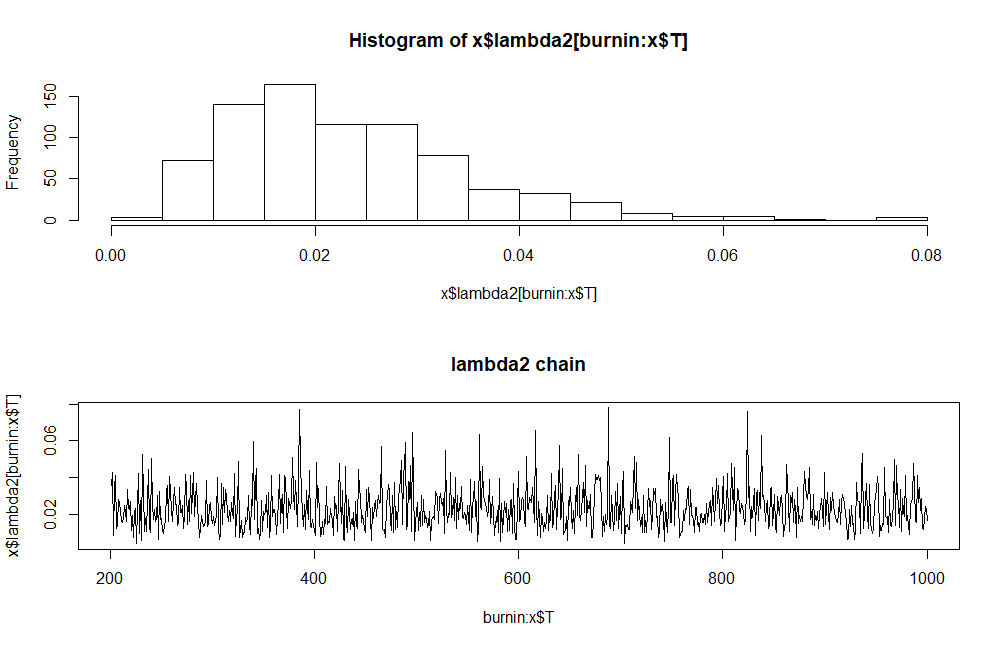
The output below shows the plot of regression coefficients which are the Beta estimates and gives a comparison of the regression parameters of the three models, namely Bayesian lasso, lasso and least square regression. We get to know that the probability of each beta coefficients is not equal to zero.



The following plot is of the initial variance parameter of Bayesian Lasso.



The following summarizes lambda2 value, which is the square of the initial lasso penalty parameter.



Thus, the mean square error estimates of all these models were evaluated and compared on their performance and also results were derived by considering the posterior estimates within the credible intervals. We finally learned about the application of ordinary least square regression, ridge regression, ordinary lasso regression and Bayesian lasso regression using this study. To conclude our work, the results from the Bayesian Lasso were strikingly similar to those from the ordinary Lasso. Although more computationally intensive, the Bayesian Lasso is easy to implement and automatically provides interval estimates for all parameters, including the error variance rather than other regression models.