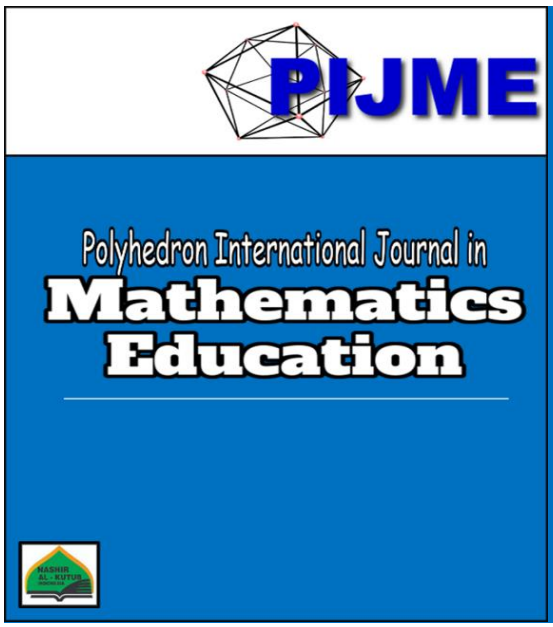
Polyhedron International Journal in Mathematics Education
Publication details, including instructions for authors and subscription information:
https://nakiscience.com/index.php/pijme



Optimizing Variable Selection with Bayesian T-Lasso Regression: A Comparison of Normal-mixture and Uniform-mixture Representations for Outlier Data Analysis

Mohammad Nasim Wafa[a], Shujauddin Shuja[b]
[a]Department of Mathematics, Faculty of Science, Herat University, Afghanistan, nasim_wafa58@yahoo.com
[b]Department of Mathematics, Faculty of Natural Science, Kabul Education University, Afghanistan, shuja.ghori@yahoo.com

To cite this article:
Wafa, M.N & Shuja, S. (2024). Optimizing Variable Selection with Bayesian T-Lasso Regression: A Comparison of Normal-mixture and Uniform-mixture Representations for Outlier Data Analysis. *Polyhedron International Journal in Mathematics Education*, *2*(1), 17-29.
To link to this article:
https://nakiscience.com/index.php/pijme
Published by:
Nasir Al-Kutub Indonesia
Residential Street Kila Rengganis, Block I, Number 11, Labuapi, Indonesia, 83361

Polyhedron International Journal in Mathematics Education
Volume 2, No 1, pp. 17-29, E-ISSN: 2987-6540
https://nakiscience.com/index.php/pijme

Original Article

Optimizing Variable Selection with Bayesian T-Lasso Regression: A Comparison of Normal-mixture and Uniform-mixture Representations for Outlier Data Analysis

Mohammad Nasim Wafa[a], Shujauddin Shuja[b]
[a]Department of Mathematics, Faculty of Science, Herat University, Afghanistan, nasim_wafa58@yahoo.com
[b]Department of Mathematics, Faculty of Natural Science, Kabul Education University, Afghanistan, shuja.ghori@yahoo.com

*Correspondence: nasim_wafa58@yahoo.com

Abstract
One of the discussed parts of the regression model was choosing the optimal model. The optimal model in regression models was chosen to determine the important explanatory variables and the negligible variables and to express the relationship between the response variable and the explanatory variables more simply with the limitations of classical variable selection processes such as step-by-step selection, and compensated regression methods. One of the compensated regression models was Lasso regression. For data collection and statistical analysis in the presence of remote observations, instead of normal distribution, T-Student distribution was used for the error of these data. In this article, we proposed a variable selection method called the T-Lasso Bayesian regression model for data analysis in the presence of outlying observations. The Bayesian t-lasso regression model, with two different representations of Laplace's prior density function for the coefficients of the regression model, was investigated, so that first the Laplace density function was discussed in the form of mixed distribution-normal scale and then in the form of mixed distribution-uniform scale. Then, by using simulation methods and real data analysis, the superiority of the Bayesian T-lasso regression method was shown by presenting the Laplace density function in a mixed-uniform scale over the normal mixed-scale display.

Article History
Received:
10 December 2024
Revised:
12 Maret 2024
Accepted:
24 April 2024
Published Online:
05 Mei 2024

Keywords:
Gibbs Sampling Algorithm;
Penalized regression;
Regression models;
Scale Mixture of uniform;
T-Lasso Regression

1. Introduction

Regression analysis is one of the most widely used methods for fitting models to data. One of the simplest and at the same time the most efficient methods for estimating regression model parameters is the method of the least squares of errors. One of the problems with the method of the least squares of errors is the ability to interpret it. On the other hand, many variables have the same behavior as other variables, or in fact, some variables are a linear combination of one or more other variables.

Hence, in dealing with such issues, a small subset of variables that have the most impact are selected and estimated. Therefore, variable selection and coefficient estimation are the most essential parts of regression modeling. Methods of estimating the lowest powers. Second, progressive variable selection, etc., does not show reliable performance when faced with data that have different characteristics. Among the damages of the model when using these methods, we can mention the lack of stability, low prediction accuracy, and incorrect selection of variables. In addition, these problems are intensified when the correlation between predictor variables is high. Contraction methods have been considered as a solution to reduce these problems, especially when the correlation between predictor variables is high. These methods estimate the regression coefficients by applying restrictions on the range of their changes. Although the existence of such limitations reduces the variance of the estimator,

Polyhedron International Journal in Mathematics Education
Volume 2, No 1, pp. 17-29, E-ISSN: 2987-6540
https://nakiscience.com/index.php/pijme

Original Article

it creates a certain amount of bias, so we can hope that the mean of the squared error will eventually decrease (Wafa, 2020; Wafa et al., 2023). Among the common contraction estimators in regression model parameter estimation, we can mention the Ridge estimator, which was introduced by Horrell and Kennard. (Hoerl & Kennard, 1970) introduced the "Ridge" regression estimator, which is the gateway to the world of "compensated estimators". It was based on the regularization method (Karapetyants & László, 2024).

Ridge regression is an introduction to the world of variable selection and estimation. This regression combats the problem of collinearity in linear models, and based on this, the compensated estimate was born. Due to the presence of all variables in the ridge estimator, its interpretation is not easily possible. Another member of this class is the Lasso estimator (Tibshirani, 1996). Instead of using compensation L2, Tibshirani minimized the second powers of the error compared to compensation function L1. This estimator led to the emergence of new estimators such as smooth truncated derivative estimator elastic net (Zou, 2006; Zou & Hastie, 2005) hard threshold (Belloni & Chernozhukov, 2013). Using the L1 compensation shrinks each coefficient to zero and makes the additional variables exactly zero. The lasso estimator performs both variable selections and shrinks the coefficients at the same time. An interesting application of Lasso estimators is in thin models (models with a large number of zero parameters). Another application of Lasso estimator is when the dimension of the parameter space is greater than the dimension of the sample space. Due to the existence of a large number of variables in high-dimensional models, the interpretation of these models is very difficult.

Therefore, the problem of variable selection plays a very important role in high-dimensional statistical modeling. Recently, many studies have been conducted in this field, for example, you can refer to (Wafa et al., 2023) and (Wafa et al., 2023; Wafa et al., 2023) pointed out. Despite the advantages of the Lasso estimator, the performance of the Lasso method as a method for choosing the optimal model is weak in the case where the observations include outlying data and the distribution of the error variable is considered normal] [5 with attention] Because outlying observations have a great effect on the fitted model and its related inferences, it is very important to use robust estimators for the presence of outlying data. Outliers are stable and considered the t-Student distribution as a suitable replacement for the normal distribution (Liu & Rubin, 1995; Shadrokh et al., 2021) in these two articles the Bayesian T-Lasso regression method is proposed in the case where the observations include outlying data or the error distribution has an abnormal behavior. In this regard, in the second part of the article, we will review the normal Lasso Bayesian regression model. In the third part, the details of T-Lasso Bayesian regression model with two different representations of Laplace density function as mixed-normal scale and mixed-uniform scale have been discussed and a hierarchical Bayes model is obtained. In the following, the Gibbs algorithm for estimating the parameters of the T-Lasso Bayesian regression model is examined.

In the fourth part, by using simulation methods and real data analysis, two methods have been studied and we will show that the Bayesian T-Lasso regression model with mixed representation-uniformity scale has a satisfactory performance in comparison with the model with mixed representation. It has a normal scale. The comparison of the models has been done based on the criteria of knowing the deviation and the mean of the squared error and the convergence of the Gibbs algorithm using the Heidelberger and Welch method. The fifth section is dedicated to the results of the article.

The Bayesian lasso regression model

Briefly, we will present the Lasso-Bayesian regression model in this section to provide context for the new material on model uncertainty presented by (Park & Casella, 2008).

$$y|\beta,\alpha^2 \sim N(X\beta,\alpha^2 I_n),$$
$$\beta_j|\alpha^2,\tau \overset{iid}{\sim} DE\left(\frac{\tau}{\alpha}\right), \qquad j=1,\dots,p$$

That here $DE\left(\frac{\tau}{\alpha}\right)$ is the double-exponential distribution with density function:

$$p\left(\beta_j\middle|\alpha^2,\tau\right)=\frac{\tau}{2\alpha}e^{-\frac{\tau|\beta_j|}{\alpha}} \qquad (1)$$

Here, we assume that the y and X columns are detrended, and therefore an intercept term is not included in the model. For a γ model with k predictor variables, the key to evaluating model uncertainty when σ^2 is known is the ability to evaluate the integral.

Polyhedron International Journal in Mathematics Education
Volume 2, No 1, pp. 17-29, E-ISSN: 2987-6540
https://nakiscience.com/index.php/pijme

Original Article

$$m_\gamma = \left(\frac{y}{\alpha^2}\right) = \int p(y|\beta_\gamma, \alpha^2) p(\beta_\gamma/\alpha^2) d\beta_\gamma$$
$$= \int (2\pi\alpha^2)^{-\frac{n}{2}} e^{-\frac{1}{2\alpha^2}(y-X_\gamma\beta_\gamma)^{\mathrm{T}}(y-X_\gamma\beta_\gamma)} \left(\frac{\tau}{2\alpha}\right)^k e^{-\tau\left|\left|\beta_\gamma\right|\right|_1/\alpha} d\beta_\gamma \qquad (2)$$

where $\left|\left|\beta\right|\right|_1$ is the L_1-norm of β.

Most of these cases and applications of the Bayesian lasso regression model ((Park & Casella, 2008; Steele & Lopez-Fernandez, 2014; Zhu et al., 2008) have focused on the composition of the scale of normal representing the bi-exponential distribution, which is embedded in it. Variables are used to create a hierarchical representation of the prior distribution. However, this model formulation adopts a simple Gibbs sampler to obtain maps of the posterior distribution of β for a fixed model.

It does not lead to a simple expression for the marginal probability. Instead of working with the scale mixture representation, we consider the direct representation of the β posterior distribution presented by Hans (2009). By breaking the density function for the bi-exponential distribution ($p\left(\beta_j\middle|\alpha^2, \tau\right) = \frac{\tau}{2\alpha} e^{-\frac{\tau\left|\beta_j\right|}{\alpha}}$) into separate positive and negative components, Hans (2009) shows that for a given set of predictor variables $p \leq n$, the posterior distribution of β is the normal-normal distribution:

$$p(\beta|\alpha^2, \tau, y) = \sum_{z\in Z_p} \omega_z N^{[z]}(\beta|\mu_z, \alpha^2\left(X^{\mathrm{T}}X\right)^{-1}$$

The amount is collected $Z_p = \{-1, 1\}^p$ that shows 2^porthants of $\mathbb{R}^p$. Urethane corresponds to a given $z\epsilon Z_p$ is defined $O_z = R_{z1} \times R_{z2} \times \ldots \times R_{zp}$ *where* R_{zp} is $[0, \infty)$ if $z_j = 1$ also if $(-\infty, 0)$ is $z_j = -1$ .Each expression and sentence as a whole contain a normalized density function for a normal distribution restricted to being in a certain orthogonal:

$$N^{[z]}(\beta|m, S) \equiv \frac{N\ (\beta|m, S)}{P(z, m\ S)} 1(\beta\epsilon O_z), \quad where \quad P(z, m\ S) = \int_{O_z} N(t|m, S) dt$$

The location vector for each term depends on the total ordinal $\mu_z = \hat{\beta}_{OLS} - \tau\sigma(X^TX)^{-1}z$, where $\hat{\beta}_{OLS}$ the least-squares estimate $(X^TX)^{-1}X^Ty$. Each term in $(p(\beta|\sigma^2, \tau, y) = \sum_{z\in Z_p} \omega_z N^{[z]}(\beta|\mu_z, \sigma^2\left(X^{\mathrm{T}}X\right)^{-1})$ also contains a weight.

$$\omega_z = \omega^{-1} \frac{P(z, \mu_z, \sigma^2\left(X^{\mathrm{T}}X\right)^{-1})}{N(0|\mu_z\sigma^2(X^{\mathrm{T}}X)^{-1})} \quad where \quad \omega = \sum_{z\epsilon Z\_p} \frac{P(z, \mu_z, \sigma^2\left(X^{\mathrm{T}}X\right)^{-1})}{N(0|\mu_z\sigma^2(X^{\mathrm{T}}X)^{-1})}$$

which makes $(p(\beta|\alpha^2, \tau, y))$ a normalized density function.

When $p > n$, the posterior distribution density function cannot be shown as $(p(\beta|\alpha^2, \tau, y))$. In this case, the probability surface (as a function of β) will be smooth in a posterior $p - n$ subspace, which means that in this subspace the posterior distribution will have exponential tails (because to the previous distribution). Along any direction not in this subspace, the posterior will have normal tails as in $(p(\beta|\alpha^2, \tau, y))$. While this complicates writing an expression for the posterior density function, it does not pose a problem to address model.

Bayesian T-Lasso regression model

Statistical inference based on normal distribution is known as vulnerable. It has robust procedures, mainly aimed at identifying outliers. After editing outliers, further analysis is often limited to least squares based on the ordinary linear model. A serious problem with this approach is that inferences from uncertainty fail in the process of elimination. In particular, the standard errors are very small. Wafa (2023) proposed a method for robust inference on regression models using the t-distribution. Its approach is to replace the normal distribution with the t- distribution in statistical models. Often, in linear regression models, the distribution of errors is considered as $\varepsilon_i \sim N(0; \sigma^2)$. In general, in addition to the normal distribution, errors can be from any other distribution. For example, T-Student's distribution, which A heavy-tailed distribution is a useful generalization of the normal distribution that can be used for statistical modeling in the presence of outlying data. Of course, the predictive consistency of Lasso regression does not require the assumption of normality of the errors, and it is sufficient that the errors have a mean of zero and (Hlavackova-Schindler, 2016) in this research, taking

Polyhedron International Journal in Mathematics Education
Volume 2, No 1, pp. 17-29, E-ISSN: 2987-6540
https://nakiscience.com/index.php/pijme

Original Article

into account the distribution of errors as), $\varepsilon_i \overset{i\cdot i\cdot d}{\sim} t_v(0;\ \sigma^2)$ to generalize the Bayesian lasso regression model for observations in the presence of outlying data and under the title of T-lasso regression model Before examining the details of Bayesian t-lasso regression model, three useful and important lemmas in this field are discussed.

Theorem1. (Showing t-Student's distribution in the form of a mixed-scale normal distribution) Suppose $y|l$ has a normal distribution with a mean of 0 and a variance of l. Also consider the distribution function l as an inverse gamma with parameters $v/2$ and $v/2$. Therefore, the marginal distribution, y, is the student's -t distribution with the degree of freedom v (Shadrokh et al., 2021).

Proof. Definition1. A random variable in the form $y \sim t_v(\mu, \sigma^2)$ say has a t distribution with the degree of freedom v if its probability density is in the form

$$f_v(y_i|\mu,\sigma^2) = \left(\frac{1}{\pi v\sigma^2}\right)^{\frac{1}{2}} \left(\frac{\Gamma\left(\frac{[v+1]}{2}\right)}{\Gamma\left(\frac{v}{2}\right)}\right)\left(1+\frac{(y_i-\mu)^2}{v\sigma^2}\right)^{-\left(\frac{v+1}{2}\right)}$$

Let μ be the location parameter and σ^2 be the scale parameter. The degree of freedom v determines the weight of the tails of the distribution. For mean$v > 1$ tistribution μ, and $v > 2$, the variance of the distribution is equal to $\frac{\sigma^2 v}{v-2}$. The special case $v = 1$ is Cauchy distribution and $v = \infty$ normal normal distribution.

Lemma: 1 to find the marginal distribution of variable y.

$$f_y(y) = \int_0^\infty f_y(l|y)dl \int_0^\infty \frac{1}{\sqrt{l}} \exp\left(-\frac{y^2}{2l}\right) l^{-\frac{v}{2}-1} \exp\left(-\frac{v}{2l}\right) dl$$

$$= \int_0^\infty \frac{1}{\sqrt{l}} \exp\, l^{-\frac{v+1}{2}-1} \exp\left(-\frac{y^2+v}{2l}\right) dl = \Gamma\left(\frac{v+1}{2}\right)\left(\frac{y^2+v}{2l}\right)^{-\frac{v+1}{2}}$$

The last line corresponds to the t-Student distribution with v degrees of freedom. It is possible to distribute t-Student $t_v(\mu, \sigma^2)$ hierarchically.

$$y|\mu,\sigma^2,l \sim N(\mu, l\sigma^2, I)$$
$$l|v \sim IG(\frac{v}{2},\frac{v}{2})$$

also showed

Lemma 2: To prove Lemma 2, it is necessary to calculate the marginal density of the random variable L as.

$$f(l) = \int_0^\infty \frac{1}{\sqrt{2\pi v}} \exp\left(-\frac{l^2}{27}\right) \times \frac{\lambda^2}{2} \exp\left(-\frac{\lambda^2 v}{2}\right) dv\ = \frac{\lambda}{2}\exp(-\lambda|l|).$$

is. By completing the square, we have the exponential expression in relation (integral):

$$f(l) = \int_0^\infty \frac{\lambda^2 e^{-\lambda|l|}}{2\sqrt{2\pi}} \cdot \frac{1}{\sqrt{v}} \exp\left\{-\frac{1}{2}\left(\frac{|l|}{\sqrt{v}} - \lambda\sqrt{v}\right)^2\right\} dv.$$

By changing the variable $u = \sqrt{v}$ and $dv = 2udu$ expression.

$$f(l) = \frac{\lambda^2 e^{-\lambda|l|}}{\sqrt{2\pi}} \int_0^\infty \exp\left\{-\frac{1}{2}\left(\frac{|l|}{\sqrt{u}} - \lambda\sqrt{u}\right)^2\right\} du \cdot$$

it will be obtained. Also, by changing the variable.

$$\eta = \frac{|l|}{u} - \lambda u, \qquad \frac{du}{d\eta} = \frac{-1+\frac{\eta}{\sqrt{\eta^2+4\lambda|l|}}}{2\lambda}$$

We have

$$f(l) = \frac{\lambda^2 e^{-\lambda|l|}}{2\lambda\sqrt{2\pi}} \int_{-\infty}^\infty \exp\left\{1 - \frac{\eta}{\sqrt{\eta^2+4\lambda|l|}}\right\} d\eta,$$

Polyhedron International Journal in Mathematics Education
Volume 2, No 1, pp. 17-29, E-ISSN: 2987-6540
https://nakiscience.com/index.php/pijme

$$\frac{\lambda^2 e^{-\lambda|l|}}{2}\int_{-\infty}^{\infty}\frac{1}{\sqrt{2\pi}}\exp\left(\frac{\eta^2}{2}\right)d\eta-\frac{\lambda^2 e^{-\lambda|l|}}{2}\int_{-\infty}^{\infty}\exp\left(-\frac{\eta^2}{2}\right)\frac{\eta}{\sqrt{\eta^2+4\lambda|l|}}d\eta=\frac{\lambda e^{-\lambda|l|}}{2}$$

that the last term follows the standard normal distribution and the integral is an odd function].

Lemma 3: We know

$$\int_{z>\frac{|x|}{\sqrt{\sigma^2}}}\lambda e^{-\lambda z}dz=e^{-\lambda\frac{|x|}{\sqrt{\sigma^2}}},$$

Therefore, the probability density function of the Laplace distribution with zero mean and variance $\sqrt{\sigma^2}/\lambda$ can be.

$$\frac{\lambda}{2\sqrt{\sigma^2}}=e^{-\lambda z}dz=ee^{-\lambda\frac{|x|}{\sqrt{\sigma^2}}}=\frac{\lambda}{2\sqrt{\sigma^2}}\int_{u>\frac{|x|}{\sqrt{\sigma^2}}}\lambda e^{-\lambda z}dz$$

$$=e^{-\lambda u}du\quad\int_{-u\sqrt{\sigma^2}<x<u\sqrt{\sigma^2}}\frac{1}{u\sqrt{\sigma^2}}\,\frac{\lambda^2}{\Gamma(2)}u^{2-1}\,e^{-\lambda u}du$$

be written and equality is proved as a result.

Theorem 2 (Showing the Laplace distribution as a mixed distribution-normal scale)

Assume that $E(1)$ denotes the standard exponential distribution with mean one, Laplace $(0,1)$ denotes the standard Laplace distribution with mean zero and variance one, and $N(\mu,\sigma^2)$ has a normal distribution with mean μ and variance σ^2. Therefore, we have:

$$V\sim 2E(1),\quad L|V\sim N(0.V)\quad\Longrightarrow\quad L\sim Laplace\ (0.1)$$

In other words, if the random variable E with standard exponential distribution is independent of the random variable Z with standard normal distribution, then.

$$L\sim\sqrt{2E}Z\sim Laplace\ (0.1).$$

Considering the scale parameter λ, we have.

$$V\sim 2/\lambda^2 E(1),\quad L|V\sim N(0.V)\quad\Longrightarrow\quad L\sim Laplace\ (0.1)$$

Proof: Refer to Lemma, 1, 2 and 3.

regression model $y_i=x_i^T\,\beta+\varepsilon_i\varepsilon_i\overset{i\cdot i\cdot d}{\sim}t_v(0;\,\sigma^2), i=1,2,3$ ... Consider n. In this model x_i^T is the $p-$dimensional vector of auxiliary variables and $\beta=(\beta_1,\beta_2\ldots,\beta_n)^T$, the vector p of the unknown guides, $\sigma^2$the scale parameter and the degree of freedom v determine the heaviness of the tails of the distribution. Considering the Laplace prior density function with mixed-scale normal representation for the vector of regression coefficients β as.

$$\prod_{j=1}^{p}\frac{1}{2\pi\tau_j^2\sigma^2}\exp(-\frac{1}{2\tau_j^2\sigma^2}\beta_j^2)\frac{\lambda^2}{2}e^{-\lambda^2/2\tau_j^2}$$

and inverse gamma density with parameters r and γ as.

$$\pi(\sigma^2)=\frac{\gamma^2}{\Gamma(r)}\left(\frac{1}{\sigma^2}\right)^{r+1}\exp\left(-\frac{\gamma}{\sigma^2}\right)$$

For the parameter, σ^2 hierarchical Bayes model as

$$y|X,\beta,v,l\sim N(X\beta,l\sigma^2,I_n),\quad l|v\sim IG\left(\frac{v}{2},\frac{v}{2}\right),\quad \beta_i|\tau_1^2,\tau_2^2,\ldots.\tau_p^2\sim N(0p,\sigma^2D_\tau),$$

$$D_\tau=Diag\left(\tau_1^2,\tau_2^2,\ldots.\tau_p^2\right),\qquad j=1,2.3\ldots.p,$$

$$\tau_j^2\sim\frac{\lambda^2}{2}e^{-\frac{\lambda^2\tau_j^2}{2}}$$

$$\sigma^2\sim G(r,\gamma)$$

we will have by using Hierarchical Bayes model, complete posterior conditional distributions for parameters of Bayesian T-lasso regression model to implement Gibbs algorithm as

$$\beta\big|y,X,\sigma^2,\tau,l\sim N_p((X'L^{-1}X,+D\tau^{-1})^{-1}y,\sigma^2\,(X'L^{-1}X,+D\tau^{-1})^{-1}),$$

Polyhedron International Journal in Mathematics Education
Volume 2, No 1, pp. 17-29, E-ISSN: 2987-6540
https://nakiscience.com/index.php/pijme

Original Article

$$l|y,X,\sigma^2 \sim \prod_{i=1}^{n} I\Gamma\left(\frac{1}{2}+\frac{v}{2},\frac{(y_i-X\beta)}{2\sigma^2}+\frac{v}{2}\right)$$

$$\sigma^2|y,X,\beta,\tau,\gamma,r \sim I\Gamma\left(\frac{n}{2}+\frac{p}{2}+r,\frac{(y-X\beta)^T((y-X\beta)^T\beta^T D^{-1}\tau\beta}{2}+\gamma\right)$$

$$\left(\frac{1}{\tau_j^2}\right)|\beta,\sigma^2,\lambda \sim \prod_{i=1}^{n} IG\left(\sqrt{\frac{\lambda^2\sigma^4}{\beta_j^2}},\lambda^2\right)$$

"Considering the Laplace prior density function with a scale-mixture representation for the regression coefficient vector β and the inverse gamma density for the parameter $\sigma 2$as $\pi(\sigma^2)=1/\sigma^2$(Wafa, 2019) , the hierarchical Bayesian model is as follows

$$y|X,\beta,v,l \sim N(X\beta,l\sigma^2,I_n), \quad l|v \sim IG\left(\frac{v}{2},\frac{v}{2}\right),$$

$$\beta|u,\sigma^2 \sim \prod_{i=1}^{n} Uniform\left(-\sqrt{\sigma^2}u_j,\sqrt{\sigma^2}u_j\right)$$

$$\mathrm{u}|\lambda \sim \prod_{i=1}^{n} Gamma(2,\lambda)$$

$$\sigma^2 \sim \pi(\sigma^2)$$

we will have Based on the hierarchical Bayes model, the posterior distribution of all parameters is equal

$$\pi(\beta,u,\lambda,\sigma^2,v,l|y,x)\ \alpha\ \pi|x,\beta\ u,\lambda,\sigma^2,v,l)\pi(\beta|u,\sigma^2)\pi(u|\lambda)\pi(l|v)\pi(\sigma^2)$$

$$\alpha \prod_{i=1}^{n}\left[\frac{1}{\sqrt{2\pi l_i\sigma^2}}\exp\left(-\frac{1}{2l_i\sigma^2}(y_i-X\beta)^2\right)\times\frac{\left(\frac{v}{2}\right)^{\frac{v}{2}}}{\Gamma\left(\frac{v}{2}\right)}l_i^{-\frac{v}{2}-1}\exp\left(\frac{-\frac{v}{2}}{l_i}\right)\right]$$

$$\times\prod_{j=1}^{p}\frac{1}{\sqrt{\sigma^2}}\mathrm{I}\left\{|\beta_j|<\sqrt{\sigma^2}\mathrm{u_j}\right\}\mathrm{e}^{-\lambda\mathrm{u_j}}\times\frac{1}{\sigma^2}$$

Is. By using Hierarchical Bayes model, it is possible to obtain complete posterior conditional distributions for the parameters of the model for the implementation of the Gibbs algorithm. By introducing $u\ =\ (u_1,u_2,\ldots u_p)$, the complete posterior conditional density functions are as

$$\beta|y,X,u,\lambda,\sigma^2,v,l \sim N_p((X'L^{-1}X)^{-1})\times\prod_{j=1}^{p} I\{|\beta_j|<\sqrt{\sigma^2}u_j\},$$

$$u\left|y,X,\beta,,\lambda,\sigma^2,v,l \sim \prod_{j=1}^{p} e^{-\lambda u_j} I\left\{u_j>\frac{|\beta_j|}{\sqrt{\sigma^2}}\right\} \sim \prod_{j=1}^{p} Exponnential\ (\lambda) I\left\{u_j>\frac{|\beta_j|}{\sqrt{\sigma^2}}\right\},\right.$$

$$\sigma^2|y,X,\beta,u,\lambda,l \sim I\Gamma\left(\frac{n-1+p}{2},(\frac{1}{2}(y-X\beta)^T L^{-1}(y-X\beta)))I(\sigma^2>\max_j(\frac{\beta_j^2}{u_j^2})\right),$$

$$l|y,X,\beta,u,\lambda,\sigma^2 \sim \prod_{i=1}^{n} I\Gamma\left(\frac{1}{2}+\frac{v}{2},\frac{y_i-X\beta}{2\sigma^2}+\frac{v}{2}\right)$$

Are that $I(0)$ is the indicator function. Using the model's hierarchical display, the posterior density function for λ under the condition of β is as

$$\pi(\lambda|\beta)\alpha\lambda^{2p}\exp\{-\lambda\sum_{j=1}^{p}|\beta_j|\pi(\lambda)$$

Is. Considering the gamma prior density with parameters a and b for λ, the conditional posterior distribution is also the gamma distribution as.

Polyhedron International Journal in Mathematics Education
Volume 2, No 1, pp. 17-29, E-ISSN: 2987-6540
https://nakiscience.com/index.php/pijme

Original Article

$$\lambda|y, X, \beta, \sigma^2 \alpha \lambda^{\alpha+2p-1} \exp\{-\lambda(b + \sum_{j=1}^{p} |\beta_j|\},$$

it will be obtained. Therefore, the adjustment parameter λ is updated along with other parameters of the model using the Gibbs algorithm and generating samples of gamma distribution with parameters $a + 2p$ and $b + \sum_{j=1}^{p} |\beta_j|$. In this article, the adjustment parameter λ is estimated as the mean of the posterior distribution and considering the values of $a = 1$ and $b = 0$ for the prior density parameters.

2. Method

The main objective of this section is to evaluate the performance of two Bayesian t-lasso regression methods with different scale-mixture representations: (1) Uniform Scale-Mixture Representation; (2) Normal Scale-Mixture Representation. The evaluation is conducted by calculating the DIC and MSE criteria for prediction errors in two simulation examples. The DIC criterion is a generalization of the AIC criterion for Bayesian model selection problems. The DIC is defined as:

$$\overline{DIC} = \overline{D(\theta)} + Pd$$

where $\overline{D(\theta)} = -2logL(\theta \mid y)$ is called the deviance and is a function of θ, the vector of the model's parameters. The first term represents the expected deviance under the posterior density function of the parameters.

$$\overline{D(\theta)} = E_{\theta|y}[D(\theta)] = E_{\theta|y}[-2logL(\theta|y)]$$

It is defined. The second component measures the number of effective parameters or P_d as:

$$P_d = \overline{D(\theta)} - D(\overline{\theta}) = E_{\theta|y}[D(\theta)] - D\big(E_{\theta|y}[\theta]\big) = E_{\theta|y}[-2logL(\theta|y)] + \ln L(\theta|y).$$

It is defined. By rearranging the Pd expression, we have: $D = D(\theta) + P_d$, so DIC can be written as:

$$DIC = D(\bar{\theta}) + 2Pd \; =$$

1) Simulated Data Under the Model: The data are generated under the model $y_i = x_i^T \beta + \epsilon_i$,
2) where the error terms ϵ_i follow a t-distribution with degrees of freedom v and scale parameter σ^2. The predictors x_i are drawn from a multivariate normal distribution with a covariance matrix Σ defined by $\sum j_k = \rho^{|j-k|}$ with $\rho = 0.7$.
3) Error Distributions: The error terms ϵ_i follow two different Student's t -distributions with scale-mixture representations (normal and uniform). The degrees of freedom considered are $v = 2{,}5{,}10{,}$, with two different values for the scale parameter σ2=9\ and $\sigma^2 = 9$ and $\sigma^2 = 25$.
4) Data Partitioning: Each simulated dataset is divided into training and test sets, each containing 50 observations.
5) Model Parameters: The coefficients β are specified as $\beta = (3, -3{,}5{,}4, -2.8, -3.2{,}0{,}0{,}0{,}0)$ with 6 non-zero and 4 zero coefficients.
6) Model Fitting and Evaluation: Models are fitted on the training data, and the Deviance Information Criterion (DIC) and Mean Squared Error (MSE) are calculated for the test set.

3. Results and Discussion

The results from 1000 simulations are summarized in Table 1, showing that the Bayesian t-lasso regression with uniform scale-mixture representation outperforms the normal scale-mixture representation across all four degrees of freedom, based on DIC and MSE. As the degrees of freedom increase and approach the normal distribution ($v = 1000$), the differences in DIC and MSE between the two methods decrease. The uniform scale-mixture representation is particularly effective for degrees of freedom $v = 5$ compared to other conditions. For variable selection, the highest posterior density (HPD) regions are computed for model parameters. Variables are excluded if the HPD regions include zero. According to Table 2, in the uniform scale-mixture representation method, the parameters $\beta_7, \beta_8, \beta_9, \beta_{10}$ can be set to zero. In contrast, the normal scale-mixture representation suggests excluding $\beta_2, \beta_5, \beta_7, \beta_8, \beta_9, \beta_{10}$. Additionally, using a standard Bayesian lasso regression with normal error distribution for the Boston dataset suggests more coefficients for exclusion, including $\beta_1, \beta_2, \beta_5, \beta_7, \beta_8, \beta_9, \beta_{10}$. Therefore, the Bayesian t-lasso regression with uniform scale-mixture representation proves more efficient in variable selection as well. Also, if the normal Bayesian lasso

Polyhedron International Journal in Mathematics Education
Volume 2, No 1, pp. 17-29, E-ISSN: 2987-6540
https://nakiscience.com/index.php/pijme

Original Article

regression model with normal error distribution is used to model Boston data, more coefficients include $\beta_1, \beta_2, \beta_5, \beta_7, \beta_8, \beta_9, \beta_{10}$ is suggested to be removed from the model. Therefore, it can be said that in the field of variable selection, T-Lasso Bayesian regression method works better with mixed representation-uniformity measure.

Table 1

Simulation Results Based on 1000 Repetitions and 100 Observations (n = 100)

Degree of Freedom (v)	$\sigma^2 = 81$	$\sigma^2 = 225$
Bayesian t-lasso (Normal Scale-Mixture)		
v=2	DIC = 450.7, MSE = 105.7	DIC = 528.1, MSE = 30.7
v=5	DIC = 449.3, MSE = 3.3	DIC = 524.8, MSE = 8.3
v=10	DIC = 437.5, MSE = 8.7	DIC = 525.9, MSE = 7.7
v=1000	DIC = 434.1, MSE = 5.5	DIC = 523.8, MSE = 5.4
Bayesian t-lasso (Uniform Scale-Mixture)		
v=2	DIC = 358.1, MSE = 17.1	DIC = 360.7, MSE = 15.7
v=5	DIC = 339.3, MSE = 3.6	DIC = 341.4, MSE = 3.4
v=10	DIC = 333.3, MSE = 1.8	DIC = 335.3, MSE = 1.7
v=1000	DIC = 335.6, MSE = 6.1	DIC = 337.8, MSE = 2.0

This table separates the values for each model and each setting of the degrees of freedom and σ^2, making it easier to compare the DIC and MSE values across different scenarios.

Degrees of Freedom (v): These are set at 5, 10, 50, and 1000, influencing the flexibility of the t-distribution.

DIC and MSE: Each cell contains two values, representing the Deviance Information Criterion (DIC) and the Mean Squared Error (MSE) for each combination of statistical model and degree of freedom. The DIC values indicate the model's complexity and fit, while the MSE values reflect the model's prediction accuracy. This table helps to compare the performance of the Bayesian Lasso regression models under different conditions, specifically how they respond to changes in the degrees of freedom and the choice of prior distribution.

Table 2

Credibility regions for the parameters of the T-Lasso Bayesian regression model with mixed display-uniformity measure

Parameter	β_1	$\beta 2$	β_3	β_4	β_5
value	(1.42, 3.30)	(0.10, 3.60)	(0.14, 5.13)	(1.34, 5.81)	(-2.85, -0.10)
parameter	β_6	β_7	β_8	β_9	β_{10}
value	(1.89, 6.81)	(-1.58, 2.33)	(-0.69, 2.58)	(-1.81, 1.60)	(-1.78, 1.44)

Real Data

In this section, the performance of Bayesian t-lasso regression with normal and uniform scale mixture representations is evaluated using the Boston dataset. This dataset was reported by Harrison Jr & Rubinfeld, (2019) to examine various factors affecting housing prices, and the dataset is publicly available through the link: Boston dataset. The dataset includes 506 observations ($n = 506$) and 13 explanatory variables (p = 13). The response variable, MEDV, represents the median value of owner-occupied homes in Boston neighborhoods. The Bayesian t-lasso regression methods with normal and uniform scale mixture representations are applied, with degrees of freedom $v = 2, 5, 10, 1000$. Additionally, the data is utilized with a normal distribution ($v = 1000$) and larger values for the degrees of freedom to observe the relationships among the explanatory variables in the correlation matrix and the heatmap. The table below represents the correlation coefficients between the variables. The coefficients indicate the strength and direction of the relationship between the variables: positive values indicate a positive relationship, and negative values indicate a negative relationship. Values closer to 1 or -1 indicate stronger relationships. This table displays the correlation coefficients between various variables in the Boston housing dataset. Each column and row represent one of the variables, and the

Polyhedron International Journal in Mathematics Education
Volume 2, No 1, pp. 17-29, E-ISSN: 2987-6540
https://nakiscience.com/index.php/pijme

Original Article

coefficients inside the table indicate the relationship between two variables. The correlation coefficients range from -1 to 1 and can be positive or negative:

Positive coefficients indicate a positive correlation, meaning that as one variable increases, the other tends to increase as well. For example, the correlation between RM (the number of rooms) and MEDV (median value of owner-occupied homes) is 0.7, showing a strong positive relationship.

Negative coefficients indicate a negative correlation, meaning that as one variable increases, the other tends to decrease. For instance, the correlation between LSTAT (percentage of lower status of the population) and MEDV is -0.74, indicating a strong negative relationship.

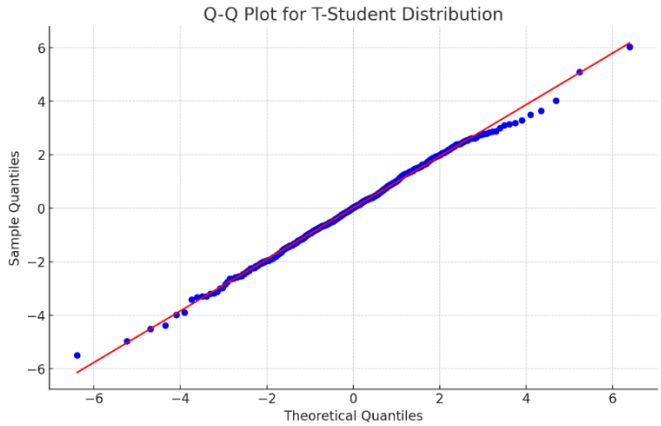
Coefficients close to zero suggest a weak or no significant relationship between the variables. For example, the correlation between CHAS (Charles River dummy variable) and ZN (proportion of residential land zoned for lots over 25,000 sq. ft.) is almost zero, indicating no significant relationship between proximity to the river and the proportion of zoned residential land. Overall, this table helps analysts understand the relationships between different variables and how they may impact house prices or other metrics in the dataset.

CRIM: Crime rate per capita by town.
RM: Average number of rooms per dwelling.
AGE: Proportion of owner-occupied units built before 1940.
DIS: Weighted distances to five Boston employment centers.
MEDV: Median value of owner-occupied homes (in $1000).

Table 3
Correlation matrix diagram

	CRM	ZN	INDS	CHAS	NOX	RM	AGE	DIS	RAD	TAX	PTRATIO	B	LSTAT	MEDV
CRIM	1	-0.2	0.4	-0.06	0.42	-0.22	0.35	-0.38	0.62	0.58	0.29	-0.38	0.45	-0.39
ZN	-0.2	1	-0.53	-0.04	-0.52	0.31	-0.57	0.66	-0.31	-0.31	-0.39	0.18	-0.41	0.36
INDUS	0.4	-0.53	1	0.06	0.76	-0.39	0.64	-0.71	0.66	0.72	0.38	-0.36	0.48	-0.48
CHAS	-0.06	-0.04	0.06	1	0.09	0.09	-0.04	0.09	-0.04	-0.12	0.05	-0.05	-0.09	0.15
NOX	0.42	-0.52	0.76	0.09	1	-0.3	0.73	-0.77	0.61	0.67	0.19	-0.39	0.59	-0.44
RM	-0.22	0.31	-0.39	0.09	-0.3	1	-0.24	0.21	-0.21	-0.29	0.36	0.13	-0.61	0.7
AGE	0.35	-0.57	0.64	-0.04	0.73	-0.24	1	-0.75	0.46	0.51	0.26	-0.27	0.64	-0.38
DIS	-0.38	0.66	-0.71	0.09	-0.77	0.21	-0.75	1	-0.49	-0.53	-0.23	0.29	-0.5	0.25
RAD	0.62	-0.31	0.66	-0.04	0.61	-0.21	0.46	-0.49	1	0.91	-0.44	-0.44	0.45	-0.38
TAX	0.58	-0.31	0.72	-0.12	0.67	-0.29	0.51	-0.53	0.91	1	-0.45	-0.53	0.54	-0.47
PTRATIO	0.29	-0.39	0.38	0.05	0.19	0.36	0.26	-0.23	-0.44	-0.45	1	0.13	-0.37	0.37
B	-0.38	0.18	-0.36	-0.05	-0.39	0.13	-0.27	0.29	-0.44	-0.53	0.13	1	-0.73	0.33
LSTAT	0.45	-0.41	0.48	-0.09	0.59	-0.61	0.64	-0.5	0.45	0.54	-0.37	-0.73	1	-0.74
MEDV	-0.39	0.36	-0.48	0.15	-0.44	0.7	-0.38	0.25	-0.38	-0.47	0.37	0.33	-0.74	1

Figure 1
Quantile diagram of t-Student distribution



First, the response variable is centered, and the explanatory variables are standardized to have a mean of zero and a variance of one. Two Bayesian Lasso regression methods are fitted to the training

Polyhedron International Journal in Mathematics Education
Volume 2, No 1, pp. 17-29, E-ISSN: 2987-6540
https://nakiscience.com/index.php/pijme

dataset, and to evaluate the performance of these methods and select the optimal model, the MSE and DIC criteria are calculated for the test dataset. The results are shown in Table 3. The results are obtained considering the student's t-distribution with different degrees of freedom $v = 2,5,10,1000$ for the model error distribution in the Bayesian Lasso regression method.

According to Table 3, the performance of the Bayesian Lasso method with the mixture-scale uniform display is better than the other method, and the DIC and MSE criteria values for different degrees of freedom show significant differences.

Table 4

Boston Data Analysis - Values of DIC and MSE for Two Methods of Bayesian Lasso Regression with Normal and Uniform Priors

Statistical Model	Degree of Freedom (ν)			
	ν = 2	ν = 5	ν = 10	ν = 1000
Bayesian Lasso Regression with Normal Prior (DIC, MSE)	(1901.2, 85.2)	(1561.7, 83.4)	(1839.9, 87.3)	(1661.3, 34.4)
Bayesian Lasso Regression with Uniform Prior (DIC, MSE)	(1901.85, 25.7)	(1889.84, 26.4)	(1879.78, 27.1)	(1867.34, 27.4)

In the Table 4,

Statistical Model: Two statistical models are evaluated:

1. Bayesian Lasso Regression with Normal Prior;
2. Bayesian Lasso Regression with Uniform Prior.

Degree of Freedom (ν): The parameter ν represents the degrees of freedom for the t-distribution models. Different degrees of freedom are evaluated, ranging from 2 to 1000.

DIC and MSE: Each cell in the table contains the values of DIC (Deviance Information Criterion) and MSE (Mean Squared Error) for each statistical model and degree of freedom. DIC is a criterion used for model comparison, balancing fit and complexity. MSE measures the average squared difference between the estimated values and the actual values.

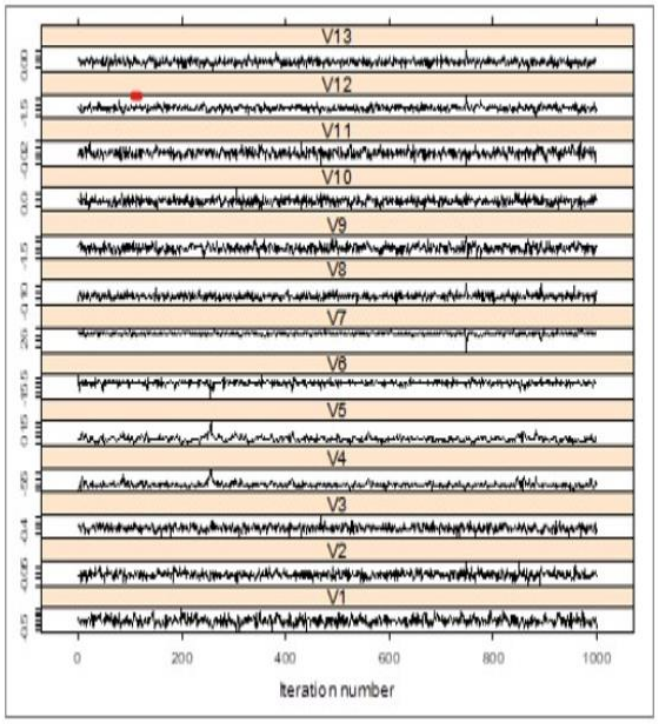
From the results of the regions of the highest posterior density for the parameters of the T-lasso Bayesian regression model with the mixed representation-uniform scale, it is concluded that the variables AGE and INDUS are suitable choices to be removed from the model. By calculating the average of the second power of the prediction error or MSPE for the model with the presence of all explanatory variables and the model with the exclusion of AGE and INDUS variables. Therefore, it can be said that the T-Lasso Bayesian regression model with mixed display-uniformity measure has performed well in selecting effective explanatory variables.

The convergence of the Markov chain of the samples obtained for the parameters of the model using the Gibbs algorithm indicates the degree of convergence of the chain obtained from the Gibbs algorithm, and the effect diagram is a good intuitive criterion for evaluating the characteristics of the convergence of the chain (Gelman, 2011) in Figure 4 for Boston data, The graph of the effect can be seen in the case where ν = 5. According to the diagram, the samples obtained from the Gibbs algorithm for the model parameters quickly traverse the posterior distribution space and have good convergence. Also, according to Heidelberger and Welch's convergence test (Heidelberger & Lewis, 1984), the Markov chain follows a stationary distribution. In general, according to the intuitive results, the T-Lasso Bayesian regression method based on the mixed-scale representation is better in terms of model selection and prediction accuracy than the T-Lasso Bayesian regression method based on the mixed-normal scale representation, and with increasing degrees of freedom T-Student distribution and the approach of model error distribution to normal distribution, the difference between the two methods decreases.

Polyhedron International Journal in Mathematics Education
Volume 2, No 1, pp. 17-29, E-ISSN: 2987-6540
https://nakiscience.com/index.php/pijme

Original Article

Figure 4
Diagram of the effect of the samples obtained from the Gibbs algorithm for the regression coefficients in the T-Lasso Bayesian regression model with a homogeneous mixture display



4. Conclusion

When linear regression models assume that errors follow a normal distribution, the traditional method of estimating parameters—known as ordinary least squares (OLS)—can be overly sensitive to outliers or unusual data points. This sensitivity can lead to less reliable results when the assumption of normality is violated. To address this issue, it's suggested to use alternative distributions that are more robust to deviations from normality. In this context, the student's t-distribution is proposed as a more robust alternative to the normal distribution for modeling errors. This distribution is less influenced by extreme values, making it a better choice when the data does not fit the normality assumption well. To further improve the robustness of regression models, a Bayesian approach called the Bayesian t-Lasso regression model has been introduced. This model extends the traditional Bayesian Lasso regression, which assumes normal errors, to situations where errors might not follow a normal distribution. The Bayesian t-Lasso model incorporates the student's t-distribution into the Lasso regression framework, offering more robust estimators under non-normal error conditions.

The Bayesian t-Lasso regression model is evaluated using two different approaches: (a) Normal-mixture-scale representation: This method combines the student's t-distribution with a normal distribution to model the errors. (b) Uniform-mixture-scale representation: This method uses a uniform distribution in combination with the student's t-distribution for a different type of prior density. Hierarchical Bayesian models and Gibbs sampling algorithms are employed to estimate the parameters of these models. The findings from simulations and real data analyses indicate that the Bayesian t-Lasso model using the Uniform-mixture-scale representation performs better in terms of mean squared error (MSE) and Deviance Information Criterion (DIC) compared to the Normal-mixture-scale representation. Future research will focus on selecting the most appropriate model for the Bayesian t-Lasso regression, especially when dealing with high-dimensional data where traditional methods may struggle.

Conflict of Interest:
The authors declare no conflict of interest.

Additional Information:
Additional information is available for this paper.

5. References

Belloni, A., & Chernozhukov, V. (2013). Least squares after model selection in high-dimensional sparse models. *Bernoulli*, *19*(2), 521-547. https://doi.org/10.3150/11-BEJ4

Gelman, A. (2011). Induction and deduction in Bayesian data analysis. *RMM*, *2*, 2011, 67–78. https://jlupub.ub.uni-giessen.de/server/api/core/bitstreams/786a421e-1af2-4baf-a8c8-30d208fb21d5/content

Heidelberger, P., & Lewis, P. A. (1984). Quantile estimation independent sequences. *Operations Research*, *32*(1), 185-209.

Hlavackova-Schindler, K. (2016). Prediction consistency of lasso regression does not need normal errors. *British Journal of Mathematics & Computer Science*, *19*(4), 10-20. https://doi.org/10.9734/BJMCS/2016/29533

Hoerl, A. E., & Kennard, R. W. (1970). Ridge regression: Biased estimation for nonorthogonal problems. *Technometrics*, *12*(1), 55-67. https://homepages.math.uic.edu/~lreyzin/papers/ridge.pdf

Karapetyants, M., & László, S. C. (2024). A Nesterov-type algorithm with double Tikhonov regularization: fast convergence of the function values and strong convergence to the minimal norm solution. *Applied Mathematics & Optimization*, *90*(1), 17. https://doi.org/10.1007/s00245-024-10163-0

Liu, C., & Rubin, D. B. (1995). ML estimation of the t distribution using EM and its extensions, ECM and ECME. *Statistica sinica*, 19-39.

Park, T., & Casella, G. (2008). The Bayesian lasso. *Journal of the American Statistical Association*, *103*(482), 681-686. https://doi.org/10.1198/016214508000000337

Shadrokh, A., Khadembashiri, Z., & Yarmohammadi, M. (2021). Regression Modeling Via T-Lasso Bayesian Method. *Journal of Advanced Mathematical Modeling*, *11*(2), 365-381. https://doi.org/10.22055/jamm.2021.35112.1859

Steele, S. E., & Lopez-Fernandez, H. (2014). Body size diversity and frequency distributions of Neotropical cichlid fishes (Cichliformes: Cichlidae: Cichlinae). *PLoS One*, *9*(9), e106336. https://doi.org/10.1371/journal.pone.0106336

Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, *58*(1), 267-288. https://webdoc.agsci.colostate.edu/koontz/arec-econ535/papers/Tibshirani%20(JRSS-B%201996).pdf

Wafa, M.N. (2019). Assessing School Students' Mathematic Ability Using DINA and DINO Models. *International Journal of Mathematics Trends and Technology-IJMTT*, *65*(12), 153-165. https://doi.org/10.14445/22315373/IJMTT-V65I12P517

Wafa, M. N., Hussaini, S.A.M & Pazhman, J. (2020). Evaluation of Students' Mathematical Ability in Afghanistan's Schools Using Cognitive Diagnosis Models. *EURASIA Journal of Mathematics, Science and Technology Education*, *16*(6), 1-12. https://doi.org/10.29333/ejmste/7834

Wafa, M. N., Zia, Z., & Frozan, F. (2023). Consistency and ability of students using DINA and DINO models. *European Journal of Mathematics and Statistics*, *4*(4), 7-13. https://doi.org/10.24018/ejmath.2023.4.4.230

Wafa, M. N., Zia, Z., & Hussaini, S. A. M. (2023). Regression Models According to Birnbaum-Saunders Distribution. *European Journal of Mathematics and Statistics*, *4*(6), 24-30. https://doi.org/10.24018/ejmath.2023.4.6.267

Zhu, Y.-N., Wu, H., Cao, C., & Li, H.-N. (2008). Correlations between mid-infrared, far-infrared, Hα, and FUV luminosities for Spitzer SWIRE field galaxies. *The Astrophysical Journal*, *686*(1), 155-162. https://doi.org/10.1086/591121

Zou, H. (2006). The adaptive lasso and its oracle properties. *Journal of the American statistical association*, *101*(476), 1418-1429. https://doi.org/10.1198/016214506000000735

Zou, H., & Hastie, T. (2005). Regularization and variable selection via the elastic net. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, *67*(2), 301-320. https://doi.org/10.1111/j.1467-9868.2005.00503.x