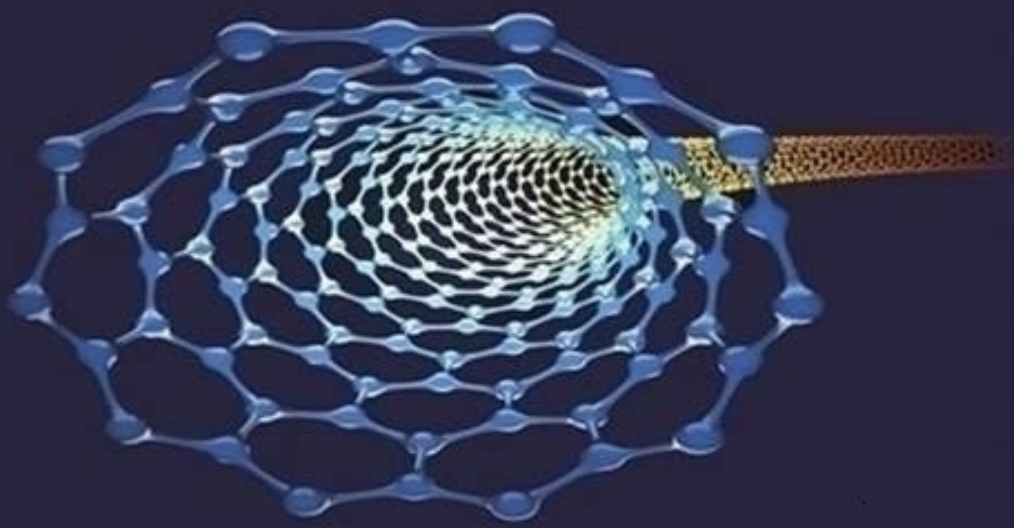
Third International Conference of Mathematics and its Applications

(TICMA 2022)



زانكۆی سەلاحەدین - هەولێر
Salahaddin University-Erbil

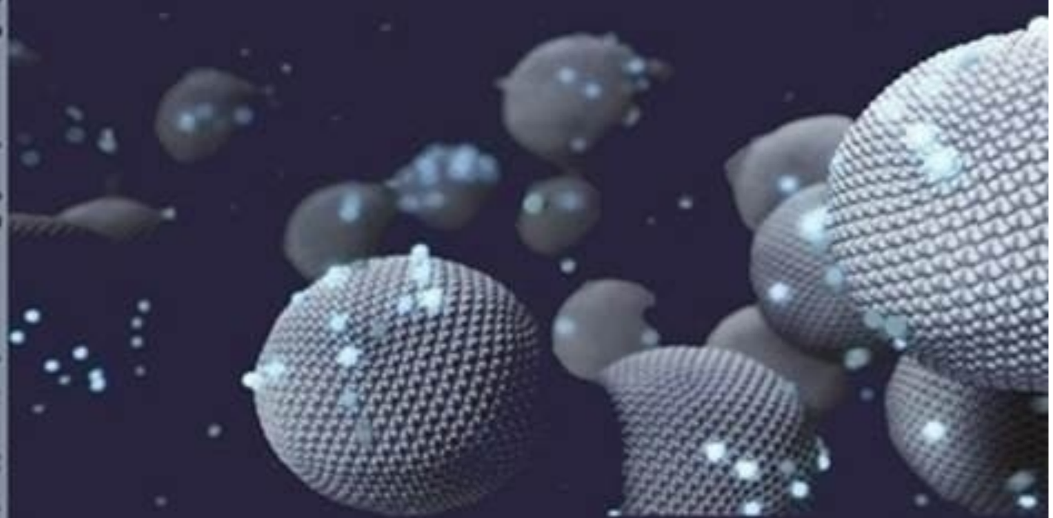
College of science / Mathematics Department

29-31 August 2022

ZANCO Journal of Pure and Applied Sciences

ZANCO

Journal of Pure and Applied Sciences



زانكۆی سەلاحەدین - هەولێر
Salahaddin University-Erbil

College of science / Mathematics Department

29-31 August 2022

Proceeding of 3rd International Conference of Mathematics and its Applications TICMA2022
https://conferences.su.edu.krd/su/ticma2022/
DOI: https://doi.org/10.31972/ticma2022

RESEARCH PAPER

Improving Prediction Accuracy of Lasso and Ridge Regression as an Alternative to LS Regression to Identify Variable Selection Problems

Pareekhan Abdulla Omer*
[1] Statistics department- College of Administration and Economics - Salahaddin University - Erbil / Iraq
Pareekhan.omer@su.edu.krd

A B S T R A C T:

This paper introduces the Lasso and Ridge Regression methods, which are two popular regularization approaches. The method they give a penalty to the coefficients differs in both of them. L1 Regularization refers to Lasso linear regression, while L2 Regularization refers to Ridge regression. As we all know, regression models serve two main purposes: explanation and prediction of scientific phenomena. Where prediction accuracy will be optimized by balancing each of the bias and variance of predictions, while explanation will be gained by constructing interpretable regression models by variable selection. The penalized regression method, also known as Lasso regression, adds bias to the model's estimates and reduces variance to enhance prediction. Ridge regression, on the other hand, introduces a minor amount of bias in the data to get long-term predictions. In the presence of multicollinearity, both regression methods have been offered as an alternative to the least square approach (LS). Because they deal with multicollinearity, they have the appropriate properties to reduce numerical instability caused by overfitting. As a result, prediction accuracy can be improved. For this study, the Corona virus disease (Covid-19) dataset was used, which has had a significant impact on global life. Particularly in our region (Kurdistan), where life has altered dramatically and many people have succumbed to this deadly sickness. Our data is utilized to analyze the benefits of each of the two regression methods. The results show that the Lasso approach produces more accurate and dependable or reliable results in the presence of multicollinearity than Ridge and LS methods when compared in terms of accuracy of predictions by using NCSS10, EViews 12 and SPSS 25.

KEY WORDS: Methodology, Regularization, Lasso regression, Ridge regression, multicollinearity.
DOI: https://doi.org/10.31972/ticma22.05

1. INTRODUCTION:

Multiple regression is frequently used to build a model for predicting future responses or to look into the link between the response variable and the predictor factors. The model's prediction accuracy is critical for the first aim, but the model's complexity is more relevant for the second goal. The least squares (LS) regression is notorious for underperforming in terms of model complexity and prediction accuracy. Ridge regression (Hoerl & Kennard, 1970), the Garotte (Breiman, 1995), Bridge regression (Frank and Friedman, 1993) and the Lasso (Tibshirani, 1996) were among the regularized regression approaches created in the last few decades to solve the faults of (LS) regression (see Van der Kooij and Meulman, 2006). The (LS) provides the coefficients that best fit to the data, with the additional criterion of finding unbiased coefficients. In this case, unbiased means that LS does not take into account which of the independent variables is more relevant than the others. It finds the coefficients for a given data set; there is just one set of betas to find Residual sum of squares (RSS). The intriguing question becomes, "Is the model with the smallest RSS truly the optimal mode?" The answer to the preceding question is (not really). It must also take into account (Bias), as implied by the word unbiased. Bias refers to how equally concerned a model is with its predictors.

This investigation will take a variety of approaches. This research will take a variety of techniques, using a variety of phrases and figures. There are two things that you should always keep in our mind. The first thing is that we always favor a model that catches the general patterns.

*** Corresponding Author: Pareekhan Abdulla Omer**
E-mail: Pareekhan.omer@su.edu.krd
Article History:
Received: 01/08/2022
Accepted: 15/09/2022
Published: 07/12/2022

The other is that we would forecast it based on new data rather than specific data. As a result, evaluation of model should be based on new data (testing set), rather than data that has already been collected (training set). Then, by adding a penalty term to the best fit produced from the trained data, regularization is a key concept used to avoid overfitting of the data, especially when the trained data are highly variable. Regularization is used to reduce volatility in the tested data, as well as to limit the influence of predictor factors on the output variable by compressing their coefficients. In statistics, the Lasso method is widely used to improve the model's prediction accuracy and interpretability. It was created in 1989 and is a regression strategy or approach that includes selection and regularization.

Lasso regression is a shrinkage-based extension of linear regression. The Lasso constrains the sum of the absolute values of the model parameters, with an upper bound of a specified constant. As a result of this constraint, the regression coefficients for some variables to shrink towards zero, i.e. (shrinkage). When there is automatic feature or variable selection, the Lasso regression is fairly simple (very easy). It's also useful when dealing with high-correlation predictors, as standard regression will usually have large regression coefficients. The Lasso regression technique can be applied in three different ways: (stepwise, backward, and forward) techniques. Feature selection is an important step in machine learning to avoid overfitting, and it's the same in regression.

If there are too many features in Lasso, some of them are completely removed, Setting the coefficients to zero completes the process. In machine learning jargon (vocab), this technique is referred to as (L1 Regularization). The Ridge regression was the most widely used technique for enhancing (improving) accuracy of prediction at the time. Ridge regression lowers prediction error by reducing overfitting by lowering the sum of the squares of the regression coefficients to be smaller than a specified value. However, it does not perform covariate selection and so does not help to the model's interpretability. The Lasso regression achieves both of these goals by reducing the total of the absolute values of the regression coefficients to be less than a fixed value, effectively driving certain coefficients to zero and removing them from the prediction process. This notion is identical to Ridge regression, which decreases the size of the coefficients as well, however Ridge regression tends to zero out many fewer coefficients.

2. METHODOLOGY:

A methodology of relationship between variables known as the regression studies, Functions are generally used to approximate the data. In the late 1880s, Francis Galton wondered if he could forecast men's height based on the height of their fathers. He proposed that men's heights are determined by the heights of their fathers, i.e., the taller the parent, the taller the son. Galton (1889) attempted to fit a straight line across the data set by plotting the heights of 14 fathers and their sons (Al-Nasser, 2017). In this case the relation between two variables Y and X can be written as:

$$y = \beta_0 + \beta_1 x_i + \varepsilon_i \qquad , \quad \varepsilon_i \sim N(0, \sigma_\varepsilon^2) \tag{1}$$

This equation refers to simple linear regression model where Y is called a dependent variable, X is called a predictor variable and ε is called a random error, other symbols β0, β1 are called the regression coefficients (parameters). In general, a multiple linear regression model studies the relationship between dependent variable and several predictors' variables or features $X_i = (X_{i1}, X_{i2}, \dots, X_{ik})$ for a given (n) samples $(x_i, y_i)_{i=1}^{n}$. This model can be written as the form:

$$Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \cdots + \beta_k X_{ik} + \varepsilon_i \qquad j = 1,2, \dots, k \quad , \quad \varepsilon_i \sim N(0, \sigma_\varepsilon^2) \tag{2}$$

Where: Y_i: the dependent variable, X_j: the jth predictor variables. β_j: the average effect on response variable Y a one-unit increase in X_j, all other predictors held constant and ε_i: the error term. The least square approach (LS), which minimizes the Residual sum of squares (RSS), is used to determine the values of these parameters $\beta_0, \beta_1, \beta_2, \dots, \beta_k$:

$$RSS = \sum_{i=1}^{n} (y_i - \hat{y}_i)^2 \tag{3}$$

Where:

y_i: represents the actual response value for the i^{th} observation, and $\hat{y}_i$: represents the predicted response value based on the multiple linear regression model. If the data consists of n observations, then the following linear regression model is considered:

$$y_i = x_i^T \beta + \varepsilon_i \tag{4}$$

where β is a vector of parameters and ε_i are scalar random errors, the matrix form (4) can be represented as (a general linear model)

$$\underline{\mathrm{Y}} = \mathrm{X}\underline{\beta} + \underline{\varepsilon} \quad , \quad \varepsilon_i \sim N(0, \sigma_\varepsilon^2) \tag{5}$$

$\underline{\mathrm{Y}}$: represents a vector of the response variable of order $(\mathrm{n} \times 1)$.

X: represents the matrix of the observation of the explanatory variables of order $(\mathrm{n} \times \mathrm{p})$, where $\mathrm{p} = \mathrm{k} + 1$

$\underline{\beta}$: vector parameters are estimated from the class $(\mathrm{p} \times 1)$.

$\underline{\varepsilon}$: random error vector of class $(\mathrm{n} \times 1)$.

In form (4) and (5), the errors are assumed to have zero mean and a constant variance: $E(\varepsilon) = 0$, $Var(\varepsilon) = \sigma^2 I_n$ (see Bager, Mohammed, & Odah, 2017), As a result, the OLS assumptions are met, and the following estimations of β are obtained:

$$\hat{\beta} = (X^T X)^{-1} X^T Y \tag{6}$$

And the fitted values of the response variable are:

$$\hat{Y} = X\hat{\beta} = X(X^T X)^{-1} X^T Y \tag{7}$$

Multicollinearity can be a concern when the predictor variables are highly correlated. This can make the model's coefficient estimations incorrect (unreliable) and have a lot of variances.

DOI: https://doi.org/10.31972/ticma2022

3. REGULARIZATION METHODS

Regularization works by adding a penalty, complexity term, or shrinkage term to the complicated model using Residual sum of squares (RSS).

$$RSS = \sum_{i=1}^{n}\left(y_i - \beta_0 - \sum_{j=1}^{k}\beta_j x_{ij}\right)^2 \qquad (8)$$

Regularization solves the overfitting problem, which has an impact on the model's accuracy. It is carried out by adding the penalty term to the best fit equation derived from the trained data. This strategy can be used to reduce the number of variables in the model while still keeping them in the model. Regularization can be used for a variety of purposes, including deciphering (understanding) simpler models such as sparse and group structure models. Ridge Regression and Lasso Regression are two major regularization techniques that are used to reduce the model's complexity. Except for the penalty term, which is different since lasso regression uses absolute weights and ridge regression uses the square of weights (see Melkumova & Shatskikh, 2017).

4. LASSO REGRESSION

Lasso was developed by Tibshirani (1996), which is widely used in the construction of prediction models. Hastie, Tibshirani, and Wainwright (2015) provide a great introduction to the Lasso's physics and use as a prediction tool (Tibshirani and Taylor, 2011). The full form of Lasso is the least Absolute Shrinkage and Selection Operation. The Lasso regression is a regularization technique. it is used over regression methods for a more accurate prediction. The "shrinkage" strategy is used in the lasso model to generate coefficients, which are then shrunk toward the center point as the mean or median. Data values are shrunk towards a central point known as the mean in shrinkage. In regularization, lasso regression is based on simple models with fewer parameters (models with fewer parameters). Because of the shrinkage process, we can get a better interpretation of the models. The shrinking procedure also allows for the detection of variables that are tightly linked to variables that correspond to the target. Penalized regression is another name for Lasso regression. In machine learning, this strategy is commonly used to pick a subset of variables. When compared to other regression models, it has a greater prediction accuracy. Model interpretation is increased by lasso regularization. The lasso regression penalizes the dataset's less important features. This dataset's coefficients are set to zero, resulting in their removal. For lasso regression, a dataset with high dimensions and correlation is well suited (Flexeder, 2010).

The Lasso method has been popularly used for variable selection problems. In a regression model, the Lasso method uses a L_1 penalty to shrink the coefficients associated with covariates towards zero and set some unimportant covariate coefficients to zero, such that important covariates are selected and unimportant covariates are left out. The Lasso method was originally developed to model the quantitative response variables (see Bak, 2017). A linear regression model (2) can be considered as:

$$y_i = \beta_0 + \sum_{j=1}^{k}\beta_j X_{ij} + \varepsilon_i \qquad \varepsilon_i \sim N(0, \sigma^2) \qquad (9)$$

Where $Y_i \in R$ is the response variable and $X_i = (X_{i1}, X_{i2}, \dots, X_{ik})$ is a vector that has (k) predictor values for subject (i) and $i = 1,2,\dots,n$, for a given sample size of n. The Lasso seeks out a model that minimizes the sum of squares of residuals $\sum_{i=1}^{n}\varepsilon_i^2$, subject to a constraint $\sum_{j=1}^{k}|\beta_j| \le t$ where $t > 0$ is a parameter that determines the amount of shrinkage applied to the coefficients and allows the model to be cleaned up (irrelevant variables from the model) by setting some βs to zero.

DOI: https://doi.org/10.31972/ticma2022

The Lasso estimate of $\hat{\beta} = (\hat{\beta}_0, \hat{\beta}_1, \dots, \hat{\beta}_k)$ is given by

$$\hat{\beta} = arg\min_{\beta}\{\sum_{i=1}^{n}(y_i - \beta_0 - \sum_{j=1}^{k}\beta_j X_{ij})^2\} \quad s.t \sum_{j=1}^{k}|\beta_j| \le t \tag{10}$$

Here, the smaller value of t will shrink coefficients more toward the origin and make more of the coefficients to be zero. Therefore, the Lasso penalty performs variable selection continuously as the t continuously increases or decreases. Note that the optimization problem in equation (2) can be rewritten in Lagrange function that the Lasso estimates of $\hat{\beta} = (\hat{\beta}_0, \hat{\beta}_1, \dots, \hat{\beta}_k)$ minimize a penalized Residual sum of squares is provided by

$$\hat{\beta}^{Lasso} = arg\min_{\beta}\{\sum_{i=1}^{n}(y_i - \beta_0 - \sum_{j=1}^{k}\beta_j X_{ij})^2 + \lambda\sum_{j=1}^{k}|\beta_j|\} = RSS + \lambda\sum_{j=1}^{k}|\beta_j| \tag{11}$$

Reversely, the larger value of the tuning parameter λ, the greater amount of shrinkage, more coefficient is set to zero, and thus a more parsimonious model is achieved. The choice of λ that optimizes the predictability of the fitted model by Lasso is obtained by the cross-validation procedure (Bak, 2017).

5. RIDGE REGRESSION

Ridge regression is not a new concept in the field of education .It has been used as an alternative prediction weighting technique to non-ordinary least squares (OLS) (see Walker, 2004) .Ridge regression is a technique for analyzing and treating multicollinearity after testing multicollinear data in multiple regression models .When predictor variables have a correlation among themselves, this is known as multicollinearity. Ridge regression seeks to reduce the standard error by adding some bias to the fisher information matrix .The reliability of regression estimates is greatly improved when the standard error is reduced .Ridge regression is a type of linear regression in which we introduce a little amount of bias, known as the Ridge regression penalty, so that we can get better long-term predictions, it's known as the L2 -norm in statistics.

When there are more predictor variables in a data set than there are observations, or when there is multicollinearity, least squares estimates are unbiased, but their variances are enormous, therefore they may be far off the true value .Ridge regression reduces standard errors by adding a degree of bias to the regression estimates .It is hoped that the net effect will be to provide more dependable estimates .The covariates (the columns of X) are super-collinear when the design matrix is high-dimensional .In regression analysis, multicollinearity refers to the situation where two (or more) covariates are highly linearly connected .As a result, the collinear subspace may not be (or may be close to not being) of full rank .Consequently, the subspace spanned by collinear may not be (or close to not being) of full rank. The function is changed in this method by including a penalty term (shrinkage term), which multiplies the lambda by the squared weight of each unique feature .As a result, the optimization function becomes:

$$\hat{\beta}^{Ridge} = arg\min_{\beta}\{\sum_{i=1}^{n}(y_i - \beta_0 - \sum_{j=1}^{k}\beta_j X_{ij})^2 + \lambda\sum_{j=1}^{k}\beta_j^2\} = RSS + \lambda\sum_{j=1}^{k}\beta_j^2 \tag{12}$$

Which is equivalent to minimization of RSS subject to, for some $c > 0$, $\sum_{j=1}^{k}\beta_j^2 < c$. constraining the sum of the squared coefficients. Writing this criterion in matrix form we have:

$$RSS(\lambda) = (Y - X\beta)^T(Y - X\beta) + \lambda\beta^T\beta$$

Then the result is the Ridge regression estimator as: $\hat{\beta}^{Ridge} = (X^T X + \lambda I)^{-1}X^T Y$ (13)

DOI: https://doi.org/10.31972/ticma2022

6. LASSO REGRESSION VERSUS RIDGE REGRESSION

Lasso regression and Ridge regression are used to reduce the model's complexity .Ridge regression is also known as L2 Regularization and Lasso linear regression is known as L1 Regularization .But first, let's distinguish between ridge and lasso regression .To obtain long-term forecasts, ridge regression introduces a modest amount of bias. By adding the penalty term, this amount of bias is known as the Ridge regression penalty .The absolute weights are contained in the penalty term in Lasso regression .As a result of the use of absolute values, the Lasso might decrease closer to the slope than ridge regression, which shrinks towards zero.

One variable is kept in Lasso linear regression, while the other correlated variables are set to zero . As a result of the loss of information, accuracy suffers .Ridge regression is frequently referred to as a Lasso regression example .As a result, it's clear that Lasso and ridge regression each have their own set of benefits .With the help of automatic variable selection for the models, Lasso eliminates the coefficients (shrinks to zero), whereas ridge regression is unable to do so .Both strategies, however, deal with over-fitting, which is a problem in realistic statistical models .The availability of data for machine learning is critical to the effectiveness of these techniques .Ridge regression is more efficient than Lasso regression, but Lasso is successful in eliminating the unwanted parameters present in the model (Melkumova and Shatskikh, 2017).

the Figure 1. helps us to understand better, where we will assume a hypothetical dataset with only two features. Using the constrain for the coefficients of Lasso and Ridge regression, the constraint regions for Lasso and Ridge regression are plotted with cyan and green colors for a two-dimensional feature space .Linear regression is responsible for the elliptical contours (eq. 4 and 5). If the coefficients are relaxed, the constrained region can expand and eventually reach the ellipse's center .When the findings of Lasso and Ridge regression resemble those of linear regression, this is the case .Otherwise, both approaches calculate coefficients by finding the first point where the elliptical contours intersect the constraint region .In contrast to the disk, the diamond (Lasso) has corners on the axes, and whenever the elliptical region hits such a point, one of the features completely vanishes. For higher dimensional feature space, there can be many solutions on the axis with Lasso regression, and thus only the most important features are selected.

Now we'll compare the shrinkage methods by examining the geometry of Lasso and Ridge regression. The estimation problem for both methods is shown in the figure below, when just two predictors are available .The circular outlines centered around the OLS estimate show locations where the RSS is constant, and the figure displays the constraint region from equations (7 and 8) .The location where the elliptical contour intersects the constraint region is found by both regression methods. Lasso has a diamond-shaped constraint region given by $|\beta_1| + |\beta_2| \leq t$. whereas Ridge regression has a circle-shaped constraint region defined by $\beta_1^2 + \beta_2^2 \leq c$. In the case of Lasso, one of the coefficients β_j is equal to zero if the contour crosses the diamond at a corner (Matthias & Emmert-Streib, 2019)

TABLE 1. Overview of regularization or penalty and methods

Methods	Regularization Term
Ridge regression	L_2 norm: $\|\beta\|_2$
Lasso regression	L_1 norm: $\|\beta\|_1$

DOI: https://doi.org/10.31972/ticma2022

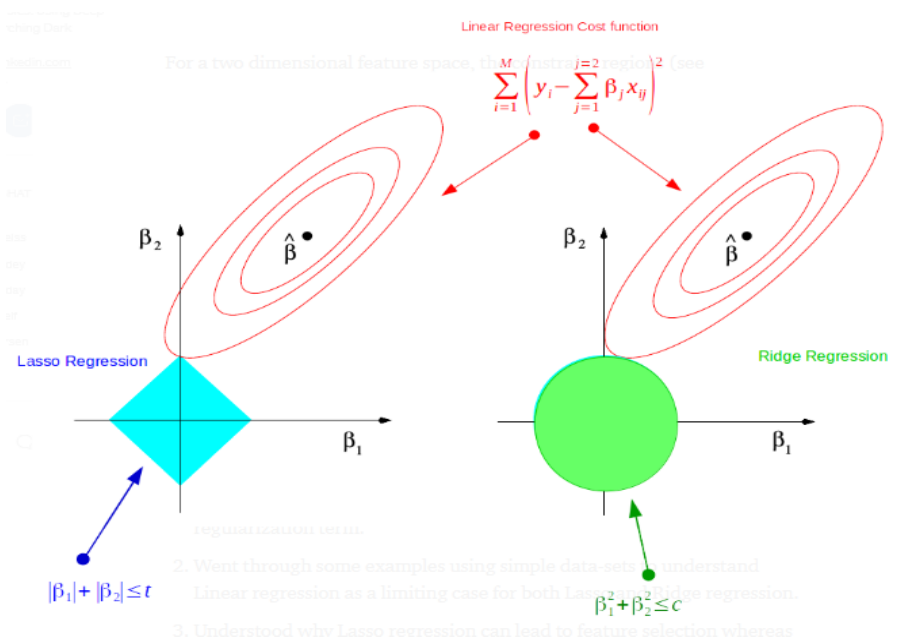


FIGURE 1. Lasso and Ridge geometrical interpretation: elliptical contours as the contours of errors and constraints for Lasso $|\beta_1| + |\beta_2| \leq t$ and Ridge $\beta_1^2 + \beta_2^2 \leq c$ (Matthias & Emmert-Streib, 2019)

7. DATA COLLECTION

The dataset of Corona virus epidemic (Covid-19) was taken for this study, which has affected on the world life totally. Especially in our region (Kurdistan), where life has changed in unexpected significantly way and millions of people have suffered from this dangerous disease. So, Influenza and the Covid-19 virus both display a similar disease presentation. In other words, they both cause respiratory disease, which can present as a variety of illnesses ranging from asymptomatic or mild to severe disease and death .Both viruses can spread through touch, droplets, and spores .As a result, everyone may prevent infection by practicing the same public health precautions, such as hand washing and proper respiratory hygiene (coughing into your elbow or into a tissue and throwing the tissue away right away) are important actions all can take to prevent infection. So, the speed of transmission is only important point of difference between the two viruses. When compared to Covid-19, influenza has a shorter median incubation period (the gap between infection and the onset of symptoms) and serial interval (the space between subsequent occurrences). Used the data of (263) case and have been taken in 2020, the information of each person gathered individually who suffered from this dangerous disease during this year. The dependent variable was (duration or number of sick daily injured and the independent variables were (Gender, Age, causes, does has other disease, in what way (he or she) got the disease, take the treatment, the result of disease (dead or alive)).

8. APPLICATION, RESULTS AND DISCUSSION

The Lasso regression analysis is a technique used in statistics and machine learning that performs both variable selection and regularization to improve the predictability accuracy and understandability of the resulting statistical model .When the subspace (which Y is projected) is close to rank deficient. It is almost impossible to separate the contribution of the individual covariates. The fit of the linear regression model to the data is frequently characterized by a significant inaccuracy in the estimates of the regression parameters corresponding to the collinear covariates, which reflects the uncertainty regarding the covariate responsible for the variance explained in Y. (Wieringen, 2015) .In regularization, we typically maintain the same number of features while reduce the coefficients' magnitudes .By using a variety of regression techniques that make use of regularization, we can reduce the magnitude of the coefficients.

Proceeding of 3rd International Conference of Mathematics and its Applications TICMA2022
https://conferences.su.edu.krd/su/ticma2022/
DOI: https://doi.org/10.31972/ticma2022

Variables	Regular LS Coeffici.	Standardized LS Coeffici.	LS Standard Error	Regular Ridge Coeffici.	Standardized Ridge Coeffici.	Ridge Standard Error	Regular Lasso Coeffici.	Standardized Lasso Coeffici.	Lasso Standard Error
Intercept	9.093963	-------	6.139416	9.19879	------	--------			
Gender- X_1	-1.32163	-0.0491	1.67572	-1.31357	-0.0488	1.667189	-1.26287	-0.046940	1.679153
Age - X_2	0.090004	0.1018	0.058112	0.08929	0.1010	0.057751	0.12730	0.143983	0.052496
Cause -X_3	0.710753	0.0451	1.004914	0.70845	0.0449	0.999360	1.182839	0.075045	0.955255
Other disease- X_4	-1.44385	-0.0460	2.207817	-1.43019	-0.0455	2.191664	-1.14771	-0.036531	2.203867
In what way- X_5	1.908793	0.1112	1.2426	1.89172	0.1102	1.23242	2.606488	0.151813	1.152546
take treatment - X_6	3.447485	0.1199	2.068082	3.41144	0.1186	2.05126	4.486400	0.155996	1.950047
Result- X_7	5.201583	0.0777	4.514616	5.17289	0.0772	4.485531	9.868086	0.147326	3.241224

TABLE 2. Least Squares vs Ridge and Lasso regression comparison

Table 2. shows the results of fitting a multiple linear regression model of LS, Ridge and Lasso regression to describe the relationship between the duration (number of sick days) and (7) of independent variables which describing them in the data collection part above, these three methods provide different results. it illustrates the Regular estimated values of the regression of each method. Additionally, it provides the estimated Standardized Regression Coefficient values for each method .The coefficients that would result from standardizing each independent and dependent variable are the standardized regression coefficients for each method .Here, standardization is described as dividing by the standard deviation of a variable and subtracting the mean from it .These standardized coefficients would yield from a regression study on these standardized variables. The formula for the standardized regression coefficient is:

$$\hat{\beta}_j^{std.} = \hat{\beta}_j \left(\frac{S_{x_j}}{S_y}\right) \quad (14)$$

Where S_y and S_{x_j} are the standard deviations for the dependent variable and the corresponding j^{th} independent variable. Also, in the table above shows the estimated error of the regression coefficient $S.E(\hat{\beta}_j)$, it is the standard deviation of the estimate, whoever the standard error of the estimate reduces, it makes the estimates more precise. Here, Lasso standard error of the estimates (yellow color column) has the values less than Ridge and LS standard error of the estimates in table 2.

TABLE 3. the testing hypotheses and dependency problem in the explanatory variables

Variables	**VIF**	**VIF**	**VIF**

DOI: https://doi.org/10.31972/ticma2022

	LS	Ridge	Lasso
Gender- X_1	1.0323	1.0214	1.89057
Age - X_2	1.3996	1.3749	8.04846 (*)
Cause -X_3	1.0456	1.0343	5.59775 (*)
Other disease - X_4	1.4436	1.4177	1.69566
In what way - X_5	1.5317	1.5017	11.0564 (*)
take treatment - X_6	1.5700	1.5380	1.77009
Result - X_7	1.1991	1.1833	14.6705 (*)

(*): represents the existence of multicollinearity when the value of VIF Lasso higher than (5).

To check the presence of dependency between the independent variables look the table 5. of correlation matrix and used VIF, in table 3. expresses three models for regression and tested their multicollinearity problem, shows that the Lasso model under investigation has this issue (problem). The Variance Inflation Factor (VIF), which assesses the inflation of the parameter estimates for all explanatory variables in the model, was used to identify the causal variables (see Farrar & Glauber, 1967), the VIF measures the presence of multicollinearity, and computed as follows:

$$VIF_j = \frac{1}{1-R_j^2} = \frac{1}{Tolerance} \ , \quad VIF = (1-R_j^2)^{-1} \ , \ j = 1,2,\dots,k \tag{15}$$

using the VIF_j is one method for detecting multicollinearity in regression data. Where the more likely of multicollinearity among the variables is indicated by a lower tolerance. When $VIF_j = 1$, it represents that the independent variables are not correlated with one another. if the value of $1 < VIF_j < 5$, It indicates that there is a moderate correlation between the variables. The difficult range for VIF_j is between 5 and 10, which indicates the range of highly correlated variables .It means $VIF_j \geq 5 \ to \ 10$, there will be multicollinearity among the regression model's predictors, and if $VIF_j > 10$ the regression coefficients for the regression matrix X are only tentatively estimated in the presence of multicollinearity (see Shrestha, N., 2020). In our results of analyzing these three methods the variables suffer from inflation in the variance of their parameters variables as showed in table (3) of (VIF Lasso), the explanatory variables (X_2, X_3, X_5, X_7) have high value of VIF, where X_2, X_3 have value between $5 < VIF_j < 10$, and X_5 , X_7 have a VIF value greater than 10 for large datasets indicates a multicollinearity problem.

DOI: https://doi.org/10.31972/ticma2022

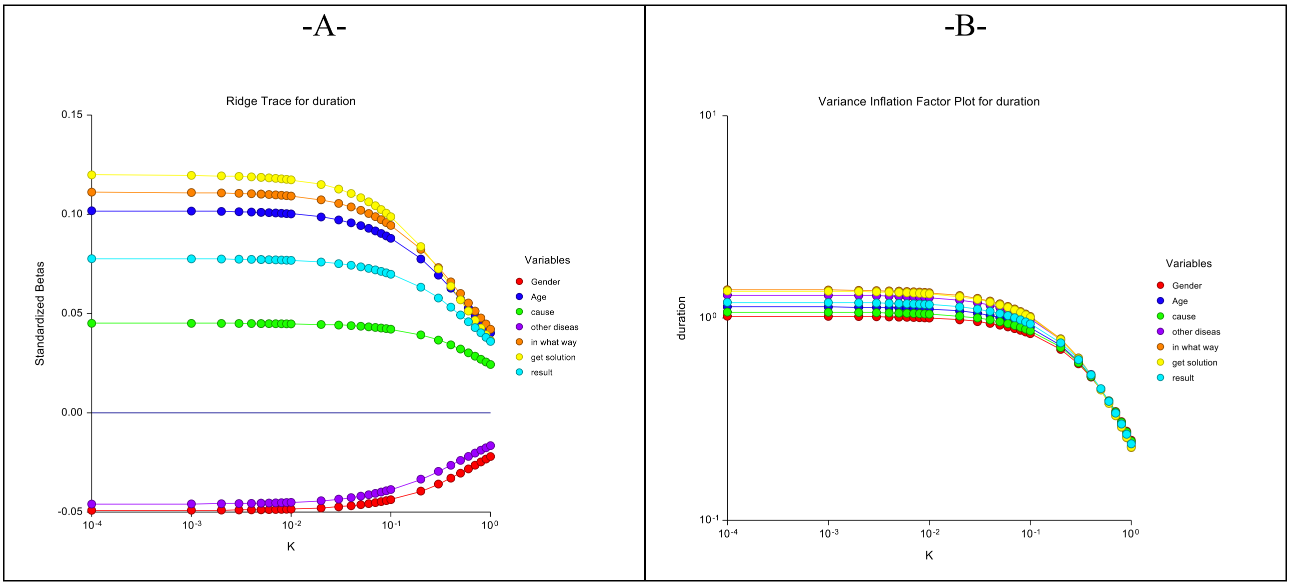


FIGURE 2. Ridge coefficient path for the data set found in NCSS

Choosing an acceptable value of K, is one of the key challenges in using Ridge regression (see Hoerl, and Kennard,1970). The Ridge Trace, a diagram created by the Ridge regression's creators, was recommended. The Ridge regression coefficients are plotted here as a function of K. The analyst selects a value of k for which the regression coefficients have stabilized while looking at the Ridge Trace (see figure 2.). Often, for low values of k, the regression coefficients frequently exhibit large fluctuations before stabilizing. Select the least value of k that results in the smallest amount of bias and causes the regression coefficients have seem to remain constant .Be aware that as k increases, eventually drive the regression coefficients to zero.

The figure (2. -A-) shows that these lines are for different explanatory variables and which of them are significant predictors of Y in above dataset. Additionally gives the standardized regression coefficients for Ridge parameter values in the range of 0.0 to 0.1. These are the regression model's coefficients when the variables are expressed in standardized form. The coefficients frequently shift considerably at first as the Ridge parameter is increased from zero but subsequently tend to stabilize. The smallest value after which the estimates shift gradually (change slowly) is a good value for the Ridge parameter. Your coefficients will be penalized by the ridge regression, and those that you estimate to be least useful (effective) will decrease the quickest.

The variance inflation factors for each of the regression model's coefficients are displayed in Figure (2. -B-). In comparison to the case in which all independent variables are uncorrelated, the variance inflation factors quantify how much the variance of the predicted coefficients is inflated. The VIFs frequently decline sharply (decrease dramatically) when the ridge parameter is increased from 0, but after that they tend to stabilize. The smallest value after which the VIFs change slowly is a suitable value for the ridge parameter. Press the alternative mouse button and choose Analysis Options to alter the range of ridge parameters investigated .When there is a multicollinearity problem between the independent variables, the Ridge regression model is based on estimation of the model parameters. If the value of (k=0), the same estimators as in the LS are obtained, where the Ridge regression coefficients are (k) .But, when the value of (k) increases away from zero, we observe the stability of the estimators as their values change. At (k=0.005), the results demonstrate that the level of the VIF of Ridge model for the explanatory variables are shown as in the figure above.

DOI: https://doi.org/10.31972/ticma2022

TABLE 4. the results of three estimation procedures

Methods	Std. Error of the Estimate	F-Ratio test	p-value	Adj R^2	MES
LS regression	13.4022	0.4850	0.8433	0.48094 %	170.0295
Ridge regression	13.2406	0.4898	0.8448	0.580893%	170.0563
Lasso regression	13.1766	3.691	0.0440(*)	0.601661%	167.9220

(*) indicate significant of Lasso regression model at the 5%

Table 4. expresses standard deviation of estimate, F-test, P-value, Adjusted R2 and MSE values. The results of applied OLS and Ridge and Lasso method are shown above, these techniques provide different outcomes, and the approach we suggest provide the smallest value by minimizing the sum squared of errors using the value of k rather than LS, which results in the value of MSE. it can be seen that the P-value of LS method and Ridge method are greater than 0.05, it means not significant statistically relationship between the variables at the 95% or higher confidence level. Consequently, the researcher should consider removing variable that is not significant from the model. The adjusted R^2 statistic is more suitable for comparing different multiple models with different numbers of independent variables. The adjusted $R^2 = 0.60166$ of Lasso regression model shows that 60.166% gives how effectively the model generalizes in comparison to other models. The value of F ratio test of measures the statistical significance of the Lasso model equal to (3.691) and MSE of Lasso has small value equal to (167.9220), F-test value is statistically significant at $(p - value \ < 0.05)$, $(0.044 < 0.05)$ and it can be observed that from the table above. Totally, the results of analyzing data by Lasso regression model obtained better than two other models.

TABLE 5. Correlations matrix

		Gender	Age	Cause	Other disease	In.what.way	take treatment	Result
Gender- X_1	Pearson Correlation	1	.046	.001	.090	.017	.006	.079
Age - X_2	Pearson Correlation	.046	1	.124*	.312**	-.116	.009	-.150*
Cause -X_3	Pearson Correlation	.001	.124*	1	.161**	-.028	.126*	-.189**
Other disease - X_4	Pearson Correlation	.090	.312**	.161**	1	-.277**	.284**	-.283**

In what way - X_5	Pearson Correlation	.017	-.116	-.028	-.277**	1	-.467**	.288**
take treatment - X_6	Pearson Correlation	.006	.009	.126*	.284**	-.467**	1	-.224**
Result - X_7	Pearson Correlation	.079	-.150*	-.189**	-.283**	.288**	-.224**	1
*. Correlation is significant at the 0.05 level (2-tailed).								
**. Correlation is significant at the 0.01 level (2-tailed).								

9. CONCLUSIONS

The statistical analysis of Lasso and Ridge regression estimation as an alternative method to LS regression estimation for the data leads to the following conclusions:

1. whoever the standard error of the estimate reduces, it makes the estimates more precise. As it seen in standard error of the estimates of Lasso has the smallest values when compared with Ridge and LS standard error of the estimates.

2. In presence of multicollinearity used VLF measurement, shows that Lasso regression method has more powerful to checking this problem between explanatory variables while the other methods haven't this power because in Lasso method found 5 explanatory variables with high dependency value of VIF among 7 of explanatory variables, Lasso method is investigated the relationship between variables for the real dataset. this problem could be improved by adding more cases in to the data, increasing the sample size of the data set.

3. The estimators used in the LS regression are the same as those found in the Ridge regression coefficients if the value of (k=0). But when the value of (k) moves away from zero, we observe that the estimators' stability increases. At (k=0.005) the results indicate the level of the VIF of Ridge model for the explanatory variables, at this value of (k) found it the best estimates of the model.

4. It observed that the proposed Lasso method is more significant than the classical LS method and Ridge method, depend on the value of p-value of the F-test and decrease value of MSE.

5. Increase the value of Adj R2 of Lasso method leads that it will be more appropriate method than the other for this data.

6. The estimation method of LS and Ridge regression method provides almost similar results, while the Lasso estimation method is able to produce consistent and more efficiency coefficients results depend on that criterion used in this study.

ACKNOWLEDGMENTS

Most of all I thanks "almighty Allah" with countless and gratitude, for all this blessing, who gave me this great opportunity to complete this research. I'm very grateful and thankful to my family for their support.

We should also thank everyone who participated in my research and to all the people I remembered who helped me.

REFERENCES

[1] AlNasser, H., 2017. *On ridge regression and least absolute shrinkage and selection operator* (Doctoral dissertation).
[2] Bak, S., 2017. *Generalized linear regression model with LASSO, group LASSO, and sparse group LASSO regularization methods for finding bacteria associated with colorectal cancer using microbiome data* (Doctoral dissertation, University of Guelph).
[3] BAGER, A.S.M., Mohammed, B.K. and Odah, M.H., 2017. Ridge Regression Analysis On The Influential Factors Of Fdi In Iraq. In *Proceedings of the INTERNATIONAL MANAGEMENT CONFERENCE* (Vol. 11, No. 1, pp. 19-25). Faculty of Management, Academy of Economic Studies, Bucharest, Romania.
[4] Flexeder, C., 2010. *Generalized lasso regularization for regression models* (Doctoral dissertation, Institut für Statistik).
[5] Farrar, D.E. and Glauber, R.R., 1967. Multicollinearity in regression analysis: the problem revisited. *The Review of Economic and Statistics*,92-107.
[6] Melkumova, L.E. and Shatskikh, S.Y., 2017. Comparing Ridge and LASSO estimators for data analysis. *Procedia engineering*, 201,746-755.
[7] Emmert-Streib, F. and Dehmer, M., 2019. High-dimensional LASSO-based computational regression models: Regularization, shrinkage, and selection. *Machine Learning and Knowledge Extraction*, 1(1),359-383.
[8] Pereira, J.M., Basto, M. and da Silva, A.F., 2016. The logistic lasso and ridge regression in predicting corporate failure. *Procedia Economics and Finance*, 39,634-641.
[9] Shrestha, N., 2020. Detecting multicollinearity in regression analysis. *American Journal of Applied Mathematics and Statistics*, 8(2),39-42.
[10] Shariff, N.S.M. and Duzan, H.M.B., 2018. An application of proposed Ridge Regression Methods to real data problem. *International Journal of Engineering & Technology*.
[11] Tibshirani, R.J. and Taylor, J., 2011. The solution path of the generalized lasso. *The annals of statistics*, 39(3),1335-1371.
[12] Flexeder, C., 2010. *Generalized lasso regularization for regression models* (Doctoral dissertation, Institut für Statistik).
[13] Walker, D.A., 2004. Ridge regression as an alternative to ordinary least squares: Improving prediction accuracy and the interpretation of beta weights. *The AIR Professional File*, 92,1-12.
[14] van Wieringen, W.N., 2015. Lecture notes on ridge regression. *arXiv preprint arXiv:1509.09169*.
[15] Van der Kooij, A.J. and Meulman, J.J., 2008. Regularization with ridge penalties, the lasso, and the elastic net for regression with optimal scaling transformations. *Submitted for publication*.