Peertechz

Engineering Group

Trends in Computer Science and Information Technology

DOI http://dx.doi.org/10.17352/tcsit.000005

DOI CC By

Eren Bas[1]*, Erol Egrioglu[1] and Vedide Rezan Uslu[2]

[1]Department of Statistics, Faculty of Arts and Science, Forecast Research Laboratory, Giresun University, Giresun, 28200, Turkey

[2]Department of Statistics, Faculty of Arts and Science, University of Ondokuz Mayis, Samsun, 55139, Turkey

Dates: Received: 02 December, 2016; **Accepted:** 30 December, 2016; **Published:** 31 December, 2016

***Corresponding author:** Eren Bas, Giresun University, Faculty of Arts and Science, Department of Statistics, Gure Campus, Giresun, Turkey, Tel: +90 454 3101400; Fax: +90 454 3101477; Email: eren.bas@giresun.edu.tr

Keywords: Ridge regression; Shrinkage parameters; Particle swarm optimization

https://www.peertechz.com

Research Article

Shrinkage Parameters for Each Explanatory Variable Found Via Particle Swarm Optimization in Ridge Regression

Abstract

Ridge regression method is an improved method when the assumptions of independence of the explanatory variables cannot be achieved, which is also called multicollinearity problem, in regression analysis. One of the way to eliminate the multicollinearity problem is to ignore the unbiased property of β . Ridge regression estimates the regression coefficients biased in order to decrease the variance of the regression coefficients. One of the most important problems in ridge regression is to decide what the shrinkage parameter *(k)* value will be. This *k* value was found to be a single value in almost all these studies in the literature. In this study, different from those studies, we found different *k* values corresponding to each diagonal elements of variance-covariance matrix of β instead of a single value of *k* by using a new algorithm based on particle swarm optimization. To evaluate the performance of our proposed method, the proposed method is firstly applied to real-life data sets and compared with some other studies suggested in the ridge regression literature. Finally, two different simulation studies are performed and the performance of the proposed method with different conditions is evaluated by considering other studies suggested in the ridge regression literature..

Introduction

The functional relation between a dependent variable and more than one independent variable is examined by multiple regression analysis. The purpose of the multiple regression analysis is the creation of the best model that can predict the dependent variable by using the independent variables. For this purpose, the most common method to create the best model is ordinary least square (OLS) estimates method. In this method, the sum of error squares to be minimal is calculated to predict the parameters of the model.

There are some valid assumptions for the implementation of the multiple regression analysis. These are; the absence of multicollinearity problem among independent variables, the variance of error term must be constant for all independent variables and the covariance between error term and independent variables must be equal to zero.

One of the major problems in multiple regression analysis is multicollinearity problem. If there is a full or high degree linear relationship among independent variables, this situation is called as multicollinearity. Besides, multicollinearity has some important effects on OLS estimates of the regression coefficients. In the presence of multicollinearity, the OLS of regression coefficients have large variance. And also, the regression coefficients can be estimated incorrectly and the standard errors of regression coefficients can be found as exaggerated in the presence of multicollinearity. If the regression coefficients can be estimated incorrect, it can be obtained incorrect results statistically.

Therefore, ridge regression method is used to obtain stable coefficient estimates for the estimation of the regression coefficients. That means, ridge regression has been suggested to overcome the multicollinearity problem.

In the literature, it is commonly accepted that if the variance inflation factors (*VIF*) values are greater than 10 there is a multicollinearity problem. This is a rule of thumb and this is not exact information. Similarly, condition number can be used to determine multicollinearity problem by using rule of thumbs. As a result of, determining of multicollinearity problem can be realized by using some criteria.

The two methods most commonly used to determine the effects of multicollinearity problem are *VIF* and condition

Citation: Bas E, Egrioglu E, Uslu VR (2016) Shrinkage Parameters for Each Explanatory Variable Found Via Particle Swarm Optimization in Ridge Regression. Trends Comput Sci Inf Technol 1(1): 031-039. DOI: http://dx.doi.org/10.17352/tcsit.000005

number methods. The diagonal elements of $Var(\hat{\beta})$ are called as *VIF* and are given by the Equation 1.

$$VIF_j = \frac{1}{(1-R_j^2)} \quad j = 1,\ldots,p \tag{1}$$

In this Equation, R_j^2 is the determination coefficient obtained from the multiple regression of X_j on the remaining $(p-1)$ regressor variables in the model.

It can be said that there is a multicollinearity problem among the relevant independent variables if these *VIF* values increase (*VIF* values ≥ 10). And also, if *VIF* values are increased, the degree of the multicollinearity increases with the increase of VIF values.

Condition number method is another method to determine the multicollinearity problem which is based on the eigenvalues of $X'X$ matrix. The formula of the condition number (*CN*) was given in Equation 2.

$$\phi = \frac{\lambda_{\max}}{\lambda_{\min}} \tag{2}$$

In this Equation, λ shows the eigenvalues of $X'X$. the relationship between condition number and multicollinearity is given in Table 1.

In summary, the determining of multicollinearity problem can be done by following two rules of thumbs. The first one is that if *VIF* values are greater than 10 multicollinearity is high. The second one is checking condition number as given in Table 1.

In addition, another problem in ridge regression is finding optimal biasing parameter *(k)* value. This *k* value is a very small constant determined by the researcher [1]. Several methods were proposed for finding it in the literature. These methods have been proposed in the studies of [2-22].

And also, there are many methods in the literature for ridge regression [23-29]. And also, [30] proposed some new methods that take care of the skewed eigenvalues of the matrix of explanatory variables. [31] Proposed an iterative approach to minimize the mean squared error in ridge regression. [32] Proposed new ridge parameters for ridge regression. [33] Proposed an optimal estimation for the ridge regression parameter. [34,35] Proposed some new estimators for estimating the ridge parameter.

This *k* value was found to be a single value in almost all these studies in the literature. But in this study, we found different *k* values corresponding to each diagonal elements of variance-covariance matrix of β instead of a single value of *k* by using a new algorithm based on particle swarm optimization.

The rest part of the paper can be outlined as below: The second section of the paper is about ridge regression. The methodology of the paper is given in Section 3. The implementation of our proposed method is given in Section 4. Two different simulation studies are performed under the title of simulation study and finally, discussions are presented in Section 6.

Table 1: Condition number and its effects.

Condition Number	Multicollinearity
CN <100	There is no serious multicollinearity
100 <*CN* <1000	Strong multicollinearity
CN >1000	Severe multicollinearity exist in the data

Ridge regression

Ridge regression is a remedy used in the presence of multicollinearity problem and it was firstly proposed by [1]. Ridge regression method has two important advantages according to OLS method. One of them is to solve the multicollinearity problem and the other one is to decrease the mean square error (MSE). The solution technique of ridge regression is similar with OLS. Besides, the difference between ridge regression and OLS is the *k* value. This *k* value is also called as biased parameter or shrinkage parameter and it takes values between 0 and 1. This *k* value is added to the diagonal elements of the correlation matrix and thus biased regression coefficients are obtained.

The OLS estimates of regression coefficients and ridge estimates of regression coefficients are shown in the Equations 3 and 4 respectively.

$$\hat{\beta} = (X'X)^{-1}X'Y \tag{3}$$

$$\hat{\beta}_R = (X'X + kI)^{-1}X'Y \tag{4}$$

As noted above, ridge regression is a biased regression method. The proof of this situation is shown in Equation 5.

$$\begin{aligned}\hat{\beta}_R &= (X'X + kI)^{-1}X'Y \\ &= (X'X + kI)^{-1}(X'X)\hat{\beta} = Z\hat{\beta}\end{aligned} \tag{5}$$

$$E(\hat{\beta}_R) = E(Z\hat{\beta}) = Z\beta$$

It is clearly seen that ridge estimates of regression coefficients $(\hat{\beta}_R)$ are biased estimates. One of the most important points to be considered in the ridge regression is the ***k*** value. There are many methods proposed in the literature to find the optimal *k* value. Ridge trace is one of these methods. Ridge trace is a plot of the elements of the ridge estimator versus *k* usually in the interval (0, 1) [1].

The other methods in the literature used to find the optimal ***k*** value were given in the Equations 6-14, respectively.

$$k = \frac{\rho\hat{\sigma}^2}{\hat{\beta}'\hat{\beta}} \quad [2] \tag{6}$$

$$k = \frac{\rho\hat{\sigma}^2}{\sum_{i=1}^{p}\lambda_i\hat{\beta}_i^2} \quad [4] \tag{7}$$

$$k = \frac{\rho\hat{\sigma}^2}{\sum_{i=1}^{p}\left\{\hat{\beta}_i^2\left[1+\left(1+\lambda_i\left(\hat{\beta}_i^2/\hat{\sigma}^2\right)^{1/2}\right)\right]\right\}} \quad [36] \tag{8}$$

Citation: Bas E, Egrioglu E, Uslu VR (2016) Shrinkage Parameters for Each Explanatory Variable Found Via Particle Swarm Optimization in Ridge Regression. Trends Comput Sci Inf Technol 1(1): 031-039. DOI: http://dx.doi.org/10.17352/tcsit.000005

$$k = \frac{\left(\lambda_{max} \hat{\sigma}^2\right)}{\left((n-p-1)\hat{\sigma}^2 + \lambda_{max} \hat{\beta}^2_{max}\right)} \quad [14](9)$$

$$k = max\left(0, \frac{p\hat{\sigma}^2}{\hat{\beta}'\hat{\beta}} - \frac{1}{n\left(VIF_j\right)_{max}}\right) \quad [19]\ (10)$$

$$k = \frac{\hat{\sigma}^2 \sum_{i=1}^{p}\left(\lambda_i \hat{\beta}_i^2\right)}{\left[s\sum_{i=1}^{p}\left(\lambda_i \hat{\beta}_i^2\right)\right]^2} \quad [37]\ (11)$$

$$k = \frac{\left\{\hat{\sigma}^2 \lambda_{max} \sum_{i=1}^{p}\left(\lambda_i \hat{\beta}_i^2\right) + \left[\sum_{i=1}^{p}\left(\lambda_i \hat{\beta}_i^2\right)\right]^2\right\}}{\lambda_{max} \sum_{i=1}^{p}\left(\lambda_i \hat{\beta}_i^2\right)} \quad [15](12)$$

$$k = max\left(\frac{\hat{\sigma}^2}{\hat{\beta}_i^2} + \frac{1}{\lambda_i}\right), i = 1, 2, \cdots, p \quad [16]\ (13)$$

$$k = \frac{p\hat{\sigma}^2}{\sum_{i=1}^{p}\left\{\hat{\beta}_i^2 / \left[\left[\left(\hat{\beta}_i^4 \lambda_i^2 / 4\hat{\sigma}^2\right) + \left(6\hat{\beta}_i^4 \lambda_i / \hat{\sigma}^2\right)\right]^{1/2} - \left(\hat{\beta}_i^2 \lambda_i / 2\hat{\sigma}^2\right)\right]\right\}}$$

[38] (14)

In this paper, for the purpose of comparing the results we just consider the methods of which a brief introduction is given as below.

[2] Suggested another method for finding **k** value which is given in Equation 15

$$k = \frac{p\hat{\sigma}^2}{\hat{\beta}'\hat{\beta}} \quad (15)$$

In this Equation $\hat{\sigma}^2$ and $\hat{\beta}$ are the OLS estimates. This method is called as *fixed point ridge regression method* (*FPRRM*).

[39] Introduced an iterative method for finding the optimal *k* value. In this method *k* is calculated in Equation 16;

$$k = \frac{p\hat{\sigma}^2(t-1)}{\hat{\beta}(t-1)'\hat{\beta}(t-1)} \quad (16)$$

In this Equation, $\hat{\sigma}^2(t-1)$ and $\hat{\beta}(t-1)$are the corresponding residual mean square and the estimate vector of regression coefficients at *(t-1)th* iteration, respectively. This method is called as *iterative ridge regression method* (*IRRM*).

And also, the generalized ridge regression estimator of Hoerl and Kennard [1, 40] is given in [41] by following Equations 17-20.

Let Λ and Q be the matrices of eigenvalues and eigenvectors of $(X'X)$. In the orthogonal version of the classical linear regression model: $Z = XQ$, $\alpha = Q'\beta$, $\hat{\alpha} = \lambda^{-1}Z'y$, $K = diag(k_1, k_2, \cdots, k_p)$, $k_i > 0$ then

$$\tilde{\beta} = Q(\Lambda + K)^{-1}\Lambda\hat{\alpha} \quad (17)$$

$\tilde{\beta}$ Is the generalized ridge estimator of β. Hoerl and Kennard [1, 40], have shown that the values of k_i which minimize the MSE of regression coefficient are given by

$$k_i = \frac{\sigma^2}{\alpha_i^2} \quad (18)$$

And the estimation of k_i values can be obtained by using Equation 19.

$$\hat{k}_i = \frac{\hat{\sigma}^2}{\hat{\alpha}_i^2} \quad (19)$$

In [41], other estimation formulas for optimum shrinkage parameters are given below.

$$\hat{k}_i = \frac{\lambda_i\hat{\sigma}^2}{(n-k)\hat{\sigma}^2 + \lambda_i\hat{\alpha}_i^2} \quad (20)$$

Methodology

Finding the optimal *k* value is an important problem in ridge regression. The *k* values recommended in the literature were given in the previous section. And also, there are some heuristic methods such as genetic algorithms to find the optimal *k* value in the literature proposed by [18, 21]. And also, [22] have found the *k* value by using particle swarm optimization (PSO). In all these methods suggested in the literature, this *k* value was found as a single value. But in this study, we found different *k* values corresponding to each explanatory variable instead of a single value of *k* by using an algorithm based on particle swarm optimization. And also, this paper is the improvement form of the study of [22].

The objective function of the paper was created by considering both mean absolute percentage error (MAPE) criterion and *VIF* values at the same time. The aim of the objective function is to find the optimal *k* values by finding the *VIF* values less than 10 and SSE (sum of square errors) minimum, at the same time. And also, we add a parameter $(\varnothing(k))$ to the second part of the objective function. This parameter can be called as penalty parameter. If the *VIF* value corresponds to any explanatory variable is bigger than 10 the value of the objective function is increased. This is an effect of the penalty parameter. This is an undesirable result.

The optimization problem in the proposed method can be given in Equation 21.

Citation: Bas E, Egrioglu E, Uslu VR (2016) Shrinkage Parameters for Each Explanatory Variable Found Via Particle Swarm Optimization in Ridge Regression. Trends Comput Sci Inf Technol 1(1): 031-039. DOI: http://dx.doi.org/10.17352/tcsit.000005

Objective function:

$$\min_{k_1,k_2,\cdots,k_p} MAPE\left(k_1,k_2,\cdots,k_p\right)+\varnothing(k_1,k_2,\cdots,k_p) \quad (21)$$

with subject to: $0 \le k_1,k_2,\cdots,k_p \le 1 \quad \left(j=1,2,\cdots,p\right)$

where $MAPE\left(k_1,k_2,\cdots,k_p\right)$ and $\varnothing\left(k_1,k_2,\cdots,k_p\right)$ can be defined in Equations 22 and 23 respectively.

$$MAPE\left(k_1,k_2,\cdots,k_p\right)=\frac{1}{n}\sum_{i=1}^{n}\left|\frac{y_i-\hat{y}_i}{y_i}\right| \quad (22)$$

$$\varnothing\left(k_1,k_2,\cdots,k_p\right)=\begin{cases} 0 & \forall\ VIF_j<10\ , j=1,2,\ldots,p \\ \sum_{j=1}^{p}VIF_j & otherwise \end{cases} \quad (23)$$

(**p** shows the number of explanatory variables.)

The optimization problem defined as in (21) was solved by using PSO in the proposed method. PSO is a popular artificial intelligence technique and it was firstly proposed by [42]. The algorithm of the proposed method is given below.

Algorithm

Step 1. The parameters such as pn, c_1, c_2 etc., are determined. These parameters are as follows:

pn: particle number of swarm

c_1 : Cognitive coefficient

c_2 : Social coefficient interval

maxt: Maximum iteration number

w: Inertia weight

Step 2. Generate random initial positions and velocities.

The initial positions and velocities are generated by uniform distribution with (0,1) parameters. Each particle has velocities up to the number of explanatory variables and each particle has positions up to the number of explanatory variables which represents $\left(k_1,k_2,\cdots,k_p\right)$ values. x_m^t Represents the position of particle ***m*** at iteration ***t*** and v_m^t represents the velocity of the particle ***m*** at iteration ***t***.

Step 3. The fitness function was defined as in (21) and the fitness values of the particles are calculated.

Step 4. *Pbest* and *Gbest* particles given in (24) and (25), respectively, are determined according to fitness values.

$$Pbest_m^t=(pm),\ m=1,2,\ \ldots,\ pn \quad (24)$$

$$Gbest^t=(pg) \quad (25)$$

Pbest is constructed by the best results obtained in the related positions at iteration ***t***. *Gbest* is the best result in the swarm at iteration ***t***.

Step 5. New velocities and positions of the particles are calculated by using the Equations given in (26) and (27).

$$v_m^{t+1}=\begin{bmatrix} w\times v_m^t+c_1\times rand_1\times \\ \left(Pbest_m^t-x_m^t\right)+c_2\times rand_2\times\left(Gbest^t-x_m^t\right)\end{bmatrix} \quad (26)$$

$$x_m^{t+1}=x_m^t+v_m^{t+1} \quad (27)$$

Where $rand_1$ and $rand_2$ are random numbers generated from *U (0,1).*

Step 6. Step 3 to Step 6 is repeated until t<*maxt.*

Step 7. The optimal $\left(k_1,k_2,\cdots,k_p\right)$ values are obtained as *Gbest*.

Implementation

The proposed algorithm was applied to two different and well known data sets in order to investigate of the proposed method. These two data sets named "Import Data" and "Longley Data" were used to evaluate the performance of the proposed method. Import data was analyzed by [43]. The variables of "Import Data" are; imports (IMPORT-Y), domestic production (DOPROD-X1), stock formation (STOCK-X2) and domestic consumption (CONSUM-X3), all measured in billions of French francs for the years 1949 through 1959. Both Import data and Longley data were solved by using fixed point method ([2]), iterative method ([39]), [22]'s method and the algorithm proposed in this paper. In the proposed algorithm, PSO parameters were chosen as $pn=30, w=0.9, c_1=c_2=2$ and $max\,t=100$. In the iterative ridge method the stopping criteria were chosen as $=10^{-6}$. The results of each method were presented in Tables 2 and 3, respectively.

As we can see from Table 2, our proposed method has minimum SSE and MAPE values. And also there is no multicollinearity problem when "Import Data" solved by our proposed method. But, there is a multicollinearity problem when "Import Data" solved by *FPRRM* and *IRRM* methods because of the *VIF* values of these methods are bigger than 10. Although, other methods can give smaller SSE and MAPE values they do not still solve the multicollinearity problem. Because it is clearly seen that some *VIF* values of these methods are greater than 10.

As we can see from Table 3, our proposed method has minimum MAPE value when compared with other methods. But SSE value of our proposed method is not the smallest one. The SSE value of OLS is smaller than our proposed methods. But, it is clearly seen that the OLS method has multicollinearity problem when "Longley Data" solved by this method. But our proposed method has no multicollinearity problem.

As a result, finding *k* values for each explanatory variable gives better results than finding a single ***k*** value. And also, our proposed has no multicollinearity problem.

Simulation study

Two different simulation studies are performed in this section of the paper in order to show the performance of the

Citation: Bas E, Egrioglu E, Uslu VR (2016) Shrinkage Parameters for Each Explanatory Variable Found Via Particle Swarm Optimization in Ridge Regression. Trends Comput Sci Inf Technol 1(1): 031-039. DOI: http://dx.doi.org/10.17352/tcsit.000005

proposed method in different levels of multicollinearity and standard deviation of error term and the superiority of the proposed method when compared with other methods.

The First Simulation Study: In this simulation study, the proposed method was compared with ridge regression methods given in [2,22,39] by a simulation study. The number of observations *(n)* was taken as 100, 500 and 1000; the standard deviation of error term (σ) was taken as 0.01 and 1 and comparisons were made for the total 6 cases. For each case, 1000 data set including multicollinearity problem was created.

The first three independent variables were generated from standard normal distribution as given in Equation 28.

$$X_i \sim N(0,1) \; i=1,2,3 \tag{28}$$

The last two independent variables were generated by using Equation 29. Thus, it is provided to arise multicollinearity problem for the data set by providing a high correlation between independent variables X_1 and X_4, X_1 and X_5.

$$X_i = U(10,20)+U(5,20)X_1+N(0,7)\, i=4,5 \tag{29}$$

The observations of dependent variable were obtained using Equation 30. So, all the coefficients in the regression model are taken as 1.

$$Y=\sum X_i + N(0,\sigma) \tag{30}$$

For each data generated in each case, $\sum VIF^2$, SSE, MAPE and *CN* values are calculated by using proposed method, the studies [2, 22, 39]. The formula of SSE is given in Equation 31.

$$SSE=\sum_{i=1}^{n}(y_i-\hat{y}_i)^2 \tag{31}$$

The most important indicator for the comparison of methods is that *VIF* and *CN* would be small. The methods [2] and [39] do not guarantee the solution of multicollinearity problem as seen in the numerical examples. The method [22] and proposed method guarantee that all *VIF* values are smaller than 10. Therefore, it is suitable to compare the proposed method with [22] method in terms of SSE and MAPE criteria.

The results of median and inter quartile range (IQR) values were given between Tables 4-9.

When all tables are examined, it is clearly seen that $\sum VIF^2$ and *CN* values of proposed method is lower than the other methods in all cases.

However, it is seen that the proposed method produces lower MAPE values compared to others despite producing higher SSE values. This is because the objective function of the proposed method may be depending to the MAPE.

The Second Simulation Study: A second simulation study was performed in the paper according to different levels of multicollinearity problem and standard deviation of error term. The regressors were generated by using Equations 32-36 given by [44].

$$w_{ij} \sim N(0,1) \;\; ; i=1,2,\ldots,n \;\; ; \; j=1,2,\ldots,6 \tag{32}$$

$$x_{ij}=(1-\rho^2)^{1/2}w_{ij}+\rho w_{i,6}; \quad i=1,2,\ldots,n \;; j=1,2,3 \tag{33}$$

$$x_{ij}=w_{ij} \;\; ,i=1,2,\ldots,n \;; \; j=4,5 \tag{34}$$

Table 2: The comparison of ***VIF*** values, SSE and MAPE obtained from OLS, FPRRM, IRRM, [22] and proposed method for Import Data.

	OLS (k=0)		[2] (k=0.0016)		[39] (k=0.0042)		[22] k=(0.0090)		Proposed Ridge Method	k values obtained from Proposed Ridge Method	
Variable	S.C.	VIF	S.C.	VIF	S.C.	VIF	S.C.	VIF	S.C. VIF		
X1	-0.34	186.11	-0.03	72.09	0.16	27.99	0.29	9.99	-0.07 9.03	k_1	0.0190
X2	0.21	1.02	0.22	1.00	0.22	1.00	0.22	0.98	0.21 1.00	k_2	0
X3	1.30	186.00	0.99	72.13	0.80	28.01	0.67	9.99	1.03 9.99	k_3	0
SSE	0.0810		0.0086		0.0095		0.0103		0.0084		
MAPE	0.1196		0.1097		0.1139		0.1185		0.1088		

S.C.: Standardized Coefficients.

Table 3: The ***VIF*** Values, SSE and MAPE values obtained from OLS, FPRRM, IRRM, [22] and proposed method for Longley Data.

	OLS (k=0)		[2] (k=0.0003)		[39] (k=0.0006)		[22] (k=0.0172)		Proposed Ridge Method		k values obtained from Proposed Ridge Method	
Variable	S.C.	VIF	S.C.	VIF	S.C.	VIF	S.C.	VIF	S.C.	VIF		
X1	0.05	135.53	-0.01	87.32	-0.01	87.32	0.25	9.99	-0.002	0.59	k_1	0.1374
X2	-1.01	1788.51	-0.25	472.15	-0.25	472.15	0.34	1.55	-0.001	0.00	k_2	0.7274
X3	-0.54	33.62	-0.43	10.95	-0.43	10.95	-0.28	2.55	-0.37	3.03	k_3	0.0140
X4	-0.20	3.59	-0.18	2.88	-0.18	2.88	-0.11	1.94	-0.15	2.17	k_4	0
X5	-0.10	399.15	-0.28	180.21	-0.28	180.21	0.16	7.32	-0.10	5.94	k_5	0.0295
X6	2.48	758.98	1.88	309.82	1.88	309.82	0.46	4.07	1.39	9.99	k_6	0
SSE	0.0045		0.0050		0.0050		0.0123		0.0065			
MAPE	0.0002		0.0753		0.0753		0.1378		0.0826			

S.C.: Standardized Coefficients.

Citation: Bas E, Egrioglu E, Uslu VR (2016) Shrinkage Parameters for Each Explanatory Variable Found Via Particle Swarm Optimization in Ridge Regression. Trends Comput Sci Inf Technol 1(1): 031-039. DOI: http://dx.doi.org/10.17352/tcsit.000005

$$e_i \sim N(0,\sigma)\ ; i = 1,2,\ldots,n \tag{35}$$

$$y_i = \sum_{j=1}^{5} \beta_j x_{i,j} + e_i \quad ; i = 1,2,\ldots,n \tag{36}$$

Where $w_{i,j}$ independent standard normal are pseudorandom numbers and ρ^2 is theoretical correlation between any two explanatory variables.

Simulation study was conducted for a total of 8 cases for

Table 4: Simulation results for n=100, $\sigma = 0.01$

Method		[39]	[2]	[22]	Proposed Method
SSE	Median	0.00001259	0.00001259	0.00001363	0.00001576
	IQR	0.00000756	0.00000756	0.00000931	0.00001270
MAPE	Median	0.00134451	0.00134451	0.00127049	0.00112317
	IQR	0.00106196	0.00106196	0.00089539	0.00059409
$\sum VIF^2$	Median	0.01140365	0.01140365	0.01139037	0.01127907
	IQR	0.01396126	0.01396126	0.01394571	0.01372089
CN	Median	34.35277075	34.35277075	34.33591354	34.27939908
	IQR	24.18839643	24.18839643	24.11194855	23.79666414

Table 5: Simulation results for n=100, $\sigma = 0.01$

Method		[39]	[2]	[22]	Proposed Method
SSE	Median	0.11927554	0.11927554	0.12470745	0.12622844
	IQR	0.07784014	0.07784014	0.08136102	0.08116543
MAPE	Median	0.13146684	0.13146684	0.12731270	0.12122859
	IQR	0.10275617	0.10275617	0.09496648	0.08890188
$\sum VIF^2$	Median	0.01265217	0.01265217	0.01055620	0.01031630
	IQR	0.01422383	0.01422383	0.01250313	0.01081028
CN	Median	36.49713949	36.49713949	34.82762683	34.19546839
	IQR	23.90916228	23.90916228	22.72993530	22.03645100

Table 6: Simulation results for n=500, $\sigma = 0.01$

Method		[39]	[2]	[22]	Proposed Method
SSE	Median	0.00006273	0.00006273	0.00006657	0.00007443
	IQR	0.00003939	0.00003939	0.00004521	0.00005199
MAPE	Median	0.00176972	0.00176972	0.00171829	0.00155416
	IQR	0.00120199	0.00120199	0.00109733	0.00082172
$\sum VIF^2$	Median	0.00047198	0.00047198	0.00047191	0.00045946
	IQR	0.00053460	0.00053460	0.00053380	0.00052792
CN	Median	35.12943563	35.12943563	35.12640617	34.87031035
	IQR	23.51132659	23.51132659	23.49173308	23.41623957

Table 7: Simulation results for n=500, $\sigma = 0.01$

Method		[39]	[2]	[22]	Proposed Method
SSE	Median	0.62581119	0.62581119	0.62641000	0.62643082
	IQR	0.40863631	0.40863631	0.40571868	0.40572352
MAPE	Median	0.17897917	0.17897917	0.17762626	0.17610858
	IQR	0.13148961	0.13148961	0.12861139	0.12632016
$\sum VIF^2$	Median	0.00047277	0.00047277	0.00044964	0.00044955
	IQR	0.00054171	0.00054171	0.00050993	0.00050497
CN	Median	35.15161545	35.15161545	34.73612619	34.78910967
	IQR	24.17568961	24.17568961	23.75550252	23.71548866

Table 8: Simulation results for n=1000, $\sigma = 0.01$

Method		[39]	[2]	[22]	Proposed Method
SSE	Median	0.00012916	0.00012916	0.00013700	0.00014808
	IQR	0.00008243	0.00008243	0.00008840	0.00009883
MAPE	Median	0.00193169	0.00193169	0.00186306	0.00173776
	IQR	0.00128107	0.00128107	0.00108223	0.00093393
$\sum VIF^2$	Median	0.00011623	0.00011623	0.00011571	0.00011366
	IQR	0.00012895	0.00012895	0.00012871	0.00012574
CN	Median	34.73295370	34.73295370	34.68869608	34.57542791
	IQR	22.68009238	22.68009238	22.68237746	22.50717173

Table 9: Simulation results for n=1000, $\sigma = 0.01$

Method		[39]	[2]	[22]	Proposed Method
SSE	Median	1.28108799	1.28108799	1.28125942	1.28152484
	IQR	0.82914889	0.82914889	0.82826388	0.82752823
MAPE	Median	0.19569205	0.19569205	0.19433497	0.19348508
	IQR	0.12670340	0.12670340	0.12694189	0.12641619
$\sum VIF^2$	Median	0.00011531	0.00011531	0.00011361	0.00011275
	IQR	0.00012654	0.00012654	0.00012256	0.00012304
CN	Median	34.91165352	34.91165352	34.64929969	34.61115844
	IQR	22.82979136	22.82979136	22.64350810	22.69768492

sample size is 100, $(n=100)$, standard deviation of the standard deviation of error term ($\sigma = 0.01, 0.1, 1, 5$) and different degrees of multiple connections ($\rho = 0.99,\ 0.999$) (Tables 10-17).

It is clearly seen that in the tables of the simulation Study 2, $\sum VIF^2$ and **CN** values of the proposed method do not change significantly when standard deviation of error term values are changed. $\sum VIF^2$ And **CN** values of the proposed method are increased dramatically when multicollinearity is increased. And also there is no a hardly ever change to be seen in the MAPE values of the proposed method with the reasonable standard deviation of error term values $(\sigma = 0.01, 0.1)$ or there is a decrease to be seen in the MAPE values of the proposed method when multicollinearity is increased.

In this simulation study, different levels of standard deviation of error term are also employed. As a result of this simulation study it is clearly seen that when standard deviation of error term value is greater than 1 and >1 the model has very big deviation from linear regression model because MAPE values are obtained about 60 and this value is not suitable. And also, it is clearly seen that in the tables of the simulation study 2, the prediction performance of the proposed is affected quite negatively when standard deviation of error term is increased.

Discussion

There are some valid assumptions to create a model in multiple regression analysis. One of them is that it should not be multicollinearity problem among independent variables. Ridge regression method is often used in the literature when there is a multicollinearity problem among independent variables.

But, ridge regression has also some problems. One of the most important problems in ridge regression is to decide what

Citation: Bas E, Egrioglu E, Uslu VR (2016) Shrinkage Parameters for Each Explanatory Variable Found Via Particle Swarm Optimization in Ridge Regression. Trends Comput Sci Inf Technol 1(1): 031-039. DOI: http://dx.doi.org/10.17352/tcsit.000005

the shrinkage parameter (k) value will be. There are many studies in the literature to find the optimal k value. In these studies, this k value was found to be a single value. But in this study, we found different k values corresponding to each explanatory variable instead of a single value of k by using a new algorithm based on particle swarm optimization. And also, the proposed method was supported by two simulation studies. Besides, it is an important novelty for ridge regression literature.

Table 10: Simulation results for n=100, $\rho=0.99, \sigma=0.01$

Method		[39]	[2]	[22]	Proposed Method
SSE	Median	0.01102	0.01102	0.00863	0.01069
	IQR	0.00670	0.00671	0.00310	0.00595
MAPE	Median	0.00089	0.00089	0.00113	0.00093
	IQR	0.00028	0.00028	0.00054	0.00032
$\sum VIF^2$	Median	3.70E-01	3.70E-01	3.33E-01	3.03E-01
	IQR	2.11E-01	2.11E-01	2.06E-01	2.13E-01
CN	Median	171.30708	171.30710	167.37020	163.00550
	IQR	57.08409	57.08410	55.21013	59.74116

Table 11: Simulation results for n=100, $\rho=0.99, \sigma=0.1$

Method		[39]	[2]	[22]	Proposed Method
SSE	Median	0.10852	0.10852	0.08735	0.10578
	IQR	0.07191	0.07191	0.05225	0.07341
MAPE	Median	0.08589	0.08589	0.09975	0.08738
	IQR	0.02448	0.02448	0.02849	0.02516
$\sum VIF^2$	Median	0.36885	0.36885	0.20309	0.14416
	IQR	0.17361	0.17361	0.12232	0.31462
CN	Median	174.00479	174.00480	150.87468	142.72940
	IQR	47.43572	47.43573	31.07611	66.63963

Table 12: Simulation results for n=100, $\rho=0.99, \sigma=1$

Method		[39]	[2]	[22]	Proposed Method
SSE	Median	1.05139	1.052158	1.009223	1.041334
	IQR	0.893263	0.893868	0.894413	0.894043
MAPE	Median	7.806875	7.806608	7.805308	7.803432
	IQR	1.835182	1.832176	1.831011	1.828457
$\sum VIF^2$	Median	0.073703	0.082648	0.14884	0.074125
	IQR	0.031235	0.036015	0.070147	0.265868
CN	Median	110.7648	114.7465	141.0569	121.504
	IQR	25.48441	25.1948	30.39541	51.88191

Table 13: Simulation results for n=100, $\rho=0.99, \sigma=5$

Method		[39]	[2]	[22]	Proposed Method
SSE	Median	1.872207	2.004527	2.009562	2.013568
	IQR	0.889942	0.999704	1.121944	1.141227
MAPE	Median	67.69349	66.27068	65.8774	65.92444
	IQR	8.280155	8.75298	8.700723	8.611794
$\sum VIF^2$	Median	8.19E-05	0.002858	0.138193	0.073933
	IQR	5.31E-05	0.009572	0.114742	0.163497
CN	Median	9.919616	49.19599	138.8491	118.3783
	IQR	4.636951	37.61656	40.51028	37.26164

Table 14: Simulation results for n=100, $\rho=0.99, \sigma=5$

Method		[39]	[2]	[22]	Proposed Method
SSE	Median	0.011187013	0.011187	0.008698	0.010818
	IQR	0.013549241	0.013549	0.004648	0.011893
MAPE	Median	0.000855509	0.000856	0.001076	0.000899
	IQR	0.000231304	0.000231	0.000559	0.000331
$\sum VIF^2$	Median	3.90E+01	3.90E+01	3.03E+01	2.96E+01
	IQR	1.36E+01	1.36E+01	1.76E+01	3.43E+01
CN	Median	1783.234141	1783.234	1623.275	1614.635
	IQR	359.552608	359.5526	483.9845	793.0598

Table 15: Simulation results for n=100, $\rho=0.99, \sigma=5$

Method		[39]	[2]	[22]	Proposed Method
SSE	Median	0.102689	0.102689	0.086116985	0.100247
	IQR	0.08151503	0.081515	0.043947845	0.07444
MAPE	Median	0.086337629	0.086338	0.100399072	0.087094
	IQR	0.020985993	0.020986	0.028808332	0.020984
$\sum VIF^2$	Median	29.72544998	29.72545	5.737783476	8.786416
	IQR	9.587968933	9.587969	12.33683362	36.05398
CN	Median	1645.428198	1645.428	1129.678979	1158.36
	IQR	379.3129664	379.313	741.150849	1433.807

Table 16: Simulation results for n=100, $\rho=0.99, \sigma=5$

Method		[39]	[2]	[22]	Proposed Method
SSE	Median	1.065452	1.061253	1.029123	1.050466
	IQR	0.564163	0.588576	0.547797	0.589277
MAPE	Median	7.797641	7.746442	7.806336	7.749347
	IQR	2.119962	2.104055	2.078705	2.111888
$\sum VIF^2$	Median	0.0345	0.430398	0.746483	0.040122
	IQR	0.027987	1.125758	17.89286	35.6086
CN	Median	273.5651	542.3747	678.4479	269.9127
	IQR	80.37178	338.2685	1084.278	1460.714

Table 17: Simulation results for n=100, $\rho=0.99, \sigma=5$

Method		[39]	[2]	[22]	Proposed Method
SSE	Median	1.804969	1.998441	1.945228	1.964092
	IQR	1.006492	1.153566	1.163403	1.240094
MAPE	Median	67.46692	66.15954	66.3527	66.40931
	IQR	9.110043	8.070038	8.292751	8.077823
$\sum VIF^2$	Median	6.79E-05	0.045357	8.733978	0.034631
	IQR	3.59E-05	0.496616	17.541	25.19872
CN	Median	9.945675	287.4393	1306.078	267.989
	IQR	4.781496	430.5148	998.2636	1204.904

In the future studies, different artificial intelligence optimization techniques can be used to find these k values for each explanatory variable.

References

1. Hoerl AE, Kennard RW (1970) Ridge regression: biased estimation for nonorthogonal problems. Technometrics 12: 55–67. Link: https://goo.gl/5ZV56T

2. Hoerl AE, Kennard RW, Baldwin KF (1975) Ridge regression: some simulations. Communications in Statistics 4: 105–123. Link: https://goo.gl/QGgP3L

Citation: Bas E, Egrioglu E, Uslu VR (2016) Shrinkage Parameters for Each Explanatory Variable Found Via Particle Swarm Optimization in Ridge Regression. Trends Comput Sci Inf Technol 1(1): 031-039. DOI: http://dx.doi.org/10.17352/tcsit.000005

3. McDonald GC, Galarneau DI (1975) A Monte Carlo evaluation of some ridge-type estimators. Journal of the American Statistical Association 70: 407–412. Link: https://goo.gl/7ZN2co

4. Lawless JF, Wang P (1976) A simulation study of ridge and other regression estimators. Communications in Statistics – Theory and Methods 14: 1589–1604. Link: https://goo.gl/WfUz0p

5. Hocking RR, Speed FM, Lynn MJ (1976) A class of biased estimators in linear regression. Technometrics 18: 425–437. Link: https://goo.gl/NEjsRY

6. Gunst RF, Mason RL (1977) Biased estimation in regression: an evaluation using mean squared error. Journal of the American Statistical Association 72: 616–628. Link: https://goo.gl/HfxIin

7. Wichern D, Curchill G (1978) A comparison of ridge estimators. Technometrics 20: 301-311. Link: https://goo.gl/U6OiUQ

8. Lawless JF (1978) Ridge and related estimation procedure Theory and Methods. Communications in Statistics 7: 139–164. Link: https://goo.gl/KceYME

9. Nordberg L (1982) A procedure for determination of a good ridge parameter in linear regression, Communications in Statistics 11: 285–309. Link: https://goo.gl/pNqtc2

10. Saleh AK, Kibria BM (1993) Performances of some new preliminary test ridge regression estimators and their properties. Communications in Statistics – Theory and Methods 22: 2747–2764. Link: https://goo.gl/4XqQNd

11. Haq MS, Kibria BMG (1996) a shrinkage estimator for the restricted linear regression model: ridge regression approach. Journal of Applied Statistical Science 3: 301–316. Link: https://goo.gl/sjCZrw

12. Kibria BM (2003) Performance of some new ridge regression estimators. Communications in Statistics – Simulation and Computation 32: 419–435. Link: https://goo.gl/3OJp6a

13. Pasha GR, Shah MA (2004) Application of ridge regression to multicollinear data. Journal of Research Science 15: 97– 106. Link: https://goo.gl/5eP2I5

14. Khalaf G, Shukur G (2005) Choosing ridge parameter for regression problem. Communications in Statistics – Theory and Methods 34: 1177–1182. Link: https://goo.gl/Nu1Xs4

15. Norliza A, Maizah HA, Robin A (2006) A comparative study on some methods for handling multicollinearity problems. Mathematika 22: 109–119. Link: https://goo.gl/Tlyqej

16. Alkhamisi MA, Shukur G (2007) A Monte Carlo study of recent ridge parameters. Communications in Statistics – Simulation and Computation 36: 535–547. Link: https://goo.gl/Mv2FMY

17. Mardikyan S, Cetin E (2008) Efficient choice of biasing constant for ridge regression. Int. J. Contemp. Math. Sciences, 3: 527–536. Link: https://goo.gl/oOgsiH

18. Prago-Alejo RJ,Torre-Trevino LM, Pina-Monarrez MR (2008) Optimal determination of k constant of ridge regression using a simple genetic algorithm. Electronics robotics and Automotive Mechanics Conference. Link: https://goo.gl/uPVi0B

19. Dorugade AV, Kashid DN (2010) Alternative method for choosing ridge parameter for regression. Applied Mathematical Sciences 4: 447–456. Link: https://goo.gl/E7MYJ5

20. Al-Hassan Y (2010) Performance of new ridge regression estimators. Journal of the Association of Arab Universities for Basic and Applied Science 9: 23–26. Link: https://goo.gl/zTjoEe

21. Ahn JJ, Byun HW, Oh KJ, Kim TY (2012) Using ridge regression with genetic algorithm to enhance real estate appraisal forecasting. Expert Systems with Applications 39: 8369–8379. Link: https://goo.gl/TM0Udi

22. Uslu VR, Egrioglu E, Bas E (2014) Finding optimal value for the shrinkage parameter in ridge regression via particle swarm optimization. American Journal of Intelligent Systems 4: 142-147. Link: https://goo.gl/U06GuG

23. Chitsaz S, Ahmed SE (2012) Shrinkage estimation for the regression parameter matrix in multivariate regression model. Journal of Statistical Computation and Simulation 82: 309-323. Link: https://goo.gl/lDIZzU

24. Firinguetti L (1997) Ridge regression in the context of a system of seemingly unrelated regression equations. Journal of Statistical Computation and Simulation 56: 145-162. Link: https://goo.gl/wWROue

25. Halawa AM, El Bassiouni MY (2000) Tests of regression coefficients under ridge regression models. Journal of Statistical Computation and Simulation 65: 341-356. Link: https://goo.gl/LUQbtW

26. Dorugade AV, Kashid DN (2010) Variable selection in linear regression based on ridge estimator. Journal of Statistical Computation and Simulation 80: 1211-1224. Link: https://goo.gl/A0WJm7

27. Golam Kibria BM (2004) Performance of the shrinkage preliminary test ridge regression estimators based on the conflicting of W, LR and LM tests, Journal of Statistical Computation and Simulation 74: 793-810. Link: https://goo.gl/TMvLYe

28. Roozbeh M, Arashi M, Niroumand HA (2011) Ridge regression methodology in partial linear models with correlated errors. Journal of Statistical Computation and Simulation 81: 517-528. Link: https://goo.gl/Y2r4Nz

29. Simpsona JR, Montgomery DC (1996) A biased-robust regression technique for the combined outlier-multicollinearity problem. Journal of Statistical Computation and Simulation 56: 1-22. Link: https://goo.gl/qgK7Fz

30. Uzuke CA, Mbegbu JI, Nwosu CR (2015) Performance of kibria, khalaf and shurkur's methods when the eigenvalues are skewed. Communications in Statistics - Simulation and Computation Link: https://goo.gl/VdvgIo

31. Wong KY, Chiu SN (2015) an iterative approach to minimize the mean squared error in ridge regression. Computational Statistics 30: 625-639. Link: https://goo.gl/rdHK1p

32. Dorugade AV (2014) new ridge parameters for ridge regression. Journal of the Association of Arab Universities for Basic and Applied Sciences 15: 94-99. Link: https://goo.gl/wpNPWf

33. Khalaf G (2013) An optimal estimation for the ridge regression parameter. Journal of Fundamental and Applied Statistics 5: 11-19.Link: https://goo.gl/bo4bde

34. Muniz G, Golam Kibria BM, Månsson K, Ghazi S (2012) On developing ridge regression parameters: a graphical investigation. Sort-Statistics and Operations Research Transactions 36: 115-138. Link: https://goo.gl/o13EFW

35. Muniz G, Golam Kibria BM (2009) On some ridge regression estimators: an empirical comparisons. Communications in Statistics - Simulation and Computation 38: 621-630. Link: https://goo.gl/wqCKbh

36. Nomura M (1988) On the almost unbiased ridge regression estimation. Communications in Statistics – Simulation and Computation 17: 729–743. Link: https://goo.gl/5kX0MM

37. Montogomery DC, Peck EA, Vining GG (2006) Introduction to Linear Regression Analysis. John Wiley and Sons. Link: https://goo.gl/M3tgXY

38. Batah FS, Ramnathan T, Gore SD (2008) The efficiency of modified jackknife and ridge type regression estimators: a comparison. Surveys in Mathematics and its Applications 3: 111–122. Link: https://goo.gl/bw8Xcf

39. Hoerl AE, Kennard RW (1976) Ridge regression: iterative estimation of the biasing parameter. Commun. Statist. Theor. Meth.5: 77-88. Link: https://goo.gl/VxoSpF

40. Hoerl AE, Kennard RW (1970) Ridge Regression: Applications

Citation: Bas E, Egrioglu E, Uslu VR (2016) Shrinkage Parameters for Each Explanatory Variable Found Via Particle Swarm Optimization in Ridge Regression. Trends Comput Sci Inf Technol 1(1): 031-039. DOI: http://dx.doi.org/10.17352/tcsit.000005

to Nonorthogorial Problems. Technometrics 12: 69-82. Link: https://goo.gl/HKnemY

41. Firinguetti L (1999) A generalized ridge regression estimator and its finite sample properties. Communications in Statistics-Theory and Methods 28: 1217-1229. Link: https://goo.gl/VTzhbj

42. Kennedy J, Eberhart R (1995) Particle swarm optimization. In Proceedings of IEEE International Conference on Neural Networks, Piscataway, NJ, USA, IEEE Press. 1942–1948

43. Chatterjee S, Hadi (2006) A Regression Analysis by Example. John Wiley and Sons. Link: https://goo.gl/Dx6iqn

44. Gibbons DG (1981) A simulation study of some ridge estimators. Journal of the American Statistical Association 76:131–139. Link: https://goo.gl/XMzqJs

Citation: Bas E, Egrioglu E, Uslu VR (2016) Shrinkage Parameters for Each Explanatory Variable Found Via Particle Swarm Optimization in Ridge Regression. Trends Comput Sci Inf Technol 1(1): 031-039. DOI: http://dx.doi.org/10.17352/tcsit.000005