Report

Regression

CSE 574 Assignment 2

Submitted By:-

Name	**UBIT name**	**Person #**
Pratik Bhat	pratiksu	50096806
Anirudha Karwa	akarwa	50097742
Andreina Uzcategui	auzcateg	50056733

1. Calculate and report the RSE for training and test data for two cases: first, without using an intercept (or bias) term, and second with using an intercept. Which one is better?

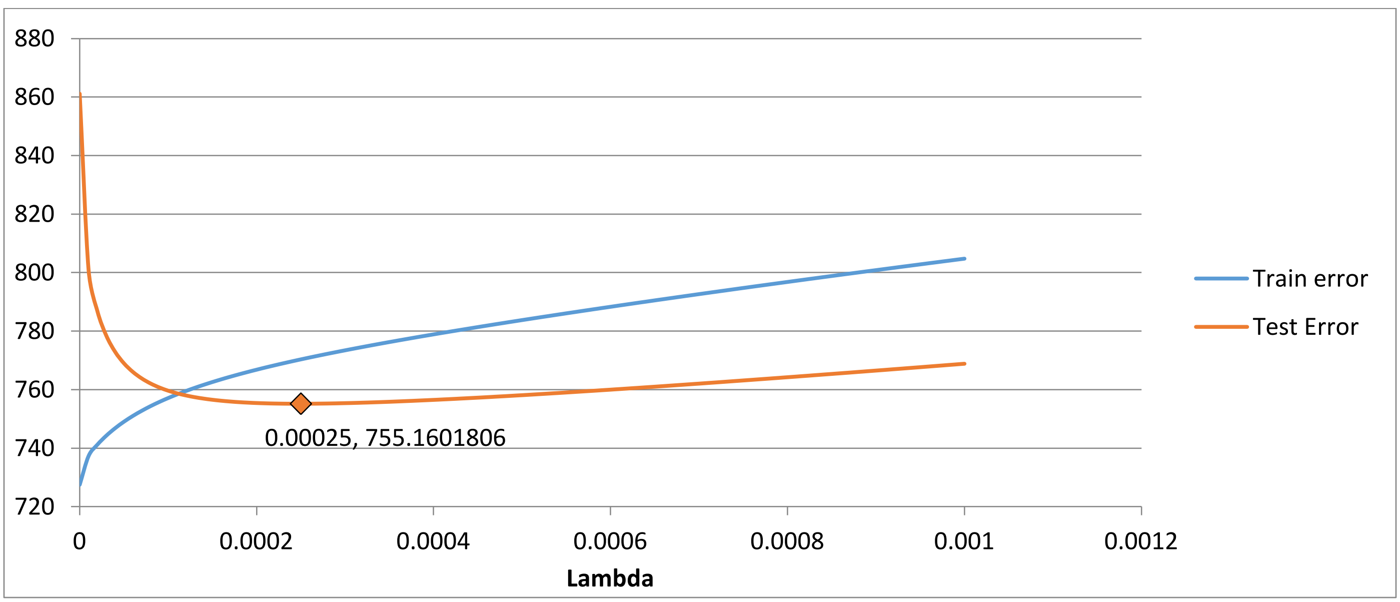
The following results are observed for RSE

	Training Error	**Test Error**
Without Intercept	2149.9	4621.2
With Intercept	727.5251	861.1434

As can be seen from the results, it is evident that using an intercept (or bias) greatly minimizes the error. Thus the approach of linear regression **with intercept** in which testing data gives minimum RSE is a better option in this case.

2. Calculate and report the RSE for training and test data using ridge regression parameters. Use data with intercept. Plot the errors on train and test data for different values of λ. Vary λ from 0 (no regularization) to 0.001 in steps of 0.00001. Compare the relative magnitudes of weights learnt using OLE (Problem 1) and weights learnt using ridge regression. Compare the two approaches in terms of errors on train and test data. What is the optimal value for λ and why?

The following plot shows RSE for train and test data for different values of **λ**



The following table compares relative magnitudes of weights learnt using OLE (Problem 1) and weights learnt using ridge regression

	Weights		
	Avg	**Min**	**Max**
OLE(with intercept)	882.8076218	-86639.45675	75914.46769
Ridge Regression	32.00730077	-202.5807345	495.6064395

Such a drastic decrease in weight values can be attributed to the usage of regularization parameter λ in the calculation of weights.

The objective function used for Ridge Regression is

$$J(\mathbf{w}) = \frac{1}{N}\sum_{i=1}^{N}(y_i - \mathbf{w}^\top \mathbf{x}_i)^2 + \lambda \mathbf{w}^\top \mathbf{w}$$

The following table compares the minimum test and train errors for both the approaches

	Train Error	**Test Error**
OLE (with intercept)	727.5251	861.1434
Ridge Regression	727.5251	755.1602

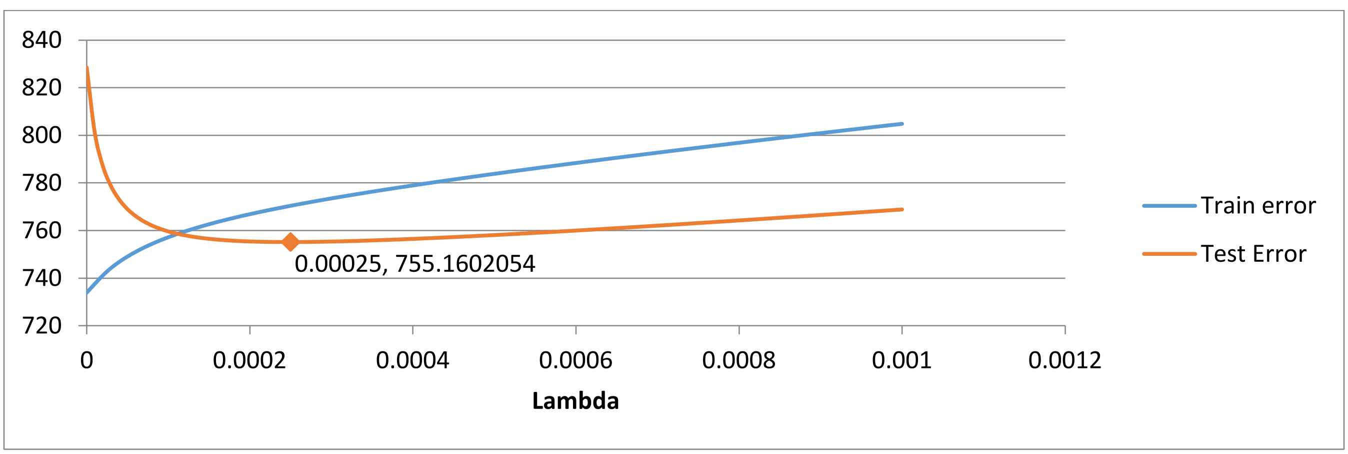
The train error for Ridge Regression is minimum at λ= 0, being 727.5251 and goes on increasing as λ increases. Hence we see that minimum train error for both the approaches is the same as OLE is equivalent to λ= 0 in Ridge Regression.

As can be observed in the above graph, the test error decreases with λ increasing until a certain point and then test error keeps on increasing. **The optimal value for λ is 0.00025 as we get the minimum testing error at that particular point.**

We can observe that test error decreases for ridge regression as compared to OLE because of the regularization parameter used while calculating weights.

3. **Plot the errors on train and test data obtained by using the gradient descent based learning by varying the regularization parameter. Compare with the results obtained in Problem 2.**

The following graph shows the RSE for train and test data using gradient descent based learning for ridge regression



In this approach we have used the gradient descent to calculate weights. This approach is used when, at times analytical calculation of $(\mathbf{X}^\top \mathbf{X})^{-1}$ tends to be computationally expensive when X is complex.

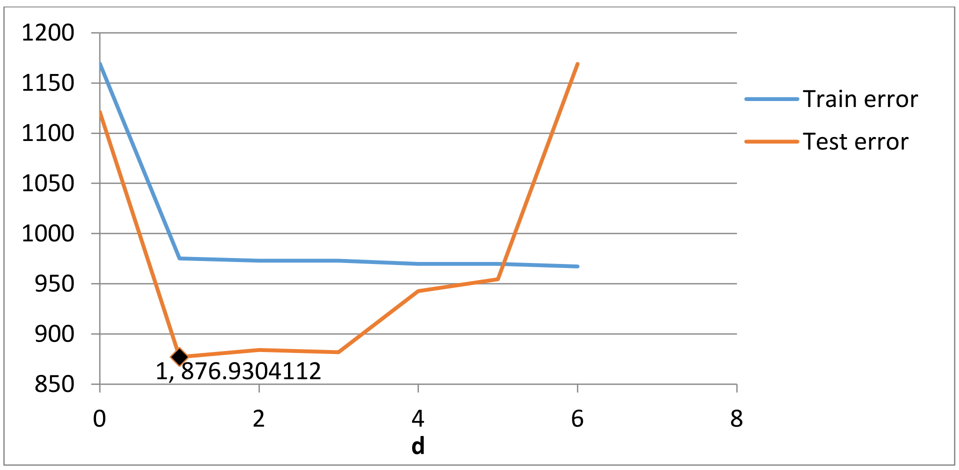
The following table compares error values calculated using analytical and gradient descent approach respectively

Method	Train Error	Test Error
Analytical	727.5251	755.1602
Gradient descent	733.6154	755.1602

We can observe that both the approaches produce almost similar results. However the gradient descent approach takes more running time depending on the number of iterations.

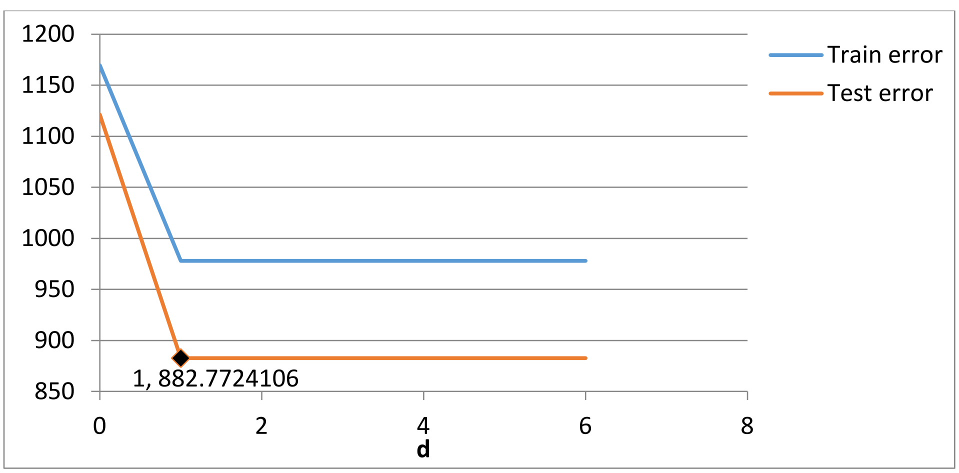
4. **Using the λ = 0 and the optimal value of λ found in Problem 2, train ridge regression weights using the non-linear mapping of the data. Vary d from 0 to 6. Note that d = 0 means using a horizontal line as the regression line, d = 1 is the same as linear ridge regression. Compute the errors on train and test data. Compare the results for both values of λ. What is the optimal value of d in terms of test error in each setting? Plot the curve for the optimal value of d for both values of λ and compare.**

The following table and graph depicts errors on train and test data for λ = 0



D	Train error	Test error
0	1169.389563	1121.2854
1	975.3366232	876.930411
2	972.9672145	883.982816
3	972.8862508	881.813533
4	969.682671	942.690606
5	969.6744464	954.445428
6	967.3602198	1169.05596

The following table and graph depicts errors on train and test data for λ = 0.00025 (Optimal)



D	Train error	Test error
0	1169.389707	1121.32832
1	977.9674663	882.772411
2	977.8257972	882.740964
3	977.8252045	882.7408
4	977.8251806	882.740794
5	977.8251804	882.740794
6	977.8251803	882.740794

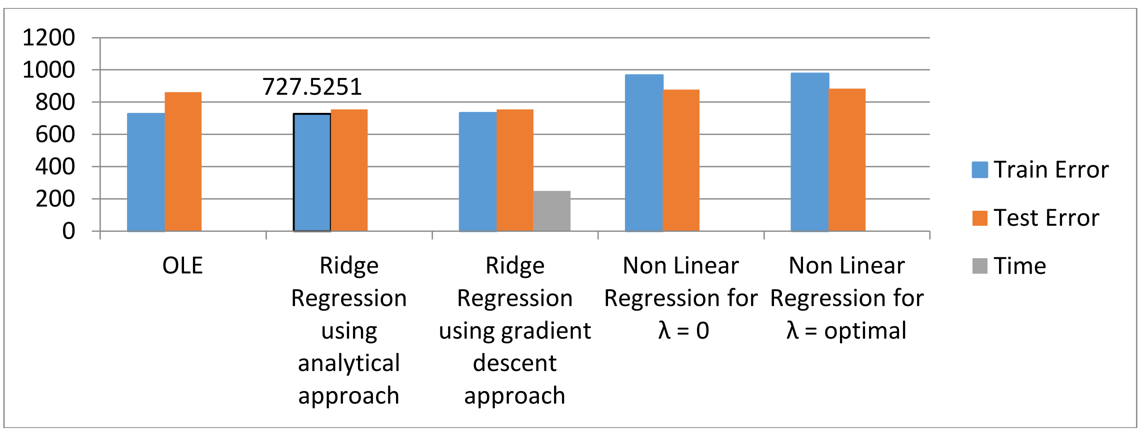
For λ = 0, the test error is minimum when d = 1 and increases as d increases. The optimal value in this setting is when d = 1 when the test error is minimum at 876.930411

For λ = 0.00025 (Optimal), the test error is almost constant at all values of d from 1 to 6. This can be verified from the above graph. Thus we can say that there are multiple optimal values of d i.e. 1-6 as test error is minimum for these values of d. However we can go for value d=1 as higher values of d will require more computation.

5. **Using the results obtained for previous 4 problems; make final recommendations for anyone using regression for predicting diabetes level using the input features. Compare the various approaches in terms of training and testing error. What metric should be used to choose the best setting?**

The following table describes the various approaches used to perform regression

	Train Error	**Test Error**	**Time**
OLE	727.5251	861.1434	0.003329
Ridge Regression using analytical approach	727.5251	755.1602	0.403656
Ridge Regression using gradient descent approach	733.6154	755.1602	248.581607
Non Linear Regression for λ = 0	967.3602	876.9304	0.028347
Non Linear Regression for λ = optimal	977.8252	882.7408	0.019964



The metrics like accuracy and running time can be used to determine the best setting. Given that the environment of predicting diabetes levels demands high accuracy (least test error), we decided that accuracy should be given more weightage. Based on our readings, we observed that both the ridge regression techniques give high accuracy. The second parameter should be running time as the environment like diabetes level prediction might demand fast output in case of emergencies. As the gradient descent approach of Ridge Regression takes more time, we will go for analytical approach.

From our analysis, we have come to the conclusion that **Ridge Regression using analytical approach** is the best method for the given scenario.