CSE 574: INTRODUCTION TO MACHINE LEARNING
ASSIGNMENT 2

Regression

Ketan Deshmukh (50097066)
Nitin Mendiratta (50097498)
Rahul Garg (50096987)

Abstract

This report explains, analyzes and compares the task of various regression techniques which is applied on Diabetes data. Eventually a model is learnt which predicts level of diabetes in patient using input parameters (physical and blood related).

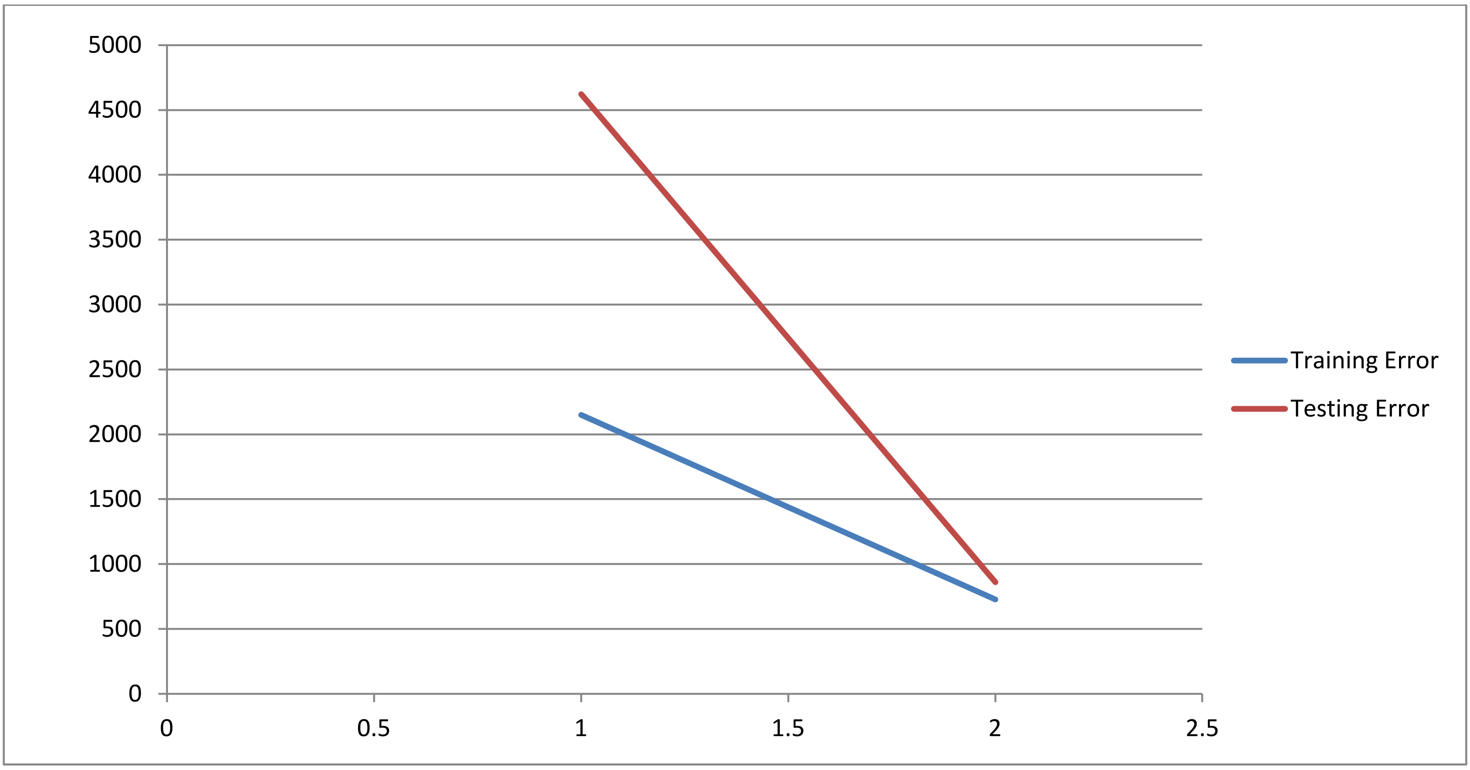
Input Data: xtrain (242 X 64) and ytrain (242 X 1) for training.
xtest (200 X 64) and ytest (200 X 1) for testing.

Problem 1: Experiment with Linear Regression

RSE	**Without Intercept**	**With Intercept**
Training Data	2149.89909911748	727.525113912334
Testing Data	4621.15486503097	861.143446940459

Graph:

Without Intercept/ With Intercept



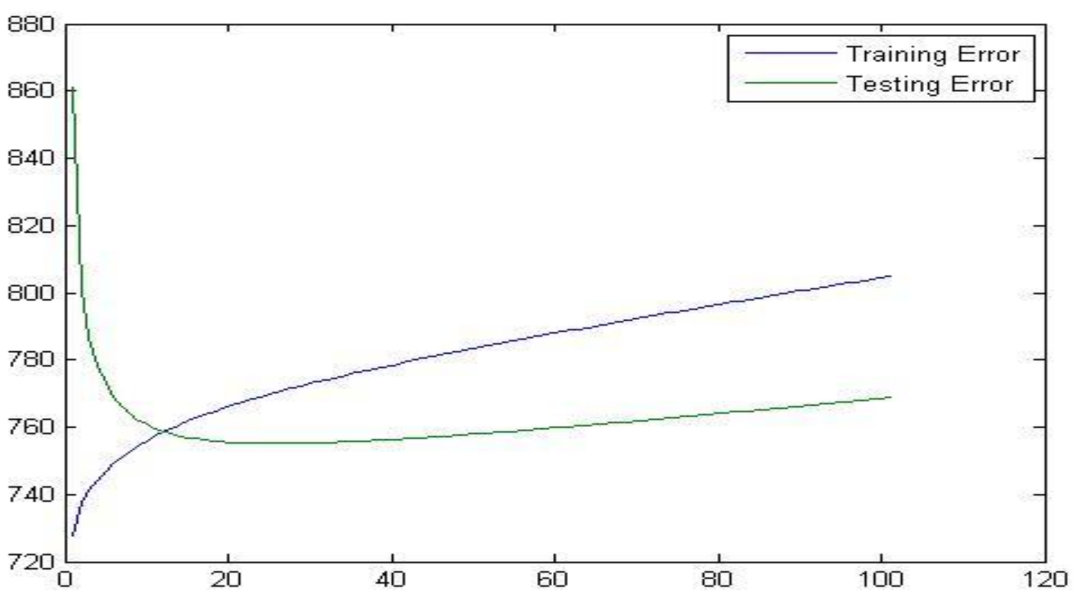
Inference:

RSE obtained in the data **with intercept** is **lesser** than the RSE obtained in the data **without intercept**. So the RSE with intercept is better.

Problem 2: Experiment with Ridge Regression

The RSE for training data using ridge regression parameters:	The RSE for testing data using ridge regression parameters:	Lambdas
727.525113912334	861.143447004592	0
737.326654021867	800.972347070264	0.00001
741.229092853148	786.821420616962	0.00002
744.287002635157	778.479010941932	0.00003
746.852344018575	772.924567180574	0.00004
749.073518549458	768.980577744128	0.00005
751.034334878216	766.060726654911	0.00006
752.789732255226	763.834379253799	0.00007
754.378726361114	762.099453672730	0.00008
755.830384386566	760.725076989025	0.00009
757.167064104183	759.622664429985	0.0001
758.406365588454	758.730115878294	0.00011
759.562393958728	758.002633757382	0.00012
760.646618244200	757.407120939874	0.00013
761.668477483077	756.918623409844	0.00014
762.635820773028	756.517994580476	0.00015
763.555234112196	756.190317559756	0.00016
764.432287752049	755.923813016765	0.00017
765.271726384653	755.709066902570	0.00018
766.077617386477	755.538474006728	0.00019
766.853467757608	755.405830291845	0.0002
767.602317348482	755.306029740172	0.00021
768.326813889992	755.234835869487	0.00022
769.029273897507	755.188707416321	0.00023
769.711732494514	755.164663860378	0.00024
770.375984463100	755.160180624132	0.00025
771.023618288437	755.173106631764	0.00026
771.656044564591	755.201598894670	0.00027
772.274519829253	755.244070190124	0.00028
772.880166668087	755.299146900053	0.00029
…	…	…

Plot of Error for training and testing data for different values of lambda:



Weights for OLE and Ridge Regression

Relative magnitude of weights for OLE:	Relative magnitude of weights for ridge regression:
148.154875997652	150.806826636244
1.27485205935122	12.2689210559462
-293.383522472091	-127.781938642610
414.725448424564	327.301795408047
272.089134325960	217.284872665057
-86639.4565039232	-16.9910412268680
75914.4674786112	-75.9060411229862
32341.6225886185	-157.273500647829
221.101214806823	127.875511498743
29299.5509907887	362.848503853670
125.230360298384	139.596007922956
94.4110833511922	66.4153081054826
-93.8628632071743	63.4646180397299
-33.7282799860075	31.4864377840353
3353.19776772331	11.9227715392270
-621.096260585102	-40.2552291566243
791.736541276843	10.4831202714049
1767.76038869445	52.3476085513086
4191.67403959680	-20.7771851233687
119.438120909469	36.9243446732568
76.6103400414815	82.8397509010744
-15.2001292868050	-5.83031285086553
82.2424593621244	61.4030546406428
-1456.66208385108	-41.9681332875373
827.386702369459	-84.2355843903930
869.290952189708	33.0174585705898
586.234495238140	11.7255643009712
427.026726482539	60.5276293965309
90.2467690203017	34.0864722312145
-17.8876224509606	16.8641515921725
141.696773821222	57.7843275507283
582.819384424240	43.2810326846951
-234.037510626391	-4.10072852030458
-256.071452266938	71.5178114109909
-385.177400568158	5.62801092642894
-33.4176737807602	30.1384158462041
-10.7350066118379	-27.7000025379077
257.107188833933	73.3064035998249
59.9554592616933	-16.8909782223469
383.728042247202	16.6707680837609
-404.158389813954	-27.8463165033148
-514.286434369122	5.79772715025340
38.3636641465535	-12.5723419251723
-44.6102888742376	50.2153205880344
-729.643530448628	-41.6773407661701
377.408336367054	-34.8399813723962
439.794290161621	21.6693107779455
308.514373313414	2.16537458114854
189.859678528048	10.1791474167418
-109.773796996158	-20.6752361960110
-1919.65707302054	-3.13335009981981
…	…

Inference:

1. The **mean weight for the OLE with intercept** corresponding for the optimal lambda value is: **882.8076**
2. The **mean weight for the ridge regression** corresponding to the optimal lambda value is: **32.0073**
3. The **relative weight of the ridge regression** is **lesser than** the **relative weight for the OLE with intercept** for optimal lambda value **(0.00025)**.

RSE	OLE(with Intercept)	Ridge Regression optimal lambda value (0.00025)
Training data	727.525113912428	770.3759844631
Testing data	861.143432654183	755.160180624132

4. Even though the RSE for training data is more in Ridge Regression than that of the OLE method with intercept, the RSE for testing data is lesser in Ridge Regression than that of the OLE method.
5. The amount of shrinkage is controlled by lambda, the tuning parameter that multiplies the ridge penalty. Large lambda means more shrinkage, and so we get different coefficient estimates for different values of lambda.
6. The **optimal value of lambda is 0.00025**. This value is chosen because it is the best value obtained due to trade off between the bias and the variance and also for this optimal value we get the minimal RSE for the test data.

This clearly shows that the Ridge Regression method is more suitable than the OLE method as it gives lesser error for the test data.

Inference of shortcomings of linear regression

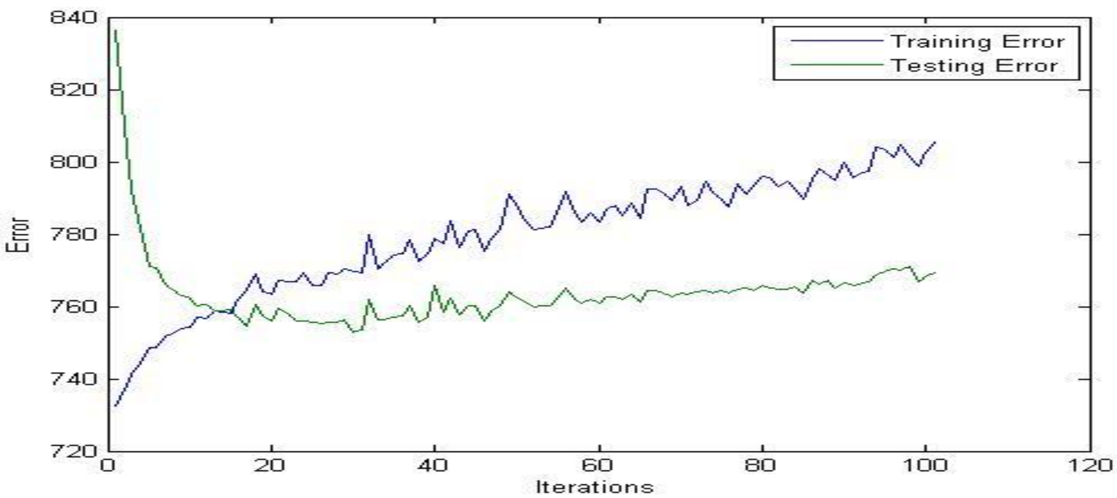
1. **Predictive ability**: It has low bias (zero bias) but suffers from high variance. So it may be worth sacrificing some bias to achieve a lower variance
2. **Interpretative ability**: with a large number of predictors, it can be helpful to identify a smaller subset of important variables. Linear regression doesn't do this.

Problem 3: Using Gradient Descent for Ridge Regression Learning

Train and Test error Using Gradient Descent for Ridge Regression Learning

Train Error	Test Error	Lambda
768.8351	753.3963	**0.00026**

Plot of errors on train and test data obtained by using the gradient descent based learning by varying the regularization parameter lambda.



Comparing results with problem 2

Error	Ridge Regression Learning using Gradient Descent (0.00026)	Ridge Regression using analytical derivation (0.00025)
Training data	768.8351	770.3759
Testing data	753.3963	755.1601

Inference:

From the above table, we get the optimal results for **Ridge Regression Learning using Gradient Descent at lambda=0.00026** since test and training error is less as compared with analytical derivation.

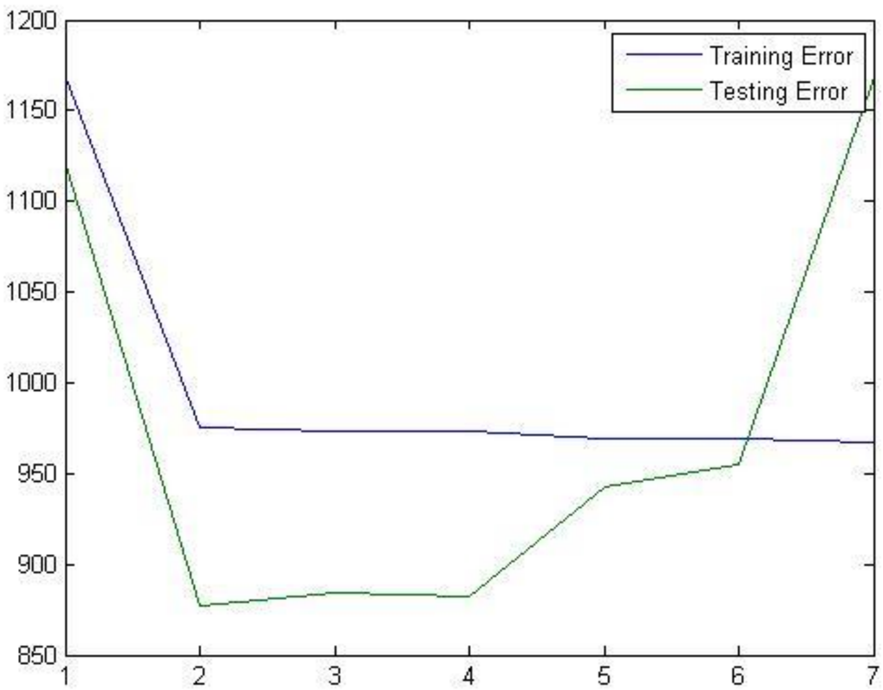
Problem 4: Non-linear Regression

For lambda =0,

Below table shows that for d =1 we get optimal result as it gives the minimum Test Error

d	Train Error	Test Error
0	1169.39	1121.285
1	975.3366	876.9304
2	972.9672	883.9828
3	972.8863	881.8135
4	969.6827	942.6906
5	969.6744	954.4454
6	967.3602	1169.056

Plot of testing and training error with zero regularization parameter by varying the d value from 0 to 6.

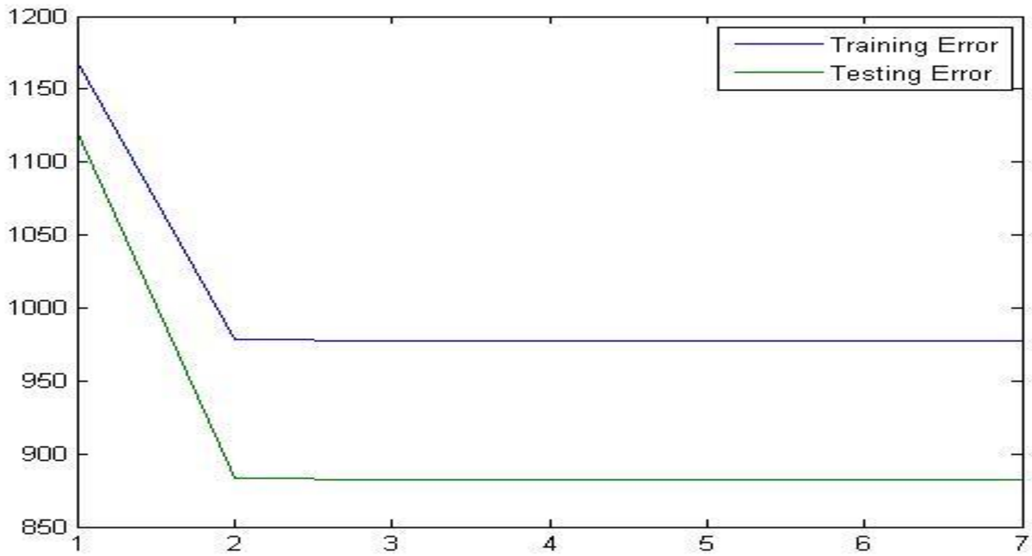


For lambda=optimal (0.00025):

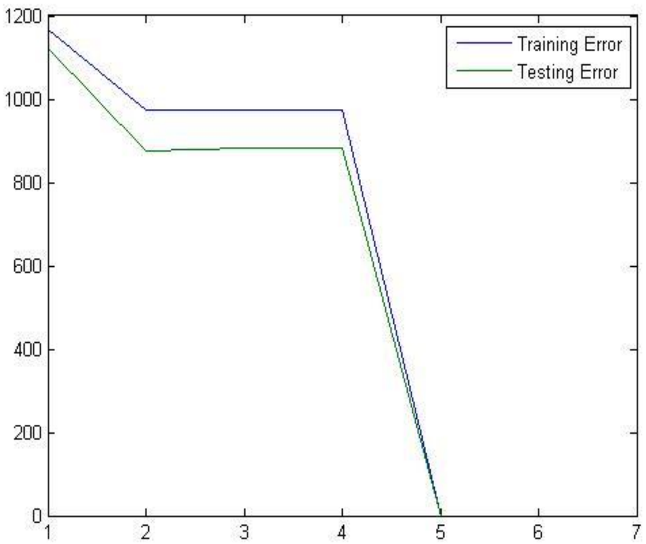
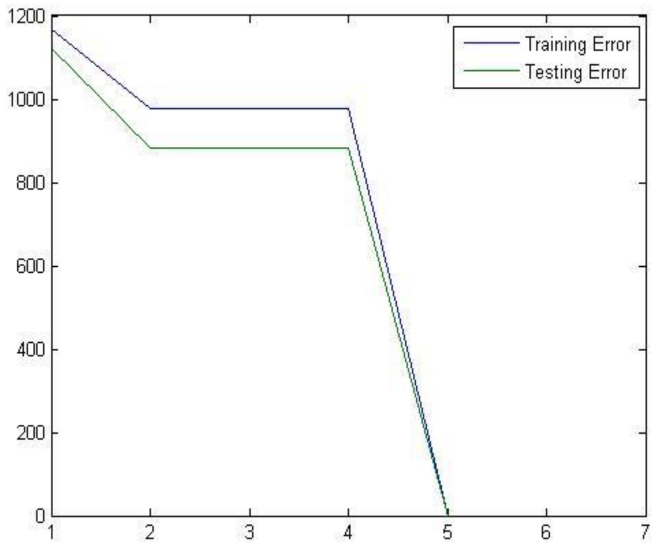
Below table shows that for d =3 we get optimal result.

d	Train Error	Test Error
0	1169.39	1121.328
1	977.9675	882.7724
2	977.8258	882.741
3	977.8252	882.7408
4	977.8252	882.7408
5	977.8252	882.7408
6	977.8252	882.7408

Plot of testing and training error with optimal regularization parameter by varying the d value from 0 to 6.



Curve for the optimal value of d=3 for both values of lambda.

Taking d=3, lambda=0	Taking d=3, lambda=0.00025
	

Problem 5: Interpreting Results

	OLE	Ridge Regression (analytical derivation)	Gradient Descent for Ridge Regression Learning	Non-linear Regression
Training Error	**727.525113912334**	770.3759844631	768.8351	972.8863
Testing Error	**861.143446940459**	755.160180624132	753.3963	972.8863

Inference:

1. Comparing all the four methods of regression, the **gradient descent for ridge regression gives the best results**. The training and testing errors are lower for the gradient descent method compared to all the other methods.
2. The **Non-linear regression method** seems to **give largest training and testing errors** compared to the other methods.
3. **Final recommendation**: We would recommend **Ridge Regression learning** using **Gradient Descent** method for anyone for predicting diabetes level using the input features since it gives the best results with least test and train errors.
4. **Metric used**: Testing error is taken as the basis to check optimality of a method.