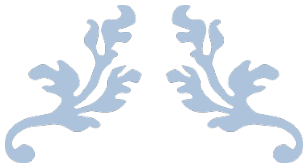


CSE 574: INTRODUCTION TO MACHINE LEARNING – ASSIGNMENT 2

# REGRESSION


TEAM MEMBERS:

Ankit Sarraf (5009 7190)

Karthick Krishna Venkatakrishan (5009 8126)

Rahul Singh (5009 7213)

## Report 1: Experiment with Linear Regression

**Without intercept**:
The RSE for training data: 2149.899
The RSE for testing data: 4621.155

**With Intercept**:
The RSE for training data: 727.525
The RSE for testing data: 861.143

**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. The explanation for such a behavior is when data with bias (intercept) is considered; one can get a better fit while doing the linear regression. Therefore, the magnitude of error in case of data with intercept is lesser.

---

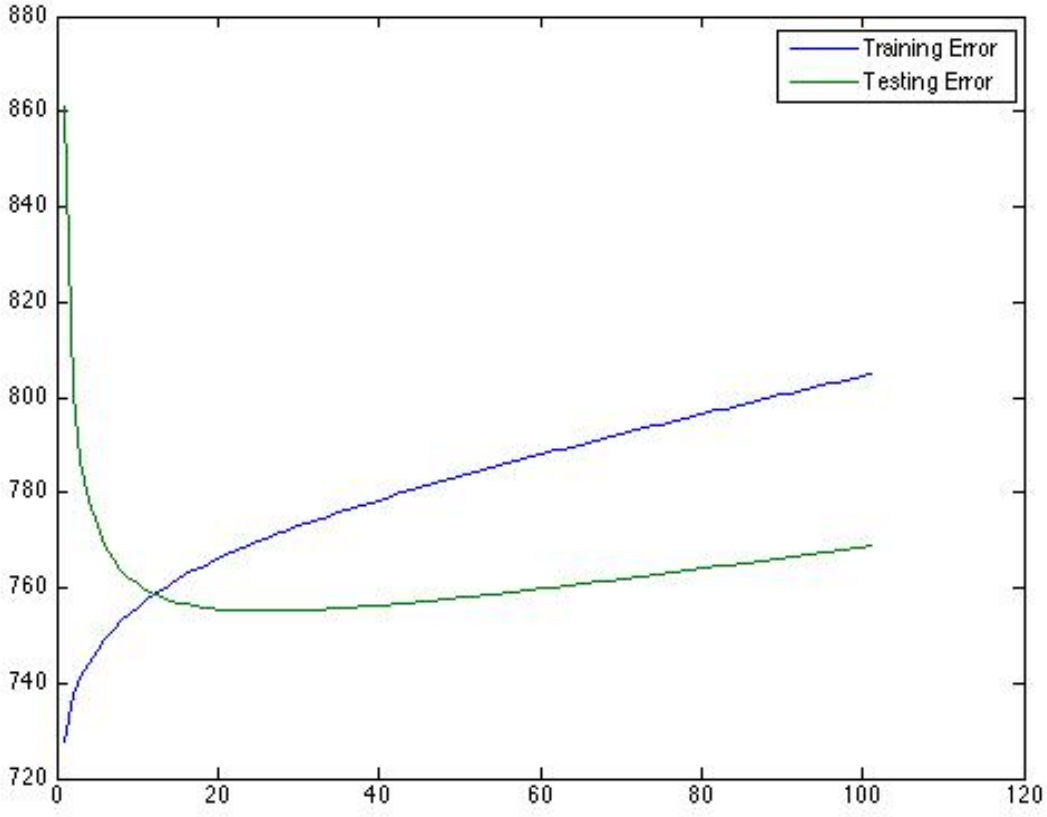
## Report 2: Experiment with Ridge Regression

| Lambda | RSE - Training data using Ridge Regression | RSE - Testing data using Ridge Regression |
|---|---|---|
| 0 | 727.5251139 | 861.1434241 |
| 1.00E-05 | 737.326654 | 800.9723471 |
| 2.00E-05 | 741.2290929 | 786.8214206 |
| 3.00E-05 | 744.2870026 | 778.4790109 |
| 4.00E-05 | 746.852344 | 772.9245672 |
| 5.00E-05 | 749.0735185 | 768.9805777 |
| 6.00E-05 | 751.0343349 | 766.0607267 |
| 7.00E-05 | 752.7897323 | 763.8343793 |
| 8.00E-05 | 754.3787264 | 762.0994537 |
| 9.00E-05 | 755.8303844 | 760.725077 |
| 0.0001 | 757.1670641 | 759.6226644 |
| 0.00011 | 758.4063656 | 758.7301159 |
| 0.00012 | 759.562394 | 758.0026338 |
| 0.00013 | 760.6466182 | 757.4071209 |
| 0.00014 | 761.6684775 | 756.9186234 |
| 0.00015 | 762.6358208 | 756.5179946 |
| 0.00016 | 763.5552341 | 756.1903176 |
| 0.00017 | 764.4322878 | 755.923813 |
| 0.00018 | 765.2717264 | 755.7090669 |
| 0.00019 | 766.0776174 | 755.538474 |
| 0.0002 | 766.8534678 | 755.4058303 |
| 0.00021 | 767.6023173 | 755.3060297 |
| 0.00022 | 768.3268139 | 755.2348359 |
| 0.00023 | 769.0292739 | 755.1887074 |
| 0.00024 | 769.7117325 | 755.1646639 |
| **0.00025** | **770.3759845** | **755.1601806** |
| 0.00026 | 771.0236183 | 755.1731066 |
| 0.00027 | 771.6560446 | 755.2015989 |
| 0.00028 | 772.2745198 | 755.2440702 |
| 0.00029 | 772.8801667 | 755.2991469 |
| 0.0003 | 773.4739908 | 755.3656348 |
| 0.00031 | 774.0568954 | 755.4424911 |
| 0.00032 | 774.6296938 | 755.5288016 |
| 0.00033 | 775.19312 | 755.6237615 |
| 0.00034 | 775.7478377 | 755.7266598 |
| 0.00035 | 776.2944486 | 755.8368662 |

| | | |
|---|---|---|
| 0.00036 | 776.833499 | 755.9538198 |
| 0.00037 | 777.3654859 | 756.0770202 |
| 0.00038 | 777.8908623 | 756.2060191 |
| 0.00039 | 778.4100417 | 756.3404141 |
| 0.0004 | 778.9234022 | 756.4798426 |
| 0.00041 | 779.4312901 | 756.6239774 |
| 0.00042 | 779.9340231 | 756.7725219 |
| 0.00043 | 780.4318929 | 756.9252067 |
| 0.00044 | 780.9251677 | 757.0817869 |
| 0.00045 | 781.4140947 | 757.2420388 |
| 0.00046 | 781.8989015 | 757.4057576 |
| 0.00047 | 782.3797985 | 757.5727558 |
| 0.00048 | 782.85698 | 757.7428608 |
| 0.00049 | 783.3306258 | 757.9159139 |
| 0.0005 | 783.8009024 | 758.0917684 |
| 0.00051 | 784.2679642 | 758.2702887 |
| 0.00052 | 784.7319544 | 758.451349 |
| 0.00053 | 785.1930063 | 758.6348325 |
| 0.00054 | 785.6512434 | 758.8206306 |
| 0.00055 | 786.1067811 | 759.008642 |
| 0.00056 | 786.5597264 | 759.198772 |
| 0.00057 | 787.0101796 | 759.3909319 |
| 0.00058 | 787.4582338 | 759.5850388 |
| 0.00059 | 787.9039762 | 759.7810147 |
| 0.0006 | 788.3474881 | 759.9787863 |
| 0.00061 | 788.7888458 | 760.1782843 |
| 0.00062 | 789.2281205 | 760.3794437 |
| 0.00063 | 789.6653789 | 760.5822027 |
| 0.00064 | 790.1006836 | 760.7865031 |
| 0.00065 | 790.5340932 | 760.9922894 |
| 0.00066 | 790.9656628 | 761.1995093 |
| 0.00067 | 791.3954441 | 761.4081129 |
| 0.00068 | 791.8234858 | 761.6180525 |
| 0.00069 | 792.2498335 | 761.8292831 |
| 0.0007 | 792.6745302 | 762.0417614 |
| 0.00071 | 793.0976165 | 762.2554461 |
| 0.00072 | 793.5191305 | 762.4702978 |
| 0.00073 | 793.9391082 | 762.6862787 |
| 0.00074 | 794.3575837 | 762.9033527 |
| 0.00075 | 794.7745888 | 763.1214849 |
| 0.00076 | 795.190154 | 763.340642 |

| | | |
|---|---|---|
| 0.00077 | 795.6043078 | 763.560792 |
| 0.00078 | 796.0170774 | 763.781904 |
| 0.00079 | 796.4284883 | 764.0039482 |
| 0.0008 | 796.8385649 | 764.226896 |
| 0.00081 | 797.2473301 | 764.4507198 |
| 0.00082 | 797.6548058 | 764.6753929 |
| 0.00083 | 798.0610126 | 764.9008894 |
| 0.00084 | 798.4659702 | 765.1271845 |
| 0.00085 | 798.8696973 | 765.3542541 |
| 0.00086 | 799.2722116 | 765.5820748 |
| 0.00087 | 799.6735301 | 765.8106239 |
| 0.00088 | 800.0736687 | 766.0398796 |
| 0.00089 | 800.4726428 | 766.2698205 |
| 0.0009 | 800.870467 | 766.500426 |
| 0.00091 | 801.2671553 | 766.7316761 |
| 0.00092 | 801.6627208 | 766.9635513 |
| 0.00093 | 802.0571764 | 767.1960327 |
| 0.00094 | 802.4505341 | 767.4291018 |
| 0.00095 | 802.8428055 | 767.6627407 |
| 0.00096 | 803.2340018 | 767.896932 |
| 0.00097 | 803.6241335 | 768.1316587 |
| 0.00098 | 804.013211 | 768.3669044 |
| 0.00099 | 804.4012439 | 768.6026528 |
| 0.001 | 804.7882417 | 768.8388882 |

Plot of training and testing data for different values of lambda (represented by the index):




**Weights for OLE and Ridge Regression**:

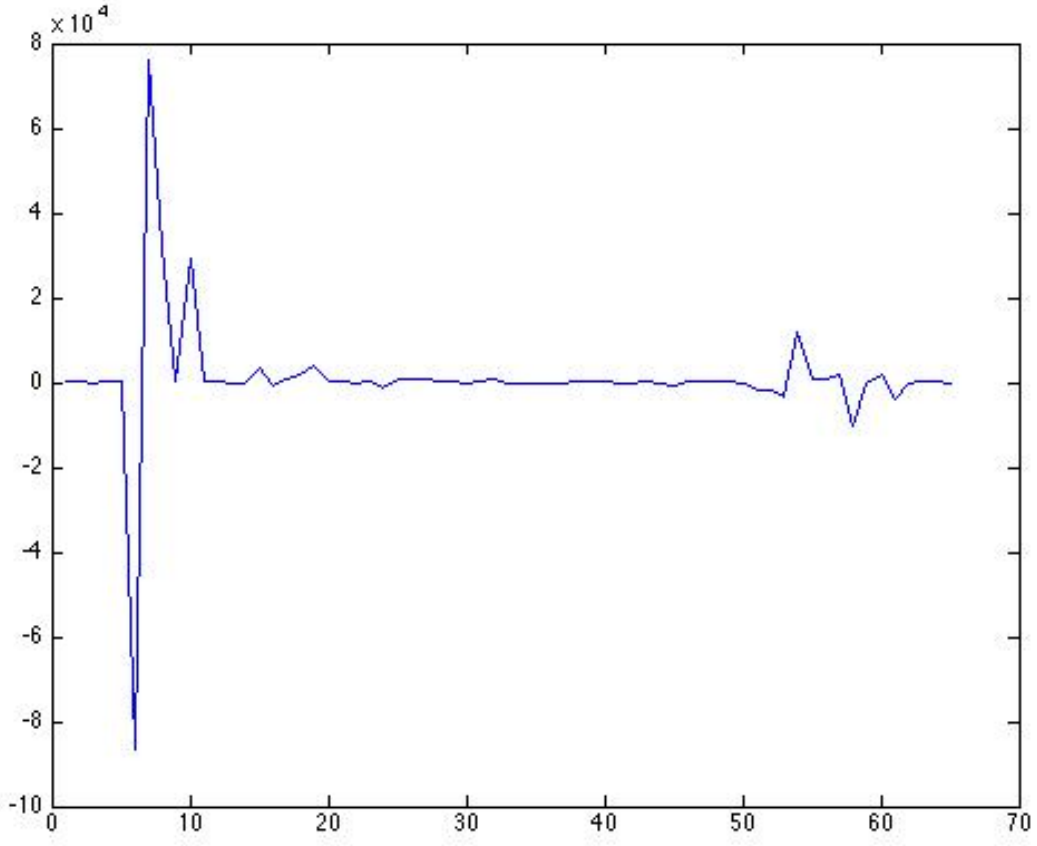
| Relative magnitude of weights for OLE | Relative magnitude of weights for ridge regression |
|---|---|
| 148.154876 | 150.8068266 |
| 1.274852078 | 12.26892106 |
| -293.3835226 | -127.7819386 |
| 414.7254475 | 327.3017954 |
| 272.0891343 | 217.2848727 |
| -86639.45621 | -16.99104123 |
| 75914.46723 | -75.90604112 |
| 32341.62249 | -157.2735006 |
| 221.1011579 | 127.8755115 |
| 29299.5509 | 362.8485039 |
| 125.230359 | 139.5960079 |
| 94.41108463 | 66.41530811 |

| | |
|---|---|
| -93.86286198 | 63.46461804 |
| -33.72827899 | 31.48643778 |
| 3353.197449 | 11.92277154 |
| -621.0961829 | -40.25522916 |
| 791.7362581 | 10.48312027 |
| 1767.76037 | 52.34760855 |
| 4191.673968 | -20.77718512 |
| 119.4381201 | 36.92434467 |
| 76.61033921 | 82.8397509 |
| -15.20013028 | -5.830312851 |
| 82.24245944 | 61.40305464 |
| -1456.662087 | -41.96813329 |
| 827.3867004 | -84.23558439 |
| 869.29096 | 33.01745857 |
| 586.2345027 | 11.7255643 |
| 427.0267304 | 60.5276294 |
| 90.24676871 | 34.08647223 |
| -17.88762379 | 16.86415159 |
| 141.6967732 | 57.78432755 |
| 582.8193796 | 43.28103268 |
| -234.0375038 | -4.10072852 |
| -256.071452 | 71.51781141 |
| -385.177402 | 5.628010926 |
| -33.41767078 | 30.13841585 |
| -10.73500665 | -27.70000254 |
| 257.1071871 | 73.3064036 |
| 59.95545989 | -16.89097822 |
| 383.7280412 | 16.67076808 |
| -404.1583922 | -27.8463165 |
| -514.2864356 | 5.79772715 |
| 38.36366489 | -12.57234193 |
| -44.61028765 | 50.21532059 |
| -729.6434944 | -41.67734077 |
| 377.408303 | -34.83998137 |
| 439.7942787 | 21.66931078 |
| 308.5143791 | 2.165374581 |
| 189.8596652 | 10.17914742 |
| -109.7737972 | -20.6752362 |
| -1919.656943 | -3.1333501 |
| -1924.633525 | 73.575854 |
| -3489.794816 | -59.44872226 |

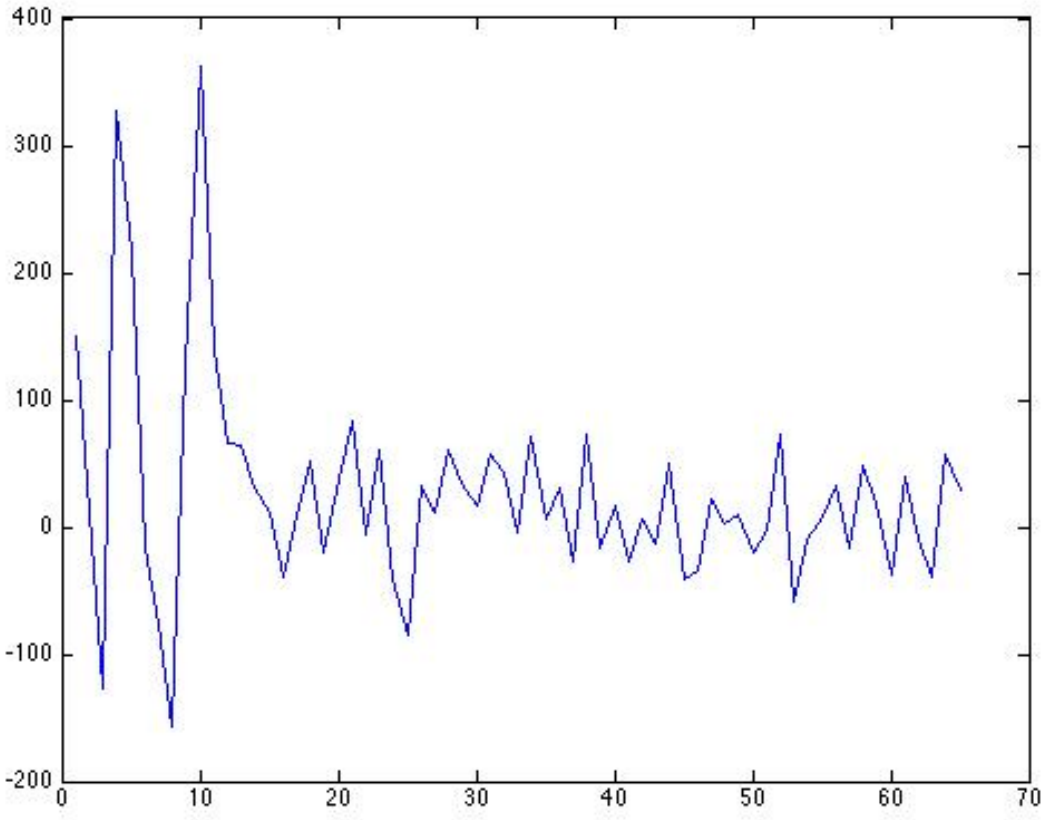
| | |
|---|---|
| 11796.96852 | -9.028338247 |
| 530.674454 | 8.371172755 |
| 543.3062412 | 31.84482717 |
| 1821.075743 | -15.94763469 |
| -10463.98071 | 47.87385561 |
| -516.6276031 | 17.72398939 |
| 2064.359125 | -37.07908572 |
| -4199.413364 | 40.27425969 |
| -140.4957053 | -9.977183838 |
| 374.1570897 | -39.03819219 |
| 51.47574926 | 56.27038742 |
| -46.44927306 | 28.69075366 |

Comparing the weights of OLE Regression and Ridge Regression

A) Weights for OLE Regression for Data with Intercept:



B) Weights for the Data with intercept using Ridge Regression




**Reason:** In Ridge Regression the higher weights are penalized. Hence they fall in the similar ranges. Where as in case of OLE Regression, the higher order weights are not penalized. Therefore, such a high variation can be observed.

The mean weight in case of OLE Regression (with data that has bias) is **882.8076**

The mean weight for the Ridge Regression is **25.7461**

So we infer that the relative weight of the ridge regression is lesser than the relative weight for the OLE with intercept, all corresponding to the optimal lambda value.

*For OLE:*
The RSE for training data: **727.525**
The RSE for testing data: **861.143**

*For Ridge Regression with the optimal lambda value (**0.00025**):*
The RSE for training data is: **770.376**
The RSE for testing data is: **755.160**

We can clearly infer that 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.

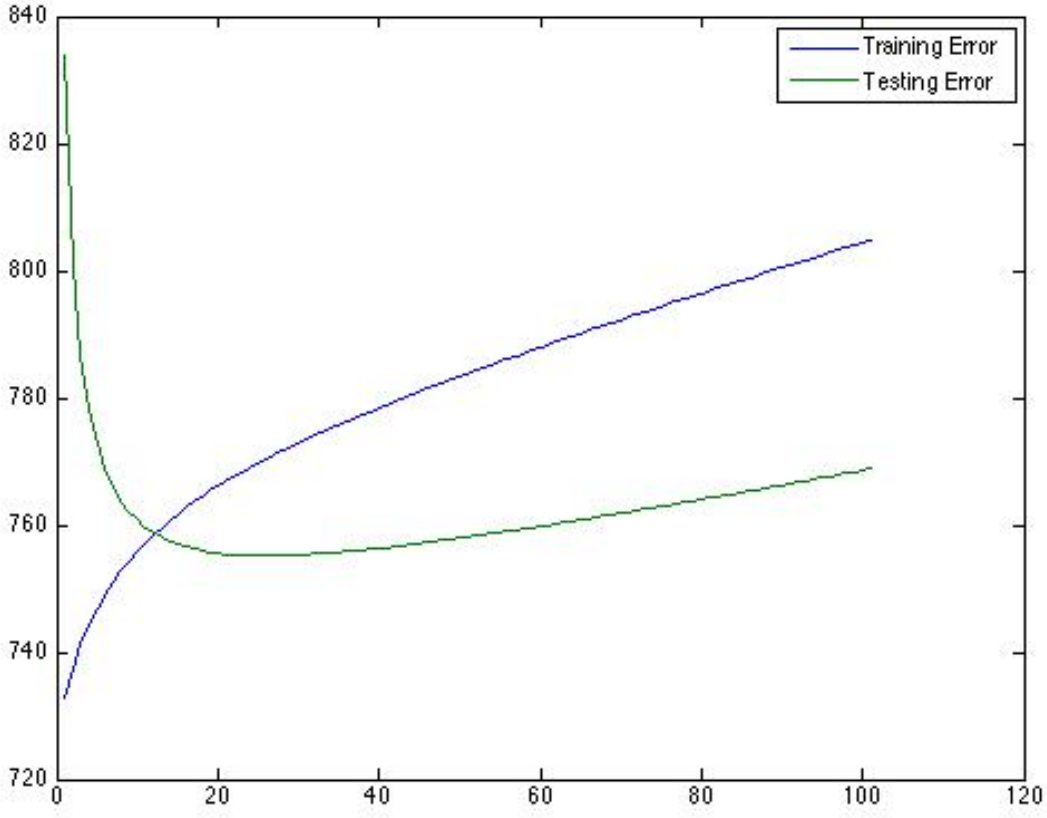
The optimal value of lambda is **0.00025**. This value is chose since it is the best value obtained due to trade off between the bias and the variance. For this optimal value we get the minimal RSE for the test data. When the value of lambda < 0.00025, due to under-fitting (high bias), high error is observed. Where as in case of lambda > 0.00025, due to over-fitting (increased variance), error increases because of behavior similar to rote learning.

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

---

## Report 3: Using Gradient Descent For Ridge Regression Learning

Plotting of training and testing data obtained by gradient descent method of ridge regression by varying the values of regularization parameter (lambda).




Training and test errors for Gradient Descent Ridge Regression Learning:

| lambda | The RSE for Training Data using Gradient Descent parameters: | The RSE for Testing Data using Gradient Descent parameters: |
|---|---|---|
| 0 | 732.7136891 | 833.9607773 |
| 1.00E-05 | 737.3741586 | 800.9847308 |
| 2.00E-05 | 741.3820506 | 786.6934061 |
| 3.00E-05 | 744.2918337 | 778.4780159 |
| 4.00E-05 | 746.8529278 | 772.9244374 |
| 5.00E-05 | 749.0737462 | 768.980135 |
| 6.00E-05 | 751.0344373 | 766.0607092 |
| 7.00E-05 | 752.7898233 | 763.8344377 |
| 8.00E-05 | 754.3786752 | 762.0994826 |

| | | |
|---|---|---|
| 9.00E-05 | 755.8304055 | 760.7250815 |
| 0.0001 | 757.1670622 | 759.6226696 |
| 0.00011 | 758.4063624 | 758.7301253 |
| 0.00012 | 759.5624007 | 758.0026337 |
| 0.00013 | 760.6466234 | 757.4071207 |
| 0.00014 | 761.6684823 | 756.9186236 |
| 0.00015 | 762.6358283 | 756.5179918 |
| 0.00016 | 763.5552392 | 756.1903162 |
| 0.00017 | 764.4322829 | 755.92382 |
| 0.00018 | 765.2717231 | 755.7090704 |
| 0.00019 | 766.0776211 | 755.5384701 |
| 0.0002 | 766.8534685 | 755.4058298 |
| 0.00021 | 767.6023107 | 755.3060358 |
| 0.00022 | 768.3268166 | 755.2348377 |
| 0.00023 | 769.0292784 | 755.1887075 |
| 0.00024 | 769.7117227 | 755.1646714 |
| **0.00025** | **770.3759844** | **755.1601785** |
| 0.00026 | 771.0236101 | 755.173103 |
| 0.00027 | 771.6560526 | 755.2015941 |
| 0.00028 | 772.2745165 | 755.2440661 |
| 0.00029 | 772.8801603 | 755.2991535 |
| 0.0003 | 773.4739991 | 755.3656317 |
| 0.00031 | 774.0569 | 755.4424838 |
| 0.00032 | 774.6296969 | 755.528801 |
| 0.00033 | 775.1931341 | 755.6237592 |
| 0.00034 | 775.7478388 | 755.7266587 |
| 0.00035 | 776.2944505 | 755.8368676 |
| 0.00036 | 776.8335006 | 755.9538193 |
| 0.00037 | 777.3654944 | 756.0770148 |
| 0.00038 | 777.8908567 | 756.2060225 |
| 0.00039 | 778.4100361 | 756.3404105 |
| 0.0004 | 778.9233985 | 756.4798456 |
| 0.00041 | 779.4312826 | 756.6239794 |
| 0.00042 | 779.9340294 | 756.7725197 |
| 0.00043 | 780.4318982 | 756.9252054 |
| 0.00044 | 780.9251643 | 757.0817817 |
| 0.00045 | 781.4140935 | 757.2420398 |
| 0.00046 | 781.8989026 | 757.4057584 |
| 0.00047 | 782.3797953 | 757.5727563 |
| 0.00048 | 782.8569822 | 757.742862 |
| 0.00049 | 783.3306271 | 757.9159112 |

| | | |
|---|---|---|
| 0.0005 | 783.8009051 | 758.091768 |
| 0.00051 | 784.2679657 | 758.2702864 |
| 0.00052 | 784.7319518 | 758.4513501 |
| 0.00053 | 785.1930144 | 758.6348334 |
| 0.00054 | 785.6512446 | 758.8206402 |
| 0.00055 | 786.106781 | 759.0086422 |
| 0.00056 | 786.5597285 | 759.1987731 |
| 0.00057 | 787.0101793 | 759.3909332 |
| 0.00058 | 787.4582358 | 759.5850381 |
| 0.00059 | 787.9039762 | 759.781014 |
| 0.0006 | 788.3474891 | 759.9787867 |
| 0.00061 | 788.7888477 | 760.1782848 |
| 0.00062 | 789.2281224 | 760.3794465 |
| 0.00063 | 789.665377 | 760.5821996 |
| 0.00064 | 790.1006886 | 760.7865021 |
| 0.00065 | 790.5340934 | 760.9922855 |
| 0.00066 | 790.965667 | 761.1995124 |
| 0.00067 | 791.3954466 | 761.4081172 |
| 0.00068 | 791.8234869 | 761.6180514 |
| 0.00069 | 792.2498322 | 761.8292831 |
| 0.0007 | 792.6745246 | 762.0417606 |
| 0.00071 | 793.0976194 | 762.2554455 |
| 0.00072 | 793.5191332 | 762.4702963 |
| 0.00073 | 793.9391052 | 762.6862746 |
| 0.00074 | 794.3575882 | 762.9033515 |
| 0.00075 | 794.7745889 | 763.121485 |
| 0.00076 | 795.1901525 | 763.3406437 |
| 0.00077 | 795.6043114 | 763.5608005 |
| 0.00078 | 796.0170772 | 763.7819056 |
| 0.00079 | 796.4284875 | 764.003947 |
| 0.0008 | 796.8385611 | 764.226896 |
| 0.00081 | 797.2473231 | 764.4507181 |
| 0.00082 | 797.6547981 | 764.6753927 |
| 0.00083 | 798.0610087 | 764.9008874 |
| 0.00084 | 798.4659731 | 765.1271835 |
| 0.00085 | 798.8697002 | 765.3542564 |
| 0.00086 | 799.2722125 | 765.5820751 |
| 0.00087 | 799.6735319 | 765.8106228 |
| 0.00088 | 800.0736698 | 766.0398785 |
| 0.00089 | 800.4726462 | 766.2698255 |
| 0.0009 | 800.8704714 | 766.5004266 |

| | | |
|---|---|---|
| 0.00091 | 801.2671508 | 766.7316763 |
| 0.00092 | 801.6627195 | 766.9635543 |
| 0.00093 | 802.0571769 | 767.1960329 |
| 0.00094 | 802.4505247 | 767.4290975 |
| 0.00095 | 802.8428052 | 767.6627392 |
| 0.00096 | 803.2340016 | 767.8969315 |
| 0.00097 | 803.62413 | 768.13166 |
| 0.00098 | 804.0132105 | 768.3669043 |
| 0.00099 | 804.4012445 | 768.6026545 |
| 0.001 | 804.7882432 | 768.8388919 |

When RSE's for the Ridge Regression and Gradient descent is observed, it can be clearly seen that both the Errors are of the same magnitude. This suggests that for the given data set the Gradient Descent gives the same results as the Ridge Regression.

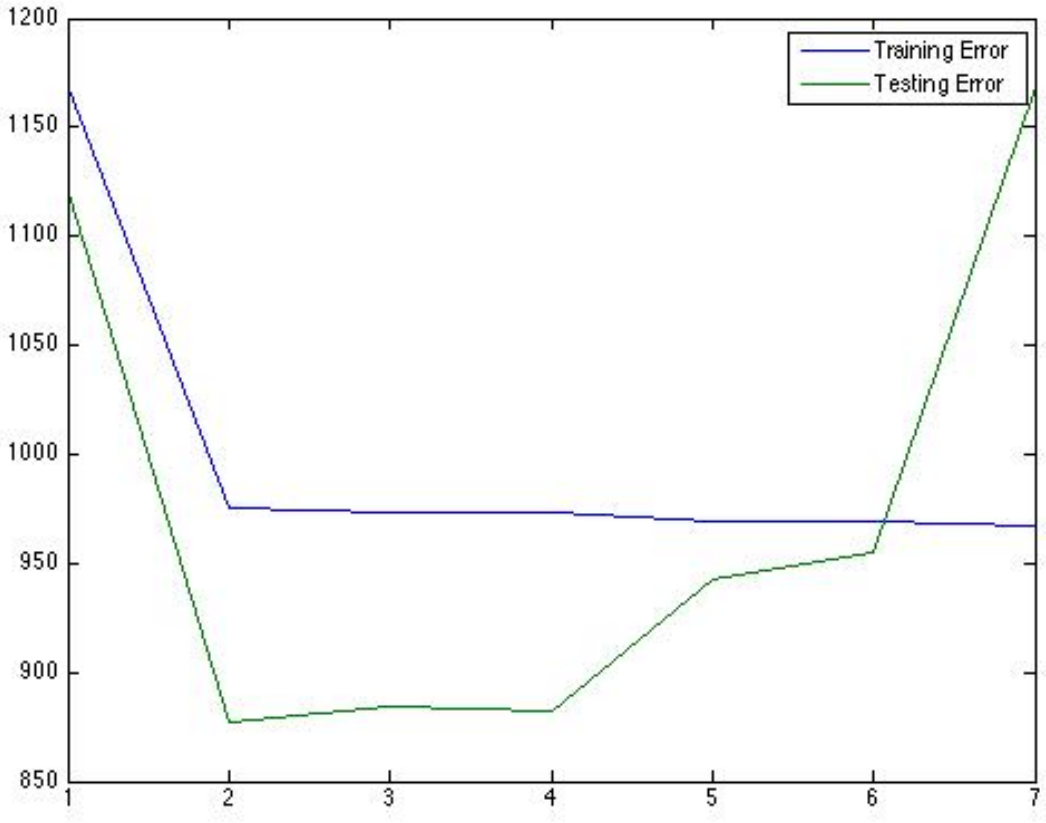
Again, the U-shaped nature of the test data is account by the Bias Variance Tradeoff as in Part 2.

---

## Report 4: Non-Linear Regression

### A) For lambda = 0:

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



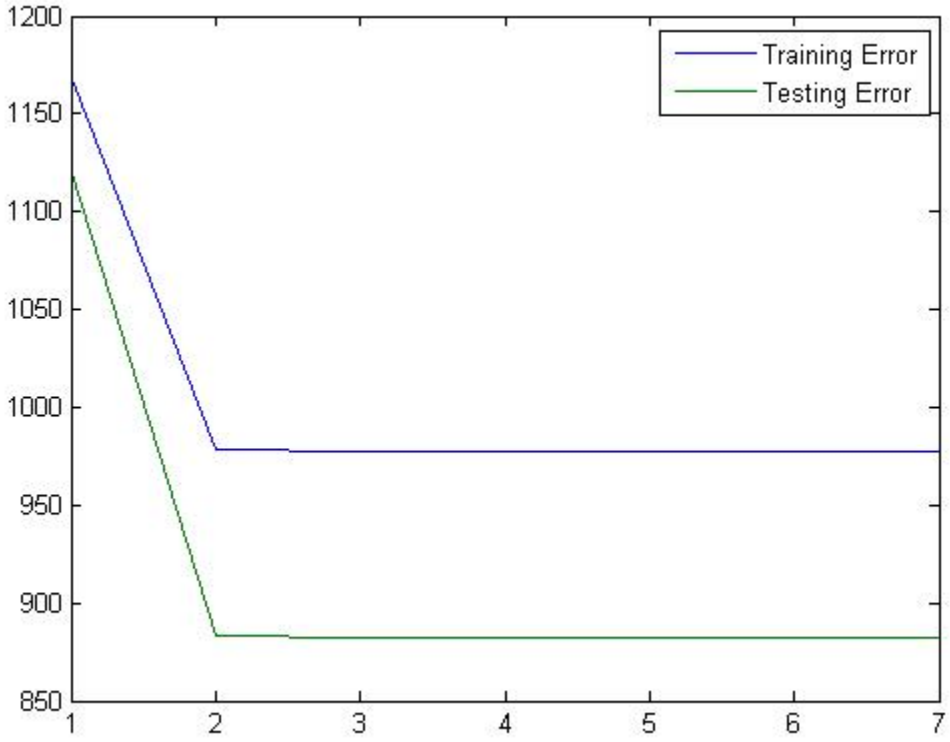

| d | Train Errors | Test Errors |
|---|---|---|
| 0 | 1169.3896 | 1121.2854 |
| 1 | 975.3366 | 876.9304 |
| 2 | 972.9672 | 883.9828 |
| 3 | 972.8863 | 881.8135 |
| 4 | 969.6827 | 942.6906 |
| 5 | 969.6745 | 954.4454 |
| 6 | 967.3602 | 1169.0560 |

*Calculated using the [~, index] = min(Error) function.

So for lambda = 0, the value of d = 6 gives the most optimal results for Training Data and the value of d = 1 produces the most optimal results for the Test Data.

**B) For lambda = lambda_optimal (0.00025):**

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




Training and testing for different values of d when lambda=optimal.

| d | Train Errors | Test Errors |
|---|---|---|
| 0 | 1169.39 | 1121.3283 |
| 1 | 977.9675 | 882.7724 |
| 2 | 977.8258 | 882.7409 |
| 3 | 977.8252 | 882.7408 |
| 4 | 977.8252 | 882.7408 |
| 5 | 977.8252 | 882.7408 |
| 6 | 977.8252 | 882.7408 |

*Calculated using the [~, index] = min(Error) function.

When observed the plot, it is evident that for lambda = 0.00025 (lambda_optimal), the value of d = 6 gives the most optimal results for Training Data and the value of d = 4 produces the most optimal results for the Test Data.

---

## Report 5: Interpreting Results:

Testing error is the major metric used to distinguish the methods. From all the above 4 methods we can conclude that:

1) The OLE with intercepts seems to be giving lesser RSE compared to the OLE without intercepts.

2) Even though the training error seems to be lesser in the OLE than the Ridge Regression, the testing error of OLE seems to more in the OLE.

3) The Ridge Regression seems to be quick and has less RSE for testing data as well. Even though the Gradient Descent Ridge Regression seems to give the same error results as that of the Ridge Regression, it takes longer time to compute than the Ridge Regression. Hence if we take time factor also into consideration, we can find that the Ridge Regression method gives the best result.

*4)* STATISTICS (When the different regressions are executed individually)
*For OLE Regression*
Without Intercept:
Training Error: 2149.8991
Testing Error: 4621.1548
With Intercept:
Training Error: 727.5251
Testing Error: 861.1434
Time Taken: 0.022352 seconds

*For Ridge Regression*
Training error: 770.3760
Testing error: 755.1602
Time Taken for calculations: 0.497997 seconds

*For gradient descent for ridge regression*
Training error: 770.3760
Testing error: 755.1602
The time taken for calculation: 49.755288 seconds

*For Non-linear regression*
For lambda = 0
Training Error: 967.3602 (d = 6)
Test Error: 882.7408 (d = 1)
For lambda = lambda_optimal (0.00025)
Training Error: 977.8252 (d = 6)
Test Error: 882.7408 (d = 4)
The time taken for calculation: 0.437013 seconds

FINAL CONCLUSION:

Non Linear Regression takes lesser time for this input dataset, but the error observed is comparatively higher than that of Gradient Descent and Ridge Regression. So when the time is the preference, OLE Regression or Non-Linear Regression can be chosen. But when the accuracy of results is of higher priority, Ridge regression or Gradient descent can be a better option.