CSE574 Introduction to Machine Learning
Programming Assignment 1

Priya Murthy: 50248887
Shefali Sharma:50247677

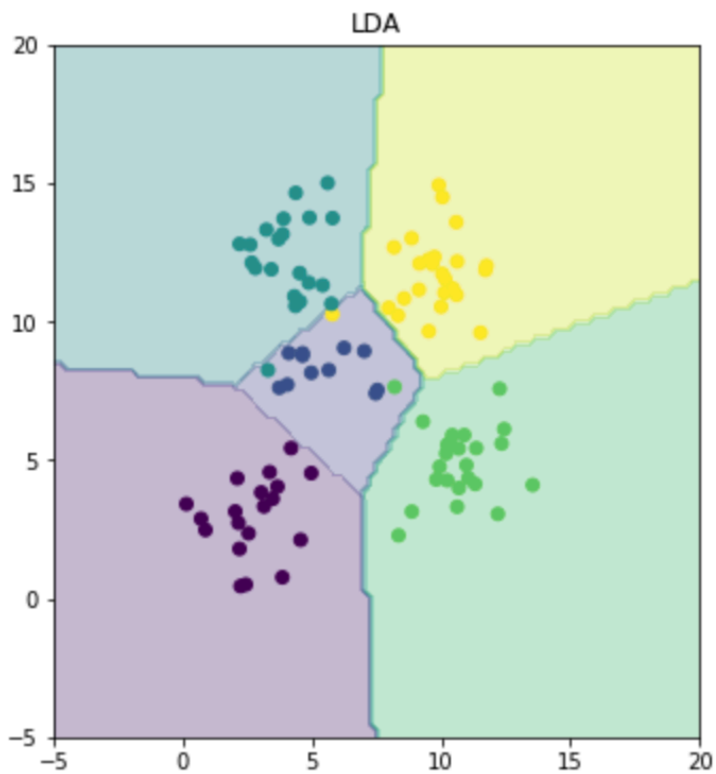
Classification and Regression

INTRODUCTION

In this assignment, we have been given two data sets: "sample.pickle" and "diabetes.pickle". We used various Classification and Regression techniques on the data.

Problem 1: Experiment with Gaussian Discriminators

Linear Discriminant Analysis:

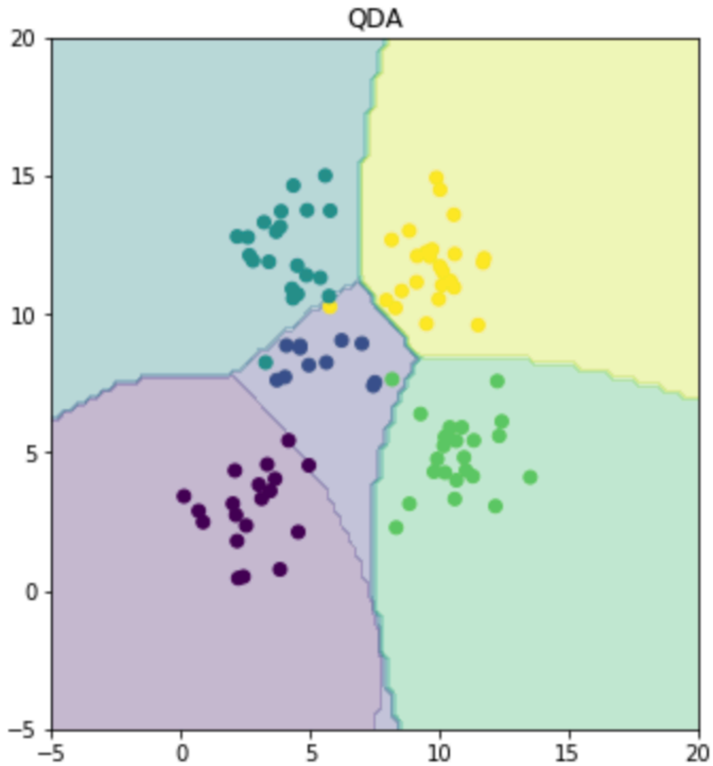


Accuracy for LDA: 97%

Observations:

- We observe discriminant boundaries for LDA which differentiate 5 classes.
- In LDA, the covariance for each class is assumed to be the same.
- Boundaries plotted are linear.

Quadratic Discriminant Analysis:



Accuracy for QDA: 96%

Observations:

- We observe discriminant boundaries for QDA which differentiate 5 classes.
- In QDA, the covariance for each class is calculated separately.
- Boundaries plotted are parabolic (quadratic decision boundaries).

Problem 2: Experiment with Linear Regression

REGRESSION METHOD	TRAIN MSE	TEST MSE
Linear Regression (without Intercept)	19099.4468446	106775.361451
Linear Regression (with Intercept)	2187.16029493	3707.84018096

Conclusion:
Looking at the above results, we can conclude that Linear Regression (with intercept) gives much better results for both Training and Test data.
This is because when linear regression line is modeled without an intercept, it has to pass through origin so it is possible that is is not the best model. Whereas if the Linear Regression line accommodates an int ercept, the model gets more closely aligned with the actual data thus giving better results.

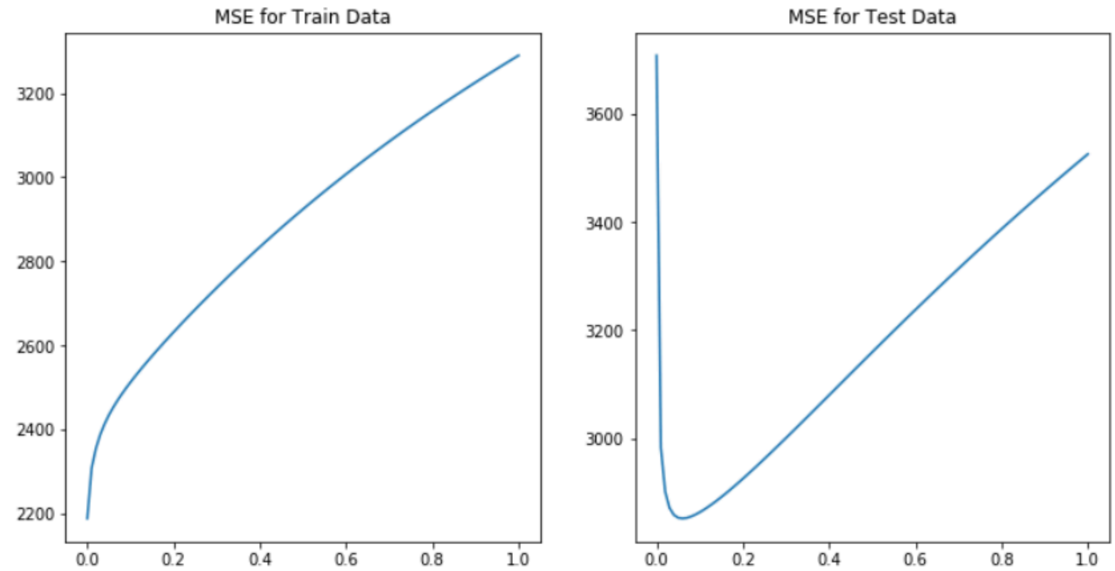
Improvement in Train error is by 88.5%, Improvement in Test error is by 96.5%.

Problem 3: Experiment with Ridge Regression

a) MSE for training and test data: (for Lambda values 0 – 1)

Sno.	Lambda	Train MSE	Test MSE
0	0	2187.1603	3707.8402
1	0.01	2306.8322	2982.4461
2	0.02	2354.0713	2900.9736
3	0.03	2386.7802	2870.9416
4	0.04	2412.119	2858.0004
5	0.05	2433.1744	2852.6657
6	0.06	2451.5285	2851.3302
7	0.07	2468.0776	2852.35
8	0.08	2483.3656	2854.8797
9	0.09	2497.7403	2858.4444
10	0.1	2511.4323	2862.7579
11	0.11	2524.6	2867.6379
12	0.12	2537.3549	2872.9623
13	0.13	2549.7769	2878.6459
14	0.14	2561.9245	2884.6269
15	0.15	2573.8413	2890.8591
16	0.16	2585.5599	2897.3067
17	0.17	2597.1052	2903.9411
18	0.18	2608.4964	2910.7394
19	0.19	2619.7484	2917.6822
20	0.2	2630.8728	2924.7532
21	0.21	2641.8789	2931.9385
22	0.22	2652.7741	2939.2259
23	0.23	2663.5643	2946.6046
24	0.24	2674.2543	2954.0651
25	0.25	2684.8481	2961.5986
26	0.26	2695.3489	2969.1976
27	0.27	2705.7596	2976.855
28	0.28	2716.0825	2984.5643
29	0.29	2726.3196	2992.3197
30	0.3	2736.4726	3000.1158
31	0.31	2746.5432	3007.9476
32	0.32	2756.5327	3015.8106
33	0.33	2766.4423	3023.7004
34	0.34	2776.2733	3031.6132
35	0.35	2786.0267	3039.5453
36	0.36	2795.7036	3047.4934
37	0.37	2805.3048	3055.4542
38	0.38	2814.8314	3063.4249
39	0.39	2824.2842	3071.4028
40	0.4	2833.6641	3079.3852
41	0.41	2842.9719	3087.3699
42	0.42	2852.2084	3095.3547
43	0.43	2861.3745	3103.3374
44	0.44	2870.4709	3111.3162
45	0.45	2879.4985	3119.2893
46	0.46	2888.4579	3127.255
47	0.47	2897.3501	3135.2117
48	0.48	2906.1757	3143.158
49	0.49	2914.9354	3151.0925
50	0.5	2923.6301	3159.014
51	0.51	2932.2604	3166.9213
52	0.52	2940.8272	3174.8133
53	0.53	2949.3311	3182.6889
54	0.54	2957.7728	3190.5472
55	0.55	2966.153	3198.3873
56	0.56	2974.4726	3206.2084
57	0.57	2982.732	3214.0096
58	0.58	2990.9322	3221.7903
59	0.59	2999.0736	3229.5499
60	0.6	3007.1571	3237.2875
61	0.61	3015.1832	3245.0028
62	0.62	3023.1527	3252.6951
63	0.63	3031.0661	3260.3639
64	0.64	3038.9242	3268.0089
65	0.65	3046.7276	3275.6295
66	0.66	3054.4769	3283.2254
67	0.67	3062.1727	3290.7961
68	0.68	3069.8156	3298.3415
69	0.69	3077.4064	3305.8611
70	0.7	3084.9454	3313.3546
71	0.71	3092.4334	3320.8219
72	0.72	3099.871	3328.2627
73	0.73	3107.2586	3335.6767
74	0.74	3114.5969	3343.0639
75	0.75	3121.8865	3350.4239
76	0.76	3129.1278	3357.7567
77	0.77	3136.3215	3365.062
78	0.78	3143.468	3372.3399
79	0.79	3150.568	3379.5901
80	0.8	3157.6218	3386.8127
81	0.81	3164.6301	3394.0074
82	0.82	3171.5933	3401.1742
83	0.83	3178.512	3408.3132
84	0.84	3185.3866	3415.4242
85	0.85	3192.2176	3422.5071
86	0.86	3199.0055	3429.5621
87	0.87	3205.7508	3436.589
88	0.88	3212.4539	3443.5878
89	0.89	3219.1153	3450.5586
90	0.9	3225.7354	3457.5014
91	0.91	3232.3147	3464.4162
92	0.92	3238.8536	3471.303
93	0.93	3245.3525	3478.1618
94	0.94	3251.8119	3484.9927
95	0.95	3258.2323	3491.7957
96	0.96	3264.6139	3498.5709
97	0.97	3270.9572	3505.3183
98	0.98	3277.2626	3512.038
99	0.99	3283.5305	3518.7301
100	1	3289.7613	3525.3946

b) Plots for errors on train and test data for different values of lambda.



c) Comparing relative magnitudes of weighs learnt using OLE and Ridge Regression

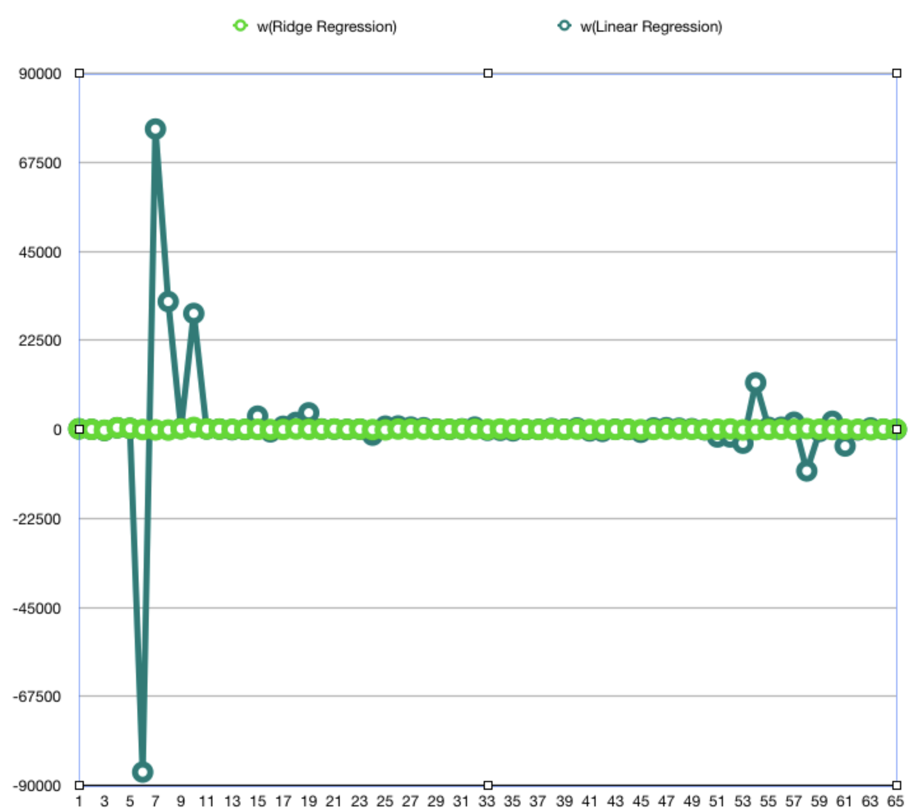


Table for the above graph values:

Weight Comparison

s.no	w(Ridge Regression)	w(Linear Regression)
1	150.4595981	1.48E+02
2	4.80776899	1.27E+00
3	-202.906114	-2.93E+02
4	421.7194576	4.15E+02
5	279.4510729	2.72E+02
6	-52.2970823	-8.66E+04
7	-128.594189	7.59E+04
8	-167.500570	3.23E+04
9	145.740681	2.21E+02
10	496.3060412	2.93E+04
11	129.9484578	1.25E+02
12	88.30438076	9.44E+01
13	11.29067689	-9.39E+01
14	1.88532531	-3.37E+01
15	-2.58364157	3.35E+03
16	-66.8944548	-6.21E+02
17	-20.6193995	7.92E+02
18	113.3930145	1.77E+03
19	17.99086827	4.19E+03
20	52.50235963	1.19E+02
21	109.6876551	7.66E+01
22	-10.7277962	-1.52E+01
23	71.67974829	8.22E+01
24	-69.3090636	-1.46E+03
25	-124.034372	8.27E+02
26	102.639818	8.69E+02
27	72.64220588	5.86E+02
28	79.24754013	4.27E+02
29	38.48319215	9.02E+01
30	32.98009446	-1.79E+01
31	92.09539122	1.42E+02
32	68.97936154	5.83E+02
33	-24.4170091	-2.34E+02
34	101.8538797	-2.56E+02
35	1.39122669	-3.85E+02
36	20.85757155	-3.34E+01
37	-29.6549013	-1.07E+01
38	130.4111599	2.57E+02
39	-16.7510879	6.00E+01
40	87.51340344	3.84E+02
41	-45.6423836	-4.04E+02
42	-30.9228849	-5.14E+02
43	-10.0713978	3.84E+01
44	31.13334896	-4.46E+01
45	-89.3352542	-7.30E+02
46	-22.7305367	3.77E+02
47	65.41116624	4.40E+02
48	55.11621318	3.09E+02
49	19.14925041	1.90E+02
50	-59.8431584	-1.10E+02
51	26.64350735	-1.92E+03
52	108.4050128	-1.92E+03
53	-137.617569	-3.49E+03
54	-83.0438356	1.18E+04
55	-20.4021477	5.31E+02
56	24.9726362	5.43E+02
57	-0.92451093	1.82E+03
58	191.9130658	-1.05E+04
59	34.78309393	-5.17E+02
60	-43.9039350	2.06E+03
61	23.2002376	-4.20E+03
62	20.8504118	-1.40E+02
63	-117.853228	3.74E+02
64	75.30611309	5.15E+01
65	60.36839226	-4.64E+01

We can see that for Ridge regression the weight values are almost similar. It is because of the regularizat ion factor of the Ridge regression, that allows this to happen.

d) Comparison of two approaches in terms of test and train MSE (Mean square error):

REGRESSION METHOD	TRAIN MSE	TEST MSE
Linear Regression (without Intercept)	19099.4468446	106775.361451
Linear Regression (with Intercept)	2187.16029493	3707.84018096
Ridge Regression (for optimal value of λ)	2451.528490643497	2851.3302134438463

Looking at the above values, we can conclude that Ridge regression performs much better on Test data t han Linear Regression with and without Intercept.
Thus, Ridge regression is better.

e) Optimum value of lambda:

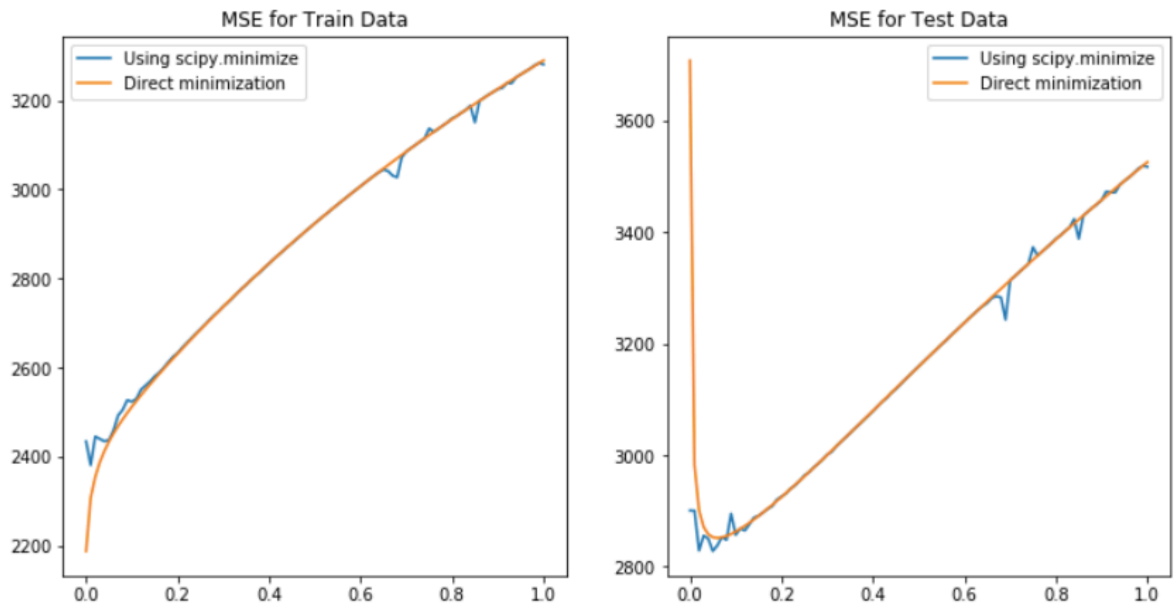
The optimal value of λ obtained is 0.06.

Test error for λ (0.06) = 2851.3302
Train error for λ (0.06) = 2451.5258

This value of lambda has been obtained by looking at the minimum Test error in case of Ridge regres sion.

Problem 4: Using Gradient Descent for Ridge Regression Learning

a) Plot the errors on train and test data obtained by using the gradient descent based learning by varyi ng the regularization parameter lambda.



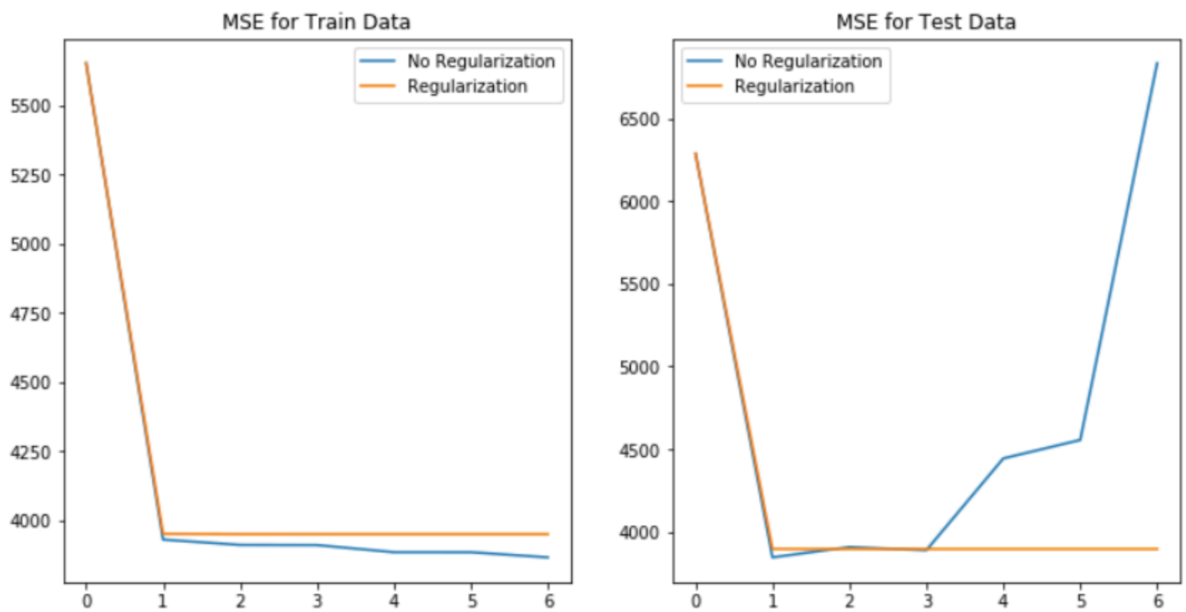
b) Comparing with Results obtained in Problem 3:
Conclusion:
Looking at the above two graphs, we can conclude that the training and test error for Ridge regression using Gradient Descent and Ridge regression are almost similar.
But in case of Ridge regression with Gradient descent, for both small values and large values of lambda, we can see that there is distortion in the graph. This is mainly due to the use of minimize and inverse function in case of gradient descent which consumes more time.

	Train MSE	Test MSE
Ridge Regression with Gradient Descent (For optimal value of λ (0.06))	2451.76715459	2824.06091919

The above figure shows value of Train and Test MSE for optimal value of lambda.

Problem 5: Non-linear Regression

a) Plots for Training and Test error in case of Non-Linear Regression



Inference from the above graphs:

i) MSE for Train Data:

From the above graph we can see that, when we have performed regularization, with increase in p values, there is not much difference in the MSE values. Whereas, when we do not have regularization i.e lambda = 0, the MSE values for train data decreases with increase in polynomial degree. This is because as the curve becomes non-linear, It tries to reduce the MSE, by adjusting according to the train data.

ii) MSE for Test Data:

From the above graph for Test data, we can see that, In case when there is no regularization, as the p value increases, there is a drastic increase in the MSE for test data. This is because, when regularization was not performed, the model gets adjusted according to the train data and with change in data (i.e with test data), the MSE increases drastically since It was made according to the test data.

When regularization is performed, the Test MSE is also consistent.

Thus, we should perform regularization as shown above.

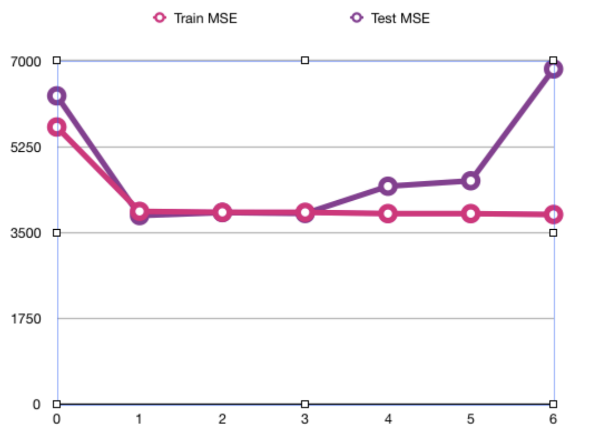
b) Plot the curve for the optimal value of p for both values of lambda and compare.

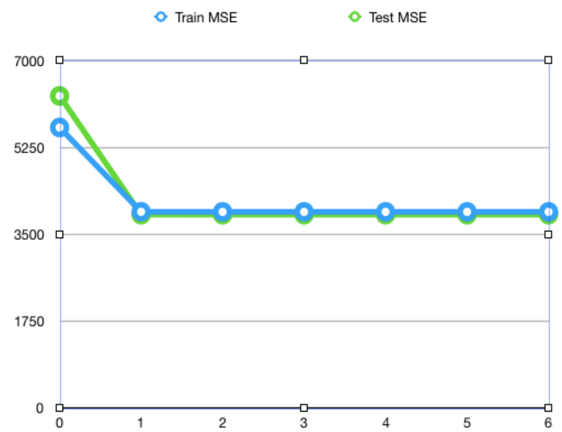
No Regularization

p	Train MSE	Test MSE
0	5650.710539	6286.404792
1	3930.915407	3845.03473
2	3911.839671	3907.128099
3	3911.188665	3887.975538
4	3885.473068	4443.327892
5	3885.407157	4554.830377
6	3866.883449	6833.459149

Regularized

p	Train MSE	Test MSE
0	5650.711907	6286.881967
1	3951.839124	3895.856464
2	3950.687312	3895.584056
3	3950.682532	3895.582716
4	3950.682337	3895.582668
5	3950.682335	3895.582669
6	3950.682335	3895.582669





The graph below for No regularization is for value of lambda = 0 and the graph for Regularization sh own above is for optimal value of lambda = 0.06.

Graph 1(No Regularization):
In this case, optimal value of p obtained for Train MSE is 5 and all the other values are almost similar for different values of p.
But for Test MSE, the optimal value for p = 1.

Graph 2(Regularization):
In this case, optimal value of p obtained for both Train MSE is 5 or 6 and Test MSE is 4,5 or 6. Also all the other values are almost similar for different values of p. This is the benefit of regularization.

Problem 6: Interpreting Results

a) Compare the various approaches in terms of training and testing error.

Approach	Train MSE	Test MSE
Linear Regression (Without Intercept)	19099.4468446	106775.361451
Linear Regression (With Intercept)	2187.16029493	3707.84018096
Ridge Regression (For optimal value of λ (0.06))	2451.528490643497	2851.3302134438463
Ridge Regression with Gradient Descent (For optimal value of λ (0.06))	2451.76715459	2824.06091919
Non Linear Regression (No Regularization)	3866.883449	3845.03473
Non Linear Regression (Regularization)	3950.682335	3895.582669

b) What metric should be used to choose the best setting?

To compare all the different approaches, we should look at the Test MSE of the approaches as it shows how good the approach performed on the Test data. Another metric that we should consider is the size of the dataset. We worked on a smaller dataset but as the size of the dataset increases, the running time becomes important. Thus, we should perform regularization on our data while working with large datasets.

Now, looking at the Test MSE, we can see that Linear Regression without Intercept has high values of Mean Squared error, for both Train and Test error.
Best approach for Train data – Linear Regression without intercept.
Best approach for Test data – Ridge regression with Gradient Descent.
Non – linear regression and Linear Regression does not perform well.

We also observed in Problem 5 that, with increase in the polynomial degree, the Train error decreases in case of no regularization but the Test error increases drastically, because of the behavior of the approach as it tries to accommodate the train data causing over fitting.