UNIVERSITY AT BUFFALO

Project Report

<u>CSE 574 – INTRO TO MACHINE LEARNING</u>

Programming Assignment 1

<u>Classification and Regression Project</u>

by

(Group 75)
Atharva Patil
Roshan Vivek Khatri
Suraj Bodapati

Problem 1: Experiment with Gaussian Discriminators

- **Output:**

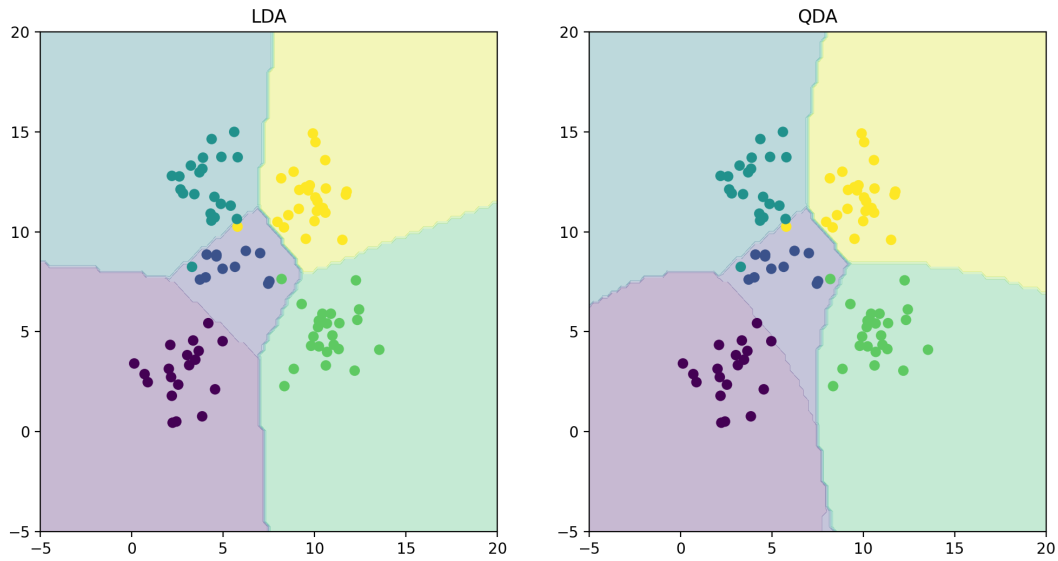


Figure 1.1: LDA and QDA

Type	Accuracy
Linear Discriminant Analysis	97.0
Quadratic Discriminant Analysis	97.0

Covariance matrix remains fixed in case of LDA, while we assign different covariance matrix in QDA. Because of this difference in Covariance matrix in LDA and QDA, we get different boundaries in LDA andQDA.

Problem 2: Experiment with Linear Regression

```
MSE training data without intercept [19099.44684457]
MSE testing data without intercept [106775.36152088]
MSE training data with intercept [2187.16029493]
MSE testing data with intercept [3707.84018113]
(base) roshankhatri@Roshans-MacBook-Air basecode %
```

a) **Training Data**

- We can observe the drop in the error while using the intercept.
- Hence, in this case, the intercept shows better results than not using intercept.
- Error training = ((19099.44684457-2187.16029493)/19099.44694457) * 100
  = 88.55%

b) **Testing Data**

- We can observe the drop in the error while using the intercept.
- Hence, in this case, the intercept shows better results than not using intercept.
- Error training = ((106775.36152088-3707.84018113)/106775.36152088) * 100
  = 96.53%

Problem 3: Experiment with Ridge Regression

Varying values of λ from 0 (no regularization) to 1 in steps of 0.01

LAMBDAS	TRAINING DATA	TESTING DATA
0.00	[2187.16029493]	[3707.84018113]
0.01	[2306.83221793]	[2982.44611971]
0.02	[2354.07134393]	[2900.97358708]
0.03	[2386.7801631]	[2870.94158888]
0.04	[2412.119043]	[2858.00040957]
0.05	[2433.1744367]	[2852.66573517]
0.06	[2451.52849064]	[2851.33021344]
0.07	[2468.07755253]	[2852.34999406]
0.08	[2483.36564653]	[2854.87973918]
0.09	[2497.74025857]	[2858.44442115]
0.10	[2511.43228199]	[2862.75794143]
0.11	[2524.60003852]	[2867.63790917]
0.12	[2537.35489985]	[2872.96228271]
0.13	[2549.77688678]	[2878.64586939]
0.14	[2561.92452773]	[2884.62691417]
0.15	[2573.84128774]	[2890.85910969]
0.16	[2585.55987497]	[2897.30665895]
0.17	[2597.10519217]	[2903.94112629]
0.18	[2608.49640025]	[2910.73937213]
0.19	[2619.74838623]	[2917.68216413]
0.20	[2630.8728232]	[2924.75322165]
0.21	[2641.87894616]	[2931.93854417]
0.22	[2652.77412633]	[2939.22592987]
0.23	[2663.56430077]	[2946.60462378]
0.24	[2674.25429667]	[2954.06505602]
0.25	[2684.84807809]	[2961.59864341]
0.26	[2695.34893502]	[2969.19763677]
0.27	[2705.75962912]	[2976.85500119]
0.28	[2716.0825067]	[2984.56432079]
0.29	[2726.31958674]	[2992.31972181]
0.30	[2736.4726296]	[3000.11580946]
0.31	[2746.54319109]	[3007.94761559]
0.32	[2756.53266482]	[3015.81055453]
0.33	[2766.44231574]	[3023.70038563]
0.34	[2776.27330654]	[3031.61318093]
0.35	[2786.02671854]	[3039.54529713]
0.36	[2795.70356824]	[3047.49335111]

0.37	[2805.30482034]	[3055.45419817]
0.38	[2814.83139806]	[3063.42491285]
0.39	[2824.28419133]	[3071.40277169]
0.40	[2833.66406312]	[3079.38523776]
0.41	[2842.97185452]	[3087.36994673]
0.42	[2852.2083886]	[3095.35469418]
0.43	[2861.3744735]	[3103.33742413]
0.44	[2870.47090474]	[3111.31621849]
0.45	[2879.49846701]	[3119.28928746]
0.46	[2888.45793552]	[3127.25496075]
0.47	[2897.35007697]	[3135.21167941]
0.48	[2906.17565032]	[3143.15798839]
0.49	[2914.93540723]	[3151.09252966]
0.50	[2923.63009243]	[3159.01403582]
0.51	[2932.26044392]	[3166.92132421]
0.52	[2940.82719309]	[3174.81329145]
0.53	[2949.33106473]	[3182.68890838]
0.54	[2957.77277699]	[3190.54721533]
0.55	[2966.15304137]	[3198.38731777]
0.56	[2974.47256259]	[3206.20838225]
0.57	[2982.73203851]	[3214.00963255]
0.58	[2990.93215999]	[3221.79034621]
0.59	[2999.07361078]	[3229.5498512]
0.60	[3007.15706742]	[3237.28752288]
0.61	[3015.1831991]	[3245.00278108]
0.62	[3023.15266757]	[3252.69508746]
0.63	[3031.06612707]	[3260.36394297]
0.64	[3038.92422416]	[3268.00888553]
0.65	[3046.72759776]	[3275.6294878]
0.66	[3054.47687898]	[3283.22535516]
0.67	[3062.17269114]	[3290.79612376]
0.68	[3069.81564971]	[3298.34145873]
0.69	[3077.40636224]	[3305.86105245]
0.70	[3084.94542842]	[3313.354623]
0.71	[3092.43344001]	[3320.82191265]
0.72	[3099.87098085]	[3328.26268646]
0.73	[3107.25862691]	[3335.67673095]
0.74	[3114.59694628]	[3343.06385289]
0.75	[3121.88649919]	[3350.42387813]
0.76	[3129.12783807]	[3357.75665047]
0.77	[3136.3215076]	[3365.0620307]
0.78	[3143.46804472]	[3372.33989556]
0.79	[3150.56797875]	[3379.59013686]

0.80	[3157.62183137]	[3386.81266063]
0.81	[3164.63011677]	[3394.00738631]
0.82	[3171.59334168]	[3401.17424594]
0.83	[3178.51200544]	[3408.31318353]
0.84	[3185.38660008]	[3415.42415428]
0.85	[3192.21761044]	[3422.50712403]
0.86	[3199.0055142]	[3429.56206859]
0.87	[3205.75078202]	[3436.58897321]
0.88	[3212.45387757]	[3443.58783202]
0.89	[3219.11525768]	[3450.55864755]
0.90	[3225.73537241]	[3457.50143021]
0.91	[3232.31466512]	[3464.41619786]
0.92	[3238.8535726]	[3471.30297539]
0.93	[3245.35252514]	[3478.16179431]
0.94	[3251.81194665]	[3484.99269234]
0.95	[3258.23225474]	[3491.79571308]
0.96	[3264.61386081]	[3498.57090566]
0.97	[3270.95717015]	[3505.3183244]
0.98	[3277.26258207]	[3512.03802854]
0.99	[3283.53048993]	[3518.7300819]
1.00	[3289.7612813]	[3525.39455263]

Using the above table, we observed that the value of MSE in testing data is lowest for λ = 0.06

Referring to the MSE values from problem 2, MSE for ridge regression is low compared to OLE regression, therefore Ridge regression is more efficient than OLE regression.

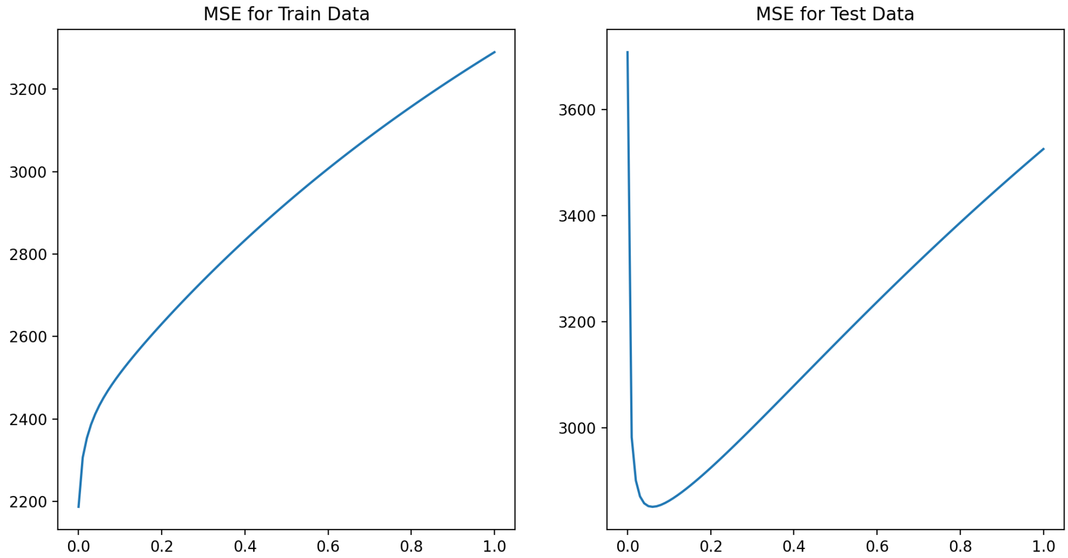


Figure 1.2: Errors observed on training and testing data for different values of Lambda

For training data, the MSE values keep increasing at a decreasing rate while in case of testing data, the MSE values first decrease and then starts to increase at a constant rate from lambda = 0.1 onwards. Thus, there is dissimilarity in trends.

Problem 4: Using Gradient Descent for Ridge Regression Learning

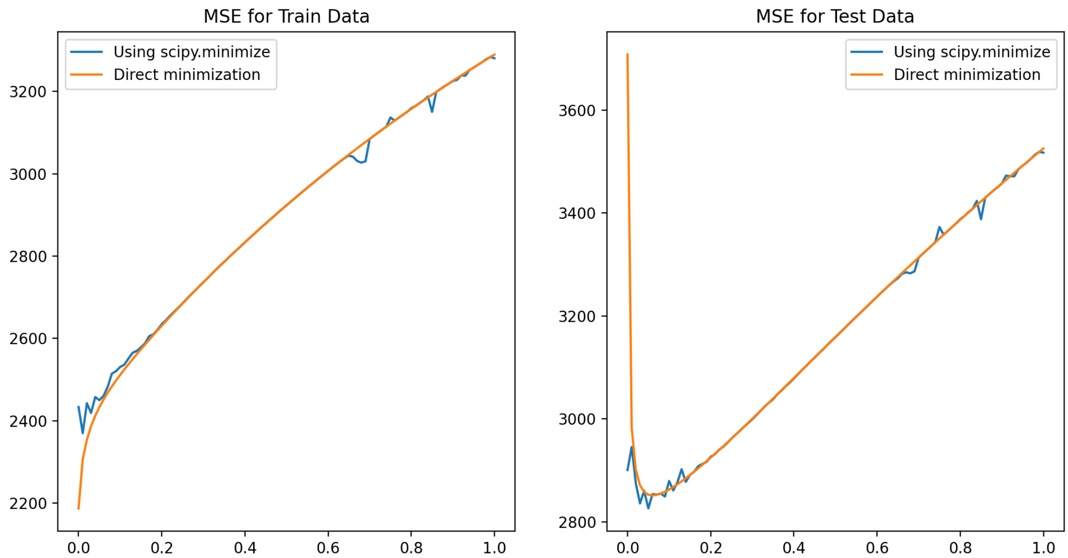


Figure 1.3: Error on training and testing data using gradient descent by varying the regularization parameter

From the two graphs, we can observe that both the graphs are similar. Whenever the lambda value is very low or very high there are a lot of irregularities in the plotting of the gradient decent. The rest of the values don't have this issue.

MSE Data for Problem 4:

```
basecode — -zsh — 80×24
(0.0, array([2433.64270284]), array([2900.55354375]))
(0.01, array([2369.89788222]), array([2945.36955486]))
(0.02, array([2442.7700224]), array([2874.62603219]))
(0.03, array([2418.97059042]), array([2835.63836365]))
(0.04, array([2457.69140609]), array([2860.29990719]))
(0.05, array([2450.16579859]), array([2825.94490202]))
(0.06, array([2460.66411221]), array([2853.84661815]))
(0.07, array([2482.32146581]), array([2852.87061317]))
(0.08, array([2514.49846491]), array([2854.7402728]))
(0.09, array([2520.58407133]), array([2849.12722276]))
(0.1, array([2530.8463871]), array([2879.28704838]))
(0.11, array([2536.40899093]), array([2860.93393745]))
(0.12, array([2551.36885717]), array([2876.5220769]))
(0.13, array([2565.30461461]), array([2902.14092714]))
(0.14, array([2569.90546222]), array([2877.53039608]))
(0.15, array([2578.57317813]), array([2890.84650119]))
(0.16, array([2588.59954045]), array([2897.11000022]))
(0.17, array([2605.81329118]), array([2908.08800732]))
(0.18, array([2610.15735546]), array([2912.33112555]))
(0.19, array([2621.32978037]), array([2916.27390094]))
(0.2, array([2634.9746352]), array([2926.52212165]))
(0.21, array([2644.30754046]), array([2931.19290291]))
(0.22, array([2655.36809597]), array([2939.82613792]))
(0.23, array([2665.35512136]), array([2945.65736566]))
```

Problem 5: Non-linear Regression

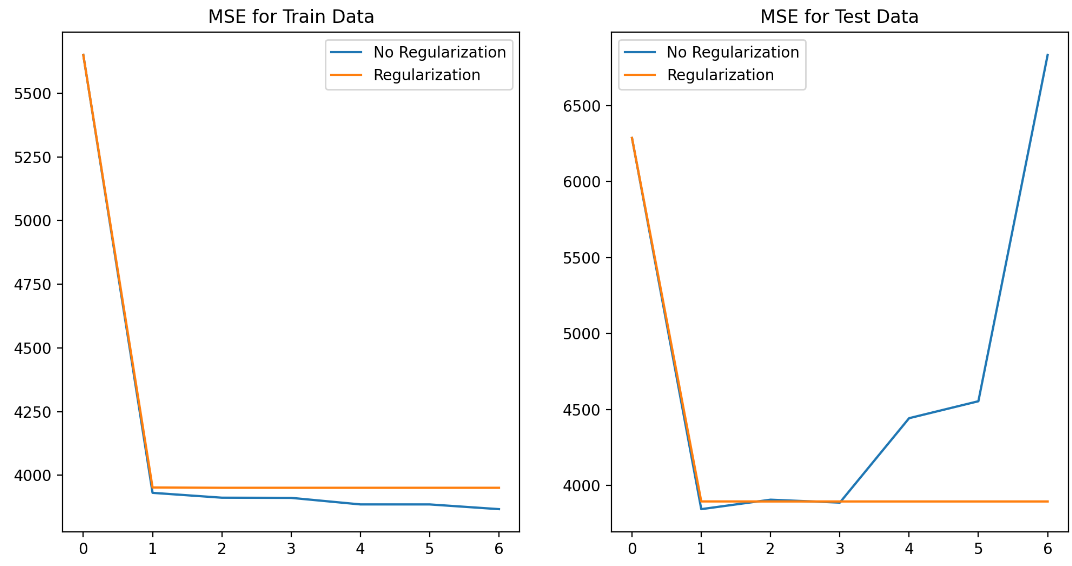


Figure 1.4: Graph 1(Errors on training data) and Graph 2(Errors on testing data)

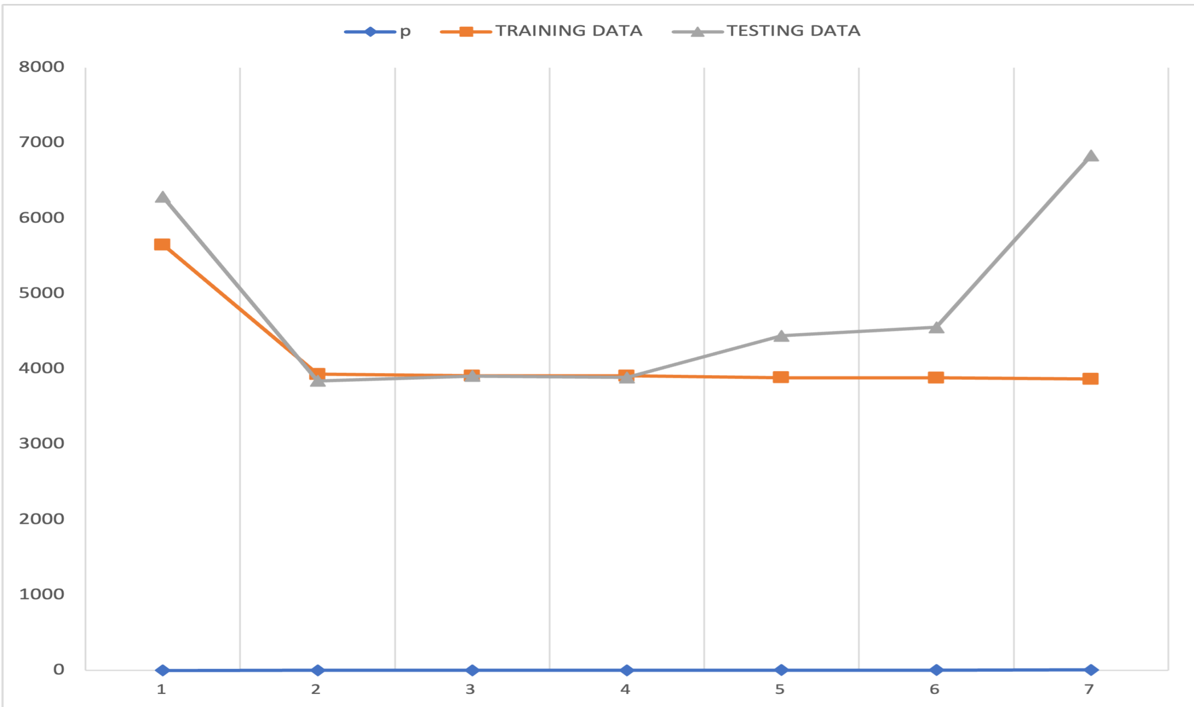
From problem 3, where there is no regularization, we choose **λ** to be 0.06.

Graph 1 shows that there is a decrease in the error. The error calculated at p = 1 in the color lines is nearly equal. For p>1, a linear graph observed along x-axis in the case of regularization. Thus, a decrease in the error is observed.

Graph 2 shows that for no regularization, when p is greater than 1,the error increases and from p = 3 onwards, growth rate becomes faster. With respect to regularization similar trend was observed on training data. It gets the minimum error at p = 1.
Relying more on training data leads to increment in error in case of test data.

Curve for the value of p varying from 0 to 6 for λ = 0 (without regularization)

p	TRAINING DATA	TESTING DATA
0	5650.71053889762	6286.404791680890
1	3930.915407315900	3845.034730173410
2	3911.8396712049500	3907.128099107940
3	3911.188664931450	3887.975538236010
4	3885.4730681122700	4443.327891813280
5	3885.40715739708	4554.830377434780
6	3866.883449446050	6833.459148719310

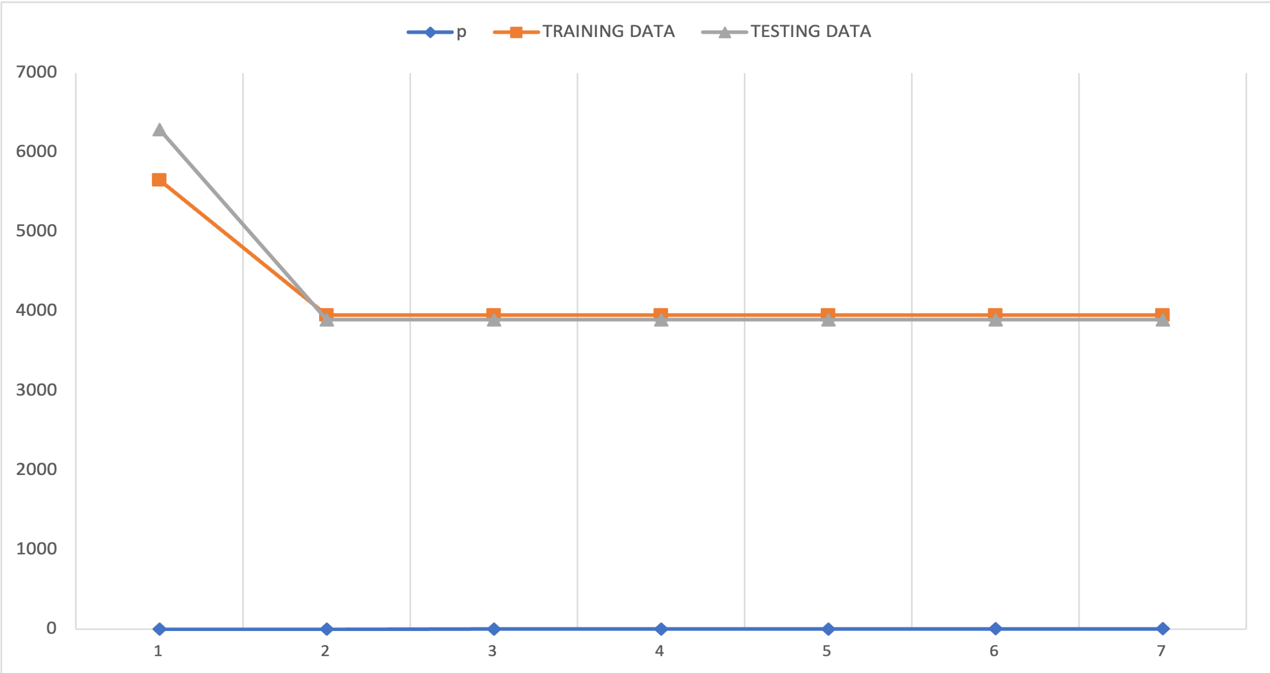


The optimal value for p in case of test error is 5 as curve is constant after that value.

The optimal value for p in case of train error is 1.

Curve for the optimal value of p varying from 0 to 6 for λ = 0.06 (with regularization)

p	TRAINING DATA	TESTING DATA
0	5650.711907032110	6286.88196694145
1	3951.839123560110	3895.8564644739600
2	3950.68731237552	3895.5840559389200
3	3950.6825315187100	3895.582715923100
4	3950.6823367953700	3895.5826682835200
5	3950.68233517702	3895.582668704420
6	3950.6823351427800	3895.5826687190900



The optimal value for p in case of test error is 4.

The optimal value for p in case of train error is 4 or 5.

PROBLEM 6 – INTERPRETING RESULTS

Type of Regression	Train MSE	Test MSE
1. With intercept (MSE Findings)	2187.1602949 3	3707.8401811 3
2. Without intercept (MSE Findings)	19099.446844 57	106775.36152 088
3. Ridge regression (MSE Findings)	2451.52849064	2851.33021344
4. Ridge regression usingGradient Descent (MSE Findings)	2450.16579859	2825.94490202
5. Non-Linear regression with regularization (MSE Findings)	3895.58	3895.58

As error obtained is minimum in calculating MSE with intercept for training data and MSE with ridgeregression in testing data, we choose these metrics as our optimal solutions.