CSE574
Project-3

Classification and Regression

Amulya Reddy Datla
UB Person number: 50560100
amulyare@buffalo.edu

Vignesh Thalur Jayachandrakumar
UB Person number: 50559900
vignesht@buffalo.edu

1 Problem-1

1.1 Accuracy

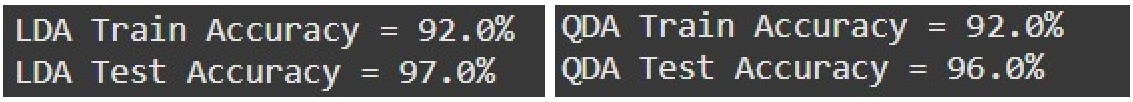
LDA Train Accuracy = 92.0%
LDA Test Accuracy = 97.0%
QDA Train Accuracy = 92.0%
QDA Test Accuracy = 96.0%

Figure 1: figure 1

The above figure shows the result obtained by the Gaussian Distributors. Accuracy obtained by Linear Discriminant Analysis is 97% for test data and 92% for train data. Accuracy obtained by Quadratic Discriminant Analysis is 96% for test data and 92% for train data.

1.2 Differences

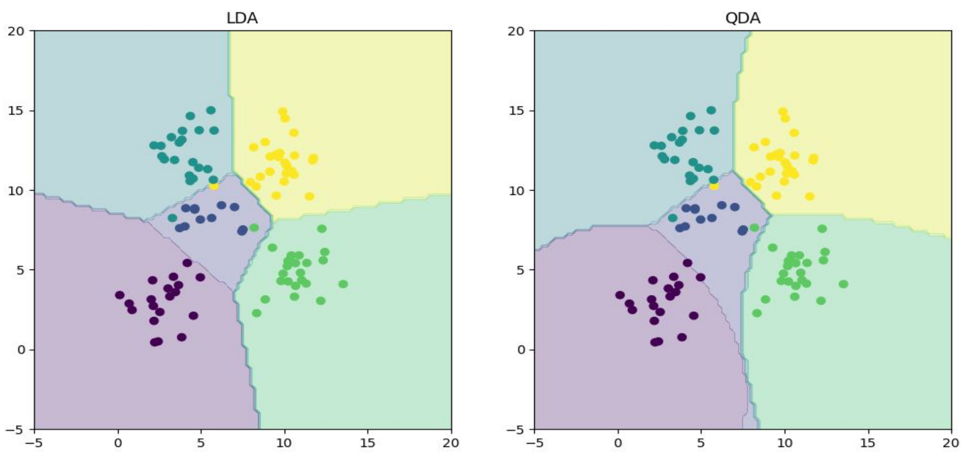

Figure 2: figure 2

The above obtained plot shows the difference between the discriminating boundaries of Linear Discriminant Analysis and Quadratic Discriminant Analysis. Linear Discriminant Analysis assumes that the covariance matrices of all classes are equal, which resulted in linear boundaries between the classes. Quadratic Discriminant Analysis assumes that each class has its own covariance matrix, leading to quadratic boundaries. LDA is less flexible as it works well when class distributions are linearly separable. QDA is more flexible as it can also perform well in cases where class distributions are more complex and non-linear. The accuracies of both LDA and QDA are comparable suggesting the data is close to being linearly separable.

2 Problem-2

2.1 Mean Square Error

```
MSE for training data without intercept: 19099.446844570695
MSE for test data without intercept 106775.36155731279
MSE for training data with intercept: 2187.160294930389
MSE for test data with intercept 3707.840181454929
```

Figure 3: figure 3

The above figure shows the result obtained by performing linear regression. Mean Square Error for training data is 19099.446 without an intercept and 2187.160 with intercept. Mean Square Error for training data is 106775.361 without an intercept and 3707.840 with intercept.

2.2 Which one is better?

Including an intercept term significantly enhances the performance of the model, as demonstrated by the substantially lower MSE values for both training and test datasets. The intercept allows the model to account for inherent offsets in the data enabling it to better capture the relationship between features and the target variable. This results in a more accurate and generalizable model making the inclusion of the intercept clearly the better choice.

3 Problem-3

3.1 Plots

The below obtained figure is the plot that shows the mean square errors on train and test data for different values of lambda. The mean square error for train data shows a monotonic increase as lambda grows from 0 to 1 indicating increasing model bias due to stronger regularization. The mean square error for test data exhibits a U-shaped curve with an optimal lambda value falling between 0.1 and 0.2, where the model achieves the best generalization performance. This indicates that a moderate amount of regularization provides the best balance between bias and variance.

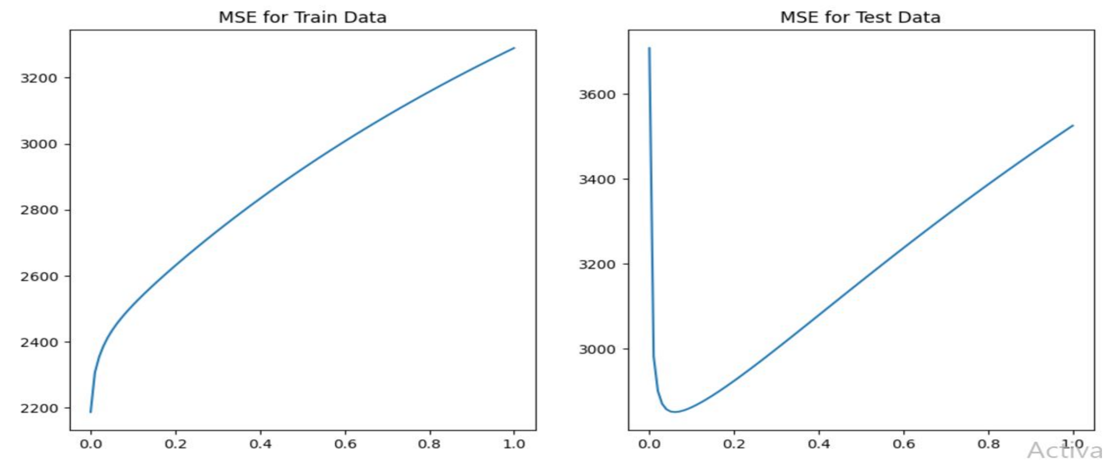

Figure 4: figure 4

3.2 Optimal Lambda

The below images show the mean squared errors (MSE) for both test and train data as lambda varies from 0 to 1 in steps of 0.01. These plots provide insights into the influence of the regularization parameter on the performance of model on both the train data and test data. The optimal value of lambda is 0.06 as the minimum MSE of test data 2851.330213 occurs at this lambda value, indicating the best generalization performance.

Lambda	MSE (Train)	MSE (Test)
0.0000	2187.160295	3707.840181
0.0100	2306.832218	2982.446120
0.0200	2354.071344	2900.973587
0.0300	2386.780163	2870.941589
0.0400	2412.119043	2858.000410
0.0500	2433.174437	2852.665735
0.0600	2451.528491	2851.330213
0.0700	2468.077553	2852.349994
0.0800	2483.365647	2854.879739
0.0900	2497.740259	2858.444421
0.1000	2511.432282	2862.757941
0.1100	2524.600039	2867.637909
0.1200	2537.354900	2872.962283
0.1300	2549.776887	2878.645869
0.1400	2561.924528	2884.626914
0.1500	2573.841288	2890.859110
0.1600	2585.559875	2897.306659
0.1700	2597.105192	2903.941126
0.1800	2608.496400	2910.739372
0.1900	2619.748386	2917.682164
0.2000	2630.872823	2924.753222
0.2100	2641.878946	2931.938544
0.2200	2652.774126	2939.225930
0.2300	2663.564301	2946.604624
0.2400	2674.254297	2954.065056
0.2500	2684.848078	2961.598643
0.2600	2695.348935	2969.197637
0.2700	2705.759629	2976.855001
0.2800	2716.082507	2984.564321
0.2900	2726.319587	2992.319722
0.3000	2736.472630	3000.115809
0.3100	2746.543191	3007.947616
0.3200	2756.532665	3015.810555
0.3300	2766.442316	3023.700386
0.3400	2776.273307	3031.613181
0.3500	2786.026719	3039.545297

0.3500	2786.026719	3039.545297
0.3600	2795.703568	3047.493351
0.3700	2805.304820	3055.454198
0.3800	2814.831398	3063.424913
0.3900	2824.284191	3071.402772
0.4000	2833.664063	3079.385238
0.4100	2842.971855	3087.369947
0.4200	2852.208389	3095.354694
0.4300	2861.374474	3103.337424
0.4400	2870.470905	3111.316218
0.4500	2879.498467	3119.289287
0.4600	2888.457936	3127.254961
0.4700	2897.350077	3135.211679
0.4800	2906.175650	3143.157988
0.4900	2914.935407	3151.092530
0.5000	2923.630092	3159.014036
0.5100	2932.260444	3166.921324
0.5200	2940.827193	3174.813291
0.5300	2949.331065	3182.688908
0.5400	2957.772777	3190.547215
0.5500	2966.153041	3198.387318
0.5600	2974.472563	3206.208382
0.5700	2982.732039	3214.009633
0.5800	2990.932160	3221.790346
0.5900	2999.073611	3229.549851
0.6000	3007.157067	3237.287523
0.6100	3015.183199	3245.002781
0.6200	3023.152668	3252.695087
0.6300	3031.066127	3260.363943
0.6400	3038.924224	3268.008886
0.6500	3046.727598	3275.629488
0.6600	3054.476879	3283.225355
0.6700	3062.172691	3290.796124
0.6800	3069.815650	3298.341459
0.6900	3077.406362	3305.861052
0.7000	3084.945428	3313.354623
0.7100	3092.433440	3320.821913
0.7200	3099.870981	3328.262686

0.6900	3077.406362	3305.861052
0.7000	3084.945428	3313.354623
0.7100	3092.433440	3320.821913
0.7200	3099.870981	3328.262686
0.7300	3107.258627	3335.676731
0.7400	3114.596946	3343.063853
0.7500	3121.886499	3350.423878
0.7600	3129.127838	3357.756650
0.7700	3136.321508	3365.062031
0.7800	3143.468045	3372.339896
0.7900	3150.567979	3379.590137
0.8000	3157.621831	3386.812661
0.8100	3164.630117	3394.007386
0.8200	3171.593342	3401.174246
0.8300	3178.512005	3408.313184
0.8400	3185.386600	3415.424154
0.8500	3192.217610	3422.507124
0.8600	3199.005514	3429.562069
0.8700	3205.750782	3436.588973
0.8800	3212.453878	3443.587832
0.8900	3219.115258	3450.558648
0.9000	3225.735372	3457.501430
0.9100	3232.314665	3464.416198
0.9200	3238.853573	3471.302975
0.9300	3245.352525	3478.161794
0.9400	3251.811947	3484.992692
0.9500	3258.232255	3491.795713
0.9600	3264.613861	3498.570906
0.9700	3270.957170	3505.318324
0.9800	3277.262582	3512.038029
0.9900	3283.530490	3518.730082
1.0000	3289.761281	3525.394553

Figure 5: figure 5

3.3 Comparison of weights

The relative magnitudes of the weights in the below image highlight the impact of regularization on the model. The L2 norm of weights learned using OLE without an intercept is extremely large at 1977655.3940, indicating potential overfitting due to the lack of a bias term. Adding an intercept reduces this to 124531.5263, reflecting better but still has weight maginitudes which are not regularized. In contrast, the weights learned in ridge regression with an optimal lambda value of 0.06 have an L2 norm of 959.3130, showing

```
L2 Norm of weights (OLE without intercept): 1977655.3940
L2 Norm of weights (OLE with intercept): 124531.5263
L2 Norm of weights (Ridge Regression, λ=0.0600): 959.3130
```

Figure 6: figure 6

the significant reduction in weight magnitudes due to regularization. This demonstrates the ability of Ridge Regression to control weights, improving model stability and generalization.

4 Problem-4

4.1 Plots

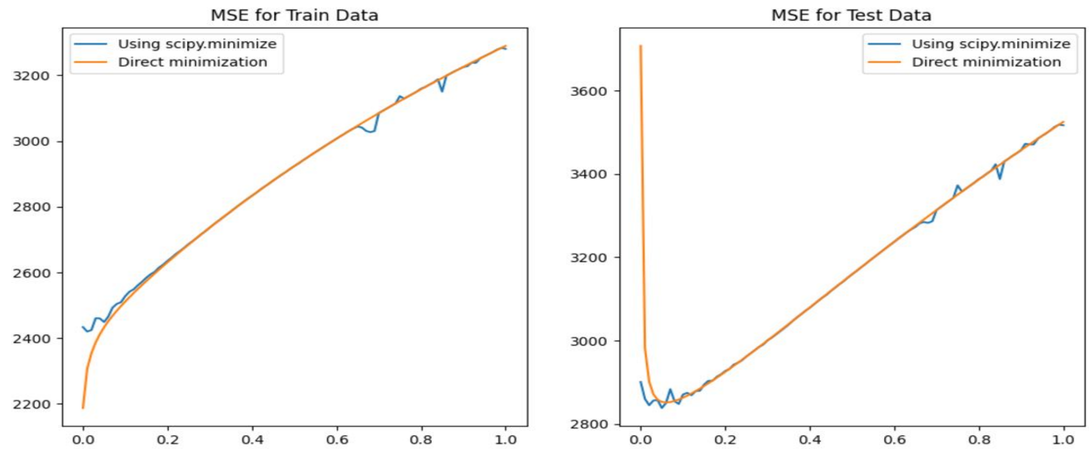

Figure 7: figure 7

The above image consists of the plot compare the Mean Squared Error (MSE) for training and test data using two optimization methods: scipy.minimize and direct minimization. For training data, both methods start at around 2200 MSE and rise to 3300, with scipy.minimize exhibiting minor oscillations while direct minimization maintains a smoother curve. For test data, direct minimization initially spikes above 3600 MSE before stabilizing, whereas scipy.minimize shows oscillations throughout. Both methods eventually converge to similar MSE values near 3500, following an upward trend. Despite differing trajectories, their final performance is comparable, with direct minimization smoother for training data but more volatile initially for test data.

4.2 Comparison

In comparing the results between the previous plot of Ridge Regression in problem 3 and the present plot of Gradient Descent Ridge Regression in problem 4, there are a few key differences. The Ridge Regression MSE curves for both training and test data show a smooth increase as the regularization parameter increases. However, in the Gradient Descent Ridge Regression, the MSE curves exhibit some fluctuations, particularly for smaller values of the regularization parameter. Despite these fluctuations, both methods converge to similar MSE values as the regularization parameter further increases. The direct minimization approach appears more stable and consistent while the gradient descent method shows slight instability or oscillations in the early stages.

5 Problem-5

```
Optimal p (No Regularization, λ=0): 1
Test MSE at Optimal p (No Regularization): 3845.034730173414
Optimal p (With Regularization, λ=0.06): 4
Test MSE at Optimal p (With Regularization): 3895.582668283526
```

Figure 8: figure 8

The above image shows the results obtained by performing non linear regression. Optimal value of p without regularization is 1 at a lambda value of 0 yielding a MSE of 3845.03 on test data. Optimal value of p with regularization is 4 at a lambda value of 0.06 yielding a MSE of 3895.58 on test data. With regularization. This shows that regularization helps increase model complexity, with a small increase in testing error helping to manage over-fitting.

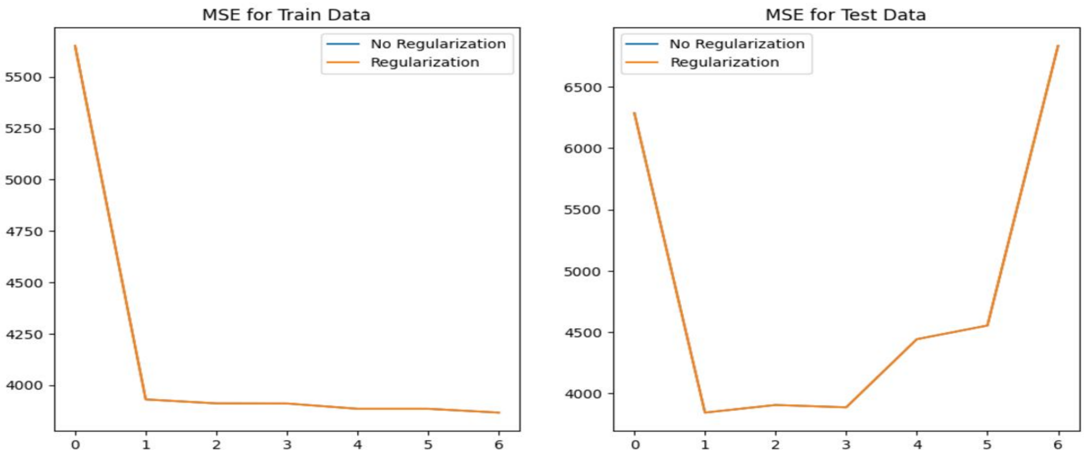

Figure 9: figure 8

The above figure is the plot obtained at a lambda value of 0. At lambda 0, both training and test data exhibit a similar pattern of decrease in MSE initially and a sharp increase after a certain point. The model performs slightly better without regularization than with regularization for both test and train data, especially on the test data, where the gap is more noticeable.

The below figure is the plot obtained at an optimal lambda value of 0.06. At lambda 0.06, regularization significantly improves the performance of the model on the test data, as seen by the lower MSE compared to no regularization. The MSE for the train data remains relatively stable across different degrees of complexity, but regularization helps avoid the sharp increase in MSE for the test data, indicating better generalization.

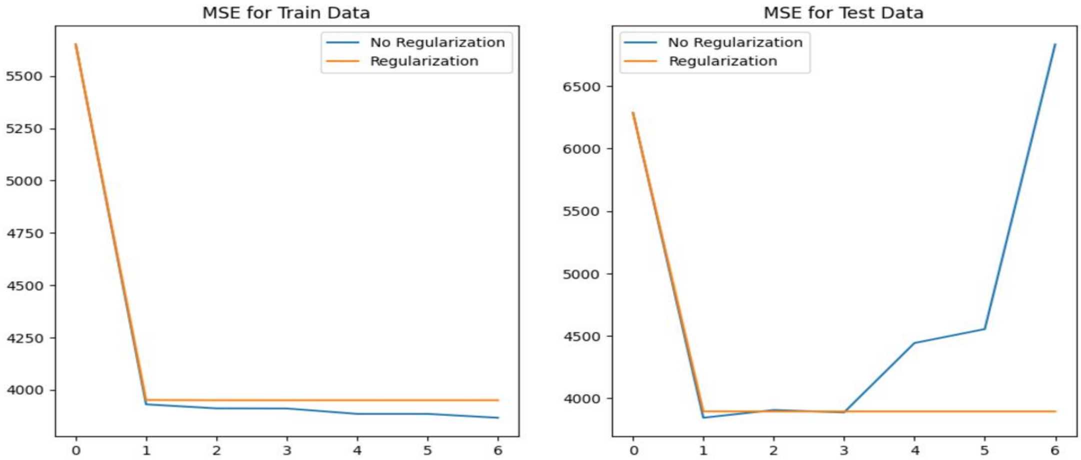

Figure 10: figure 9

6 Problem-6

Method	**MSE(train)**	**MSE(test)**
Linear Discriminant Analysis (LDA)	0.153	0.14
Quadratic Discriminant Analysis (QDA)	0.153	0.15
Linear Regression (without intercept)	19099.45	106775.36
Linear Regression (with intercept)	2187.16	3707.84
Ridge Regression (Optimal $\lambda = 0.06$)	2451.53	2851.33
Gradient Descent Ridge Regression ($\lambda = 0.06$)	2465.84	2850.42
Gradient Descent Ridge Regression (Optimal $\lambda = 0.05$)	2448.98	2837.93
Non-linear Regression (Optimal $p = 4$, with regularization)	3950.68	3895.58

Ridge Regression and Gradient Descent Ridge Regression with optimal regularization strikes the best balance between training and testing errors, improving generalization by penalizing large coefficients. In contrast, Ordinary Least Squares suffers from overfitting with low training error but high test error. Non-linear Regression with regularization performs well with complex patterns, but without regularization, it risks overfitting. The best metric for model selection is Mean Squared Error on the test set, and Ridge Regression with regularization provides the most reliable results for accurate predictions. The metric that can be selected is Mean square Error. Considering the values obtained for the test data Gradient Descent Ridge Regression can be used for best setting .