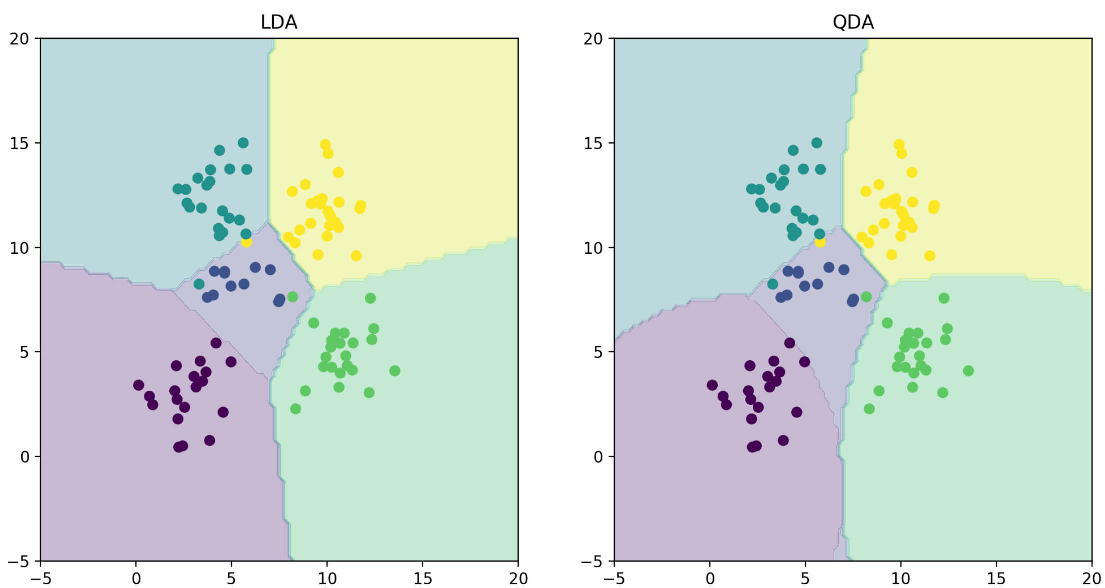
CSE 574 Introduction to Machine Learning
Programming Assignment 2

Team 66:
Weiyi Jiang 50207995
Shih-Chia Chen 50207079

I. Problem 1: Experiment with Gaussian Discriminators

- Report the accuracy of LDA and QDA:
  LDA Accuracy = 97.0%
  QDA Accuracy = 97.0%

- Figure 1.



- Explain why there is a difference in the two boundaries.
  The difference between the two boundaries is whether the boundaries are curved or straight. We can observe from the above figure that the boundaries in LDA are straight lines but the lines in QDA are curves. The reason why boundaries in LDA are straight is that the covariance matrix for all the classes are the same. However, in QDA, each class has its own covariance matrix.

II. Problem 2: Experiment with Linear Regression

- Report the MSE for training and test data for two cases:
  First, without using an intercept (or bias) term, and second with using an intercept

	Without intercept	With intercept
Training data	106775.361596	3707.8401817
Test data	19227.6796395	2087.6538161

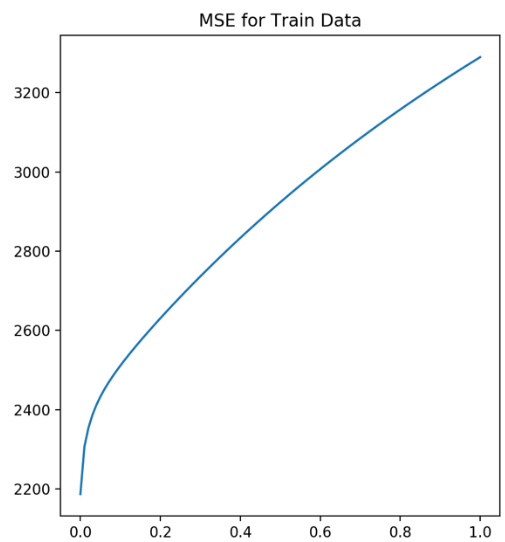
- Which one is better?
  The MSE value without intercept are much higher than with intercept in both training data and test data. The less MSE value indicates that the error are less than the higher MSE value. Since using bias means the line does not need to go through point (0,0). Therefore, using a bias is better.

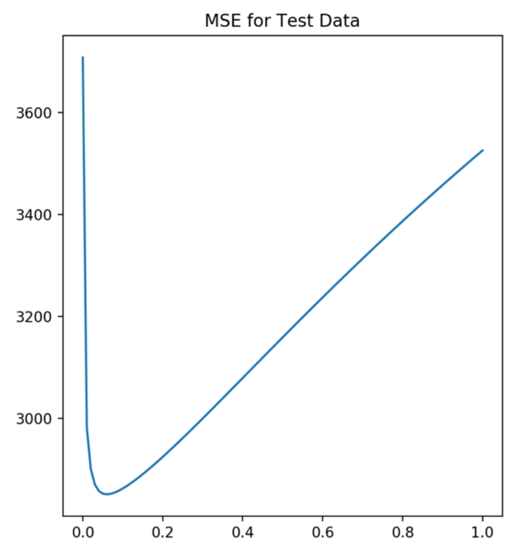
III. Problem 3: Experiment with Ridge Regression

- Report the MSE for training and test data using ridge regression parameters, and use data with intercept.

lambda	Training MSE	Test MSE	lambda	Training MSE	Test MSE
0	[2187.16029493]	[3707.8401817]	0.51	[2932.26044392]	[3166.92132421]
0.01	[2306.83221793]	[2982.44611971]	0.52	[2940.82719309]	[3174.81329145]
0.02	[2354.07134393]	[2900.97358708]	0.53	[2949.33106473]	[3182.68890838]
0.03	[2386.7801631]	[2870.94158888]	0.54	[2957.77277699]	[3190.54721533]
0.04	[2412.119043]	[2858.00040957]	0.55	[2966.15304137]	[3198.38731777]
0.05	[2433.1744367]	[2852.66573517]	0.56	[2974.47256259]	[3206.20838225]
0.06	[2451.52849064]	[2851.33021344]	0.57	[2982.73203851]	[3214.00963255]
0.07	[2468.07755253]	[2852.34999406]	0.58	[2990.93215999]	[3221.79034621]
0.08	[2483.36564653]	[2854.87973918]	0.59	[2999.07361078]	[3229.5498512]
0.09	[2497.74025857]	[2858.44442115]	0.6	[3007.15706742]	[3237.28752288]
0.1	[2511.43228199]	[2862.75794143]	0.61	[3015.1831991]	[3245.00278108]
0.11	[2524.60003852]	[2867.63790917]	0.62	[3023.15266757]	[3252.69508746]
0.12	[2537.35489985]	[2872.96228271]	0.63	[3031.06612707]	[3260.36394297]
0.13	[2549.77688678]	[2878.64586939]	0.64	[3038.92422416]	[3268.00888553]
0.14	[2561.92452773]	[2884.62691417]	0.65	[3046.72759776]	[3275.6294878]
0.15	[2573.84128774]	[2890.85910969]	0.66	[3054.47687898]	[3283.22535516]
0.16	[2585.55987497]	[2897.30665895]	0.67	[3062.17269114]	[3290.79612376]
0.17	[2597.10519217]	[2903.94112629]	0.68	[3069.81564971]	[3298.34145873]
0.18	[2608.49640025]	[2910.73937213]	0.69	[3077.40636224]	[3305.86105245]
0.19	[2619.74838623]	[2917.68216413]	0.7	[3084.94542842]	[3313.354623]
0.2	[2630.8728232]	[2924.75322165]	0.71	[3092.43344001]	[3320.82191265]
0.21	[2641.87894616]	[2931.93854417]	0.72	[3099.87098085]	[3328.26268646]
0.22	[2652.77412633]	[2939.22592987]	0.73	[3107.25862691]	[3335.67673095]
0.23	[2663.56430077]	[2946.60462378]	0.74	[3114.59694628]	[3343.06385289]
0.24	[2674.25429667]	[2954.06505602]	0.75	[3121.88649919]	[3350.42387813]
0.25	[2684.84807809]	[2961.59864341]	0.76	[3129.12783807]	[3357.75665047]
0.26	[2695.34893502]	[2969.19763677]	0.77	[3136.3215076]	[3365.0620307]
0.27	[2705.75962912]	[2976.85500119]	0.78	[3143.46804472]	[3372.33989556]
0.28	[2716.0825067]	[2984.56432079]	0.79	[3150.56797875]	[3379.59013686]
0.29	[2726.31958674]	[2992.31972181]	0.8	[3157.62183137]	[3386.81266063]
0.3	[2736.4726296]	[3000.11580946]	0.81	[3164.63011677]	[3394.00738631]
0.31	[2746.54319109]	[3007.94761559]	0.82	[3171.59334168]	[3401.17424594]
0.32	[2756.53266482]	[3015.81055453]	0.83	[3178.51200544]	[3408.31318353]
0.33	[2766.44231574]	[3023.70038563]	0.84	[3185.38660008]	[3415.42415428]

0.34	[2776.27330654]	[3031.61318093]	0.85	[3192.21761044]	[3422.50712403]
0.35	[2786.02671854]	[3039.54529713]	0.86	[3199.0055142]	[3429.56206859]
0.36	[2795.70356824]	[3047.49335111]	0.87	[3205.75078202]	[3436.58897321]
0.37	[2805.30482034]	[3055.45419817]	0.88	[3212.45387757]	[3443.58783202]
0.38	[2814.83139806]	[3063.42491285]	0.89	[3219.11525768]	[3450.55864755]
0.39	[2824.28419133]	[3071.40277169]	0.9	[3225.73537241]	[3457.50143021]
0.4	[2833.66406312]	[3079.38523776]	0.91	[3232.31466512]	[3464.41619786]
0.41	[2842.97185452]	[3087.36994673]	0.92	[3238.8535726]	[3471.30297539]
0.42	[2852.2083886]	[3095.35469418]	0.93	[3245.35252514]	[3478.16179431]
0.43	[2861.3744735]	[3103.33742413]	0.94	[3251.81194665]	[3484.99269234]
0.44	[2870.47090474]	[3111.31621849]	0.95	[3258.23225474]	[3491.79571308]
0.45	[2879.49846701]	[3119.28928746]	0.96	[3264.61386081]	[3498.57090566]
0.46	[2888.45793552]	[3127.25496075]	0.97	[3270.95717015]	[3505.3183244]
0.47	[2897.35007697]	[3135.21167941]	0.98	[3277.26258207]	[3512.03802854]
0.48	[2906.17565032]	[3143.15798839]	0.99	[3283.53048993]	[3518.7300819]
0.49	[2914.93540723]	[3151.09252966]	1	[3289.7612813]	[3525.39455263]
0.5	[2923.63009243]	[3159.01403582]			



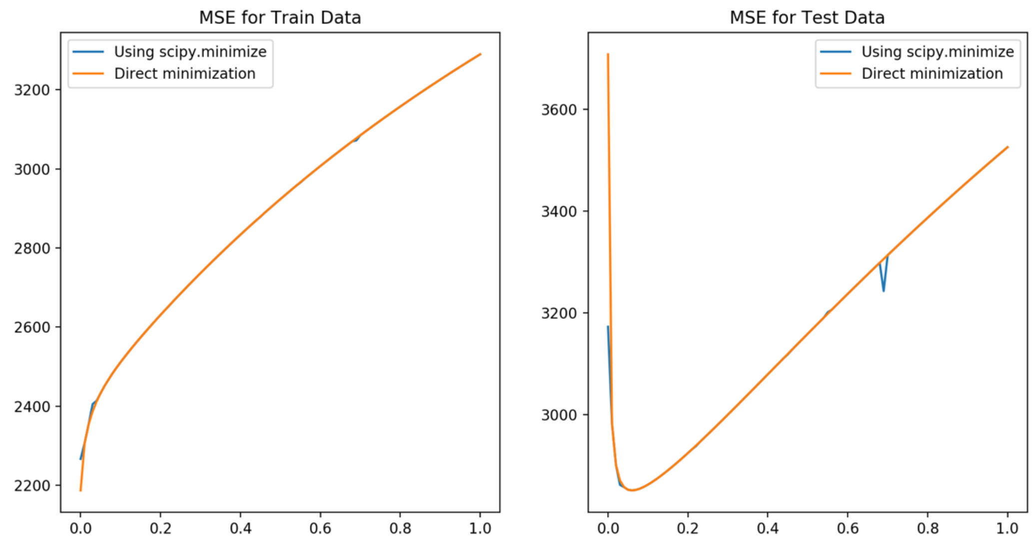


- Compare the relative magnitudes of weights learnt using OLE (Problem 2) and weights learnt using ridge regression. Compare the two approaches in terms of errors on train and test data. What is the optimal value for lambda and why?

  The optimal value for lambda is 0.06, since when lambda is 0.06, we can get the smallest MSE value for test data. For the training data, the MSE value in lambda = 0.06 is bigger than the smallest MSE, but it is still relatively small.

IV. Problem 4: Using Gradient Descent for Ridge Regression Learning

- Compare with the results obtained in Problem 3.
  Assigning maxiter = 100. When the iteration times is bigger, the regularized squared error curve will become more smooth, and fit the Direct minimization curve.

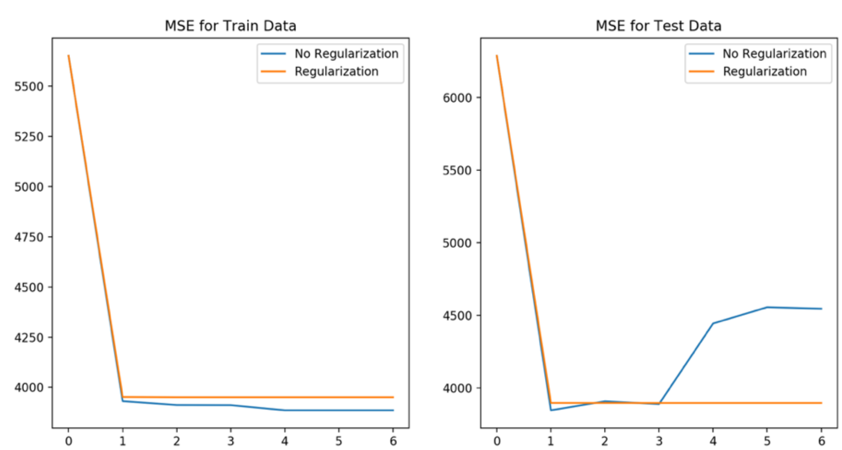


- Figure: The errors on train and test data obtained by using the gradient descent based learning by varying the regularization parameter lambda.

V. Problem 5: Non-linear Regression

- Using the lambda = 0 and the optimal value of lambda found in Problem 3, train ridge regression weights using the non-linear mapping of the data. Vary p from 0 to 6. Compute the errors on train and test data. Compare the results for both values of lambda. What is the optimal value of p in terms of test error in each setting?

From the figure, we can tell in the training data, the MSE values of regularized are always greater than the no regularized. However, for the test data, at first, the MSE values without regularization are smaller or equal to the MSE with regularization, but when p comes to 3, the MSE without regularization becomes much greater then the regularization. This means that there might be an overfitting problem.
Therefore, the optimal value of p in test data is 1, and in training data is 6.



VI. Interpreting Results

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

Problem	Training MSE	Test MSE
2 (No Intercept)	106775.361596	19227.6796395
2 (Intercept)	3707.8401817	2087.6538161
3 (optimal)	2451.52849064	2851.33021344
4 (optimal)	2249.81088315	2851.33212778
5 (optimal, no regularization)	3950.68233514	3895.58266828
5 (optimal, regularization)	3885.4071574	3845.03473017

- What metric should be used to choose the best setting?

From the metric, we can know that regression with intercept is much better than regression without intercept, in both training and testing MSE, since the MSE with intercept are much lower than the without intercept.
For problem 3 and 4, the training MSE is smaller than problem 2, but in the test data, both of the 3 and 4's MSE are bigger than problem 2. In problem 3 and 4, since it has iterated 100 times in the problem 4, we can find a smaller MSE in training data. However, it takes more time to iterate 100 times.
In problem 5, the MSE with regularization in training data is smaller, but it does mean in the testing data will have the same result.