CSE 574 INTRODUCTION TO MACHINE LEARNING

PROGRAMMING ASSIGNMENT 2

Classification and Regression

By Group 38-

Tulika Sengupta (tulikase@buffalo.edu)

Darshan Prakash Mhatre(dmhatre@buffalo.edu)

Divyansh Bhatnagar(dbhatnag@buffalo.edu)

Problem 1: Experiment with Gaussian Discriminators

1.1 Accuracy

LDA Accuracy = 97%
QDA Accuracy = 96%

1.2 QDA and LDA Plots

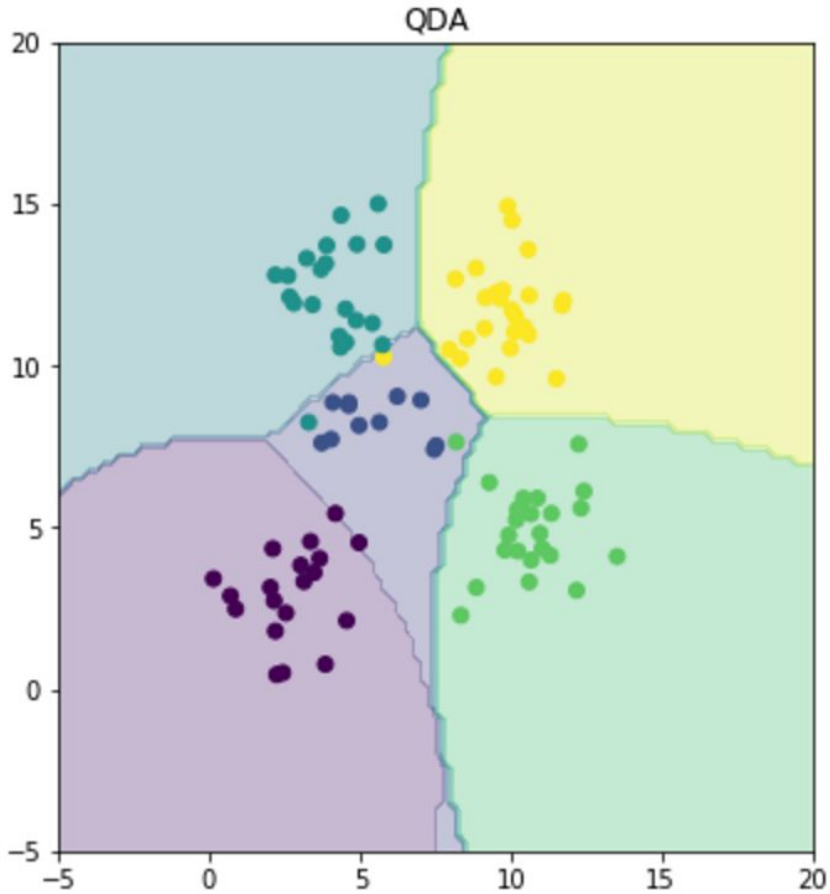

Figure 1: QDA

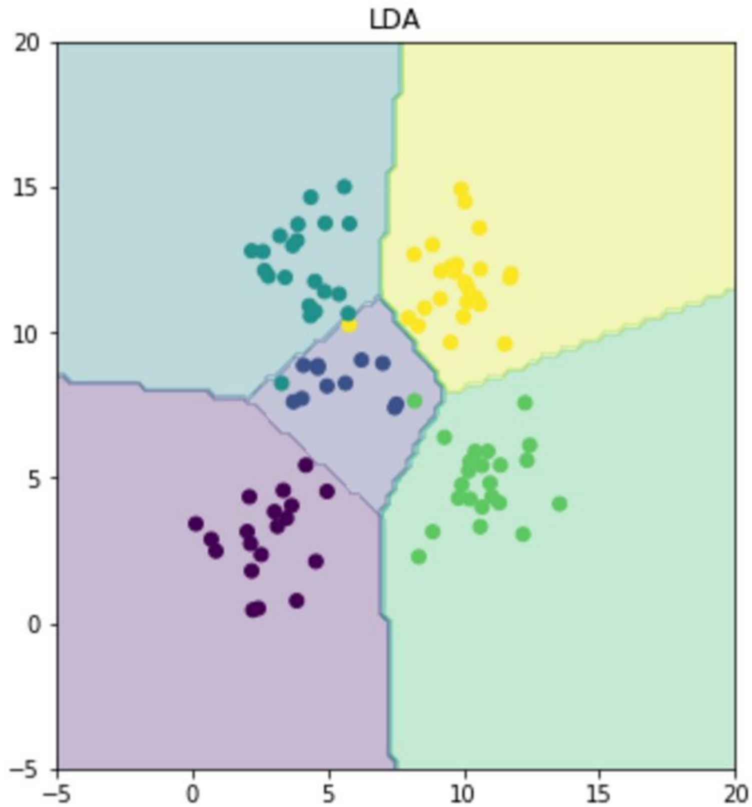

Figure 2: LDA

The clear difference which we can observe between the boundaries plotted by QDA and LDA is because QDA is more flexible classifier. This is because, in LDA covariance matrix is common for all k classes whereas, in QDA we calculate different covariance matrix for different classes.

Problem 2: Experiment with Linear Regression

2.1 Observation

	MSE for test data	MSE for training data
Without intercept	**106775.36155789**	**19099.44684457**
With intercept	**3707.84018132**	**2187.16029493**

Table 1: MSE for training and test data

In general, linear regression model without intercept creates two problems. Firstly, it looses the interpretation of R square values. Secondly, we might get biased slope estimators.

In our linear regression model, the intercept is very significant. Hence, in every case, be it test data or training data, we are getting better results in terms of MSE for linear regression model with intercept.

Problem 3: Experiment with Ridge Regression

3.1 MSE for Training and Test Data

Lambda	MSE Training data	MSE Testing Data
0.0	**2187.16029493**	3707.84018132
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
0.1	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

Table 2: Lambda and MSE

3.2 MSE Plots

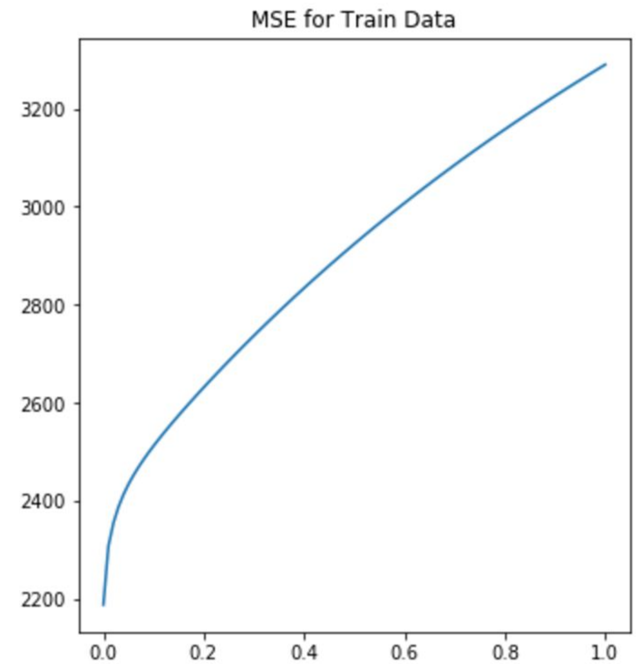

Figure 3: MSE for Train Data

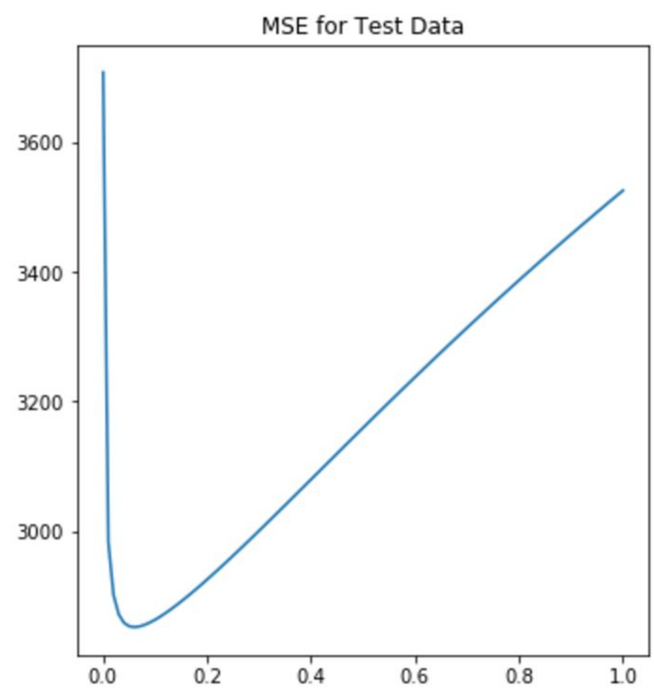

Figure 4: MSE for Test Data

3.3 Comparison of Relative Magnitudes

OLE	Ridge
-4.12E+02	1.50E+02
-3.46E+02	2.17E+01
5.79E+02	-3.91E+01
5.89E+01	1.90E+02
-1.36E+06	1.31E+02
1.19E+06	1.29E+01
5.07E+05	-1.26E+01
-1.35E+03	-1.12E+02
4.48E+05	9.94E+01
4.78E+02	2.03E+02
-1.41E+02	1.13E+02
-9.19E+02	2.79E+01
-3.96E+02	5.56E+01
-7.26E+04	3.59E+01
-8.95E+04	9.32E+00
-3.24E+03	-1.48E+01
1.41E+03	2.08E+00
3.92E+04	2.62E+01

Table 3: Weights of OLE and Ridge Regression

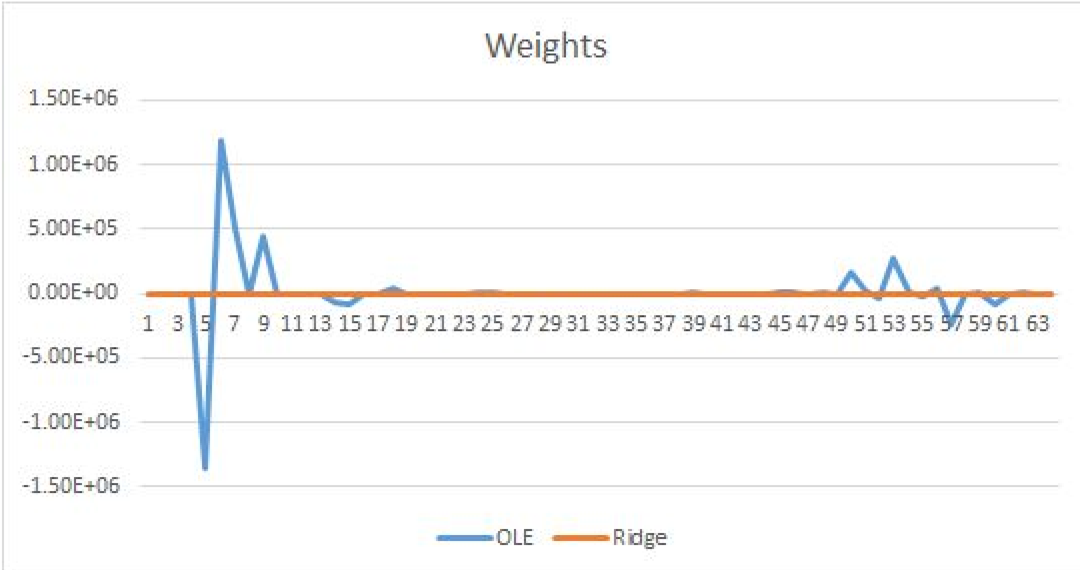

Figure 5: Weights comparison between OLE and Ridge Regression

The weight magnitudes in OLE was much higher as compared to the weights learnt in Ridge Regression. This is because of the regularization factor that was introduced in case of Ridge Regression. Regularization helps in keeping the weights in check, thereby not allowing the weights to attain very high values.

3.4 Comparison of the two approaches in terms of errors on train and test data

Regression Method	**Test MSE**	**Train MSE**
Linear Regression Without Intercept	106775.36155789	19099.44684457
Linear Regression With Intercept	3707.84018132	2187.16029493
Ridge Regression(optimal)	2851.33021344	2187.16029493

Table 4: Comparison of Ridge Regression and Linear Regression

As can be seen from the table, OLE makes much more number of mistakes as compared to Linear Regression on the training and the test data. Hence, Ridge Regression is a better method than the OLE method.

3.5 Optimal Value of Lambda

Optimal Lamda for Test Data: 0.059999999999999998
Corresponding MSE: 2851.33021344

Optimal Lamda For Training Data: 0.0
Corresponding MSE: 2187.16029493

This value of lambda is optimal because the mean squared error for the given test and training data is least at lambda = 0.059 and lambda = 0.0 respectively.

Problem 4: Using Gradient Descent for Ridge Regression Learning

4.1 Comparison with Problem 3

In this part, unlike in problem 3 where we calculated just the mean squared error, here we calculated the regularized squared error and the gradient of the error with respect to the weights, using the regularization parameter lambda.
In problem 3, we are computing the inverse of the covariance matrix to compute the weights. Although this technique is faster than Problem 4 technique because computing the minimization might take some time to converge, computing the inverse of a matrix is not always easy. The matrix of a singular matrix does not exist. For these problems, using the gradient descent is a better option as we get satisfactory results with this method as well.

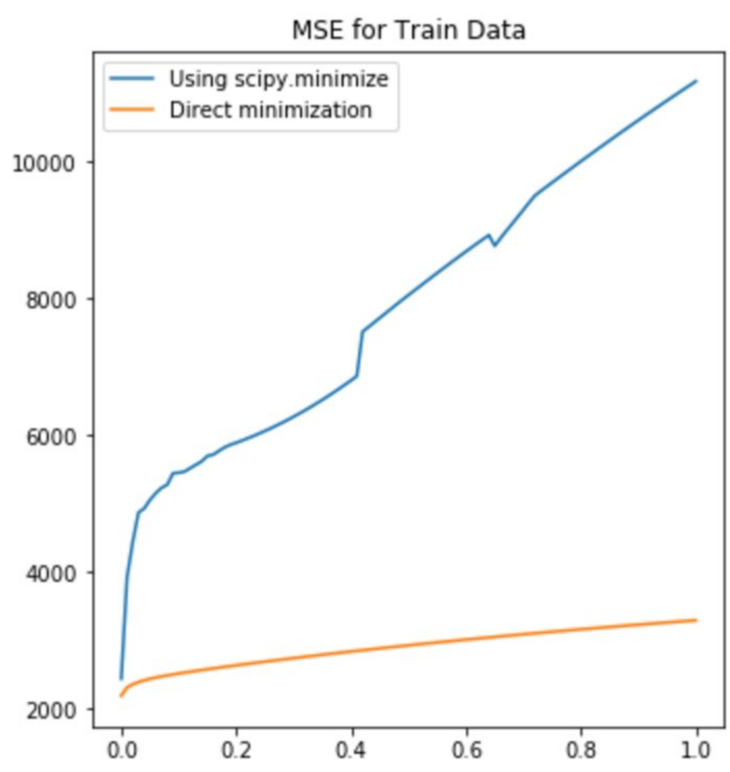

Figure 6: MSE for Train Data

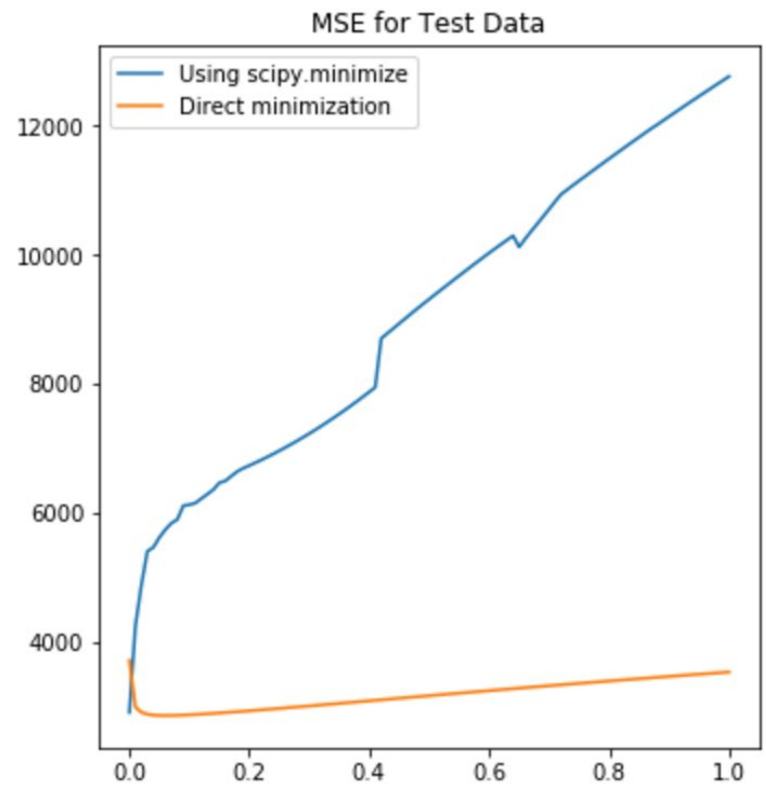

Figure 7: MSE for Test Data

MSE TRAIN	MSE TEST
2187.16029493	3707.84018132
2306.83221793	2982.44611971
2354.07134393	2900.97358708
2386.7801631	2870.94158888
2412.119043	2858.00040957
2433.1744367	2852.66573517
2451.52849064	2851.33021344
2468.07755253	2852.34999406
2483.36564653	2854.87973918
2497.74025857	2858.44442115
2511.43228199	2862.75794143
2524.60003852	2867.63790917
2537.35489985	2872.96228271
2549.77688678	2878.64586939
2561.92452773	2884.62691417
2573.84128774	2890.85910969
2585.55987497	2897.30665895
2597.10519217	2903.94112629
2608.49640025	2910.73937213
2619.74838623	2917.68216413
2630.8728232	2924.75322165
2641.87894616	2931.93854417
2652.77412633	2939.22592987
2663.56430077	2946.60462378
2674.25429667	2954.06505602
2684.84807809	2961.59864341
2695.34893502	2969.19763677
2705.75962912	2976.85500119
2716.0825067	2984.56432079
2726.31958674	2992.31972181
2736.4726296	3000.11580946
2746.54319109	3007.94761559
2756.53266482	3015.81055453
2766.44231574	3023.70038563
2776.27330654	3031.61318093
2786.02671854	3039.54529713
2795.70356824	3047.49335111
2805.30482034	3055.45419817
2814.83139806	3063.42491285
2824.28419133	3071.40277169
2833.66406312	3079.38523776
2842.97185452	3087.36994673
2852.2083886	3095.35469418
2861.3744735	3103.33742413

Table 5: MSE for Train and Test data

Problem 5: Non-linear Regression

5.1 MSE Plots

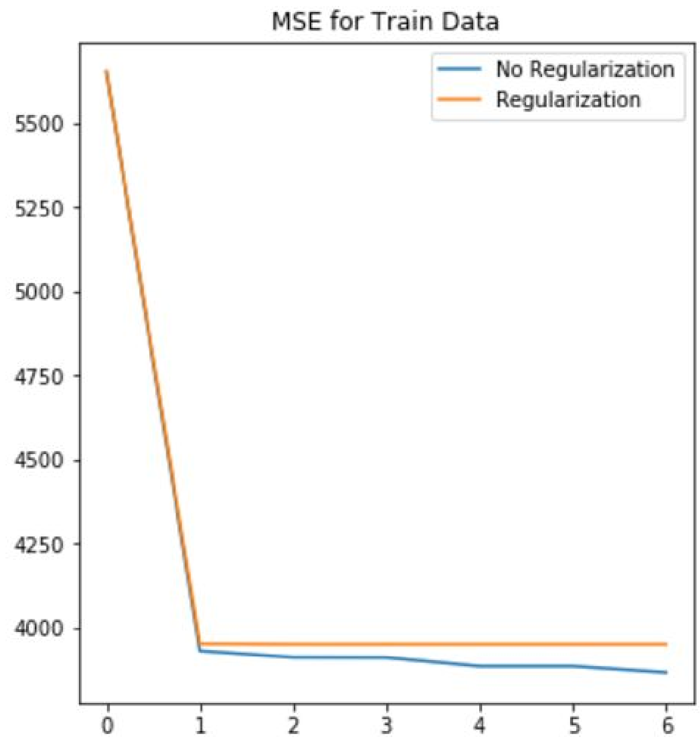

Figure 8: MSE for train data

Above plots show the MSE values obtained for different order of polynomials for the input values for training data. It shows two different plots one with Regularization and other without Regularization.

For the above plot we have used the optimal value of lambda obtained in problem 3. which is '0'

From the above plot we can say that error value decreases as we increase the degree of polynomial.

For without regularization model, we obtain minimal error at p=6, But, we should know that cost of computation increases as degree of polynomial increases.

For model with regularization, we obtain optimal value, at p=1 i.e. at linear regression model itself.

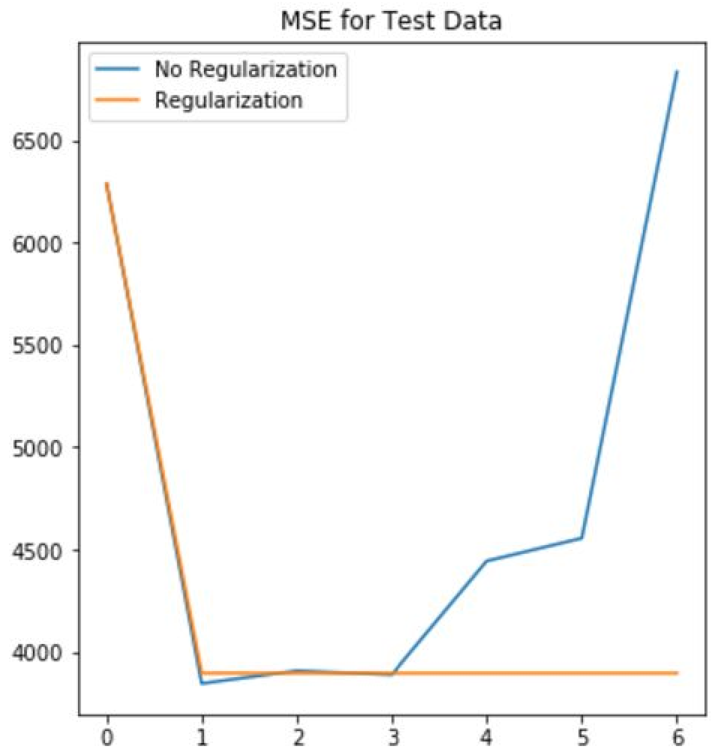

Figure 9: MSE for Test data

Above plots show the MSE values obtained for different order of polynomials for the input values for test data. It also shows two different plots one with Regularization and other without Regularization.

Here, based on the plots of train and test data for model without regression, we can see the case of overfitting the model. Because, at p =6 we got optimal value for training data, but on the test data for the same value of p we are getting very high error value.

This, is not the case with model having regularization implementation. In this case, we get optimal value at p=6 but MSE is almost the same for range of p from 1 to 6. So, by considering the cost of computation involved with higher order of polynomials, we can simply consider p =1 to be our optimal solution.

Thus, we can conclude that, model with regularization(lambda=optimal lambda obtained in problem3) will give better results as compared to one without regularization(lambda=0).

Problem 6: Interpreting Results

6.1 Comparison of various approaches

Regression Method	Test MSE	Train MSE
Linear Regression Without Intercept	106775.36155789	19099.44684457
Linear Regression With Intercept	3707.84018132	2187.16029493
Ridge Regression(optimal)	2851.33021344	2187.16029493
Gradient Descent(optimal)	3584.14496815	2212.12005345
Non Linear Method Without Regularization for optimal p	3845.03473017	3866.88344945
Non Linear Method With Regularization for optimal p	3895.58	3950.68

Table 6: Comparison of MSE for different regression methods

In order to choose the best using regression for predicting diabetes level, we can consider the accuracy metrics. The most accurate technique will be the one which produces the minimum mean squared error.

We can observe from the table above that Ridge Regression and Gradient Descent produces the best results. But as discussed earlier, using Ridge Regression is not suitable in cases when finding the inverse of the matrix is difficult. Hence in those cases, we can choose Gradient Descent.

Also, as we can see from the table, the mean squared error is maximum in case of linear regression without intercept, hence that will be least preferred method for predicting diabetes level.

But again, we can't suggest best setting solely on the basis of MSE metrics. Because, if we go solely by MSE metrics then it will give 'Non linear model without regularization' as best fit. But it is not true as we explained in problem 5, that this is not the case of best fit but the case of overfit.