Assignment-1
Report

Pankaj Kumar
2019262

Q1 In **Question 1**, I have implemented linear Regression

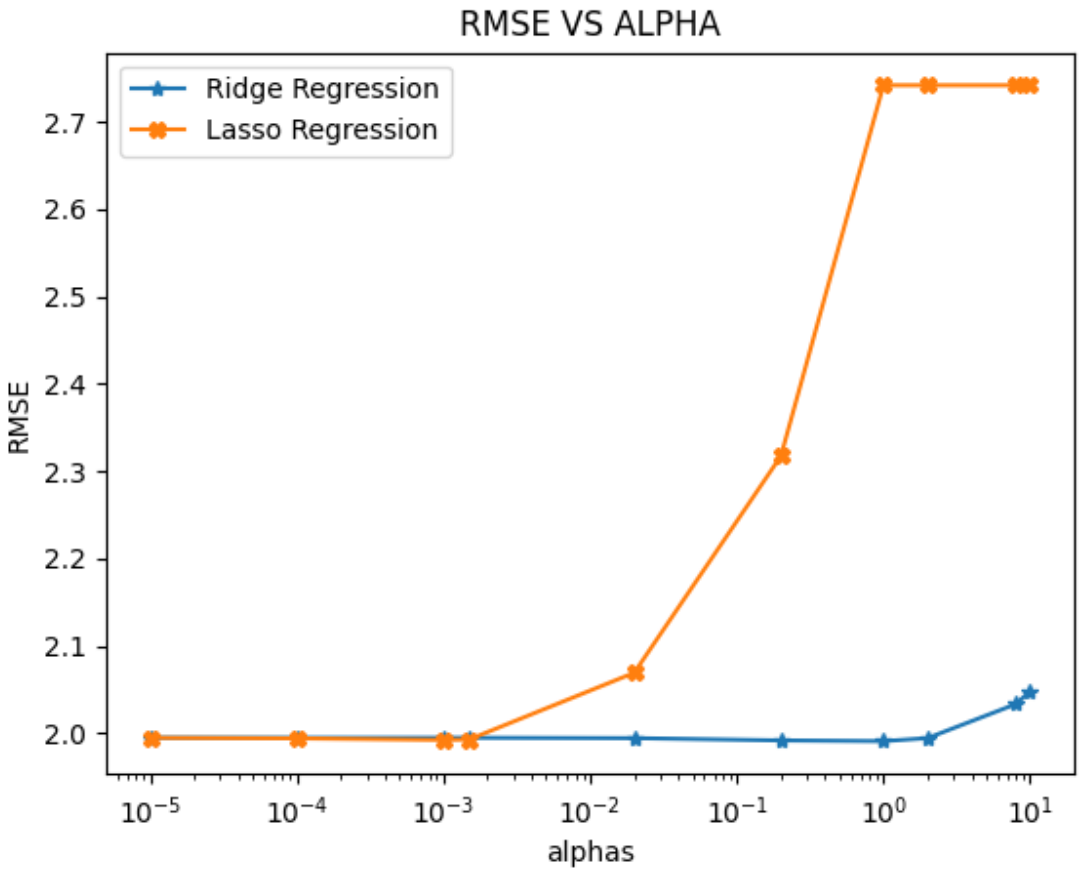
- First I have loaded the Data set as a global
- Then I have pre-processed the data where I replaced the variable as M->1, F->2, I->3
  And returning the X having all rows but column-1 and Y as all rows and last column value and setting the random seed to 0 for no duplication.
- Then I split the dataset into test and train
- Then I am finding the thetas using the gradient descent on the training dataset
- And using that thetas I am predicting the training and test data result
- And then I am finding the error using the RMSE

```
[ 6.0007302  -0.49682318  3.60964062  3.41228451  1.86449253  2.7047675
 -5.47335997 -0.21353216  5.17930516]
Train RMSE =  2.4944261820577593  Test RMSE =  2.132948408019076
```

In **Question 1b**, I am plotting the graph showing the effect of value on the testing data's RMSE

For both Ridge and Lasso Regression and for Ridge and Lasso regression I am using the sklearn's library to implement it.

- Here I have take the alpha values as [1e-05,0.0001,0.001,0.0015, 0.02,0.2,1,2,8,10]
- Now for each alpha that belongs to the above values, I am fitting the training data for both Ridge and Lasso Regression
- Then I am predicting the test data for both ridge and lasso
- Now for each case, I am calculating the RMSE and Coefficient (Intercept, Coefficient) for both Ridge as well as Lasso
- Then I have to find the minimum RMSE error and print its corresponding alpha and coefficient, the same goes for lasso Regression as well.
- After calculating all these values I have plotted the graph RMSE vs Alpha for both Ridge as well as Lasso Regression.
- And we can check the best alpha from the graph.




```
Best Ridge alpha would be :  1
Best Ridge Coef would be :  [  4.45386088  -0.39589125   2.31092303   7.91316727   7.13274689
   7.04046292 -17.76128733  -6.56347534  11.04776136]
Best Lasso alpha would be :  0.001
Best Lasso Coef would be :  [  4.26650567  -0.38202805   0.         11.235931     8.32045844
   8.5550278  -19.78870552  -8.65524114   9.78886033]
```

In **Question 1c**, I have found the best alpha value and model coefficient for both Ridge and Lasso regression using Sklearn's Grid Search Function having the same alpha value as in Q1,b

- For each alpha in the list, I have used GridSearchCV to find the grid of lasso and ridge.
- After finding the grid of Ridge and Lasso, I'm printing the best alpha among the list and best model coefficients (Intercepts and Coefficient)

```
Best Ridge alpha would be :  1
Best Rdige Coef would be :  [  4.42465431  -0.40405838   2.43139224   7.58801632   8.26251436
   7.14700751 -17.82383629  -6.83075109  10.43930842]
Best Lasso alpha would be :  0.0015
Best Lasso Coef would be :  [  4.45452358  -0.39478179   0.         10.41111624   8.92782902
   7.89465262 -18.95119982  -7.28050993   9.80482372]
```

Q2 In **Question 2**, I have implemented the Logistic Regression

- First I have loaded the Dataset as in global
- Then I have pre-processed the data where I have replaced the values of 0 to its median in Columns ‘Pregnancies’, ‘Glucose’,’BloodPressure’,’ SkinThickness’,’ Insulin’,’ BMI’
  And then return X having all the rows and column-1 and Y having all the rows and last column.

```
Confusion Matrix SGD :  [[89 10]
 [27 28]]
Accuracy SGD :  0.7597402597402597
Precisiom SGD :  0.7368421052631579
Recall SGD :  0.509090909090909
F1 SGD :  0.6021505376344085
```

2b We saw that smaller value of learning rate takes time to converge.
On other hand for Higher learning rate ,model/cost overshoots
We have to choose the learning rate in accordance with the cost function (as it overshoots for some cases)as it depend on the rate of cost function
The learning rate is directly proportional to speed the minimization of working of the cost function.(but sometime overshoots there we should picks the best learning rate by trying the different value)

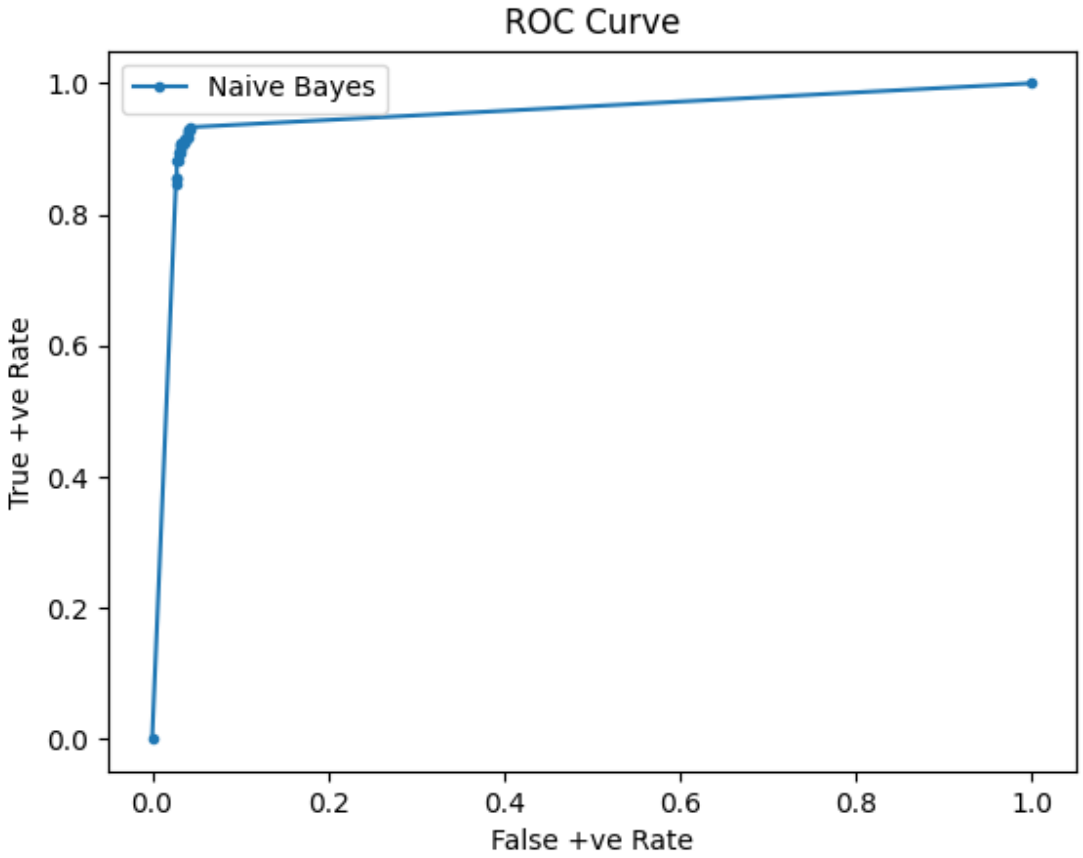
2c

```
Confusion Matrix :  [[89 10]
 [27 28]]
Accuracy :  0.7597402597402597
Precisiom :  0.7368421052631579
Recall :  0.509090909090909
F1 :  0.6021505376344085
```

Q3 For Pre-processing we have to extract the dataset for trouser and pullover and then binarizes the dataset

c)In Question 3c, I have plot the ROC curve, confusion matrix, and accuracy, precision, recall using inbuilt sklearn's library. After preprocessing the data, I have Gaussian fit the training data. Then I have calculated the precision, recall, thresholds using the precision_recall_curve and finds the confusion matrix.
Then using the probability of ns and lr. I have plot the graph for the same

```
0.911 0.973
[[973  27]
 [151 849]]
```



Q4

1 a We have to add the parameter and after manipulating we have to run the model for the both output dataset if the output of both data set are different we can say that it is suspicious.

C Wi = β0 +β1Xi +ui we can add the feature

2

L2 regularization encourages small coefficients as it tries to penalizes the count of non-zero coefficients values therefore if we increase the coefficients this will only happen when we have sufficient gain in fitting the data i,e its ability. As this situation convert problem from convex to non-convex.

Thus we can say that it promotes smaller coefficients.

3

Refrence

[Difference between Batch Gradient Descent and Stochastic Gradient Descent | by Aerin Kim]
[What is the difference between Gradient Descent and Stochastic Gradient Descent?]
[Precision-Recall Curves. Sometimes a curve is worth a thousand… | by Doug Steen]
[Gradient Descent in Python. When you venture into machine learning… | by Sagar Mainkar]