SAMET ÖZGÜL

705201014

BBL536E Data Science

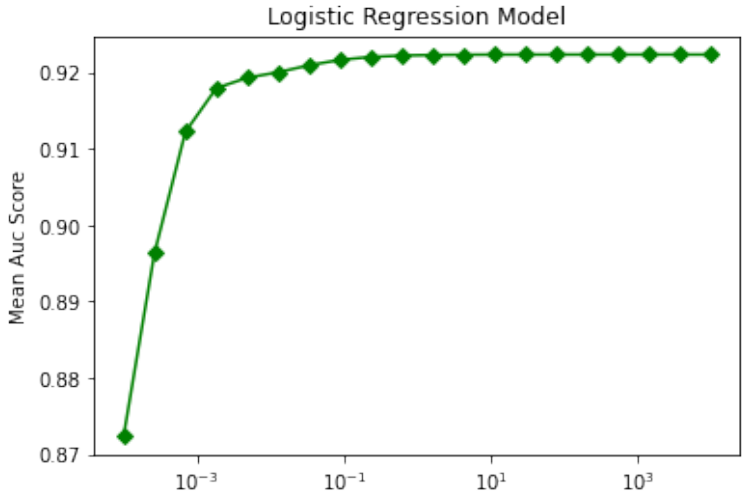
Question 1

Firstly, I read data from file using pandas read_excel command. The preprocessing steps are perform one by one. The data is splitted X, y_1 and y_1. X represents our features, y_1 and y_2 represent outputs. I create parameters for gridsearhcv to find best parameters. After find best parameters with loops, cross_val_score helps to find mean and standart deviation values. The result is close enough the given in question.

-	Mean Absolute Error	*	Mean Square Error	/
Output	RandomForest	RidgeRegression	RandomForest	RidgeRegression
Y1	0.3264245017186668±0.04777957958275208	2.0684062874150566±0.24478405555575394	0.24860697545245825±0.10011504913900393	8.589288373986665±1.791267376161846
Y2	1.0078042905869906±0.1584577568858271	2.2830741086533677±0.2724593537075571	2.7932144110587758±0.8433308541147456	10.481908726214302±2.523616365288306

Question 2

The result is close like in homework. The algorithm use 20 different C for logistic regression. In this graph, increasing of C value gives best results. The auc score is came from cross_val_score.



Firstly, I calculate average classification score for logistic regression with C=1. The result is close the given result. Also I get score for neural network. The model parameter come from task-3.

```
Best parameters of Random regressor {'max_depth': 150, 'min_samples_leaf': 3, 'min_samples_split': 2, 'n_estimators': 1000}
Best parameters for Neural Network {'alpha': 0.1, 'hidden_layer_sizes': (10, 10, 10, 10, 10)}
classification report for logistic regression
              precision    recall  f1-score   support

           0       0.93      0.98      0.95     29269
           1       0.67      0.38      0.49      3681

    accuracy                           0.91     32950
   macro avg       0.80      0.68      0.72     32950
weighted avg       0.90      0.91      0.90     32950

classification report for neural network
              precision    recall  f1-score   support

           0       0.94      0.96      0.95     29269
           1       0.62      0.55      0.58      3681

    accuracy                           0.91     32950
   macro avg       0.78      0.75      0.77     32950
weighted avg       0.91      0.91      0.91     32950

classification report for Random forest
              precision    recall  f1-score   support

           0       0.94      0.97      0.95     29269
           1       0.65      0.48      0.55      3681

    accuracy                           0.91     32950
   macro avg       0.79      0.72      0.75     32950
weighted avg       0.90      0.91      0.91     32950
```