A20474413
Jorge Gonzalez Lopez

Assignment 5

S484 Spring 2021
April 27, 2021

Question 1

a)

- **Answer**

```
0.2471959533758522
```

- **Explanation**

  When training the model using the Adaptive Boosting Technique, the weights for the training are initialized to 1s and used to train the model successively.

b)

- **Answer**

```
0.1601657094959913
```

- **Explanation**

  With each iteration the weights of the model are updated, and the model generally keeps getting better.

c)

- **Answer**

```
8.744448964925766e-08
```

- **Explanation**

  The model converges at iteration 17 as it reaches the accuracy threshold stablished.

d)

- **Answer**

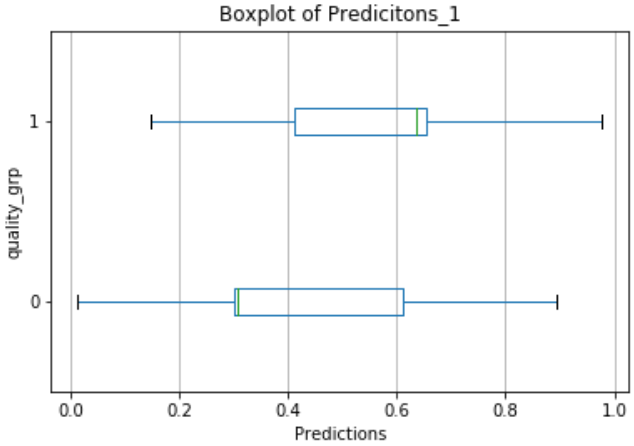
```
AUC = 0.5217044935894776
```

- **Explanation**

  To compute the AUC of the test dataset, the ensemble Predictions Probabilities of the test data (computed during the boosting of the model) are used.

e)

- **Figure**



Question 2

a)

- **Answer**

The input features are:
['alcohol', 'citric_acid', 'free_sulfur_dioxide', 'residual_sugar', 'sulphates']

- **Explanation**

The Forward selection method has been used. Therefore, by selecting the features with the lowest p value (if it is less than 0.05), all the features have to be used to train the model.

	0	1	2	3	4	5	6
0	0	Intercept	1	-2250.999215	None	None	None
0	1	Intercept + alcohol	2	-1901.627969	698.742	1	5.61252e-154
0	1	Intercept + citric_acid	2	-2244.669505	12.6594	1	0.000373678
0	1	Intercept + free_sulfur_dioxide	2	-2250.791593	0.415245	1	0.519319
0	1	Intercept + residual_sugar	2	-2238.436637	25.1252	1	5.37273e-07
0	1	Intercept + sulphates	2	-2248.194233	5.60996	1	0.0178586
0	2	Intercept + alcohol + citric_acid	3	-1894.249813	713.499	2	1.16333e-155
0	2	Intercept + alcohol + free_sulfur_dioxide	3	-1887.539331	726.92	2	1.41703e-158
0	2	Intercept + alcohol + residual_sugar	3	-1894.706814	712.585	2	1.83728e-155
0	2	Intercept + alcohol + sulphates	3	-1896.740118	708.518	2	1.40355e-154
0	3	Intercept + alcohol + free_sulfur_dioxide + ci...	4	-1881.252688	739.493	3	5.7295e-160
0	3	Intercept + alcohol + free_sulfur_dioxide + re...	4	-1884.719246	732.56	3	1.82635e-158
0	3	Intercept + alcohol + free_sulfur_dioxide + su...	4	-1878.849457	744.3	3	5.19768e-161
0	5	citric_acidIntercept + alcohol + free_sulfur_d...	5	-1873.631344	754.736	4	4.88799e-162
0	5	residual_sugarIntercept + alcohol + free_sulfu...	5	-1874.122122	753.754	4	7.97459e-162
0	6	citric_acidIntercept + alcohol + free_sulfur_d...	6	-1869.823683	762.351	5	1.61187e-162

b)

- **Answer**

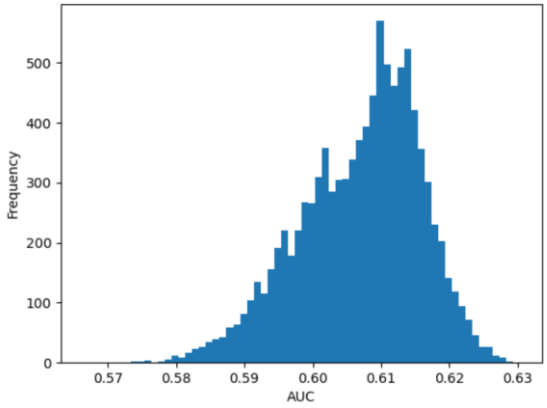
```
AUC = 0.606763204846272
```

- **Explanation**

It has been computed with the roc_auc_score function over the y_test and the predictions that the model returned of X_test.

c)

- **Figure**



- **Explanation**

To generate the histogram, a logistic regression model has been trained with 10 000 bootstrap samples and the auc has been computed and stored for all of them.

d)

- **Answer**

```
2.5th percentile:  0.5878594249201278
97.5th percentile:  0.6217501348491763
```

e)

- **Explanation**

Due to the fact that the value 0.5 does not fall within the confidence limits ([0.558, 0.653]), then it can be concluded that the AUC of the testing data is significatively different from 0.5 and therefore, be confident that the model is not the random model.