A20474413

Jorge Gonzalez Lopez **Assignment 5**

Question 1

S484 Spring 2021

April 27, 2021

# 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.

# Answer

0.1601657094959913

# Explanation

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

# Answer

8.744448964925766e-08

# Explanation

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

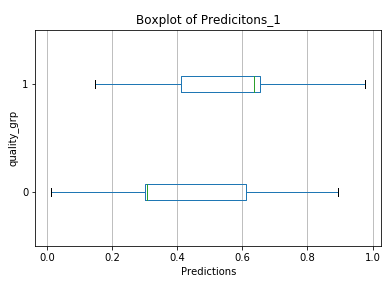
# Answer

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.

# Figure



Question 2

# 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.

# 

# 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.

# 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.

# Answer

2.5th percentile: 0.5878594249201278

97.5th percentile: 0.6217501348491763

# 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.

# 