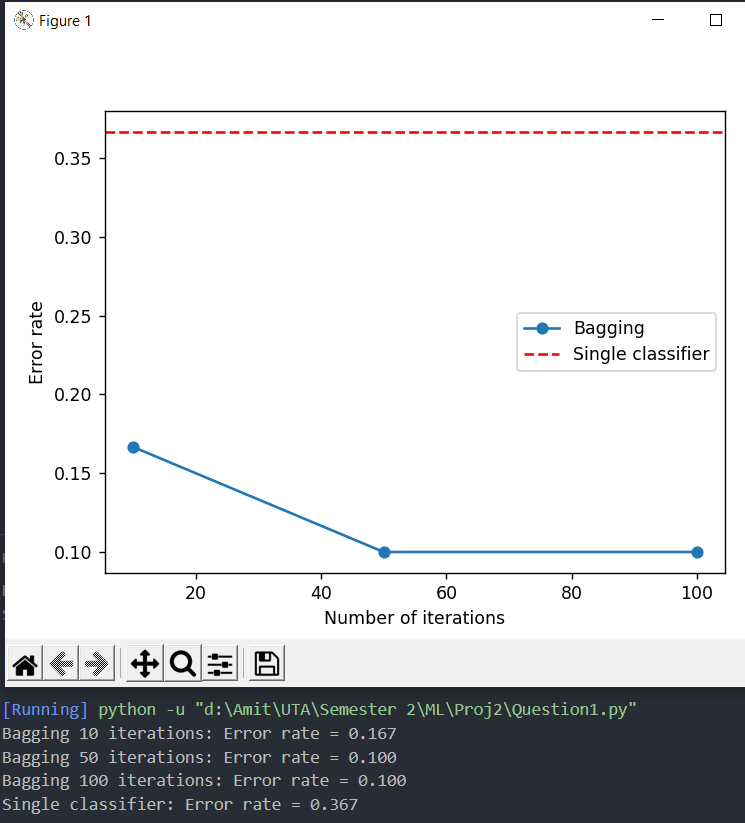
**Machine Learning : CSE 6363 Project 2**

Student Name : Amit Munna Gupta

Student ID : 1002066302

Q1

Bagging, also known as bootstrap aggregation, is the ensemble learning method that is commonly used to reduce variance within a noisy dataset. In bagging, a random sample of data in a training set is selected with replacement—meaning that the individual data points can be chosen more than once.

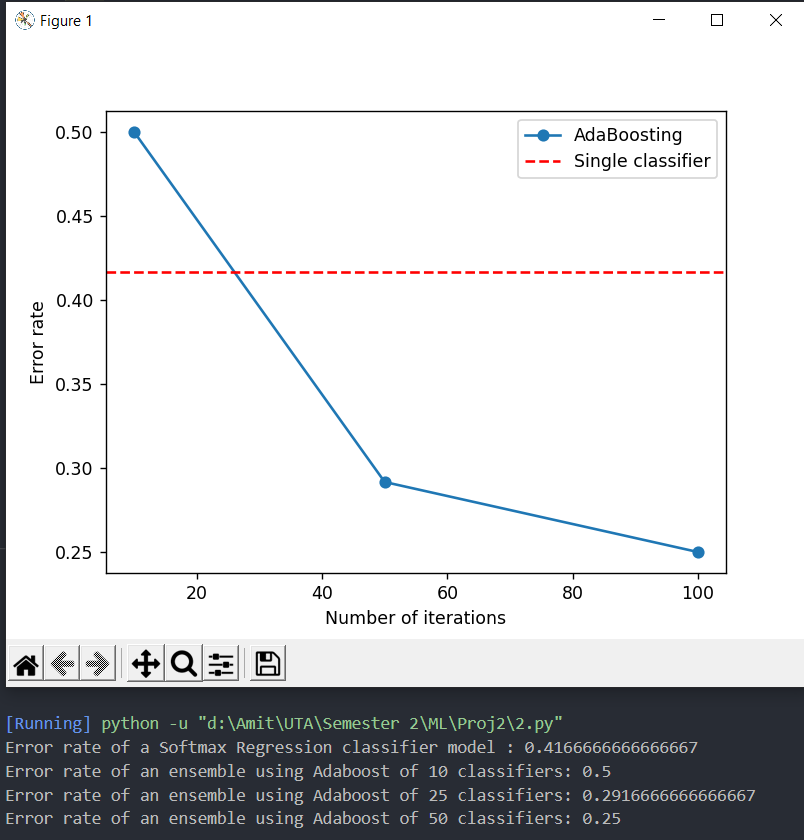


Observations:

* The single classifier has the highest error rate of 0.367, indicating that it performs the worst among all the models listed here.
* As the number of iterations in bagging increases, the error rate decreases. The bagging with 10 iterations has an error rate of 0.167, which is lower than the single classifier.
* The bagging with 50 and 100 iterations have even lower error rates of 0.100, indicating that they perform the best among all the models listed here.

Q2

Adaboost is a technique in Machine Learning used as an Ensemble Method.

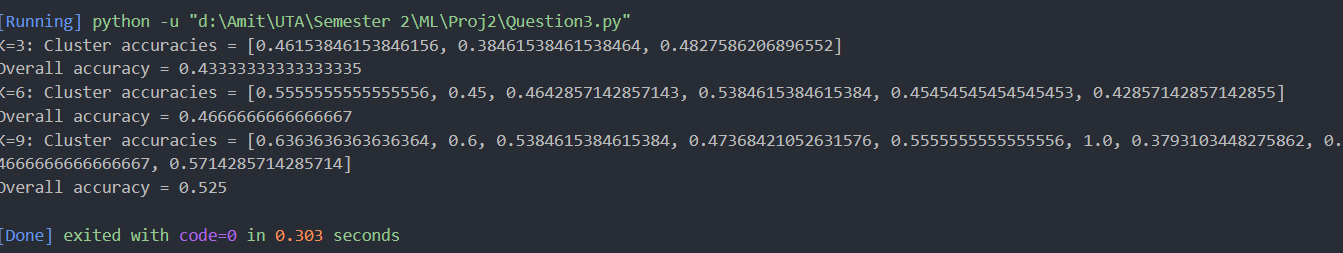


Observations:

* The Softmax Regression classifier model has an error rate of 0.4167, which means that it misclassifies 41.67% of the instances in the dataset.
* The Adaboost ensemble of 10 classifiers has a higher error rate of 0.5, indicating that it performs worse than the Softmax Regression model.
* As the number of classifiers in the Adaboost ensemble increases, the error rate decreases. The ensemble of 25 classifiers has an error rate of 0.2917, which is lower than both the Softmax Regression model and the Adaboost ensemble of 10 classifiers. The ensemble of 50 classifiers has an even lower error rate of 0.25, indicating that it performs the best among all the models listed here.
* The improvement in performance with the increasing number of classifiers in the Adaboost ensemble suggests that the base classifiers used in the ensemble are diverse enough to contribute to the overall performance of the model. The ensemble approach helps to combine the strengths of individual classifiers and reduce their weaknesses, resulting in a more robust and accurate model.

*Comparing bagging and Adaboost for* ***my dataset****, we can observe that the bagging approach seems to have a greater impact on performance improvement with increasing iterations, as seen by the larger decrease in error rate from 10 to 50 iterations compared to the Adaboost ensemble from 10 to 25 classifiers.*

Q3



Observations:

* The overall accuracy for all the values of k is around 0.4-0.5, which is not very high.
* For k=3, the cluster accuracies are relatively close to each other, with the highest accuracy being 0.48. However, the overall accuracy is still only 0.43.
* For k=6 and k=9, there are some clusters with higher accuracies (up to 0.57), but there are also clusters with lower accuracies (as low as 0.37). The overall accuracy for k=9 is slightly higher than for k=3 and k=6.
* Overall, the results suggest that the dataset may not have clear boundaries between the different classes, making it difficult for the K-means algorithm to accurately classify the data points into different clusters.