

IE615: Data Analytics for Operations Research

Assignment 3

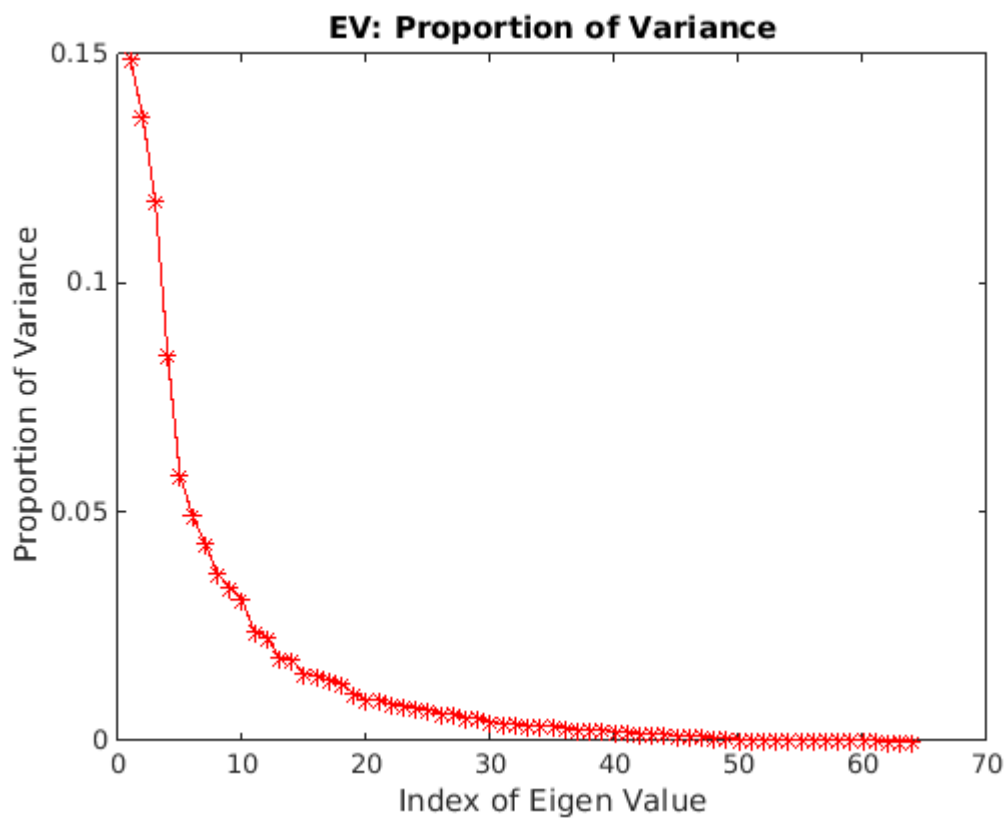
Arun R
163190013

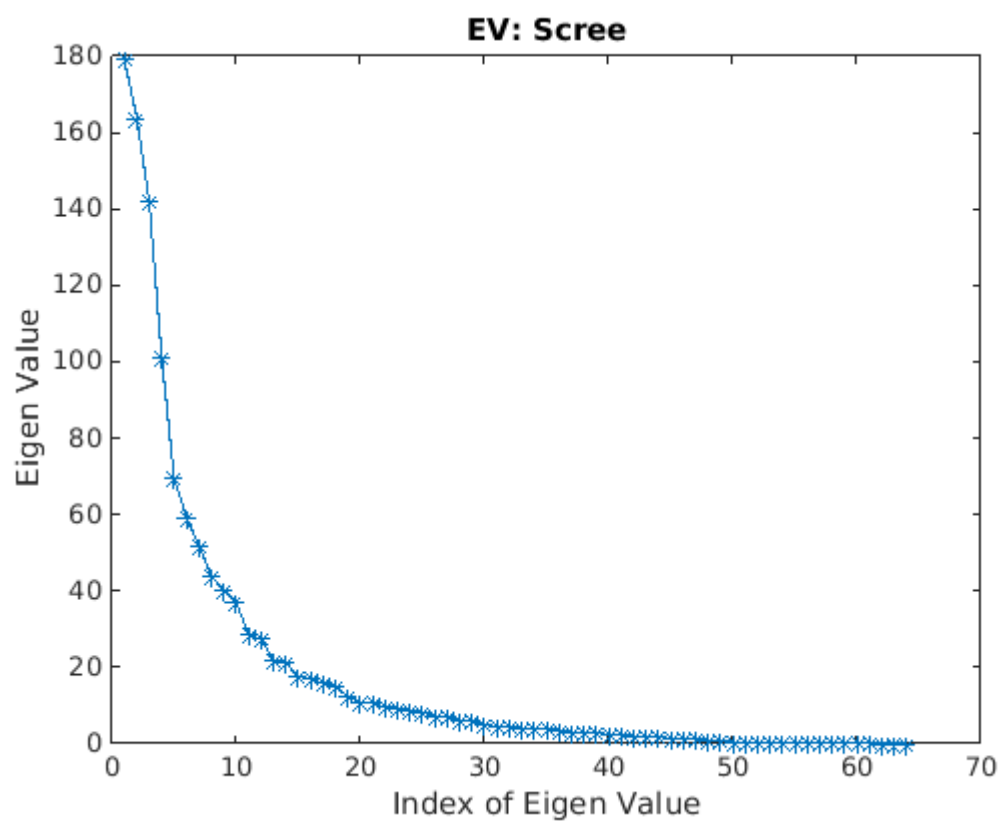
In this report, I will explain how to run the code and show some of the results.. There are four folders (one folder for each question) in the zip file that I have submitted.

2) Question 2: See Folder “Question 2”

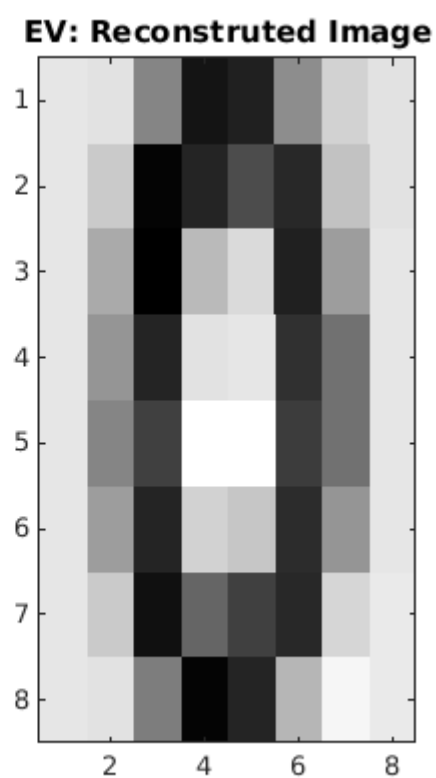
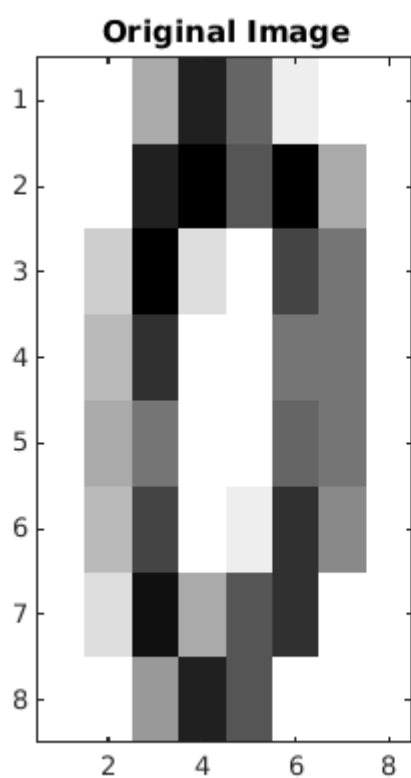
The folder has two matlab files. Running “PCA_EV.m” gives the output of PCA with eigen values. Running “PCA_SVD.m” gives the output of PCA with SVD. Since the principal components are same in SVD and EV, the output from both the programs are same.

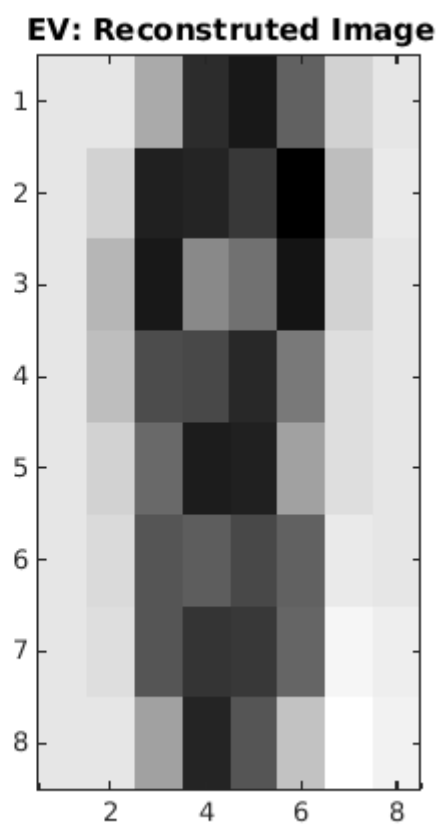
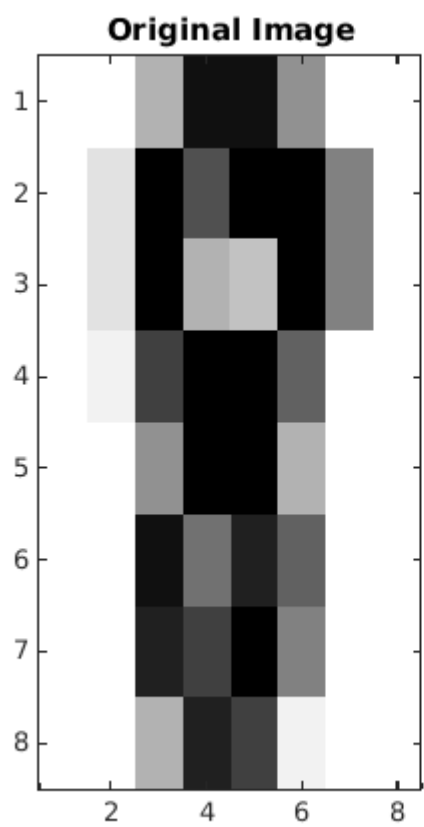
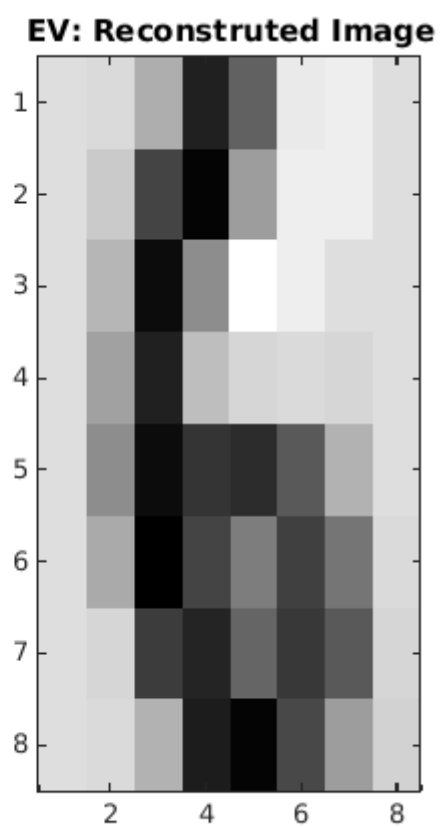
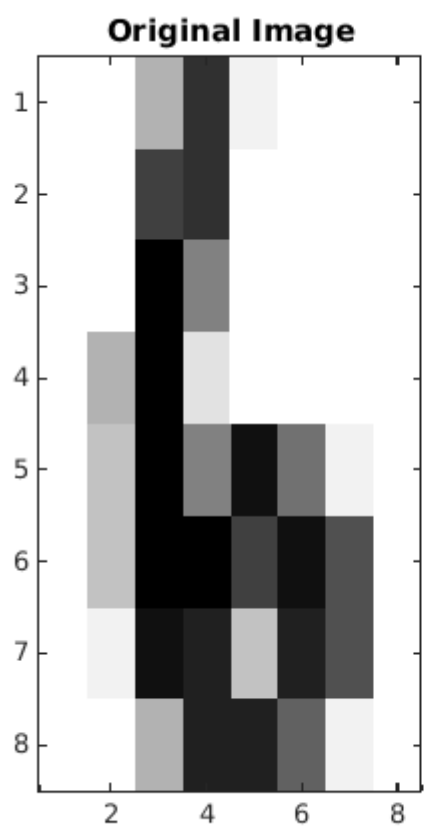
SCREE and proportion of variance plot is shown in both programs. To view the original image and reconstructed image, enter the serial number of data when prompted.





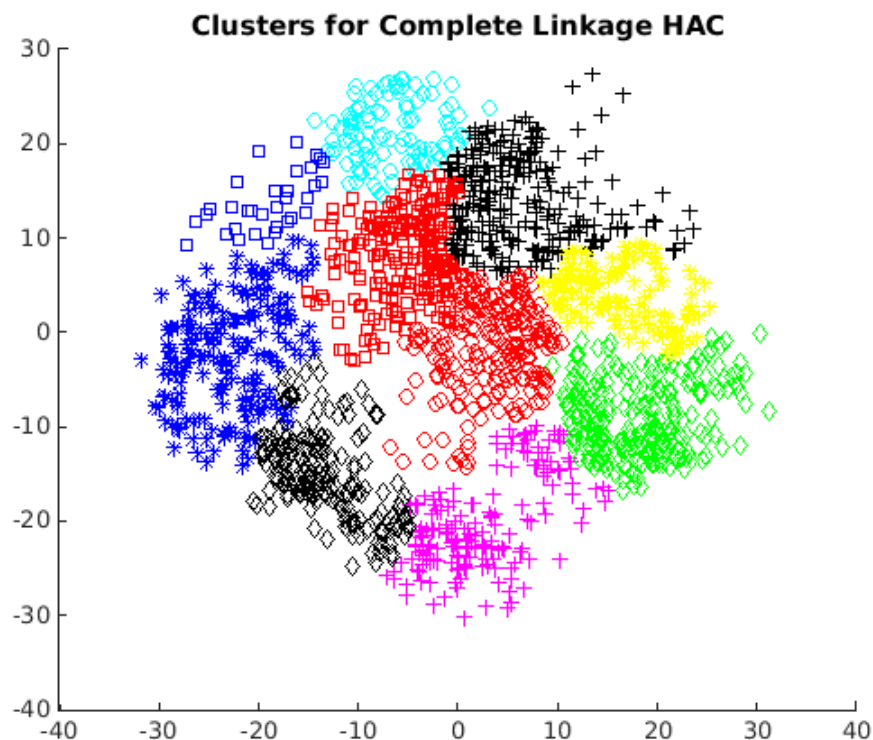
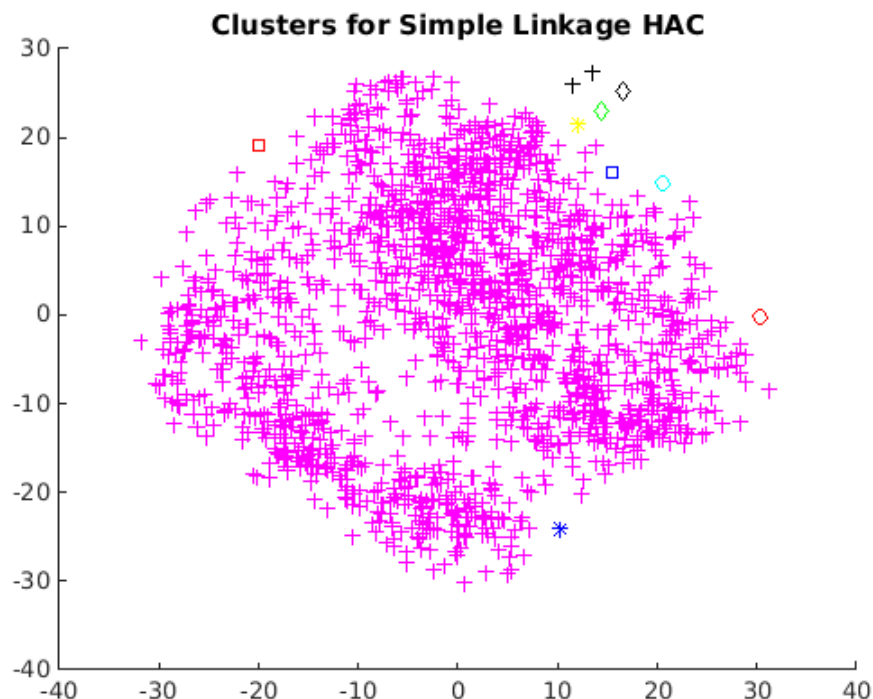
Some images:

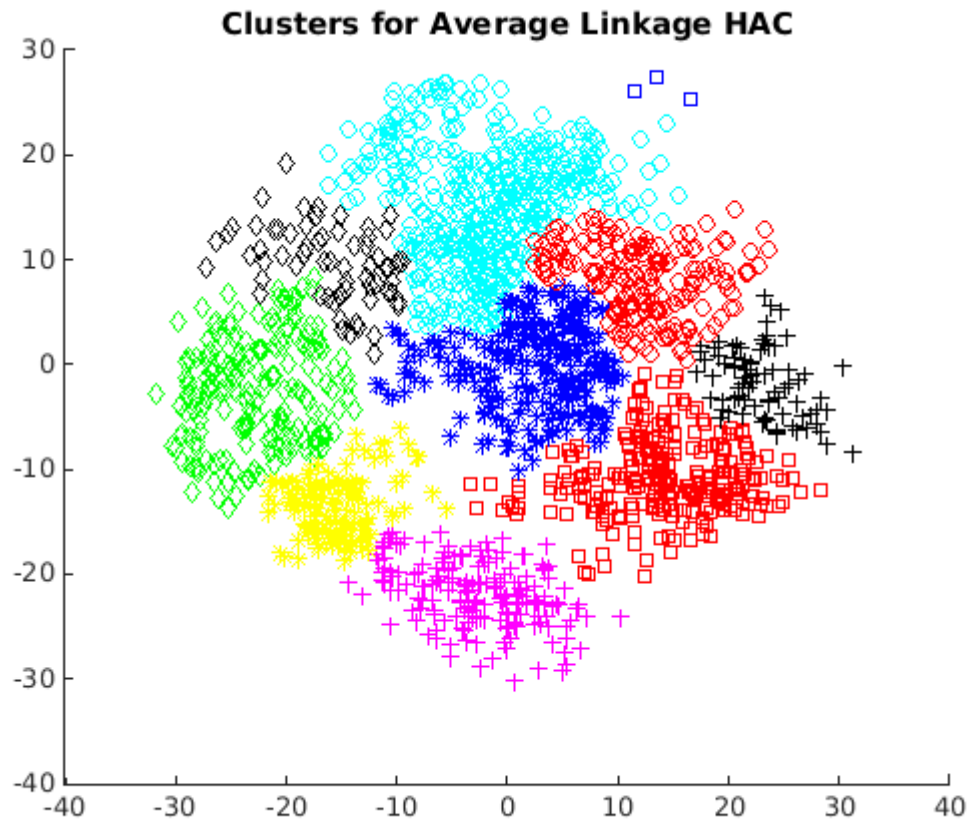




3) Question 3: See Folder “Question 3”

The folder has four matlab files. “Dendrogram.m” is the main folder that has to be executed. The program will ask you to enter a number (1 for simple linkage, 2 for complete linkage and 3 for average linkage) for the linkage function that has to be used in HAC. “simpleLinkage.m”, “completeLinkage.m” and “averageLinkage.m” are the subroutines used to compute the three linkage functions. After execution, the clusters will be displayed.



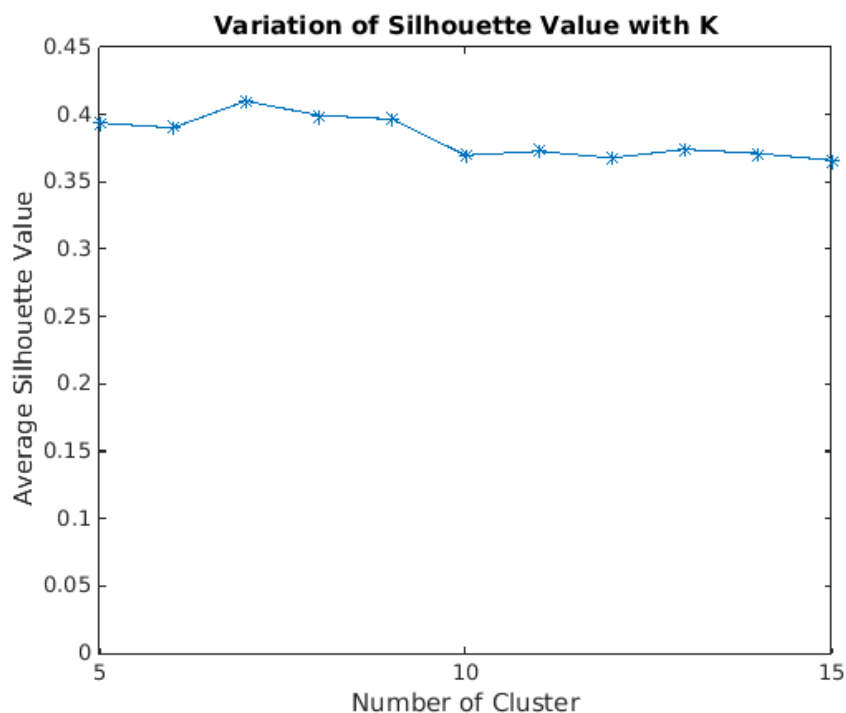


Note: The clusters represented by blue star and blue square are different. Similarly clusters represented with black diamond and black plus are different, and so are the clusters represented by red circle and red square.

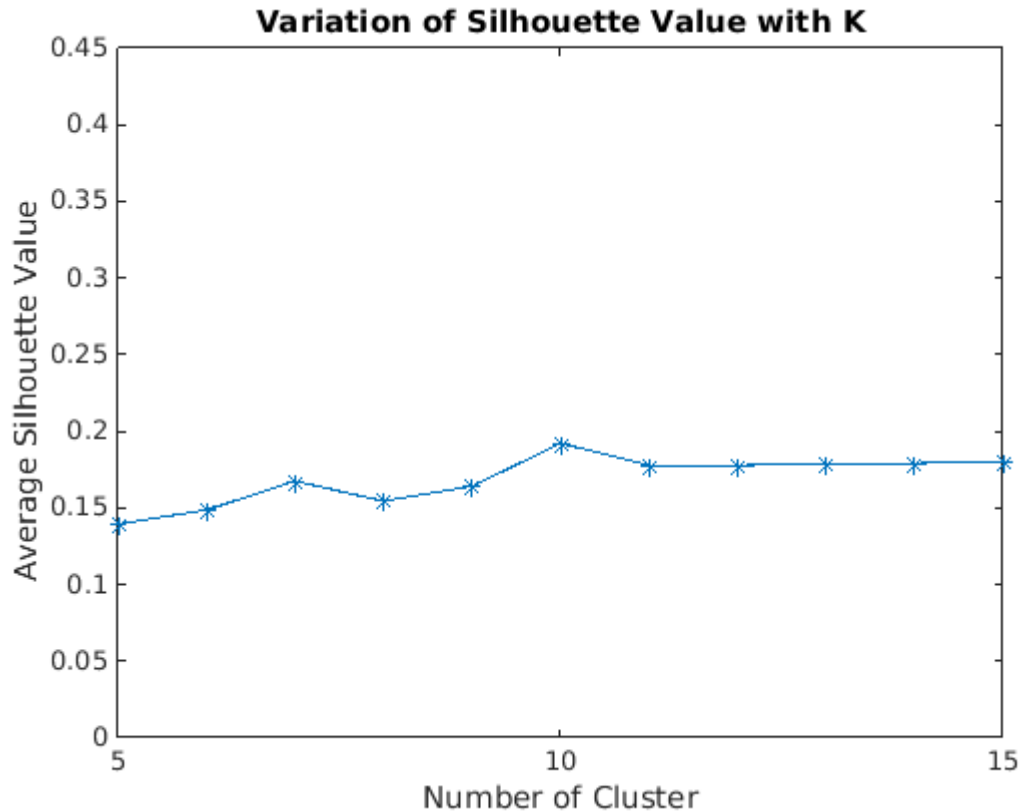
4) Question 4: See Folder “Question 4”

The folder has one matlab file. On executing this program, you can see the graph showing the variation of silhouette value with number of clusters (K).

Silhouette obtained using 2 principal components:



Silhouette obtained using all 64 attributes:



1) Question 1: See Folder “Question 1”

The folder has three matlab files. Run “AdaBoost.m” to get the output of AdaBoost and run “AlphaBoost.m” to get the output for AlphaBoost. The third file named “WeakLearner.m” is used as a subroutine in AdaBoost and AlphaBoost.

Note: Irrespective of T and alpha my code always gives the prediction as -1 to all data. Reason: The first weak learner predicts -1 for all data, and it has the lowest epsilon (in other words, highest weightage). So, this weak learner dominates all the weak learners that are learned subsequently. Hence, the ensemble learner gives the same prediction as the first weak learner. For this ensemble learner,

Number of True Positive: 0
Number of True Negative: 500
Number of False Positive: 0
Number of False Negative: 268
Prediction = undefined
Recall = 0
F1-score = undefined