A2) Files used-
bayes.py
gaussian.py

The predicted classes are

7% - Regular
10%- Spam
19%- Spam

A3) Files used -
em.py
gaussian.py

The EM algorithm was executed for 300 to 10K iterations and the cluster mean was found to stabilize around 3K iterations at

{**0**: 21.74, **1**: 14.58, **2**: 4.95}

Results

Cluster 0
[19.83, 20.06, 20.57, 20.96, 21.05, 22.1, 22.15, 22.37, 22.67, 23.46, 23.72, 24.21, 24.26, 26.02, 27.68]

Cluster 1
[9.62, 9.84, 10.15, 10.5, 10.71, 10.93, 11.63, 12.08, 12.37, 12.94, 13.04, 13.39, 13.57, 13.7, 14.52, 14.66, 14.68, 14.72, 14.85, 15.15, 15.23, 16.58, 16.79, 17.07, 17.12, 17.34, 17.7, 17.77, 17.97, 18.69, 18.81, 19.01, 19.22, 19.65]

Cluster 2
[0.64, 1.31, 1.43, 1.56, 1.91, 2.15, 2.19, 2.22, 2.32, 2.33, 2.78, 2.97, 3.07, 3.28, 3.67, 3.68, 3.89, 3.99, 4.26, 4.26, 4.57, 4.57, 4.68, 5.16, 5.2, 5.21, 5.32, 5.33, 5.36, 5.47, 5.51, 5.62, 5.72, 5.9, 6.17, 6.23, 6.51, 6.6, 6.65, 6.9, 6.91, 7.26, 7.29, 7.48, 7.54, 7.7, 7.92, 8.2, 8.42, 8.43, 8.6]

The elements of each cluster were fed into the first program to retrieve the cluster labels and it was found that Clusters 0 and 1 are **SPAM** whereas Cluster 2 **is REGULAR**

Only 3 elements in Cluster 2 were wrongly classified as SPAM as anything above CAPS letter percentage of 8.39 is classified as SPAM