Task3, student number: s1643102

Report for Gaussian classification

3.2

my_gaussian_system:	Elapsed time is 4.889569 seconds.
Number of test samples:	7800
Number of errors:	1219
Accuracy:	84.37%

Table 3.2

3.3

In my_improved_gaussian_classify I implemented Gaussian classification using K-means clustering. Firstly, I divided the train data into k different clusters, each cluster having a single cluster centre. I set the maximum number of iterations for finding the clusters to 50, which I find a reasonable number. After dividing the train data, each test sample is allocated to their nearest cluster (based on the distance from the cluster centre).

For each cluster I calculate the Gaussian distribution for every class restricted to that cluster. Essentially, I calculate the posterior probability of the test samples (from that cluster) belonging to the different classes. Each test sample is assigned to the class with the highest Gaussian probability for its cluster.

I apply k-means clustering for different number of clusters to obtain the most accurate implementation. The results are shown in Table 3.3.1.

Epsilon	Number of clusters	Number of test samples	Number of errors	Accuracy
0.01	2	7800	1067	86.32%
0.01	3	7800	1152	85.23%
0.01	4	7800	1216	84.41%
0.01	5	7800	1212	84.46%
0.01	6	7800	1202	84.59%
0.01	7	7800	1266	83.77%
0.01	8	7800	1283	83.55%
0.01	9	7800	1293	83.42%
0.01	10	7800	1352	82.67%

Table 3.3.1

We see that the highest accuracy is obtained for 2-means clustering. We have improved the accuracy by 1.95%.

I have also tried using different values for epsilon. The results are shown in Table 3.3.2.

Epsilon	Number of clusters	Number of test samples	Number of errors	Accuracy
0.02	2	7800	1026	86.85%
0.03	2	7800	1013	87.01%
0.04	2	7800	1020	86.92%
0.05	2	7800	1018	86.95%
0.06	2	7800	1032	86.77%
0.07	2	7800	1049	86.55%
0.08	2	7800	1052	86.51%
0.09	2	7800	1062	86.38%

Table 3.3.2

We see that the highest accuracy is obtained for 2-means clustering with epsilon = 0.03. We have managed to improve the accuracy by 2.64%.