EE 527: Machine Learning Laboratory

Assignment 8

Due date: 27 March 2023

[1.A] Download the Binary Classification dataset from the following URL

https://archive.ics.uci.edu/ml/datasets/MiniBooNE+particle+identification#

Perform Naïve-Bayes Classification by generating Likelihood $P(X|\omega)$ using **(a)** Single Gaussian, and **(b)** Gaussian Mixture Model (GMM)

Obtain **class-wise F1** scores for both cases. Additionally, vary the number of modes of the GMM and plot the class-wise F1 scores.

[1.B] Download the **Iris dataset** for multiclass classification and repeat the above experiment. Obtain **class-wise F1 Scores** for each class **and the Overall Accuracy.** vary the number of modes of the GMM and plot the overall accuracy.

[2] Application of K-Means Clustering in image segmentation. Consider the R-G-B values of each pixel of the input image as 3-dimensional feature vector. Initialize K-Means through data labels or cluster centroids. Perform K-means iterations till convergence and report the K cluster centroids. Revisit the Image and replace each pixel colour (R-G-B) with the nearest cluster centroid (rounded) color values. Repeat this experiment with different values of K and visualize the results.

[3.A] Write a function $generatePointFromRandomCluster(\ )$ that randomly generates a point in $\boldsymbol{R}_i$. The point must lie within any one of the following $N = 17$ circles $\{\boldsymbol{C}_i\ ;\ i = 1, \ldots N\}$ that are specified in the $(Center - X\ ,\ Center - Y,\ Radius)$ format. Any particular call to this function randomly chooses a circle and generates a point within it.

(0, 0, 10) ; (0, 50, 15) ; (50, 0, 15) ; (0,-50, 15) ; (-50, 0, 15) ; (35, 35, 15); (35,-35, 15) ; (-35, 35, 15) ; (-35,-35, 15) ; (0, 100, 20) ; (100, 0, 20) ; (0,-100, 20) ; (-100, 0, 20) ; (70, 70, 20) ; (70,-70, 20) ; (-70, 70, 20) ; (-70,-70, 20).

[3.B] The data point $x_i \in \boldsymbol{C}_i\ (i = 1 \ldots N)$ at each instant is obtained using the function $generatePointFromRandomCluster(\ \ )$. Perform incremental clustering with default variance $v_d = 10$ and Chebychev inequality threshold $\lambda = 3$. As the number of clusters increase beyond 50, drop the cluster with lowest weight value (π), so that at any given iteration, there are a maximum of $K = 50$ clusters. Plot the clusters as differently coloured ellipses after every 100 iterations.