DA311 Machine Learning Lab

Assignment 3

Date: August 22nd, 2023

(A) Given 600 training images of 3 characters in a folder named **TrainCharacters.rar**. Each of the 200 training images of each class is of size 128 × 128. You are given 300 test images (100 in each class) of size 128 × 128 in a separate folder **TestCharacters.rar**. (Please download the dataset from the following link: `https://github.com/tsharma12/IITG-DA311-Machine-Learning-Lab/tree/main/Week2_Dataset`)

Assume the samples to be generated from a multi-dimensional Gaussian distribution,

$$\mathcal{N}\left(\mu, \sum\right) = \frac{1}{(2\pi)^{n/2}|\sum|^{1/2}} \exp\left(-\frac{1}{2}(x-\mu)^T \sum^{-1}(x-\mu)\right), \tag{1}$$

having class-specific mean vectors $\boldsymbol{\mu_i}$. Implement the nearest mean classifier by considering the covariance matrix of each class to be the identity matrix. Report the individual character accuracy as well as the average accuracy for each of the models.

Note: If you happen to encounter memory storage issues during simulation, you may consider resizing the images to a more manageable size (say **32 × 32**) for the feature computation.

(B) Give 4 examples of images from the test set that are mis-classified by each classifier designed in **Question (A)**. You need to display both the true label and the predicted class for each image.

(C) Load the data.csv (Download) file and display using a scatter plot. Find the principal components and visualize them as vectors over the input data. Also, plot the projection of each data point onto the principal axes.

(D) Load the diabetes.csv (Download) file and reduce them to 2 dimensions using PCA. Also, visualize the samples along the two principal directions.