DA311 Machine Learning Lab

Assignment 2

Date: August 17th, 2023

(1) 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 $\boldsymbol{\mu_i}$. Consider each of the modeling schemes for computing the covariance matrix.

- (*i*) The samples of a given character class are modeled by a separate covariance matrix $\sum_{\boldsymbol{i}}$.
- (*ii*) The samples across all the characters are pooled to generate a common diagonal covariance matrix $\sum$.
- (*iii*) The samples of a given character are separately modeled by a diagonal covariance matrix $\sum_i$. The diagonal entries of the matrix correspond to the variances of the individual features. The features are assumed to be independent; hence their cross variances are forced to zero.

For each scenario above in (*i*), (*ii*), and (*iii*), build a generative Bayesian classifier using the training images and categorize the characters contained in the test folder. The mean and the covariance matrices are to be estimated from the training data using the Maximum Likelihood techniques, as discussed in class. Report the individual character accuracy as well as the averaged accuracy for each of the models.

Note: Employ the **128 × 128** pixel intensity values directly as features in this task. 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. However, note that to beat the curse of dimensionality, you must add a regularization term of the form $\boldsymbol{\lambda I}$ in the covariance matrix computation.