EE 527: Machine Learning Laboratory

Assignment 6

Due date: 24 Feb 2023

---

1. PCA & Eigen Faces

Consider the faces dataset, which consists of 15 subjects with 11 faces per subject. Set aside any 1 face per identity for testing and treat the rest of the 150 faces as the training set.

Perform PCA on the training data and visualize the top-K eigen faces. Transform the data (train & test) by projecting the mean centered samples onto these K vectors obtained using training set.

Find the best match for each of the test sample based on Euclidean distance criterion to the training vectors in the transformed K-dimensional space. Display both the true query face and the predicted face. Experiment with different values of K.

2. Linear Discriminant Analysis (LDA)

Generate $\mathbf{n}$ points each in two hyper-spheres in $\mathbf{d}$-dimensional space. The hyperspheres have a radius of $\mathbf{r}$ and are centered at $\mathbf{C_1}$ and $\mathbf{C_2}$. Collect these points into arrays $\mathbf{S_1}$ and $\mathbf{S_2}$ respectively. (Take $\mathbf{n}$=1000, $\mathbf{d}$=20)

(a) Compute centroids ($\mathbf{m_1}$, $\mathbf{m_2}$) and covariance matrices **(**$\mathbf{C_1}$, $\mathbf{C_2}$) of both arrays.

(b) Compute the optimal direction vector $\hat{\boldsymbol{w}}$ (unit vector) for **LDA**.

(c) Project the vector data in arrays $\mathbf{S_1}$ and $\mathbf{S_2}$ to generate the respective array of scalars $\mathbf{zS_1}$ and $\mathbf{zS_2}$.

(d) Plot the normalized histograms of $\mathbf{zS_1}$ and $\mathbf{zS_2}$ in two different colours (**red** & **blue**).

Carry out the above steps for each of the following cases by choosing $\mathbf{r}$, $\mathbf{C_1}$ & $\mathbf{C_2}$ appropriately and report observations.

Case 1: $||\mathbf{C_1} - \mathbf{C_2}||_2 > \mathbf{2r}$.

Case 2: $||\mathbf{C_1} - \mathbf{C_2}||_2 = \mathbf{2r} - \boldsymbol{\varepsilon}$.

Case 3: $\mathbf{r} < ||\mathbf{C_1} - \mathbf{C_2}||_2 < \mathbf{2r}$.

Case 4: $||\mathbf{C_1} - \mathbf{C_2}||_2 < \mathbf{r}$.