Assignment-2 (Due is on 20 th March 2019)

Dimensionality reduction techniques and logistic regression

PART-1 PCA

1. Given the multivariate Gaussian data with mean $\underline{0}$ and covariance $\mathbf{I}$, generation of multivariate Gaussian data with required mean vector and co-variance matrix.

$$\underline{m} = \begin{bmatrix} 0.5 & 0.5 \end{bmatrix} \qquad \mathbf{C} = \begin{bmatrix} 0.9 & 0.1 \\ 0.1 & 0.9 \end{bmatrix}$$

$$\mathbf{x} = \begin{bmatrix} 0.67 & 0.71 & 0.48 & 0.72 & 0.29 & 0.89 & -1.07 & -2.94 & 0.33 & 1.38 \\ -1.20 & 1.63 & 1.03 & -0.3 & -0.78 & -1.15 & -0.81 & 1.44 & -0.75 & -1.71 \end{bmatrix}$$

(a) Compute eigenvectors and eigenvalues of the matrix $\mathbf{C}$.

$\mathbf{E}=$

$\mathbf{D}=$

(b) Compute $\mathbf{y} = \mathbf{E} \times \mathbf{D}^{\frac{1}{2}} \times \mathbf{x}$

$\mathbf{y}=$

(c) Compute co-variance matrix (estimate) $\hat{\mathbf{C}} = E[\mathbf{y}.\mathbf{y}^T] - \mu_y.\mu_y^T$

$\hat{\mathbf{C}} =$

Observation : $\hat{\mathbf{C}} \simeq \mathbf{C}$

2. Representation of the following data using PCA basis.

$$\mathbf{x} = \begin{bmatrix} 0.59 & 1.67 & 1.42 & 1.11 & 1.64 & 0.39 & 1.42 & 1.43 & 1.45 & 1.45 & 0.70 & 1.21 \\ 0.59 & 1.73 & 1.10 & 1.09 & 1.79 & 0.66 & 1.52 & 1.52 & 1.54 & 1.54 & 0.81 & 1.31 \end{bmatrix}$$

(a) Compute the estimate of the covariance matrix (scatter matrix).
$\mathbf{C}_x = E[\mathbf{x}.\mathbf{x}^T] - \mu_x.\mu_x^T$

$\mathbf{C}_x =$

(b) Compute the eigenvalues and corresponding eigenvectors (PCA basis), where $\lambda_1 > \lambda_2$

$\lambda_1 =$		$\lambda_2 =$
$e_1 =$		$e_2 =$

(c) Construct the eigenmatrix: $\mathbf{E} = [e_1\ e_2]$

$\mathbf{E}=$

(d) Obtain the co-efficients of PCA basis: $\mathbf{y} = \begin{bmatrix} y_1 \\ y_2 \end{bmatrix} = E^T.\mathbf{x}$

$\mathbf{Y}=$

(e) Compute: $\hat{\underline{\mathbf{x}}}_1 = y_1\underline{e}_1 + y_2\underline{e}_2$ and $\hat{\underline{\mathbf{x}}}_2 = y_1\underline{e}_1$
Compare $\underline{x}, \hat{\underline{x}}_1 and \hat{\underline{x}}_2$, and comment.

$\underline{\mathbf{x}}=$	$\hat{\underline{\mathbf{x}}}_1 =$	$\hat{\underline{\mathbf{x}}}_2 =$
Comments:		

(f) i. Plot the points $\mathbf{x}$ on the graph sheet.
ii. Join the co-ordinate [0 0] and $\underline{e}_1$ to form the vector $\underline{v}_1$.(Line L_1)
iii. Join the co-ordinate [0 0] and $\underline{e}_2$ to form the vector $\underline{v}_2$.(Line L_2)
iv. Draw the perpendicular lines from the points to the line L_1.
v. Measure the distance from the origin to the points on the line L_1.

$\hat{\underline{y}}_1 =$

Compare y_1 and $\hat{y}_1$.

Comments:

(g) What are dimensionality reduced data?

(h) Distance matrix of the original data and the data obtained using projected data.

$D_1 =$	$D_2 =$

(OR) Load Image.mat and explore dimensionality reduction using PCA. Obtain the distance matrix of the original data and the data obtained using projected data.

PART-2 LDA

$$DATA_1 = \begin{bmatrix} 1.08 & 0.75 & 0.85 & 0.94 & 0.40 & 1.25 & 1.19 & 0.99 & 0.69 & 1.32 \\ 0.08 & -0.19 & -0.11 & 0.00 & -0.09 & -0.21 & 0.07 & 0.04 & -0.02 & 0.02 \end{bmatrix}$$

$$DATA_2 = \begin{bmatrix} 0.01 & -0.01 & 0.09 & -0.05 & -0.45 & 0.07 & -0.33 & -0.06 & -0.33 & -0.24 \\ 0.85 & 1.05 & 0.93 & 1.41 & 1.45 & 1.20 & 0.88 & 1.08 & 1.10 & 1.01 \end{bmatrix}$$

1. Compute the centroid of the individual classes, and overall centroid.

$C_1 =$	$C_2 =$	$C =$

2. Compute the between-class and within-class scatter matrices.

$$S_B = \frac{1}{2}\sum_{i=1}^{2}(C - C_i)(C - C_i)^T; \qquad S_W = \frac{1}{20}\sum_{i=1}^{2} n_i.S_i$$

where S_i is the scatter matrix of the i^{th} class. (Refer Part-1 (2))

3. Compute the eigenvalues and the eigenvectors corresponding to $S_W^{-1}.S_B$. (LDA Basis) with $\lambda_1 > \lambda_2$

$\lambda_1 =$	$\lambda_2 =$	$e_1 =$	$e_2 =$

4. Construct the matrix $\mathbf{E} = [e_1\ e_2]$

$\mathbf{E} =$

5. Obtain the co-efficients of the LDA basis: $\underline{\mathbf{y}} = \mathbf{E}^T.\underline{\mathbf{x}}$

$\mathbf{y} =$

6. Compute: $\hat{\underline{\mathbf{x}}}_1 = y_1\underline{e}_1 + y_2\underline{e}_2$ and $\hat{\underline{\mathbf{x}}}_2 = y_1\underline{e}_1$
Compare $\underline{x}, \hat{\underline{x}}_1 and \hat{\underline{x}}_2$, and comment.

$\underline{\mathbf{x}}$=	$\hat{\underline{\mathbf{x}}}_1 =$	$\hat{\underline{\mathbf{x}}}_2 =$
Comments:		

(a) Plot the points $\mathbf{x}_1$ (using blue) and $\mathbf{x}_2$ (using green) on the graph sheet.

(b) Join the co-ordinate [0 0] and $\underline{e}_1$ to form the vector $\underline{v}_1$.(Line L_1)

(c) Join the co-ordinate [0 0] and $\underline{e}_2$ to form the vector $\underline{v}_2$.(Line L_2)

(d) Draw the perpendicular lines from the points to the line L_1.

(e) Measure the distance from the origin to the points on the line L_1.

$\hat{\underline{y}}_1 =$

Compare y_1 and $\hat{y}_1$.

Comments:

7. What are dimensionality reduced data?

8. Distance matrix of the original data and the data obtained using projected data.

$D_1 =$	$D_2 =$

(OR) Load Image.mat. Each column is the data. There are 40 classes. First 10 columns belong to Class 1. Last 10 columns belong to Class 40. Explore dimensionality reduction using LDA.

PART-3 KLDA

$$DATA_1 = [\underline{x}_{11}\ \underline{x}_{12}] = \begin{bmatrix} 1.08 & 0.75 \\ 0.08 & -0.19 \end{bmatrix} \qquad DATA_2 = [\underline{x}_{21}\ \underline{x}_{22}] = \begin{bmatrix} 0.01 & 0.09 \\ 0.85 & 0.93 \end{bmatrix}$$

1. Compute the Gram-matrix using the Gaussian kernel function with variance=1.

Use: $K(\underline{\alpha}, \underline{\beta}) = exp(-\frac{(\underline{\alpha}-\underline{\beta})^T(\underline{\alpha}-\underline{\beta})}{2\sigma^2})$

$$\mathbf{G} = \begin{bmatrix} K(\underline{x}_{11}, \underline{x}_{11}) & K(\underline{x}_{12}, \underline{x}_{11}) & K(\underline{x}_{21}, \underline{x}_{11}) & K(\underline{x}_{22}, \underline{x}_{11}) \\ K(\underline{x}_{11}, \underline{x}_{12}) & K(\underline{x}_{12}, \underline{x}_{12}) & K(\underline{x}_{21}, \underline{x}_{12}) & K(\underline{x}_{22}, \underline{x}_{12}) \\ K(\underline{x}_{11}, \underline{x}_{21}) & K(\underline{x}_{12}, \underline{x}_{21}) & K(\underline{x}_{21}, \underline{x}_{21}) & K(\underline{x}_{22}, \underline{x}_{21}) \\ K(\underline{x}_{11}, \underline{x}_{22}) & K(\underline{x}_{12}, \underline{x}_{22}) & K(\underline{x}_{21}, \underline{x}_{22}) & K(\underline{x}_{22}, \underline{x}_{22}) \end{bmatrix}$$

	Class 1 data in HDS	Class 2 data in HDS
G=		

2. Compute the centroid of individual classes and the overall centroid.

$C_1 =$	$C_2 =$	$C =$

3. Compute the between-class and within-class scatter matrices.

$$S_B = \frac{1}{2}\sum_{i=1}^{2}(C - C_i)(C - C_i)^T; \qquad S_W = \frac{1}{20}\sum_{i=1}^{2} n_i.S_i$$

where S_i is the scatter matrix of the i^{th} class. (Refer Part-1 (2))

4. Compute the eigenvalues and the eigenvectors corresponding to $S_W^{-1}.S_B$. (KLDA Basis) with $\lambda_1 > \lambda_2 > \lambda_3 > \lambda_4$

$\lambda_1 =$	$\lambda_2 =$	$\lambda_3 =$	$\lambda_4 =$
$e_1 =$	$e_2 =$	$e_3 =$	$e_4 =$

5. Construct the matrix $\mathbf{E} = [e_1\ e_2\ e_3\ e_4]$

E=

6. Obtain the co-efficients of the KLDA basis: $\underline{\mathbf{y}} = \mathbf{E}^T.\underline{\mathbf{x}}$

y=

7. Compute $\hat{\underline{\mathbf{x}}}_1 = y_1\underline{e}_1 + y_2\underline{e}_2 + y_3\underline{e}_3 + y_4\underline{e}_4$, $\hat{\underline{\mathbf{x}}}_2 = y_1\underline{e}_1 + y_2\underline{e}_2 + y_3\underline{e}_3$, $\hat{\underline{\mathbf{x}}}_3 = y_1\underline{e}_1 + y_2\underline{e}_2$ and $\hat{\underline{\mathbf{x}}}_4 = y_1\underline{e}_1$

$\hat{\underline{\mathbf{x}}}_1 =$	$\hat{\underline{\mathbf{x}}}_2 =$	$\hat{\underline{\mathbf{x}}}_3 =$	$\hat{\underline{\mathbf{x}}}_4 =$
Comments:			

8. Obtain the projected data for the following data using KLDA basis (e_1 only)

$$DATA_1 = \begin{bmatrix} 1.08 & 0.75 & 0.85 & 0.94 & 0.40 & 1.25 & 1.19 & 0.99 & 0.69 & 1.32 \\ 0.08 & -0.19 & -0.11 & 0.01 & -0.09 & -0.21 & 0.07 & 0.04 & -0.02 & 0.02 \end{bmatrix}$$

$$DATA_2 = \begin{bmatrix} 0.01 & -0.01 & 0.09 & -0.05 & -0.45 & 0.07 & -0.33 & -0.06 & -0.33 & -0.24 \\ 0.85 & 1.05 & 0.93 & 1.41 & 1.45 & 1.20 & 0.88 & 1.08 & 1.10 & 1.01 \end{bmatrix}$$

(a) For the arbitrary vector $\underline{v} = \begin{bmatrix} v_1 \\ v_2 \end{bmatrix}$, compute $\begin{bmatrix} K(\underline{x}_{11}, \underline{v}) \\ K(\underline{x}_{12}, \underline{v}) \\ K(\underline{x}_{21}, \underline{v}) \\ K(\underline{x}_{22}, \underline{v}) \end{bmatrix}$

(b) Project the data $\underline{v}$ to $\underline{u} = [e_1^T].\underline{v}$

9. Plot the following
Subplot 1: Plot $DATA_1$ in red color (using 'r'), and plot $DATA_2$ in green color (using 'g')
Subplot 2: Plot the projected data $\underline{u}$ with $\underline{e}_1$ as the x-axis.
What are the dimensionality reduced data using KLDA.

10. Distance matrix of the original data and the data obtained using the projected data (with $\underline{e}_1$, $\underline{e}_2$, $\underline{e}_3$ and $\underline{e}_4$)

$D_1 =$

$D_2 =$

(OR) load Image.mat and explore dimensionality reduction using KLDA. Obtain the distance matrix obtained using the original data and using the projected data.

PART-1 Multi-Class Logistic Regression

$$DATA_1 = \begin{bmatrix} 1.08 & 0.75 & 0.85 & 0.94 & 0.40 & 1.25 & 1.19 & 0.99 & 0.69 & 1.32 \\ 0.08 & -0.19 & -0.11 & 0.01 & -0.09 & -0.21 & 0.07 & 0.04 & -0.02 & 0.02 \end{bmatrix}$$

$$DATA_2 = \begin{bmatrix} 0.01 & -0.01 & 0.09 & -0.05 & -0.45 & 0.07 & -0.33 & -0.06 & -0.33 & -0.24 \\ 0.85 & 1.05 & 0.93 & 1.41 & 1.45 & 1.20 & 0.88 & 1.08 & 1.10 & 1.01 \end{bmatrix}$$

In $DATA_1$ and $DATA_2$, the first 5 columns correspond to training data, and last 5 columns correspond to testing data.

1. Formulate the matrix $\mathbf{T}$ as follows:
   $$\mathbf{T} = \begin{bmatrix} 1 & 1 & 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 & 1 & 1 \end{bmatrix} \rightarrow \text{Two-class problem}$$
   $t_{nk} \rightarrow n^{th}$ row belongs to the n^{th} class, and k^{th} column belongs to the k^{th} data in the Train data.

2. The probability that the k^{th} vector belongs to the 1^{st} class is given as
   $$\frac{e^{w_1^T \phi(x_k)}}{e^{w_1^T \phi(x_k)} + e^{w_2^T \phi(x_k)}} = y_1(x_k) = y_{1k}$$
   Similarly, the probability that the k^{th} vector belongs to the 2^{nd} class is given as
   $$\frac{e^{w_2^T \phi(x_k)}}{e^{w_1^T \phi(x_k)} + e^{w_2^T \phi(x_k)}} = y_2(x_k) = y_{2k}$$

3. Initialize the weight vectors $w_1 = [0\,0]^T$ and $w_2 = [0\,0]^T$

4. Compute the cross-entropy
   $$\mathbf{J} = -\sum_{n=1}^{2}\sum_{k=1}^{5} t_{nk}.ln(y_{nk})$$

5. Update the weight vector using the following.
   $$w_1(t+1) = w_1(t) + \eta \sum_{k=1}^{5} (t_{1k} - y_{1k})\phi(x_k)$$
   $$w_2(t+1) = w_2(t) + \eta \sum_{k=1}^{5} (t_{2k} - y_{2k})\phi(x_k)$$

6. Compute the cross-entropy using the newly updated weights

$$\mathbf{J} = -\sum_{n=1}^{2}\sum_{k=1}^{10} t_{nk}.ln(y_{nk})$$

7. Repeat steps 5 and 6 for 5 iterations.

8. Plot cross-entropy versus number of iterations

9. Obtain y_{1k} and y_{2k} (computed using the latest $\underline{w}_1$ and $\underline{w}_2$ and the testing data.

$Vector_k$	y_{1k}	y_{2k}	Class Index
$\begin{bmatrix} 1.25 \\ -0.21 \end{bmatrix}$			
$\begin{bmatrix} 1.19 \\ 0.07 \end{bmatrix}$			
$\begin{bmatrix} 0.99 \\ 0.04 \end{bmatrix}$			
$\begin{bmatrix} 0.69 \\ -0.02 \end{bmatrix}$			
$\begin{bmatrix} 1.32 \\ 0.02 \end{bmatrix}$			
$\begin{bmatrix} 0.07 \\ 1.20 \end{bmatrix}$			
$\begin{bmatrix} -0.33 \\ 0.88 \end{bmatrix}$			
$\begin{bmatrix} -0.06 \\ 1.08 \end{bmatrix}$			
$\begin{bmatrix} -0.33 \\ 1.10 \end{bmatrix}$			
$\begin{bmatrix} -0.24 \\ 1.01 \end{bmatrix}$			