Linear Discriminant Analysis (LDA)

Dimensionality Reduction Techniques (LDA)

- **Linear Discriminant Analysis (LDA) :**
- Linear Discriminant Analysis (LDA) is used to **solve dimensionality reduction** for data with higher attributes.
- **History :** The original dichotomous discriminant analysis was developed by Sir Ronald Fisher in **1936**.
- **Introduction :**
- Pre-processing step for pattern-classification and machine learning applications.
- Used for feature extraction.
- Linear transformation that maximize the separation between multiple classes.

Dimensionality Reduction Techniques (LDA)

- **Linear Discriminant Analysis (LDA) :**
- To reduce the dimensions of a d-dimensional data set by projecting it onto a (k)-dimensional subspace (**where k < d**).
- **Feature space data is well represented:-**
- **Compute eigen vectors** from dataset
- Collect them in **scatter matrix**
- Generate ***k*-dimensional data from d-dimensional dataset**.

Dimensionality Reduction Techniques (LDA)

- **Linear Discriminant Analysis (LDA) :**
- **Scatter Matrix :**
- Within class scatter matrix
- In between class scatter matrix
- Maximize the between class measure & minimize the within class measure.

$$S_W = \sum_{i=1}^{c} S_i$$

$$S_B = \sum_{i=1}^{c} N_i (\boldsymbol{m}_i - \boldsymbol{m})(\boldsymbol{m}_i - \boldsymbol{m})^T$$

Dimensionality Reduction Techniques (LDA)

- **Linear Discriminant Analysis (LDA) :**
- **LDA Steps :**

- Compute the d-dimensional mean vectors.
- Compute the scatter matrices
- Compute the eigenvectors and corresponding eigenvalues for the scatter matrices.
- Sort the eigenvalues and choose those with the largest eigenvalues to form a d×k dimensional matrix
- Transform the samples onto the new subspace.

Example

① AKS

Q) factory ABC produces a very expensive and high quality chip rings that their qualities are measured in terms of weight in mg and diameter in cm. Dataset is:

wt in (mg)	Diameter in (cm)	Result
4	1	Pass
2	4	Pass
2	3	Pass
3	6	Pass
4	4	Pass
9	10	fail
6	8	fail
9	5	fail
8	7	fail
10	8	fail

$\Rightarrow$

$X_1 = (x_1, x_2) = \{(4,1), (2,4), (2,3), (3,6), (4,4)\}$

$X_2 = (x_1, x_2) = \{(9,10), (6,8), (9,5), (8,7), (10,8)\}$

$X_1 \Rightarrow$ samples for pass class

$X_2 \Rightarrow$ Samples for fail class

Step-1 Compute within class scatter matrix ② AKS
(S_W)

$$S_W = S_1 + S_2$$

where, S_1 ⇒ is the covariance matrix for X_1 for class pass
S_2 ⇒ is the covariance matrix for X_2 for class fail.

$$S_1 = \frac{1}{n} \sum_{x \in X_1} (x - \mu_1)(x - \mu_1)^T$$

where μ_1 is the mean of X_1 of class yes which is computed as

$$\mu_1 = \left\{ \frac{4+2+2+3+4}{5}, \frac{1+4+3+6+4}{5} \right\}$$

$$\mu_1 = \{3.00 \quad 3.60\}$$

Similarly $\mu_2 = \{8.4 \quad 7.60\}$

Now : $x_1 = \begin{Bmatrix} 4 & 2 & 2 & 3 & 4 \\ 1 & 4 & 3 & 6 & 4 \end{Bmatrix}$ $\mu_1 = \begin{Bmatrix} 3 \\ 3.6 \end{Bmatrix}$

$(x_1 - \mu_1) = \begin{Bmatrix} 1 & -1 & -1 & 0 & 1 \\ -2.6 & 0.4 & -0.6 & 2.4 & 0.4 \end{Bmatrix} \Rightarrow$ Mean Subtracted sample for x_1 for class yes.

Now, for each x, we are going to calculate $(x-\mu_1)(x-\mu_1)^T$. so we will get "5" such matrice in this case.

① $\begin{bmatrix} 1 \\ -2.6 \end{bmatrix} \begin{bmatrix} 1 & -2.6 \end{bmatrix} = \begin{bmatrix} 1 & -2.6 \\ -2.6 & 6.76 \end{bmatrix}$ — ①

② $\begin{bmatrix} -1 \\ 0.4 \end{bmatrix} \begin{bmatrix} -1 & 0.4 \end{bmatrix} = \begin{bmatrix} 1 & -0.4 \\ -0.4 & 0.16 \end{bmatrix}$ — ②

④ AKS

$$\begin{bmatrix} -1 \\ 0.6 \end{bmatrix} \begin{bmatrix} -1 & -0.6 \end{bmatrix} = \begin{bmatrix} 1 & 0.6 \\ 0.6 & 0.36 \end{bmatrix} \quad — (3)$$

$$\begin{bmatrix} 0 \\ 2.4 \end{bmatrix} \begin{bmatrix} 0 & 2.4 \end{bmatrix} = \begin{bmatrix} 0 & 0 \\ 0 & 5.76 \end{bmatrix} \quad — (4)$$

$$\begin{bmatrix} 1 \\ 0.4 \end{bmatrix} \begin{bmatrix} 1 & 0.4 \end{bmatrix} = \begin{bmatrix} 1 & 0.4 \\ 0.4 & 0.16 \end{bmatrix} \quad — (5)$$

Adding (1) (2) (3) (4) and (5) and taking average we get covariance matrix S_1.

$$S_1 = \begin{bmatrix} 0.8 & -0.4 \\ -0.4 & 2.6 \end{bmatrix}$$

Similarly for x_2 for class false, the covariance (5)
matrix is given by

$$S_2 = \begin{bmatrix} 1.84 & -0.04 \\ -0.04 & 2.64 \end{bmatrix} \;\&\; \mu_2 = [8.4 \quad 7.6]$$

$$\therefore \quad S_W = S_1 + S_2$$

$$= \begin{bmatrix} 0.8 & -0.4 \\ -0.4 & 2.6 \end{bmatrix} \begin{bmatrix} 1.84 & -0.04 \\ -0.04 & 2.64 \end{bmatrix}$$

$$= \begin{bmatrix} 2.64 & -0.44 \\ -0.44 & 5.28 \end{bmatrix} \Rightarrow \text{within class scatter Matrix } S_W$$

We have calculated, $x_1, x_2, \mu_1, \mu_2, S_1, S_2$ and S_W.
Now we need to calculate $S_B \Rightarrow$ Between class scatter matrix.

STEP:2: Calculate S_B.

$$S_B = (\mu_1 - \mu_2)(\mu_1 - \mu_2)^T$$

$$= \left[\begin{pmatrix} 3 \\ 3.6 \end{pmatrix} - \begin{pmatrix} 8.4 \\ 7.6 \end{pmatrix}\right] \left[\begin{pmatrix} 3 \\ 3.6 \end{pmatrix} - \begin{pmatrix} 8.4 \\ 7.6 \end{pmatrix}\right]^T$$

$$= \begin{pmatrix} -5.4 \\ -4 \end{pmatrix} \begin{pmatrix} -5.4 & -4 \end{pmatrix}$$

$$= \begin{pmatrix} 29.16 & 21.6 \\ 21.6 & 16.00 \end{pmatrix} \Rightarrow S_B$$

Now $S_W = \begin{bmatrix} 2.64 & -0.44 \\ -0.44 & 5.28 \end{bmatrix}$, $S_B = \begin{pmatrix} 29.16 & 21.6 \\ 21.6 & 16.00 \end{pmatrix}$

STEP-3: find the Best LDA Projection Vector (7) (6)

We find this using Eigen Vector having largest Eigen value

We calculate this using formula:

$$S_W^{-1} S_B V = \lambda V \qquad (9)$$

λ → eigen value (scalar quantity)

V → eigen vector for corresponding eigen value. (vector quantity)

$$\Rightarrow \left| S_W^{-1} S_B V - \lambda V \right| = 0$$

I → Identity matrix.

$$\left| S_W^{-1} S_B - \lambda I \right| = 0$$

$$\left| \begin{pmatrix} 0.384 & 0.032 \\ 0.032 & 0.192 \end{pmatrix} \begin{pmatrix} 29.16 & 21.6 \\ 21.6 & 16.0 \end{pmatrix} - \lambda \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \right| = 0$$

$$\Rightarrow \begin{vmatrix} 11.89-\lambda & 8.81 \\ 5.08 & 3.76-\lambda \end{vmatrix} = 0.$$

$$\Rightarrow (11.89-\lambda)(3.76-\lambda) - 5.08 \times 8.81 = 0$$

$$\Rightarrow 44.7 - 11.89\lambda - 3.76\lambda + \lambda^2 - 44.7 = 0$$

$$\Rightarrow \lambda^2 - 15.76\lambda = 0$$

$$\Rightarrow \lambda(\lambda - 15.76) = 0$$

$$\Rightarrow \lambda = 0 \quad \text{or} \quad \lambda = 15.76$$ → two eigen values we get.

Put highest eigen value in eq (a) on page 7.

Note: we take bigger eigen values to reduce dimention

$$\Rightarrow \underbrace{\begin{bmatrix} 11.89 & 8.81 \\ 5.08 & 3.76 \end{bmatrix}}_{(S_W^{-1} \cdot S_B)} \underbrace{\begin{bmatrix} v_1 \\ v_2 \end{bmatrix}}_{\text{Eigen vector we need to find}} = \underbrace{15.65}_{\text{highest Eigen value.}} \begin{bmatrix} v_1 \\ v_2 \end{bmatrix}$$

We get after solving above eq:

$$\begin{bmatrix} v_1 \\ v_2 \end{bmatrix} = \begin{bmatrix} 0.91 \\ 0.39 \end{bmatrix} \text{ — } \boxed{\text{Solution}}$$

<u>Note:</u> Inverse of Matrix:

$$\begin{bmatrix} a & b \\ c & d \end{bmatrix}^{-1} = \frac{1}{ad-bc} \begin{bmatrix} d & -b \\ -c & a \end{bmatrix}$$

$$S_W = \begin{bmatrix} 2.64 & -0.44 \\ -0.44 & 5.28 \end{bmatrix}, \quad S_W^{-1} = \frac{1}{13.74} \begin{bmatrix} 5.28 & 0.44 \\ 0.44 & 2.64 \end{bmatrix} = \begin{bmatrix} 0.384 & 0.032 \\ 0.032 & 0.192 \end{bmatrix}$$

STEP:4 Dimention Reduction:

$$Y = W^T X,\} \longrightarrow \text{input data samples}$$

W

Projection matrix

OR

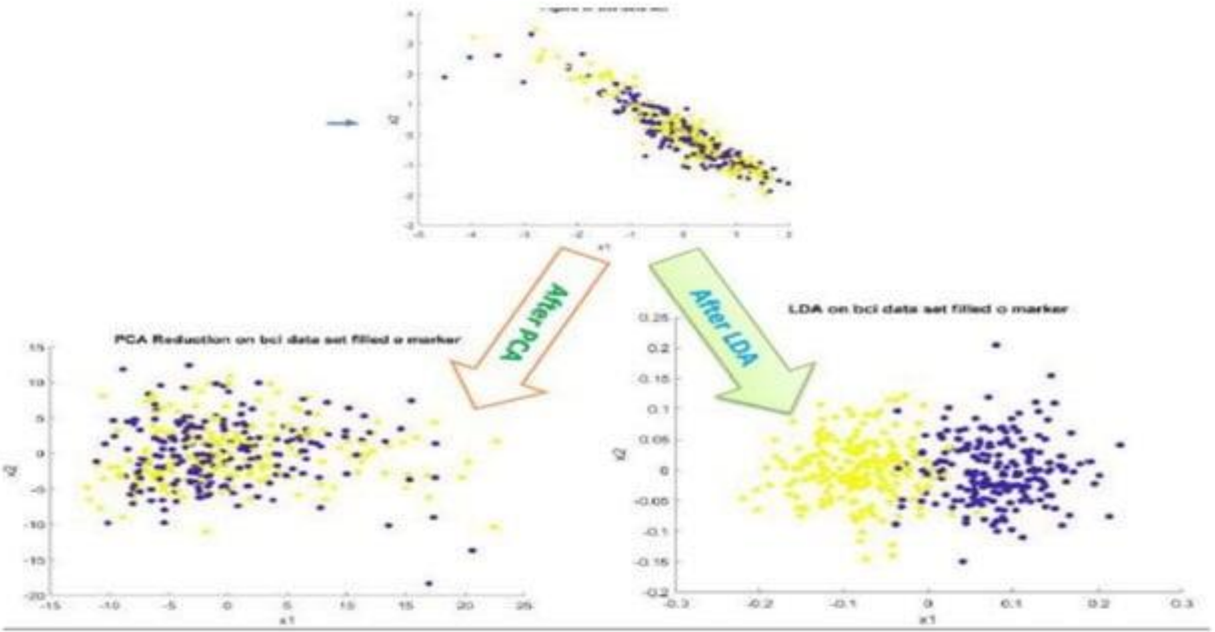
higest eigen Vector

$$\therefore \quad Y = \begin{bmatrix} 0.91 & 0.39 \end{bmatrix}_{1\times 2} \begin{bmatrix} 4 & 2 & 2 & 3 & 4 \\ 1 & 4 & 3 & 6 & 4 \end{bmatrix}_{2\times 5}$$

$$Y = \begin{bmatrix} 4.03 & 3.38 & 2.99 & 5.07 & 5.2 \end{bmatrix}$$

$\therefore$ we see 2-D data sample is changed into 1-D data samples.

Example : After PCA and LDA


After PCA
After LDA
PCA Reduction on bci data set filled o marker
LDA on bci data set filled o marker
x1
x2