Student ID:

Question 1 (12 marks)

(a) For the give data {(-3, 3); (-1, 1); (0, 0); (1, -1); (3, -3)}; please illustrate the steps to calculate the following (handwritten):
(b) The covariance matrix;
(c) Eigen values;
(d) Eigen vectors;
(e) Apply PCA on the dataset and represent the data after the dimension reduction. (Please illustrate the calculation steps one by one).

Solutions:

$$X = \begin{bmatrix} -3 & -1 & 0 & 1 & 3 \\ 3 & 1 & 0 & -1 & -3 \end{bmatrix};$$

(a) $\mu = \begin{bmatrix} 0 \\ 0 \end{bmatrix}; \tilde{X} = X - \mu = \begin{bmatrix} -3 & -1 & 0 & 1 & 3 \\ 3 & 1 & 0 & -1 & -3 \end{bmatrix}$

$$cov(X) = \frac{1}{n-1}\tilde{X}\tilde{X}^T = \frac{1}{4}\begin{bmatrix} -3 & -1 & 0 & 1 & 3 \\ 3 & 1 & 0 & -1 & -3 \end{bmatrix}\begin{bmatrix} -3 & 3 \\ -1 & 1 \\ 0 & 0 \\ 1 & -1 \\ 3 & -3 \end{bmatrix} = \begin{bmatrix} 5 & -5 \\ -5 & 5 \end{bmatrix}$$

(b) $det\begin{vmatrix} 5-\lambda & -5 \\ -5 & 5-\lambda \end{vmatrix} = 0 \;\Rightarrow\; (5-\lambda)^2 - 25 = 0 \;\Rightarrow\; \lambda^2 - 10\lambda = 0 \;\Rightarrow\; \lambda_1 = 0;\ \lambda_2 = 10$

(c) when $\lambda_1 = 0;\ \overrightarrow{v_1} = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix},$

$$\begin{bmatrix} 5-0 & -5 \\ -5 & 5-0 \end{bmatrix}\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix} \;\Rightarrow\; \overrightarrow{v_1} = \frac{1}{\sqrt{2}}\begin{bmatrix} 1 \\ 1 \end{bmatrix};$$

when $\lambda_1 = 10;\ \overrightarrow{v_2} = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix},$

$$\begin{bmatrix} 5-10 & -5 \\ -5 & 5-10 \end{bmatrix}\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix} \;\Rightarrow\; \overrightarrow{v_2} = \frac{1}{\sqrt{2}}\begin{bmatrix} -1 \\ 1 \end{bmatrix};$$

(d) after the reduction, new $X = \frac{1}{\sqrt{2}}[-1 \quad 1] \times \begin{bmatrix} -3 & -1 & 0 & 1 & 3 \\ 3 & 1 & 0 & -1 & -3 \end{bmatrix} = [3\sqrt{2} \quad \sqrt{2} \quad 0 \quad -\sqrt{2} \quad -3\sqrt{2}]$

Question 2 (10 marks)

Based on the lecture slides, please do following LDA analysis;

Fro give data: Class 1, {(3,2);(2,3);(4,4);(3,1);(3,5);(3,3)}

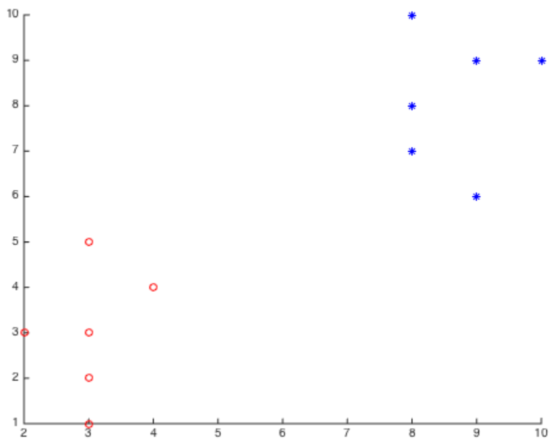
Class 2, {(9,9);(10,9);(8,7);(8,10);(9,6);(8,8)}

(a) Plot the data in the image;
(b) Calculate the class mean and covariance matrix for these two classes;
(c) Calculate the Within-class scatter matrix and Between-class scatter matrix;
(d) Write the generalized eigen value problem for the LDA;
(e) Compute the projection vector.
(You can use matlab to do the calculation or do the exercise by hand except (a). If you use matlab, please copy the codes here and also list the results.)

Solutions:

(a)

```
% samples for class 1&2
x1 = [3,2;2,3;4,4;3,1;3,5;3,3];
x2 = [9,9;10,9;8,7;8,10;9,6;8,8];

% plot the data
scatter  (x1(:,1),x1(:,2),'ro');hold on;
scatter  (x2(:,1),x2(:,2),'b*');

% class means
Mu1 = mean(x1)';
Mu2 = mean(x2)';

% covariance matrix of the first & second class
S1 = cov(x1);
S2 = cov(x2);

% within-class scatter matrix
Sw = S1+S2;

% between-class scatter matrix
Sb = (Mu1-Mu2)*(Mu1-Mu2)';

% computing the LDA projection
invSw = inv(Sw);
invSw_by_Sb = invSw*Sb;

% getting the projection vector
[V, D] = eig(invSw_by_Sb);

% the projection vector
W=V(:,1);
```



(b) Class mean of class 1: $\begin{bmatrix}3\\3\end{bmatrix}$

Class mean of class 2: $\begin{bmatrix}8.6667\\8.1667\end{bmatrix} = \begin{bmatrix}52/6\\49/6\end{bmatrix}$

Covariance matrix of class 1: $\begin{bmatrix}0.4 & 0.2\\0.2 & 2\end{bmatrix}$

Covariance matrix of class 2: $\begin{bmatrix}0.6667 & 0.0667\\0.0667 & 2.1667\end{bmatrix}$

(c) Within-class scatter matrix: $\begin{bmatrix}1.0667 & 0.2667\\0.2667 & 4.1667\end{bmatrix}$

Between-class scatter matrix: $\begin{bmatrix}32.1111 & 29.2778\\29.2778 & 26.6944\end{bmatrix}$

(d) The generalized eigen value problem for the LDA is to find the eigenvector of $S_b {S_w}^{-1}$.

(e) The required projection vector: $\begin{bmatrix}0.9842\\0.1771\end{bmatrix}$

Question 3 (12 marks)
For the given data; please calculate the Minkowski Distance with r=1, 2, infinite.

points	x	y
P1	0	3
P2	3	1
P3	5	1
P4	3	2
P5	2	4

Solutions:

L1	P_1	P_2	P_3	P_4	P_5
P_1	0	5	7	4	3
P_2	5	0	2	1	4
P_3	7	2	0	3	6
P_4	4	1	3	0	3
P_5	3	4	6	3	0

L2	P_1	P_2	P_3	P_4	P_5
P_1	0	$\sqrt{13}$	$\sqrt{29}$	$\sqrt{10}$	$\sqrt{5}$
P_2	$\sqrt{13}$	0	2	1	$\sqrt{10}$
P_3	$\sqrt{29}$	2	0	$\sqrt{5}$	$\sqrt{18}$
P_4	$\sqrt{10}$	1	$\sqrt{5}$	0	$\sqrt{5}$
P_5	$\sqrt{5}$	$\sqrt{10}$	$\sqrt{18}$	$\sqrt{5}$	0

$L_{infinite}$	P_1	P_2	P_3	P_4	P_5
P_1	0	3	5	3	2
P_2	3	0	2	1	3
P_3	5	2	0	2	3
P_4	3	1	2	0	2
P_5	2	3	3	2	0

Question 4 (16 marks)
For the given data, x=(2,3,4,3,1,1,1), y=(2,4,3,5,1,2,2), please calculate the following distances;

(a) Cosine distance
(b) Correlation distance
(c) L2 norm distance
(d) Please illustrate the property of scale invariant for correlation distance.

Solutions:
(a) x=(2,3,4,3,1,1,1), y=(2,4,3,5,1,2,2);
$x \cdot y = 2 \times 2 + 3 \times 4 + 4 \times 3 + 3 \times 5 + 1 \times 1 + 1 \times 2 + 1 \times 2 = 48$
$\|x\| = \sqrt{2^2 + 3^2 + 4^2 + 3^2 + 1^2 + 1^2 + 1^2} = \sqrt{41}$
$\|y\| = \sqrt{2^2 + 4^2 + 3^2 + 5^2 + 1^2 + 2^2 + 2^2} = \sqrt{63}$
$cos(x, y) = \frac{x \cdot y}{\|x\|\|y\|} = \frac{48}{\sqrt{41}\sqrt{63}} = 0.9445$
(b) $\bar{x} = \frac{1}{7}(2 + 3 + 4 + 3 + 1 + 1 + 1) = \frac{15}{7}$
$\bar{y} = \frac{1}{7}(2 + 4 + 3 + 5 + 1 + 2 + 2) = \frac{19}{7}$

$$S_x = \sqrt{\frac{1}{7-1}[(2-\frac{15}{7})^2 + (3-\frac{15}{7})^2 + (4-\frac{15}{7})^2 + (3-\frac{15}{7})^2 + (1-\frac{15}{7})^2 + (1-\frac{15}{7})^2 + (1-\frac{15}{7})^2]}$$
$$= \sqrt{\frac{1}{7-1}[\frac{1+36+169+36+64+64+64}{49}]} = \sqrt{\frac{1}{6} \times \frac{434}{49}} = \sqrt{\frac{31}{21}}$$

$$S_y = \sqrt{\frac{1}{7-1}[(2-\frac{19}{7})^2 + (4-\frac{19}{7})^2 + (3-\frac{19}{7})^2 + (5-\frac{19}{7})^2 + (1-\frac{19}{7})^2 + (2-\frac{19}{7})^2 + (2-\frac{19}{7})^2]}$$
$$= \sqrt{\frac{1}{7-1}[\frac{25+81+4+256+144+25+25}{49}]} = \sqrt{\frac{1}{6} \times \frac{560}{49}} = \sqrt{\frac{40}{21}}$$

$$S_{xy} = \frac{1}{7-1}[(2-\frac{15}{7}) \times (2-\frac{19}{7}) + (3-\frac{15}{7}) \times (4-\frac{19}{7}) + (4-\frac{15}{7}) \times (3-\frac{19}{7}) + (3-\frac{15}{7}) \times (5-\frac{19}{7}) + (1-\frac{15}{7}) \times (1-\frac{19}{7}) + (1-\frac{15}{7}) \times (2-\frac{19}{7}) + (1-\frac{15}{7}) \times (2-\frac{19}{7})]$$
$$= \frac{1}{7-1}[\frac{5+54+26+96+96+40+40}{49}] = \frac{17}{14}$$

$Correlation(x, y) = \frac{S_{xy}}{S_x S_y} = 0.7242$

(c) $L_2(x, y) = [\sum(x-y)^2]^{\frac{1}{2}} = [0+1+1+4+0+1+1]^{\frac{1}{2}} = 2\sqrt{2}$
(d) assume the original data is x and y, after scale translation, y are scaled by a scalar α, then
$y_{new} = \alpha y,$

$$\overline{y}_{new} = \alpha \overline{y}$$

$$s_{xy_{new}} = \frac{\sum_{i=1}^{n} (x_i - \overline{x})(\alpha y_i - \alpha \overline{y})}{n-1} = \alpha s_{xy}$$

$$s_{y_{new}} = \sqrt{\frac{\sum_{i=1}^{n} (\alpha y_i - \alpha \overline{y})^2}{n-1}} = \alpha s_y$$

$$correlation(x, y_{new}) = \frac{s_{xy_{new}}}{s_{x_{new}} \times s_{y_{new}}} = \frac{s_{xy}}{s_x \times s_y}$$

Therefore, the correlation distance is scale invariant.