Mansoura University		Pattern Recognition
Faculty of Engineering		CSE397
Biomedical Engineering Program		Fall-2017
Full Mark is 50		Time allowed is **TWO Hours**

Exam is **FIVE** questions and **ONE** page. Attempt **ALL** questions, and assume **ANY MISSING** data, with **REASONABLE** and **CLEAR** assumptions

Q1) Briefly discuss (*i*) the difference between supervised and unsupervised learning; (*ii*) what are parametric and nonparametric learning; (*iii*) the difference classification and clustering; and (*iv*) the importance of data dimensionality reduction, its advantages and disadvantages, and discuss its types **[8 Points]**

Q2) A given object is to be classified based on an observed feature x into two classes ω_1 and ω_2. If the class-conditioned probabilities for the x feature are

$$p(x|\omega_1) = \alpha_1 \begin{cases} \frac{x}{2}; & 0 \le x \le 1 \\ 1 - \frac{x}{2}; & 1 \le x \le 2 \\ 0; & otherwise \end{cases} \qquad p(x|\omega_2) = \alpha_2 \begin{cases} x - 1; & 1 \le x \le 3 \\ 5 - x; & 3 \le x \le 5 \\ 0; & otherwise \end{cases}$$

(*a*) Determine the values of α_1 and α_2; (*b*) If the prior probabilities for the two classes are 0.4 and 0.6 respectively, find the Bayes' decision boundary and the Bayes classification error. **[12 points]**

Q3) For the Bayes' classifier, the discriminant functions ($g_i(x)$) are the posterior probabilities, i.e., $P(\omega_i|x) = \frac{p(x|\omega_i)P(\omega_i)}{p(x)}$; where **x** is a feature vector. For independent features, each with the same variance (σ^2), $g_i(\mathbf{x})$ can be expressed as $g_i(\mathbf{x}) = W_i^T\mathbf{x} + w_{i0}$. Derive an expression for W_i^T and w_{i0}. ***Hint:*** *The n-dimensional Gaussian distribution is given by* $N(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ *or* $f_X(X) = \frac{1}{(2\pi)^{n/2}|\Sigma|^{1/2}} exp\left(-\frac{1}{2}(X-\mu)^T\Sigma^{-1}(X-\mu)\right)$; *where μ is the mean value of **X**, and Σ is the* (n x *n*) *covariance matrix.* **[10 Points]**

Q4) Suppose you are given training samples for 2D data sets as in **Table 1**. Use the PCA analysis to find the eigenvectors and the percentage of total variance for each principal components. **[10 Points]**

Table 1: Training samples for Q4

	Samples									
Feature	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**
X_1	2.5	0.5	2.2	1.9	3.1	2.3	2.0	1.0	1.5	1.2
X_2	2.4	0.7	2.9	2.2	3.0	2.7	1.6	1.1	1.6	0.9

Q5) A 2D sample space with the samples (**A_1 to A_8**) is given in **Table 2**. Use the *k*-means clustering algorithm to cluster this data into three clusters (i.e., *k*=3). Find the final class centroids and the total classification error of the *k*-means clustering. Assume the initial class centroids (seeds) are A_1, A_4, and A_7 (***Hint:** use only two iterations and the Euclidian distance to determine the class membership*). **[10 Points]**

Table 2: Samples data for Q5

	Samples							
Feature	A_1	A_2	A_3	A_4	A_5	A_6	A_7	A8
F_1	2	2	8	5	7	6	1	4
F_2	10	5	4	8	5	4	2	9

Wish you all the best

Dr. Fahmi Khalifa

Q1) Briefly discuss (*i*) the difference between supervised and unsupervised learning; (*ii*) what are parametric and nonparametric learning; (*iii*) the difference classification and clustering; and (*iv*) the importance of data dimensionality reduction, its advantages and disadvantages, and discuss its types **[8 Points]**

Answer

(i) **Supervised:** Patterns whose class is known a-priori are used for training.
Unsupervised: The number of classes is (in general) unknown and no training patterns are available.

(ii) **Parametric Learning:** Assume model, find optimum parameters from data.
Nonparametric Learning: Can't assume a model for distribution densities and might be a very complicated model with large number of parameters.

(iii)

Classification	Clustering
learning function that classifies an input sample into one of several predefined classes	common descriptive task where one seeks to identifies a finite set of categories (classes) to describe the data
known categories (Supervised)	unknown categories (Unsupervised)

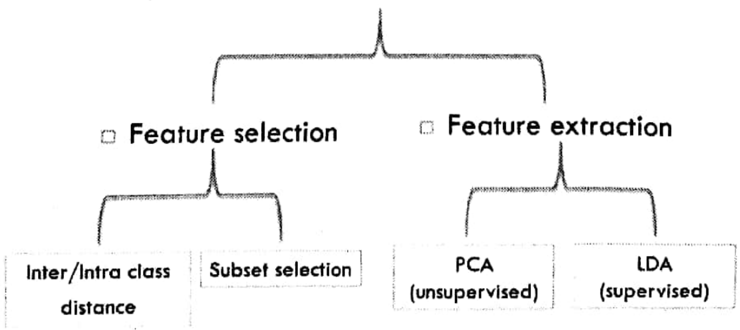
(iv)

a) As data dimensionality increases, it becomes very difficult to extract meaningful conclusions from a given data set. Thus, dimensionality reduction has the following importance and advantages

- Reduces time complexity: Less computation
- Reduces space complexity: Less parameters
- Simpler models are more robust on small datasets
- More interpretable; simpler explanation
- Data visualization (beyond 2 attributes, it gets complicated)

b) However; data dimensionality reduction has two main disadvantages (1) loss information, and (2) increased error in the resulting recognition system

c) Types of data dimensionality reduction



	Mansoura University		Pattern Recognition		
	Faculty of Engineering		CSE397		
	Biomedical Engineering Program		Fall-2017		
	Full Mark is **50**		Time allowed is **TWO Hours**		

Exam is **FIVE** questions and **ONE** page. Attempt **ALL** questions, and assume **ANY MISSING** data, with **REASONABLE** and **CLEAR** assumptions

Q2) A given object is to be classified based on an observed feature x into two classes ω_1 and ω_2. If the class-conditioned probabilities for the x feature are

$$p(x|\omega_1) = \alpha_1 \begin{cases} \frac{x}{2}; & 0 \le x \le 1 \\ 1 - \frac{x}{2}; & 1 \le x \le 2 \\ 0; & otherwise \end{cases} \qquad p(x|\omega_2) = \alpha_2 \begin{cases} x-1; & 1 \le x \le 3 \\ 5-x; & 3 \le x \le 5 \\ 0; & otherwise \end{cases}$$

(a) Determine the values of α_1 and α_2; (b) If the prior probabilities for the two classes are 0.4 and 0.6 respectively, find the Bayes' decision boundary and the Bayes classification error. **[12 points]**

Solution

$P(\omega_1) = 0.4 \qquad P(\omega_2) = 0.6$

$$\int_{-\infty}^{\infty} pdf = 1$$

First

$$\int_0^2 p(x|\omega_1)\,dx = 1 \Rightarrow \int_0^1 \alpha_1 \frac{x}{2}\,dx + \int_1^2 \alpha_1\left(1 - \frac{x}{2}\right)dx = 1$$

$$\alpha_1\left[\frac{x^2}{4}\Big|_0^1 + x - \frac{x^2}{4}\Big|_1^2\right] = 1 \Rightarrow \alpha_1\left[\frac{1}{4} + (2-1) - \left(1 - \frac{1}{4}\right)\right] = 1$$

$$\therefore \alpha_1 = 2$$

Second

$$\int_3^5 p(x|\omega_2)\,dx = 1$$

$$\Rightarrow \int_1^3 \alpha_2(x-1)\,dx + \int_3^5 \alpha_2(5-x)\,dx = 1$$

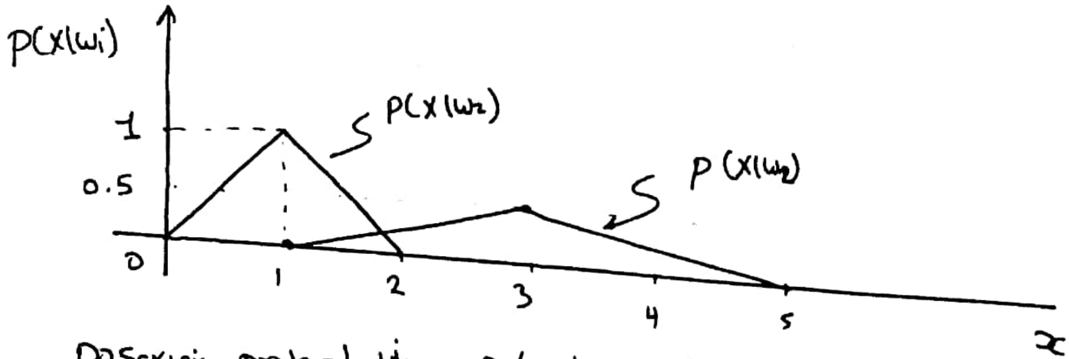
$$\alpha_2\left[\frac{x^2}{2} - x\Big|_1^3 + 5x - \frac{x^2}{2}\Big|_3^5\right] = 1 \Rightarrow$$

$$\alpha_2\left[\left(\frac{9}{2} - 3\right) - \left(\frac{1}{2} - 1\right) + \left(5(5) - \frac{25}{2}\right) - \left(5(3) - \frac{9}{2}\right)\right] = 1 \qquad \alpha_2 = \frac{1}{4}$$

$$p(x|\omega_1) = \begin{cases} x & 0 \le x \le 1 \\ 2-x & 1 \le x \le 2 \\ 0 & \text{elsewhere} \end{cases} \qquad p(x|\omega_2) = \begin{cases} \frac{1}{4}(x-1) & 1 \le x \le 3 \\ \frac{1}{4}(5-x) & 3 \le x \le 5 \\ 0 & \text{elsewhere} \end{cases}$$

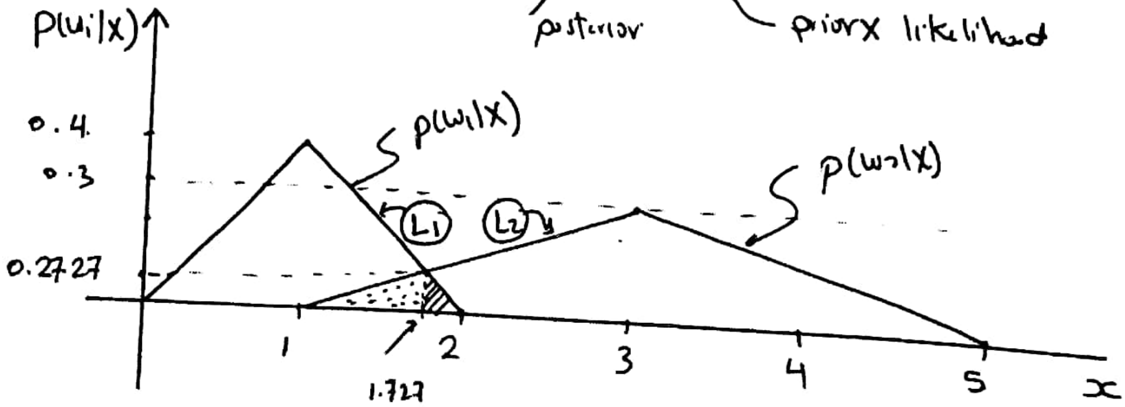
Mansoura University		Pattern Recognition
Faculty of Engineering		CSE397
Biomedical Engineering Program		Fall-2017
Full Mark is **50**		Time allowed is **TWO Hours**

Exam is **FIVE** questions and **ONE** page. Attempt **ALL** questions, and assume **ANY MISSING** data, with **REASONABLE** and **CLEAR** assumptions



posterior probability $p(w_i|x) = p(w_i)\,p(x|w_i)$

posterior — prior x likelihood



a) Bayes decision Rule:

Choose w_1 if $p(w_1|x) > p(w_2|x)$

From the last figure, decision threshold (boundary is a threshold in 1D case) can be obtained by Equating L_1 and L_2

$$0.4(2-x) = 0.6\left(\frac{1}{4}(x-1)\right)$$

$$8 - 4x = \frac{6}{4}x - \frac{6}{4}$$

$$\left(4 + \frac{6}{4}\right)x = 8 + \frac{6}{4}$$

$$22x = 38 \Rightarrow x = 1.727.$$

$\therefore$ Choose w_1 if $X \leqslant 1.727$, otherwise Choose w_2

b) probability of error is the dotted and shaded area in the last Figure.. Can be calculated by integration or by visual inspection

P_{error} = Area of two triangles

$= \frac{1}{2}(0.727)(0.2727) + \frac{1}{2}(0.2727)(0.2727)$

$= 0.0992 \quad + 0.0372$

$= 0.1363$

Q3) For the Bayes' classifier, the discriminant functions ($g_i(x)$) are the posterior probabilities, i.e., $P(\omega_i|\boldsymbol{x}) = \frac{p(x|\omega_i)P(\omega_i)}{p(x)}$; where **x** is a feature vector. For independent features, each with the same variance (σ^2), $g_i(\mathbf{x})$ can be expressed as $g_i(\mathbf{x}) = W_i^T\,\mathbf{x} + w_{i0}$. Derive an expression for W_i^T and w_{i0}. ***Hint:*** *The n-dimensional Gaussian distribution is given by* $\mathbf{N(\mu, \Sigma)}$ *or* $f_X(X) = \frac{1}{(2\pi)^{n/2}|\Sigma|^{1/2}} exp\left(-\frac{1}{2}(\boldsymbol{X}-\boldsymbol{\mu})^T\boldsymbol{\Sigma}^{-1}(\boldsymbol{X}-\boldsymbol{\mu})\right)$; *where* μ *is the mean value of* **X**, *and* Σ *is the* (n x n) *covariance matrix.* **[10 Points]**

Solution

$g_i(x) = g_j(x)$

$\frac{1}{\sigma^2}\mu_i^t x - \frac{1}{2\sigma^2}\mu_i^t\mu_i + \ln p(w_i) = \frac{1}{\sigma^2}\mu_j^t \cancel{x} - \frac{1}{2\sigma^2}\mu_j^t\mu_j + \ln p(w_j)$

$\cancel{\frac{1}{\sigma^2} x (\mu_i^t}$

$\frac{1}{\sigma^2}\mu_i^t x - \frac{1}{2\sigma^2}\mu_i^t\mu_i + \ln p(w_i) - \frac{1}{\sigma^2}\mu_j^t X + \frac{1}{2\sigma^2}\mu_j^t\mu_j - \ln p(w_j) = 0$

$$\frac{1}{\sigma^2} x(\mu_i^t - \mu_j^t) - \frac{1}{2\sigma^2}(\mu_i^t \mu_i - \mu_j^t \mu_j) + \ln p(w_i) - \ln p(w_j) = 0$$

multiplying by σ^2

$$(\mu_i^t - \mu_j^t) x - \frac{1}{2}(\mu_i^t \mu_i - \mu_j^t \mu_j) + \sigma^2 \ln \frac{p(w_i)}{p(w_j)} = 0$$

$$\therefore (\mu_i - \mu_j)^t x - \frac{1}{2}(\mu_i + \mu_j)(\mu_i - \mu_j)^t + \sigma^2 \ln \frac{p(w_i)}{p(w_j)} = 0$$

$$(\mu_i - \mu_j)^t \left[x - \frac{1}{2}(\mu_i + \mu_j) + \frac{\sigma^2}{(\mu_i - \mu_j)^t(\mu_i - \mu_j)} \ln \frac{p(w_i)}{p(w_j)} \cdot (\mu_i - \mu_j)\right] = 0$$

$$(\mu_i - \mu_j)^t \left[x - \frac{1}{2}(\mu_i + \mu_j) + \frac{\sigma^2}{\|\mu_i - \mu_j\|^2} \ln \frac{p(w_i)}{p(w_j)} (\mu_i - \mu_j)\right] = 0$$

$$\therefore w^T(x - x_0) = 0 \qquad + \quad w^T = (\mu_i - \mu_j)^t$$

$$\therefore x_0 = \cancel{x}\ \frac{1}{2}(\mu_i + \mu_j) + \frac{\sigma^2}{\|\mu_i - \mu_j\|^2} \ln \frac{p(w_i)}{p(w_j)} \cdot (\mu_i - \mu_j)$$

Q4) Suppose you are given training samples for 2D data sets as in **Table 1**. Use the PCA analysis to find the eigenvectors and the percentage of total variance for each principal components. and the percentage of total variance for each principal components. **[10 Points]**

Table 1: Training samples for Q4

	Samples									
Feature	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**
X_1	2.5	0.5	2.2	1.9	3.1	2.3	2.0	1.0	1.5	1.2
X_2	2.4	0.7	2.9	2.2	3.0	2.7	1.6	1.1	1.6	0.9

Solution

1. Get μ_1 and μ_2: $\mu_1 = 1.81$ $\mu_2 = 1.91$

2. Recenter data $X_1 = X - \mu$

$X_1 =$

x_1	0.69	-1.31	0.39	0.09	1.29	0.49	0.19	-0.81	-0.31	-0.71
x_2	0.49	-1.21	0.99	0.29	1.09	0.79	-0.31	-0.81	-0.31	-1.01

Mansoura University		Pattern Recognition
Faculty of Engineering		CSE397
Biomedical Engineering Program		Fall-2017
Full Mark is 50		Time allowed is **TWO Hours**

Exam is **FIVE** questions and **ONE** page. Attempt **ALL** questions, and assume **ANY MISSING** data, with **REASONABLE** and **CLEAR** assumptions

3) $Cov(X) = \frac{1}{N-1}(X-\mu)^T(X-\mu) = \frac{1}{N-1} X_1 X_1^T$, $N = 10$

$$Cov(X) = \begin{bmatrix} 0.6166 & 0.615 \\ 0.615 & 0.7166 \end{bmatrix} = \Sigma$$

4) Find the eigenvalues

$$\det(\Sigma - \lambda I) = \text{zero}$$

$$\begin{vmatrix} 0.6166-\lambda & 0.615 \\ 0.615 & 0.7166-\lambda \end{vmatrix} = 0 \Rightarrow (0.6166-\lambda)(0.7166-\lambda) - (0.615)^2 = 0$$

$\lambda_1 = 1.2840$

$\lambda_2 = 0.0490$

5) Find the eigenvectors $\Sigma X = \lambda_i X$ $i=1,2$ $\Rightarrow (\Sigma - \lambda I)X = 0$

$$\begin{bmatrix} 0.6166-\lambda_i & 0.615 \\ 0.615 & 0.7166-\lambda_i \end{bmatrix}\begin{bmatrix} x_{i1} \\ x_{i2} \end{bmatrix} = 0 \qquad X_1 = \begin{bmatrix} 0.678 \\ 0.735 \end{bmatrix} \qquad X_2 = \begin{bmatrix} 0.733 \\ -0.678 \end{bmatrix}$$

6) percentages of total variance (PoV) $= \frac{\lambda_i}{Trace(\Sigma)}$, $Trace(\Sigma) = \sum_{i=1}^{n} a_i$

$\Rightarrow$ for λ_1: $PoV = \frac{\lambda_1}{\lambda_1 + \lambda_2} = 96.3\%$

$\Rightarrow$ for λ_2: $PoV = \frac{\lambda_2}{\lambda_1 + \lambda_2} = 3.7\%$

Q5) A 2D sample space with the samples (**A_1 to A_8**) is given in **Table 2**. Use the k-means clustering algorithm to cluster this data into three clusters (i.e., k=3). Find the final class centroids and the total classification error of the k-means clustering. Assume the initial class centroids (seeds) are A_1, A_4, and A_7 (***Hint: use only two iterations and the Euclidian distance to determine the class membership***). **[10 Points]**

Table 2: Samples data for Q5

	Samples							
Feature	**A_1**	**A_2**	**A_3**	**A_4**	**A_5**	**A_6**	**A_7**	**A8**
F_1	2	2	8	5	7	6	1	4
F_2	10	5	4	8	5	4	2	9

Mansoura University		Pattern Recognition
Faculty of Engineering		CSE397
Biomedical Engineering Program		Fall-2017
Full Mark is **50**		Time allowed is **TWO Hours**

Exam is **FIVE** questions and **ONE** page. Attempt **ALL** questions, and assume **ANY MISSING** data, with **REASONABLE** and **CLEAR** assumptions

Solution:

a)

d(a,b) denotes the Eucledian distance between a and b. It is obtained directly from the distance matrix or calculated as follows: $d(a,b)=sqrt((x_b-x_a)^2+(y_b-y_a)^2))$

seed1=A1=(2,10), seed2=A4=(5,8), seed3=A7=(1,2)

epoch1 – start:

A1:

d(A1, seed1)=0 as A1 is seed1

d(A1, seed2)= $\sqrt{13}$ >0

d(A1, seed3)= $\sqrt{65}$ >0

→A1 ∈ cluster1

A2:

d(A2,seed1)= $\sqrt{25}$ = 5

d(A2, seed2)= $\sqrt{18}$ = 4.24

d(A2, seed3)= $\sqrt{10}$ = 3.16 ← smaller

→ A2 ∈ cluster3

A3:

d(A3, seed1)= $\sqrt{36}$ = 6

d(A3, seed2)= $\sqrt{25}$ = 5 ← smaller

d(A3, seed3)= $\sqrt{53}$ = 7.28

→ A3 ∈ cluster2

A4:

d(A4, seed1)= $\sqrt{13}$

d(A4, seed2)=0 as A4 is seed2

d(A4, seed3)= $\sqrt{52}$ >0

→ A4 ∈ cluster2

A5:

d(A5, seed1)= $\sqrt{50}$ = 7.07

d(A5, seed2)= $\sqrt{13}$ = 3.60 ← smaller

d(A5, seed3)= $\sqrt{45}$ = 6.70

→ A5 ∈ cluster2

A6:

d(A6, seed1)= $\sqrt{52}$ = 7.21

d(A6, seed2)= $\sqrt{17}$ = 4.12 ← smaller

d(A6, seed3)= $\sqrt{29}$ = 5.38

→ A6 ∈ cluster2

A7:

d(A7, seed1)= $\sqrt{65}$ >0

d(A7, seed2)= $\sqrt{52}$ >0

d(A7, seed3)=0 as A7 is seed3

→ A7 ∈ cluster3

A8:

d(A8, seed1)= $\sqrt{5}$

d(A8, seed2)= $\sqrt{2}$ ← smaller

d(A8, seed3)= $\sqrt{58}$

→ A8 ∈ cluster2

end of epoch1

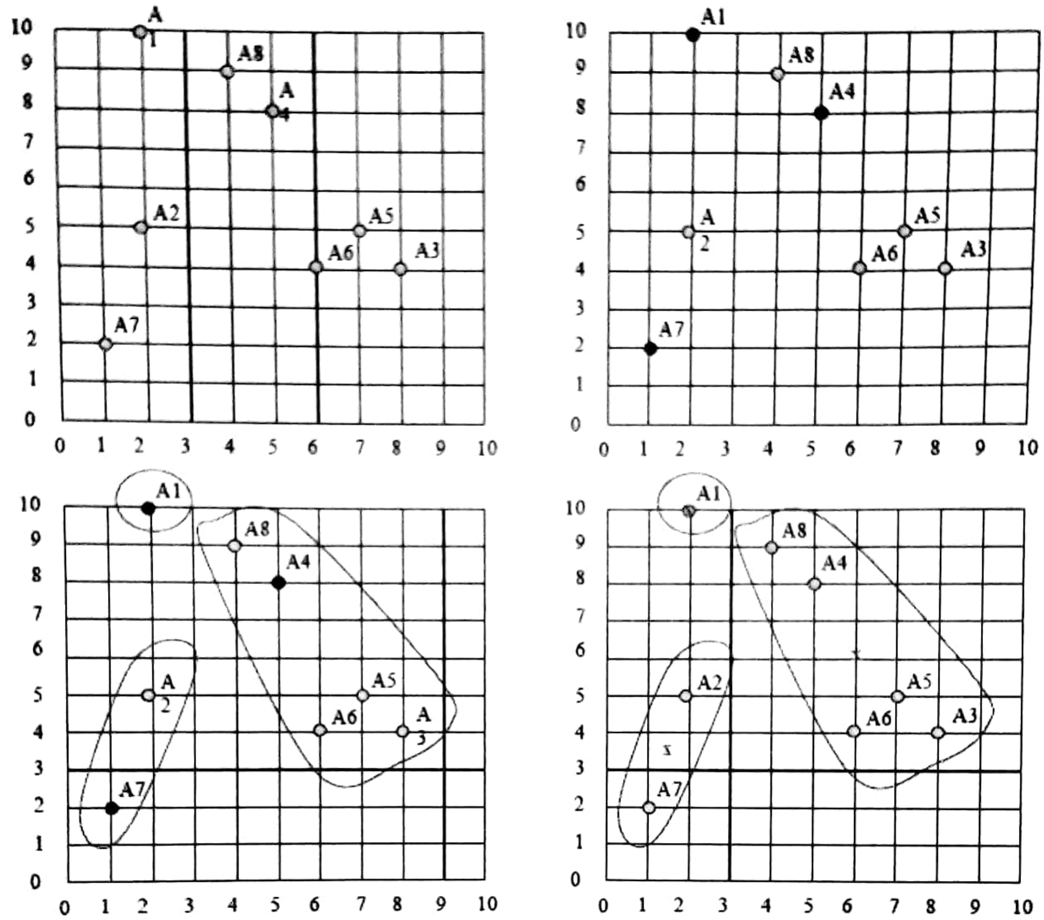
new clusters: 1: {A1}, 2: {A3, A4, A5, A6, A8}, 3: {A2, A7}

b) centers of the new clusters:

C1= (2, 10), C2= ((8+5+7+6+4)/5, (4+8+5+4+9)/5) = (6, 6), C3= ((2+1)/2, (5+2)/2) = (1.5, 3.5)

Mansoura University		Pattern Recognition
Faculty of Engineering		CSE397
Biomedical Engineering Program		Fall-2017
Full Mark is **50**		Time allowed is **TWO Hours**

Exam is **FIVE** questions and **ONE** page. Attempt **ALL** questions, and assume **ANY MISSING** data, with **REASONABLE** and **CLEAR** assumptions



We would need two more epochs. After the 2nd epoch the results would be:
1: {A1, A8}, 2: {A3, A4, A5, A6}, 3: {A2, A7}
with centers C1=(3, 9.5), C2=(6.5, 5.25) and C3=(1.5, 3.5).
After the 3rd epoch, the results would be:
1: {A1, A4, A8}, 2: {A3, A5, A6}, 3: {A2, A7}
with centers C1=(3.66, 9), C2=(7, 4.33) and C3=(1.5, 3.5).