Mansoura University		CSE 397: Pattern Recognition	
Faculty of Engineering		Spring-2018: **Self Study**	
Biomedical Eng. Program		Exam date: May 3rd, 2018	
Full Mark is **50**		Time Allowed is **120 Minutes**	
Exam is **FIVE** Questions and **ONE** page. Please attempt **ALL** questions Assume **ANY MISSING** data, with **REASONABLE** and **CLEAR** assumptions			

Q1) Draw a block diagram of a generic pattern recognition system and briefly explain the function of each block in your pipeline? **[10 Points]**

Q2) Consider the data of four adults, indicating their weight and their health status. Devise a simple classifier that can properly classify all four patterns. How a fifth adult of weight 76 kg is classified using this classifier? **[5 Points]**

Table 1: Data for Q2

None	Yes
50	Unhealthy
60	Healthy
70	Healthy
80	Unhealthy

Q3) Using a naïve Bayes' classifier and the training data in **Table 2**, classify a given person as male or female based on height, weight, and foot size of 6', 130 *lbs*, 8", respectively. Assuming no covariance between the features and the data is Gaussian. [***Hint Gaussian probability density function with a mean μ and standard deviation σ is given as:*** $\frac{1}{\sqrt{\sigma^2 2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$] **[10 Points].**

Table 2: Training Data for Q3

sex	Height (feet)	Weight (lbs)	Foot size (inches)
M	6	180	12
M	5.92 (5'11")	190	11
M	5.58 (5' 7")	170	12
M	5.92 (5'11")	165	10
F	5	100	6
F	5.5 (5' 6")	150	8
F	5.42 (5' 6')	130	7
F	5.75 (5' 9')	150	9

Q4) Suppose you are given a training sample for 2D data sets as in **Table 3**. Use the PCA analysis to find the eigenvectors and the percentage of total variance for each principal components. **[15 Points]**

Table 3: 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) Suppose you have a 2D sample space with the samples given in **Table 3**. Using the *k*-means clustering algorithm with K=2, find the final class centroids and the total classification error of the *k*-means clustering. Assume the initial class distribution is $C_1 = \{1,2,4\}$ and $C_2 = \{3,5\}$ (***Hint: use only two iterations and the Euclidian distance to determine the class membership***). **[10 Points]**

Table 4: Training samples for Q5

	Samples				
Feature	**1**	**2**	**3**	**4**	**5**
F_1	0	0	1.5	5	5
F_2	2	0	0	0	2

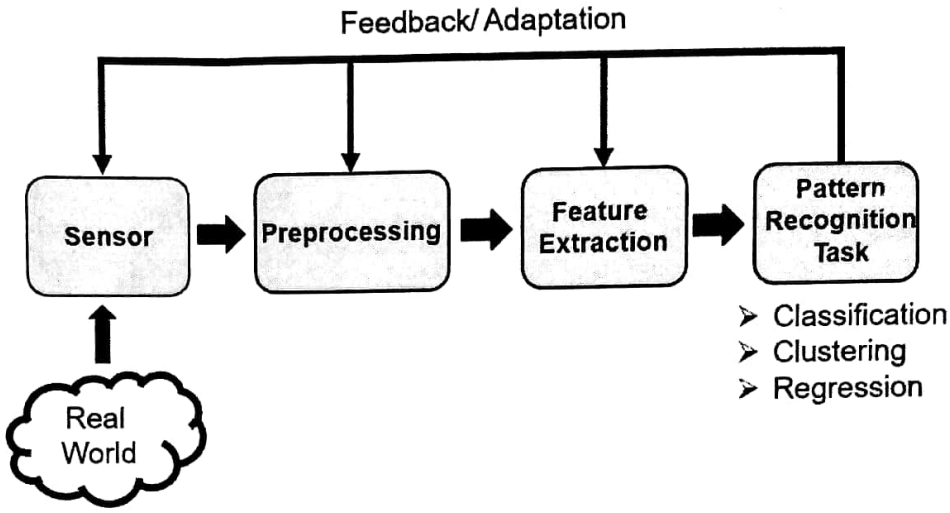
Best of Luck

Dr. Fahmi Khalifa

Mansoura University		CSE 397: Pattern Recognition	
Faculty of Engineering		Spring-2018: **Self Study**	
Biomedical Eng. Program		Exam date: May 3rd, 2018	
Full Mark is **50**		Time Allowed is **120 Minutes**	
Exam is **FIVE** Questions and **ONE** page. Please attempt **ALL** questions Assume **ANY MISSING** data, with **REASONABLE** and **CLEAR** assumptions			

Q1) Draw a block diagram of a generic pattern recognition system and briefly explain the function of each block in your pipeline? **[10 Points]**

Answer



1. The sensing/acquisition stage uses a transducer such as a camera or a microphone to acquire signals (e.g., an image) , which should be of sufficient quality that distinguishing "features" can be adequately measured.
2. Pre-processing is often used to remove noise, handle missing data, and detect and handle outlier data.
3. Features extraction is to obtain some attributes that distinguish between classes. The quality of the features is related to their ability to discriminate examples from different classes.
   - Examples from the **same class** should have similar feature values, while examples from **different classes** should have different feature values (Discriminating Features)
   - Good features should have small **intra-class** variations and large **inter-class** variations
4. The goal of the pattern recognition task is to (i) classify new data (test data) to one of the classes, characterized by a decision region or (i) to cluster all data samples into a distinct (predetermined) number of classes or (iii) to be used in regression analysis

---

Q2) Consider the data of four adults, indicating their weight and their health status. Devise a simple classifier that can properly classify all four patterns. How a fifth adult of weight 76 kg is classified using this classifier? **[5 Points]**

Solution

With this limited data set, the decision boundaries between the two classes should be set at 55 and 65 kg. There is some arbitrariness as to what happens on the actual boundary values, but we could choose for example:

For a weight, x, in kg

If $55 \leq x < 75$ that person is classified as healthy.

The fifth person, at 76kg, should be classified as unhealthy.

Mansoura University		CSE 397: Pattern Recognition
Faculty of Engineering		Spring-2018: **Self Study**
Biomedical Eng. Program		Exam date: May 3rd, 2018
Full Mark is **50**		Time Allowed is **120 Minutes**
Exam is **FIVE** Questions and **ONE** page. Please attempt **ALL** questions Assume **ANY MISSING** data, with **REASONABLE** and **CLEAR** assumptions		

Q3) Using a naïve Bayes' classifier and the training data in **Table 2**, classify a given person as male or female based on height, weight, and foot size of 6', 130 *lbs*, 8", respectively. Assuming no covariance between the features and the data is Gaussian. [***Hint Gaussian probability density function with a mean μ and standard deviation σ is given as:*** $\frac{1}{\sqrt{\sigma^2 2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$] **[10 Points].**

Table 2: Training Data for Question # 3

sex	Height (feet)	Weight (lbs)	Foot size (inches)
M	6	180	12
M	5.92 (5'11")	190	11
M	5.58 (5' 7")	170	12
M	5.92 (5'11")	165	10
F	5	100	6
F	5.5 (5' 6")	150	8
F	5.42 (5' 6')	130	7
F	5.75 (5' 9')	150	9

Solution

The means and variances of the features are first found

SEX	Mean (height)	Variance (height)	Mean (weight)	Variance (weight)	Mean (footsize)	Variance (footsize)
Male	5.855	0.035	176.25	0.0122	11.25	0.091
Female	5.417	0.097	132.5	0.0558	7.5	1.666

Assume the prior probabilities are equal, i.e., P(male) = P(female) = 0.5

For the classification as male the posterior (equn.4.14 and 4.15) is given by

$$posterior(male) = \frac{P(male)P(height \mid male)\ P(weight \mid male)P(footsize \mid male)}{evidence}$$

For the classification as female the posterior is given by

$$posterior(female) = \frac{P(female)P(height \mid female)\ P(weight \mid female)P(footsize \mid female)}{evidence}$$

Each of the features are distributed according to a Gaussian, so for height

$$p(\text{height}|\text{male}) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(\frac{-(6-\mu)^2}{2\sigma^2}\right) \approx 1.5789$$

using the appropriate values of μ and σ2.

Similarly

P(weight | male) = 0.00000598; P(footsize | male) = 0.0013112

Using these, the Posterior Numerator (male) = 0.0000000061984

P(height | female) = 0.223436; P(weight | female) = 0.016789; P(footsize | female) = 0.28669

Using these, Posterior Numerator (female) = 0.00053778

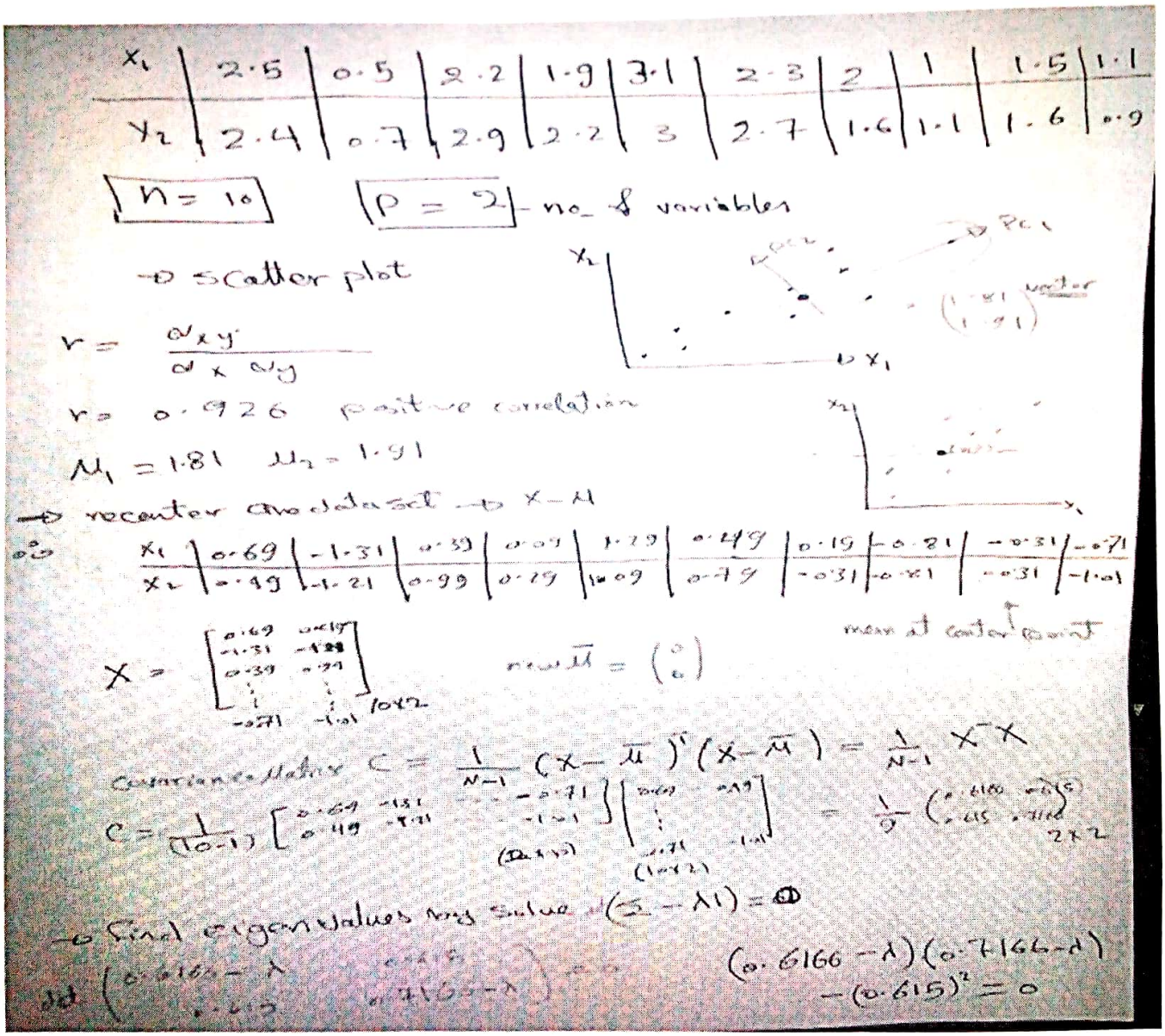
Since posterior (female) > posterior (male) , then we predict that the sample is female.

Mansoura University		CSE 397: Pattern Recognition	
Faculty of Engineering		Spring-2018: **Self Study**	
Biomedical Eng. Program		Exam date: May 3rd, 2018	
Full Mark is **50**		Time Allowed is **120 Minutes**	
Exam is **FIVE** Questions and **ONE** page. Please attempt **ALL** questions Assume **ANY MISSING** data, with **REASONABLE** and **CLEAR** assumptions			

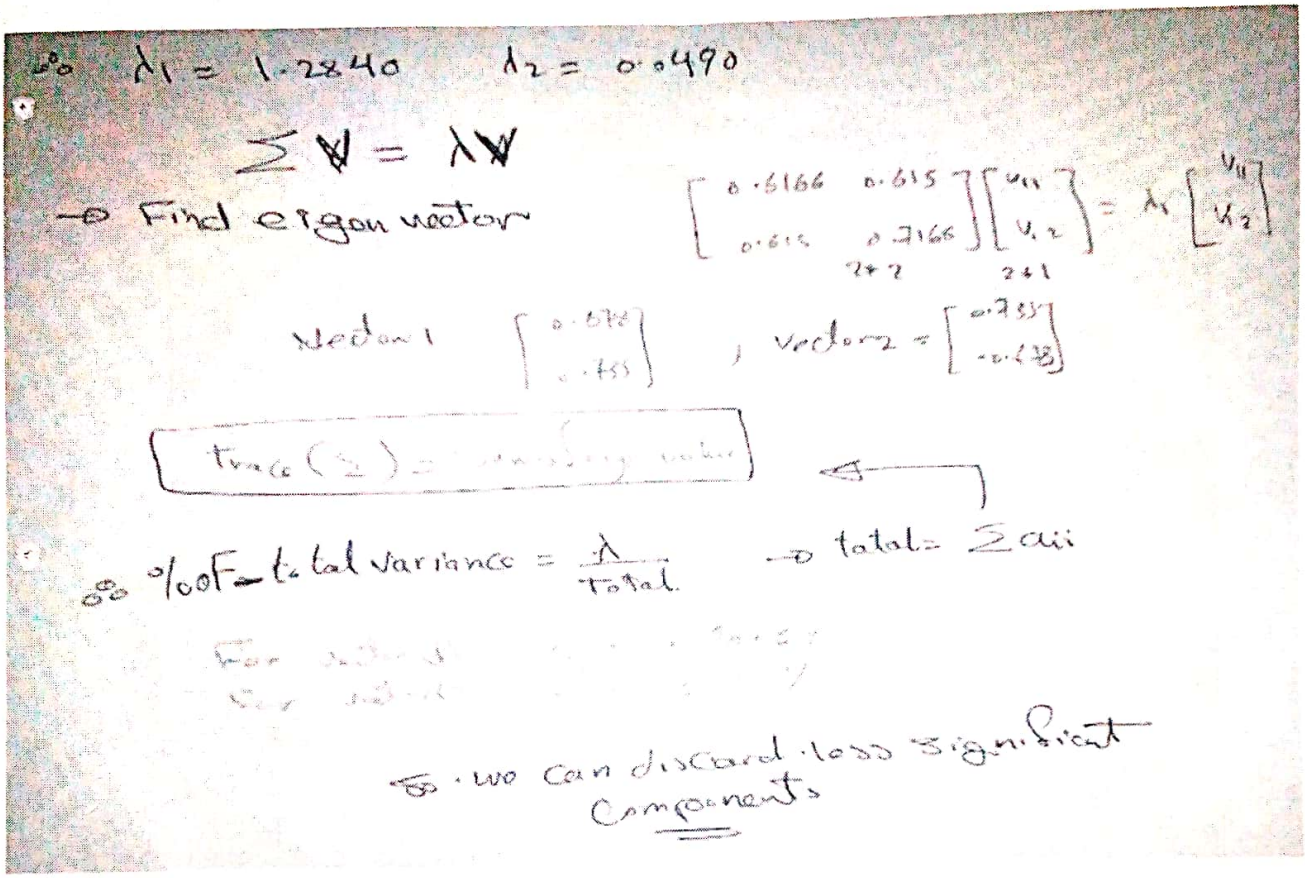
Q4) Suppose you are given a training sample for 2D data sets as in **Table 3**. Use the PCA analysis to find the eigenvectors and the percentage of total variance for each principal components. **[15 Points]**

Table 3: 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



Q6) Suppose you have a 2D sample space with the samples given in **Table 3**. Using the *k*-means clustering algorithm with K=2, find the final class centroids and the total classification error of the K-means clustering. Assume the initial class distribution is $C_1 = \{1,2,4\}$ and $C_2 = \{3,5\}$ (***Hint: use only two iterations***). **[10 Points]**

Solution

Table 3: Training samples for Q6

	Samples				
Feature	**1**	**2**	**3**	**4**	**5**
F_1	0	0	1.5	5	5
F_2	2	0	0	0	2

2. Random distribution of samples:

$C_1 = \{x_1, x_2, x_4\}$ and $C_2 = \{x_3, x_5\}$

3. Centriods:

$M_1 = \{(0+0+5)/3, (2+0+0)/3\} = \{1.66, 0.66\}$

$M_2 = \{(1.5+5)/2, (0+2)/2\} = \{3.25, 1.00\}$

Within-cluster variations:

$e_1^2 = [(0-1.66)^2+(2-0.66)^2] + [(0-1.66)^2 + (0-0.66)^2]+[(5-1.66)^2+ (0-0.66)^2] = 19.36$

$e_2^2 = [(1.5-3.25)^2 + (0-1)^2] + [(5-3.25)^2 + (2-1)^2] = 8.12$

Total square error:

$E^2 = e_1^2 + e_2^2 = 19.36 + 8.12 = 27.48$

2. Reassign all samples:

$d(M_1, x_1) = 2.14$	and	$d(M_2, x_1) = 3.40$	$\Rightarrow$	$x_1 \in C_1$
$d(M_1, x_2) = 1.79$	and	$d(M_2, x_2) = 3.40$	$\Rightarrow$	$x_2 \in C_1$
$d(M_1, x_3) = 0.83$	and	$d(M_2, x_3) = 2.01$	$\Rightarrow$	$x_3 \in C_1$
$d(M_1, x_4) = 3.41$	and	$d(M_2, x_4) = 2.01$	$\Rightarrow$	$x_4 \in C_2$
$d(M_1, x_5) = 3.60$	and	$d(M_2, x_5) = 2.01$	$\Rightarrow$	$x_5 \in C_2$

3. New clusters:

$C_1 = \{x_1, x_2, x_3\}$ and $C_2 = \{x_4, x_5\}$

New centroids:

$M_1 = \{0.5, 0.67\}$
$M_2 = \{5.0, 1.0\}$

Errors:

$e_1^2 = 4.17, \quad e_2^2 = 2.00 \rightarrow E^2 = 6.17$