King's College London

This paper is part of an examination of the College counting towards the award of a degree. Examinations are governed by the College Regulations under the authority of the Academic Board.

Degree Programmes	MSc, MSci
Module Code	7CCSMPNN
Module Title	Pattern Recognition
Examination Period	May 2016 (Period 2)

Time Allowed	Three hours
Rubric	ANSWER QUESTION ONE AND ANY THREE OUT OF THE REMAINING FIVE QUESTIONS. ANSWER EACH QUESTION ON A NEW PAGE OF YOUR ANSWER BOOK AND WRITE ITS NUMBER IN THE SPACE PROVIDED. All questions carry equal marks. If more than three questions other than Question One are answered, clearly indicate which answers you would like to be marked on the front page of the answer booklet, otherwise, Question One and the first three subsequent questions answered, in exam paper order, will be marked.
Calculators	Calculators may be used. The following models are permitted: Casio fx83 / Casio fx85.
Notes	Books, notes or other written material may not be brought into this examination

PLEASE DO NOT REMOVE THIS PAPER FROM THE EXAMINATION ROOM

TURN OVER WHEN INSTRUCTED

1. Question one is compulsory

a. What is *Pattern Recognition*?

[3 marks]

b. Draw a block diagram of the key components of a pattern recognition system.

[5 marks]

c. Sketch a diagram to show a two-dimensional two-class dataset of the following cases. Use *cross* to indicate class 1 and *circle* to indicate class 2.

i. Linearly separable

ii. Nonlinearly separable

iii. Multi-modal

[9 marks]

d. Given the Minimum Error Rate Classifier for class ω_i with the discriminant function

$$g_i(\mathbf{x}) = \frac{P(\omega_i)}{\sqrt[d/2]{(2\pi)}\sqrt{|\mathbf{\Sigma}_i|}} \exp\big(-\frac{1}{2}(\mathbf{x}-\boldsymbol{\mu_i})^T\mathbf{\Sigma}_i^{-1}(\mathbf{x}-\boldsymbol{\mu_i})\big),$$

determine if the following discriminant functions can be used to achieve minimum error rate classification. Explain your answer.

i. $g_i(\mathbf{x}) + 5$

ii. $\dfrac{25}{3} \times g_i(\mathbf{x})$

iii. $\sin(g_i(\mathbf{x}))$

iv. $-\dfrac{1}{2}(\mathbf{x}-\boldsymbol{\mu_i})^T\mathbf{\Sigma}_i^{-1}(\mathbf{x}-\boldsymbol{\mu_i}) - \dfrac{d}{2}\ln(2\pi) - \dfrac{1}{2}\ln(|\mathbf{\Sigma}_i|) + \ln(P(\omega_i))$

where $\mathbf{x}$ denotes a sample, ω_i denotes class i, $P(\omega_i)$ is the prior probability, $\boldsymbol{\mu_i}$ is the mean vector, $\mathbf{\Sigma}_i$ is the covariance matrix, $\sin$ is the sinusoidal function and $\ln$ is the natural logarithm function.

[8 marks]

2. a. Draw a block diagram of a *multiclass discriminant function classifier* and explain how it works for classification.

[10 marks]

b. Describe the *Sequential Multiclass Perceptron Learning Algorithm* using pseudocode.

[7 marks]

c. Consider the training dataset shown in Table 1.

Class	Feature vector
1	$[-2, 6]$
1	$[-1, -4]$
1	$[3, -1]$
2	$[-3, -2]$
3	$[-4, -5]$

Table 1: Training dataset.

Determine the class of a new feature vector $\mathbf{x} = [-2, 0]$ with the k-nearest-neighbour classifier using Euclidean distance and the following values of k:

i. $k = 1$

ii. $k = 3$

iii. $k = 5$

[8 marks]

3. **a.** Name three feature extraction techniques

[3 marks]

b. Consider the dataset in Table 2.

Class	Feature vector $\mathbf{x}^T$
1	$[1, 2]$
1	$[2, 1]$
1	$[3, 3]$
2	$[6, 5]$
2	$[7, 8]$

Table 2: Training dataset.

i. Using *Linear Discriminant Analysis* (LDA), given a projection vector $\mathbf{w}^T = [1, 2]$, find the projection of the feature vectors. Is the dataset in the projection space linearly separable? Explain your answer.

[8 marks]

ii. Using *Fisher's method*, determine which of the following projection weights is more effective in the context of *Linear Discriminant Analysis (LDA)*. Explain your answer.

- $\mathbf{w}^T = [-1, 5]$
- $\mathbf{w}^T = [2, -6]$

[14 marks]

4. A diagram of 3-layer partially connected neural network is shown in Figure 1.

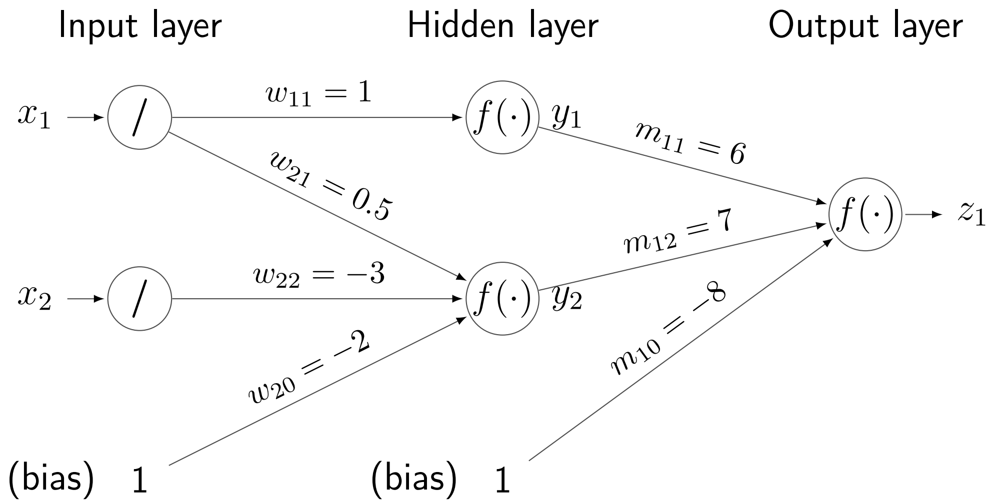

Figure 1: A diagram of 3-layer partially connected neural network.

a. Given that the output of the neural network in Figure 1 can be expressed as

$$z_1 = f\Big(\mathbf{W}_{kj} f\left(\mathbf{W}_{ji}\mathbf{x} + \mathbf{W}_{j0}\right) + \mathbf{W}_{k0}\Big)$$

where $\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$ and $f(\cdot)$ is any activation function, determine the matrices $\mathbf{W}_{ji}$, $\mathbf{W}_{kj}$, $\mathbf{W}_{j0}$ and $\mathbf{W}_{k0}$.

[8 marks]

QUESTION 4 CONTINUES ON NEXT PAGE

b. A classifier is designed as $y = \text{sgn}(z_1)$ where $\text{sgn}(z_1) = \begin{cases} 1 & \text{if } z_1 \geq 0 \\ -1 & \text{if } z_1 < 0 \end{cases}$, $y = 1$ represents class 1 and $y = -1$ represents class 2. Considering a sample $\mathbf{x} = \begin{bmatrix} 2 \\ -6 \end{bmatrix}$ and the activation function $f(\cdot)$ as a logarithmic sigmoid function, i.e., $f(n) = \dfrac{1}{1+e^{-n}}$, determine which class the sample belongs to.

[5 marks]

c. Consider the activation function in all the input, hidden and output units as a linear function. The neural network is trained with the stochastic backpropagation algorithm using the cost $J = \frac{1}{2}\|z_1 - t\|^2$ where $\|\cdot\|$ denotes the l_2 (or Euclidean) norm operator. Given the only training sample $\mathbf{x} = \begin{bmatrix} 1 \\ 5 \end{bmatrix}$ and its target output $t = 0.8$, determine the updated value of w_{20} at the next iteration using the learning rate $\eta = 0.01$.

[12 marks]

5. a. What are *Linguistic Hedges* in the context of fuzzy set?

[4 marks]

b. Name three categories of *Linguistic Hedges* in the context of fuzzy set? Briefly describe them.

[9 marks]

c. Consider a 2-input single-output Sugeno fuzzy inference system with the following 4 rules:

Rule 1: **IF** x is $Small$ and y is $Small$ **THEN** z is $-x+y+1$

Rule 2: **IF** x is $Small$ and y is $Large$ **THEN** z is $-y+3$

Rule 3: **IF** x is $Large$ and y is $Small$ **THEN** z is $-x+3$

Rule 4: **IF** x is $Large$ and y is $Large$ **THEN** z is $x+y+2$

where the membership functions are define as $\mu_{x_{Small}}(x) = 1 - \dfrac{1}{1+e^{-x}}$, $\mu_{x_{Large}}(x) = \dfrac{1}{1+e^{-x}}$, $\mu_{y_{Small}}(y) = 1 - \dfrac{1}{1+e^{-y+5}}$ and $\mu_{y_{Large}}(y) = \dfrac{1}{1+e^{-y+5}}$.

Given that the fuzzy "AND" operation is the "product" operation, determine the output of the Sugeno fuzzy inference system for $x = 1$ and $y = 8$.

[12 marks]

6. A support vector machine (SVM) classifier is employed to classify the following samples:

$$\text{Class 1: } \mathbf{x}_1 = \begin{bmatrix} 1 \\ 2 \end{bmatrix}.$$
$$\text{Class 2: } \mathbf{x}_2 = \begin{bmatrix} 7 \\ 8 \end{bmatrix}, \mathbf{x}_3 = \begin{bmatrix} 10 \\ 15 \end{bmatrix}.$$

a. Identify the support vectors by inspection.

[7 marks]

b. Design an SVM classifier to correctly classify all given samples.

[15 marks]

c. What is the margin given by the hyperplane obtained in Question 6.b?

[3 marks]