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 2017

Time Allowed	Three hours
Rubric	ANSWER FOUR OF SIX QUESTIONS. All questions carry equal marks. If more than four questions are answered, the answers to the first four questions in exam paper order will count.
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

Answer each question on a new page of your answer book and write its number in the space provided.

PLEASE DO NOT REMOVE THIS PAPER FROM THE EXAMINATION ROOM

DO NOT TURN OVER UNTIL INSTRUCTED TO DO SO

1. **a.** Give a brief definition of each of the following terms:

i. Minimum Error Rate classifier

ii. Minimum Loss classifier

[4 marks]

b. Write down the equation for the Likelihood Ratio Test for minimum error, and explain how this equation is used to determine the class of an exemplar.

[4 marks]

Table 1, below, shows samples that have been recorded from two univariate class-conditional probability distributions:

Table 1

class	samples of x									
ω_1	0.82	1.42	1.53	0.21	0.72	1.73	0.49	1.03	0.69	2.12
ω_2	2.54	1.36	3.77	2.86	0.92	6.01				

c. Apply k_n nearest neighbours to the data shown in Table 1 in order to estimate the density of $p(x|\omega_1)$ at location $x = 1.5$, using $k = 3$.

[4 marks]

d. The class-conditional probability distribution for class 2 in Table 1, $p(x|\omega_2)$, is believed to be a Normal distribution. Calculate the parameters of this Normal distribution using Maximum-Likelihood Estimation, use the "unbiased" estimate of the covariance.

[4 marks]

QUESTION 1 CONTINUES ON NEXT PAGE

e. Using the number of samples given in Table 1 estimate the prior probabilities of each class, *i.e.*, calculate P(ω_1) and P(ω_2).

[2 marks]

f. Using the Likelihood Ratio Test and the results from question 1.c, question 1.d, and question 1.e determine the class in which a new sample with value $x = 1.5$ should be placed.

[3 marks]

g. Using the Likelihood Ratio Test *with losses* and the results from question 1.c and question 1.d determine in what class a new sample with value $x = 1.5$ should be placed, when $P(\omega_1) = 0.4P(\omega_2)$ and the loss, λ_{ij}, associated with choosing class i when the actual class is j, is as follows: $\lambda_{11} = 0$, $\lambda_{12} = 2$, $\lambda_{22} = 0$, $\lambda_{21} = 1$.

[4 marks]

2. **a.** In augmented feature space, a dichotomizer is defined using the following linear discriminant function $g(\mathbf{x}) = \mathbf{a}^T\mathbf{y}$ where $\mathbf{a}^T = (-2, 1, 2, 1, 2)$ and $\mathbf{y} = \left(1, \mathbf{x}^T\right)^T$. Determine the class of the following feature vectors, $\mathbf{x}_1 = (1, 1, 1, 1)^T$ and $\mathbf{x}_2 = (0, -1, 0, 1)^T$.

[4 marks]

b. One method of learning the parameters of a linear discriminant function is by using the perceptron learning algorithm. Write pseudo-code for the *sequential* perceptron learning algorithm.

[4 marks]

Table 2

$\mathbf{x}^T$	$class$
$(5, 1)$	1
$(5, -1)$	1
$(7, 0)$	1
$(3, 0)$	-1
$(2, 1)$	-1
$(1, -1)$	-1

c. Determine if the dataset shown in Table 2 is linearly separable.

[2 marks]

QUESTION 2 CONTINUES ON NEXT PAGE

d. Apply the Sequential Perceptron Learning Algorithm to determine the parameters of a linear discriminant function that will correctly classify the data shown in Table 2. Assume an initial value of $\mathbf{a} = (w_0, \mathbf{w}^T)^T = (-25, 5, 2)^T$, and use a learning rate of 1.

[6 marks]

e. A support vector machine (SVM) is to be used to classify the data shown in Table 2.

i. Identify the support vectors by inspection.

[3 marks]

ii. Design a SVM classifier to classify all given points.

[6 marks]

3. a. Give brief definitions of the following terms:

i. an artificial neural network

ii. action potential

iii. firing rate

[6 marks]

Table 3, below, shows a linearly separable dataset.

Table 3

$\mathbf{x}^T$	$class$
$(0, 2)$	1
$(2, 1)$	1
$(-3, 1)$	0
$(-2, -1)$	0
$(0, -1)$	0

b. Is the dataset shown in Table 3:

i. Univariate or multivariate?

ii. Continuous or discrete?

[2 marks]

c. Re-write the dataset shown in Table 3 using augmented vector notation.

[2 marks]

QUESTION 3 CONTINUES ON NEXT PAGE

d. A linear threshold neuron has a transfer function which is a linear weighted sum of its inputs and an activation function that applies a threshold (θ) and the Heaviside function. For a linear threshold neuron with parameter values $\theta = -2$, $w_1 = 0.5$ and $w_2 = 1$ calculate the output of the neuron in response to each of the samples in the dataset given in Table 3.

[4 marks]

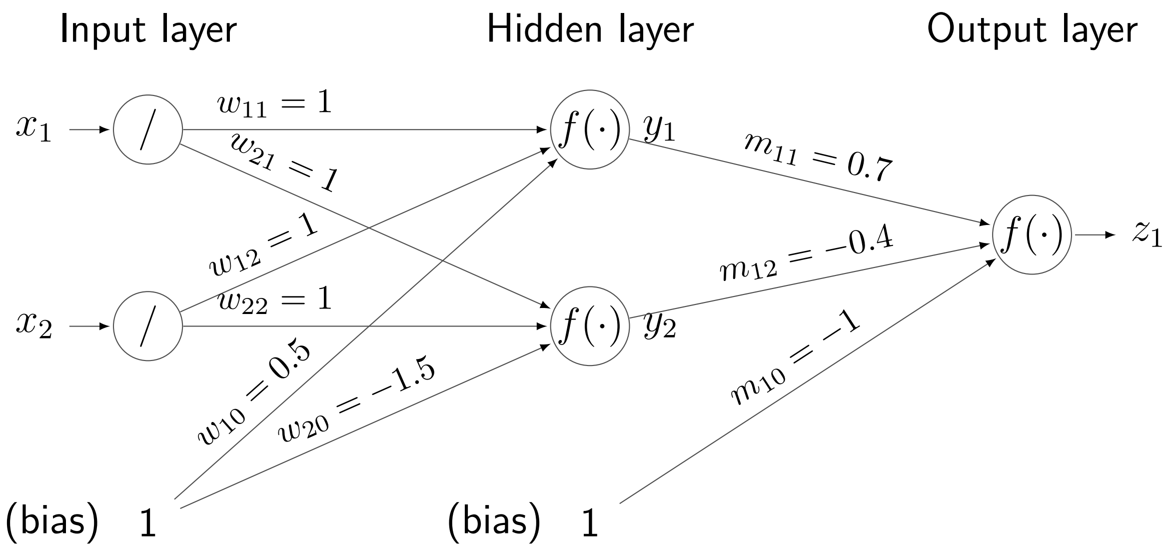
e. Plot the data given in Table 3 and on the same axes plot the decision boundary defined by the linear threshold neuron with parameter values $\theta = -2$, $w_1 = 0.5$ and $w_2 = 1$.

[4 marks]

f. Apply the sequential delta learning algorithm to find the parameters of a linear threshold neuron that will correctly classify the data in Table 3. Assume initial values of $\theta = -2$, $w_1 = 0.5$ and $w_2 = 1$, and a learning rate of 1.

[7 marks]

4. a. A diagram of 2-2-1 fully connected multilayer neural network is shown below, where "/" denotes a linear function, $f(\cdot) = \text{sgn}(\cdot)$ denotes the sign function, $\text{sgn}(a) = \begin{cases} 1 & \text{if } a \geq 0 \\ -1 & \text{if } a < 0 \end{cases}$.



Show that this neural network is able to solve the XOR problem as defined in Table 4.

Table 4

Class	$\mathbf{x}^T = [x_1, x_2]^T$
1	$[-1, -1]$
2	$[-1, +1]$
2	$[+1, -1]$
1	$[+1, +1]$

[10 marks]

QUESTION 4 CONTINUES ON NEXT PAGE

b. A radial basis function (RBF) neural network is to be used to correctly classify the XOR dataset shown in Table 4.

i. There are two hidden units that both employ a Gaussian activation function with a standard deviation of $\sigma_j = \frac{\rho_{\max}}{\sqrt{2n_H}}$. The centres of these hidden units are given by the first and the last input patterns (*i.e.*, the two input patterns in class 1). Compute the outputs at the hidden units for all four input patterns and list them in a table.

[7 marks]

ii. Show that the dataset when represented by the hidden units is linearly separable.

[3 marks]

iii. The RBF network has one output unit which employs a linear activation function, with a threshold. Write down the equation that would be used to determine the weights of this output neuron using the least squares method. Hence, using the answer you obtained for question 4.b.i write an expression in matrix form that could be used to calculate the weights of the output neuron (you do not need to calculate the pseudo-inverse).

[5 marks]

5. a. Give a brief definition of each of the following terms.

i. Feature Selection

ii. Feature Extraction

iii. Dimensionality Reduction

[6 marks]

b. Write pseudo-code for applying Oja's learning rule to find the first principal component of a multi-dimensional data set.

[4 marks]

c. Consider the following 2-dimensional dataset:

$$\mathbf{x}_1 = \begin{pmatrix} 1 \\ 2 \end{pmatrix}, \mathbf{x}_2 = \begin{pmatrix} 3 \\ 5 \end{pmatrix}, \mathbf{x}_3 = \begin{pmatrix} 5 \\ 4 \end{pmatrix}, \mathbf{x}_4 = \begin{pmatrix} 8 \\ 7 \end{pmatrix}, \mathbf{x}_5 = \begin{pmatrix} 11 \\ 7 \end{pmatrix}.$$

i. Calculate the equivalent zero-mean dataset.

[4 marks]

ii. Apply two epochs of the batch version of Oja's learning rule to the zero-mean data calculated in answer to question 5.c.i. Use a learning rate of 0.01 and an initial weight vector of [-1,0].

[7 marks]

iii. Using the answer obtained to question 5.c.ii, project the zero-mean data onto the first principal component.

[4 marks]

6. a. Apart from clustering, name one other application for unsupervised learning in pattern recognition.

[1 marks]

Table 6, below, shows unlabelled 2-dimensional feature vectors which are to be clustered into two classes.

Table 6

i	$\mathbf{x}_i^T$
1	$(1, 0)$
2	$(0, 2)$
3	$(1, 3)$
4	$(3, 0)$
5	$(3, 1)$

b. Apply the k-means algorithm to the dataset shown in Table 6. Use the Euclidean distance criterion and start with initial cluster centres of $\mu_1^T = (3, 2)$ and $\mu_2^T = (4, 0)$. Determine the final cluster centres and the cluster to which each training sample is allocated.

[8 marks]

c. Apply hierarchical agglomerative clustering to the dataset shown in Table 6, in order to find two clusters. Use the Euclidean distance as the similarity metric and determine the distance between clusters using single-link clustering (where the distance between clusters is the minimum distance between constituent samples). Determine the cluster to which each training sample is allocated.

[8 marks]

QUESTION 6 CONTINUES ON NEXT PAGE

d. Apply the competitive learning algorithm (without normalisation) to the dataset shown in Table 6. Perform five iterations of the algorithm with the samples chosen in the order of $\mathbf{x}_1$, $\mathbf{x}_2$, $\mathbf{x}_3$, $\mathbf{x}_4$, $\mathbf{x}_5$. Start with initial cluster centres of $\mathbf{w}_1^T = (1, 1)$ and $\mathbf{w}_2^T = (2, 2)$. Use a learning rate of $\eta = 0.5$. Determine the final cluster centres and the cluster to which each training sample is allocated.

[8 marks]