...g. CSE 8th Semester (21)

19 October 2016

ISLAMIC UNIVERSITY OF TECHNOLOGY (IUT)
ORGANISATION OF ISLAMIC COOPERATION (OIC)
Department of Computer Science and Engineering (CSE)

...ESTER FINAL EXAMINATION — SUMMER SEMESTER, 2015-2016

...RATION: 3 Hours — FULL MARKS: 150

CSE 4835: Pattern Recognition

...rogrammable calculators are not allowed. Do not write anything on the question paper.

There are **8 (eight)** questions. Answer any **6 (six)** of them.

Figures in the right margin indicate marks.

The following figure represents a feed-forward neural network with one hidden layer. 13

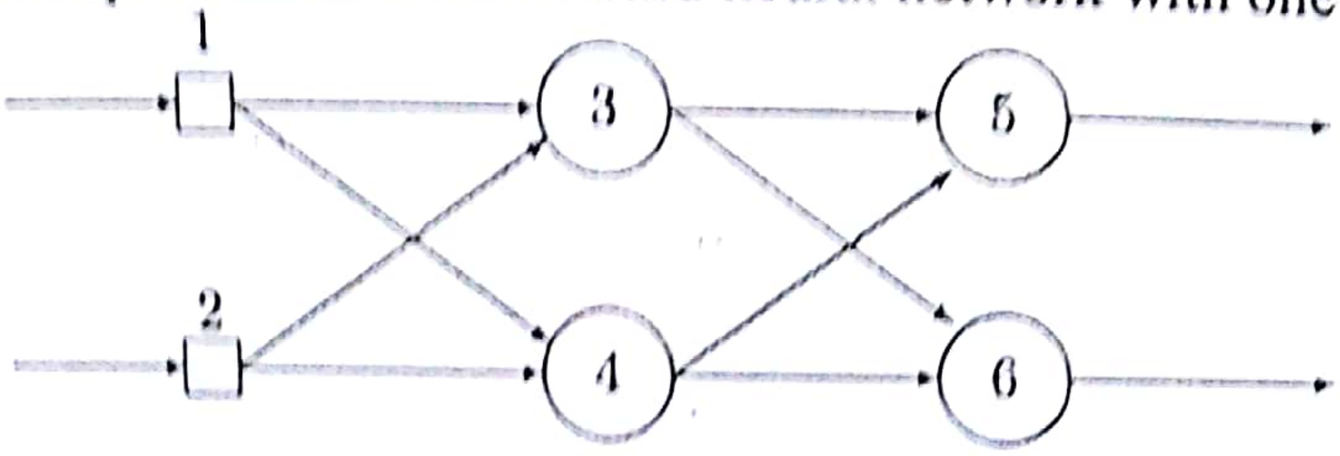

Figure 1: A feed-forward neural network.

A weight on connection between nodes i and j is denoted by w_{ij}, such as w_{13} is the weight on the connection between nodes 1 and 3. The following table lists all the weights in the network.

$w_{13} = -2$	$w_{35} = 1$
$w_{23} = 3$	$w_{45} = -1$
$w_{14} = 4$	$w_{36} = -1$
$w_{24} = -1$	$w_{46} = 1$

Each of the nodes 3, 4, 5 and 6 uses the following activation function:

$$\varphi(v) = \begin{cases} 1 & \text{if } v \geq 0 \\ 0 & \text{otherwise} \end{cases}$$

where v denotes the weighted sum of a node. Each of the input nodes (1 and 2) can only receive binary values (either 0 or 1). Calculate the output of the network (y_5 and y_6) for each of the following input patterns.

Pattern:	P_1	P_2	P_3	P_4
Node 1:	0	1	0	1
Node 2:	0	0	1	1

b) The XOR function (exclusive or) returns true only when one of the arguments is true and another is false. Otherwise, it returns always false. Do you think it is possible to implement this function using a single unit? How about using a network of several units? 7

c) What is a training set and how is it used to train neural networks? 5

a) What is the primary objective of a Support Vector Machine (SVM)? Explain with appropriate example. 5

b) In case of a Support Vector Machine, show that $m = 2/|w|$. Here, m represents the margin and w represents the weight vector. Furthermore, explain how you can find the solution vector w and bias b in SVM. 13

c) Suppose out of a sample of 100 patients, half of them are actually sick and the other half are healthy. Now, a doctor X correctly identifies only 80% of the sick persons and altogether he identifies 60 persons as sick. Another doctor Y finds only 20% of the total sample as healthy which is correctly diagnosed. Calculate the confusion matrix for both doctors X and Y.
Further, draw the ROC graph for X and Y and comment on the performance of them.

3. a) Why is Principal Component Analysis (PCA) used for? In which situation PCA is better than Linear Discriminant Analysis (LDA)?
b) The dataset given in the following table represents the height and weight of 20 women. Show the step-by-step calculation to derive the new dataset by applying PCA. Assume that the feature vector consists of both eigenvectors.

Table: Dataset for Question 3.(b)

height	weight
57	93
58	110
60	99
59	111
61	115
60	122
62	110
61	116
62	122
63	128
62	134
64	117
63	123
65	129
64	135
66	128
67	135
66	148
68	142
69	155

4. a) In Parzen window method, can we use a Gaussain window function instead of a rectangular window function $\varphi(u)$ for calculating k_n? If, yes, then define a Gaussian window function and explain the corresponding effects on the final density estimate of $p_n(x)$.
b) Why is a margin included in the weight space of possible solutions? Explain with figures.
c) Suppose the following criterion function $J(w)$ is to be minimized using Gradient Descent technique:

$$J(w)=\frac{1}{2}\sum_{y\in M}\frac{(w^t y-b)^2}{\|y\|^2}$$

where, w is the solution weight vector along with bias b and y is a sample from the misclassified set M. Provide the iterative update equation for finding the desired solution vector w. What are the stopping conditions?
d) How many ways can you devise multicategory classifiers employing linear discriminant functions? For each of the designs, state the limitations with appropriate illustrations.

Define the criterion function J in Linear Discriminant Analysis (LDA) for a two-class problem. How are the inter-class separation and intra-class scatterness incorporated in this function? Explain them. 7

Extend the design of the criterion function J for multi-class problem. How do you find the desired projection lines maximizing J? Explain which lines you have picked for projection. 6

In a human activity recognition (HAR) system, how can you incorporate LDA in order to improve your system classification performance? Describe with necessary illustrations. Assume that your system will classify 5 different human activities. Clearly define your system input data. 12

Consider a two-class one-feature classification problem with the following Guassian class-conditional densities: $P(x \mid \omega_1) \equiv N(0,1)$ and $P(x \mid \omega_2) \equiv N(1,2)$.

Assume $P(\omega_1) = P(\omega_2) = 1/2$, and a 0-1 loss function.

i. Derive the Bayes Decision Boundary Equation for classification. 7

ii. Is this classifier a linear machine? Justify your answer. 4

iii. Now keep $P(\omega_1) = P(\omega_2) = 1/2$, but change the loss function as 6

$\lambda_{11} = \lambda_{22} = 0, \lambda_{12} = 2, \lambda_{21} = 1$. Approximately show the effect on class regions and classification errors.

a) Prove that the minimum distance to class-member classifier is nonlinear. 8

a) Define the log-likelihood function $l(\theta)$ in Maximum Likelihood Estimation (MLE) method. Show how $l(\theta)$ is maximized to find the best estimates of $\theta = \{\mu, \sigma^2\}$ for the univariate case. Find the estimated values too. 1+8

b) In nonparametric density estimate techniques you can use k-nearest neighbor method to estimate $p_n(x)$. But how can you estimate the likelihood probability of a class $p(x|\omega)$ using the same technique? 5

c) What are the differences between Maximum Likelihood Estimation (MLE) and Bayes Parameter Estimation methods? 6

d) Derive the equation for finding the distance r from a sample point x to the decision boundary plane H. How can you find the position of the origin by simply seeing the equation of a decision boundary? 3+2

a) Suppose Islamic University of Technology (IUT) wants to implement a biometric security system, which will authenticate and allow an authorized person to enter through the university main gate. All unauthorized personnel are not allowed to enter and the security team is alerted. If the biometric validation is to be done based on finger print (Right hand thumb only) and you have been asked to design such system, then briefly describe the components of the pattern classification system as per your design. [Note: Cover all aspects as much as possible including risks] 15

b) Define class model. How does a model change with the change of a classification problem but with the same input data? Explain with examples. 1+5

c) Define the generalized linear discriminant function $g(x)$ for linear machine and fit this discriminant function into a simple two-layer feed-forward neural network (NN) model with appropriate figures. 1+3