Course Code : CST 414-2 **MQNR/MW – 19 / 9658**

Seventh Semester B. E. (Computer Science and Engineering) Examination

Elective – II

MACHINE LEARNING

Time : 3 Hours] [Max. Marks : 60

Instructions to Candidates :—

(1) Assume suitable data wherever necessary.
(2) Each question carries marks as indicated against them.

1. (a) Consider the following set of training examples :

Instance	A1	A2	Classification
1	T	T	Yes
2	T	T	Yes
3	T	F	No
4	F	F	Yes
5	F	T	No
6	F	T	No

(i) What is the **entropy** of this collection of training examples with respect to the target function classification ? Comment on entropy of this example.

(ii) What is the **information Gain** of A1 relative to these training examples ?

(iii) What is the **information Gain** of A2 relative to these training examples ?

(iv) Construct a decision tree for given dataset. 7 (CO 1)

(b) Calculate the size of instance space and hypotheses space for dataset in Q. 1 (a). More generally, how does the number of possible instances and hypotheses grow with the addition of a new attribute A that takes on k possible values ? 3 (CO 1)

OR

(c) Why do we need "Confusion matrix" for evaluation of a machine trained for a pattern recognition task ? In a two class problem, There are 60 samples of each class. A classifier classifies 45 of the As as As, end all 60 of the Bs as Bs. Form an Integer and Fractional Confusion matrix for these results. Also calculate the accuracy of the classifier. 3 (CO 1)

2. (a) Write distance weighted K–Nearest Neighbor algorithm for approximating discrete valued and real valued target function. How can we determine the best value of k ? 5 (CO 2)

(b) Consider training a two input perceptron. Give an upper bound on the number of training examples sufficient to assure with 90% confidence that the learned perceptron will have true error of at most 5%. Does this bound seem realistic ? 5 (CO 2)

OR

(c) Describe Radial Basis Function (RBF) network. Solve XOR classification problem using RBF network. 5 (CO 2)

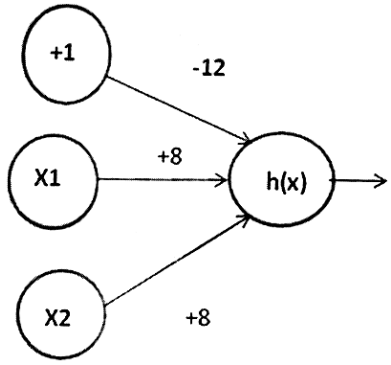
3. (a) Fill in the Blanks :—

(i) The classification boundary realized by the perceptron is a _____.

(ii) A perceptron has two inputs x_1 and x_2 with weights w1 and w2 a bias weight of w_0. The activation function of the perceptron is h(x). The output y of the perceptron is given by : y = _____.

(iii) We provide a training input x to a perceptron learning rule. The desired output is t and the actual output is o. If learning rate is η, the weight update performed by the learning rule is given by the equation _____.

(iv) The neural network given below takes two binary valued inputs x_1, $x_2 \in \{0, 1\}$ and the activation function is the binary threshold function ($h(x) = 1$ if $x > 0$; 0 otherwise). It computes Logical function _____.



(v) The back propagation learning algorithm is also known as generalized _____ Rule. 5 (CO 2)

(b) Describe the Recurrent Neural network with neat sketch and its applications. 5 (CO 2)

4. (a) Diabetic Retinopathy is a disease that affects 80% people who have diabetes for more than 10 years. 5% of the Indian population has been suffering from diabetes for more than 10 years. What is the joint probability of finding an Indian suffering from Diabetes for more than 10 years and also has Diabetic Retinopathy ? 2 (CO 3)

(b) You are given the following set of training examples. Each attribute can take on one of three nominal values : a, b or c.

A1	A2	A3	Class
a	c	a	C1
c	a	c	C1
a	a	c	C2
b	c	a	C2
c	c	b	C2

How would a Naive Bayes classifier classify the example : $< A1 = a,\ A2 = c,\ A3 = b >$? 4 (CO 3)

(c) Describe brute force MAP learning algorithm. 4 (CO 3)

5. (a) Assume the following dataset is :

Given (2, 2), (4, 4), (5, 5), (6, 6), (9, 9), (0, 4), (4, 0).

K – means is run with k = 3, to cluster the dataset. Use Manhattan distance as similarity measure.

K – means initial clusters C1, C2 and C3 are as follows :

C1 : {(2, 2), (4, 4), (6, 6)}

C2 : {(0, 4), (4, 0)}

C3 : {(5, 6), (9, 9)}

Now, K – means is run for a single iteration : evaluate the new clusters and predict their centroids ? 5 (CO 3)

(b) Perform a hierarchical clustering of the data in Fig. 5 (b) using the single linkage algorithm and Euclidean distance. Show the distance matrices and the dendrogram.

Sample	x	y
1	0·0	0·0
2	0·5	0·0
3	0·0	2·0
4	2·0	2·0
5	2·5	8·0
6	6·0	3·0
7	7·0	3·0

Fig. 5 (b) 5 (CO 3)

6. (a) What do you mean by ensemble Learning ? Describe Bagging and Boosting methods with suitable example. Why Boosting is preferred over bagging ? 5 (CO 4)

(b) In a support vector machine, suppose we only have four training examples in two dimensions, positive example at x1 = [0, 0], x2 = [2, 2] and negative examples at x3 = [h, 1], x4 = [0, 3], where we treat $0 \leq h \leq 3$ as a parameter :

(1) How large can $h \geq 0$ be so that the training points are still linearly separable ?

(2) Does the orientation of the maximum margin decision boundary change as a function of h when the points are separable (Yes/No) ? Comment. 5 (CO 4)