Course Code : CST 404 – 2 / CST 414 – 2 **EVFU/MW – 18 / 6107**

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

Elective – II

MACHINE LEARNING

Time : 3 Hours] [Max. Marks : 60

Instructions to Candidates :—

(1) All questions carry marks as indicated against them.
(2) Assume suitable data wherever necessary.
(4) Illustrate your answers with neat sketches wherever necessary.

1. Attempt any **Two** questions :—

(a) Consider the following set of Training examples :—

Instance	X	Y	Z	Class
1	1	1	1	I
2	1	1	0	I
3	0	0	1	II
4	1	0	0	II

(i) What is the entropy of this collection of training examples with respect to the target function class ?

(ii) What is the information gain of attribute X , Y and Z relative to these training examples ?

(iii) What is over fitting in decision tree learning ? List the methods used to handle the problem of over fitting. 5 (CO 1)

(b) Apply Candidate Elimination Algorithm on following examples, for classifying poisonous mushrooms.

Features :—

- color {red, brown, gray}
- size {small, large}
- shape {round, elongated}

- land {humid, dry}
- air humidity {low, high}
- texture {smooth, rough}

The set of training Examples of D :
((red, small, round, humid, low, smooth) , poisonous)
((red, small, elongated, humid, low, smooth) , poisonous)
((gray, large, elongated, humid, low, smooth) , not – poisonous)
((red, small, elongated, humid, high, rough) , poisonous) 5 (CO 1)

(c) Calculate the size of hypothesis space in the above Q. 1 (b) dataset for mushrooms classification. 5 (CO 1)

2. Attempt any **Two** questions :—

(a) Write distance weighted K – NN algorithm for discrete valued and real valued target function. 5 (CO 2)

(b) Prove that the Conjuction of Boolean literals are PAC lernable. 5 (CO 2)

(c) What is the VC dimension for following cases :— 5 (CO 2)

(i) K – Nearest Neighbor classifier when k = 1 ? Why ? 1

(ii) Linear Support Vector Machines in d – dimensional space ? Why ? 1

(iii) Square in 2 – dimensional space. 2

(iv) Rectangle in 2 – dimensional space. 1

3. Attempt any **Two** questions :—

(a) Given a two input neuron with vector W = [3 ; 2] and input vector X = [– 5 ; 7] , we would like to have an output of 0.5.

(i) What is the bias weight if linear transfer function is used ?

(ii) What is the bias weight if unipolar sigmoid transfer function is used ? 5 (CO 2)

(b) Derive Gradient Descent rule in a single neuron with activation as **unipolar sigmoid** function. 5 (CO 2)

(c) What are the values of weights w1 , w2 and w3 for the perceptron whose decision surface is illustrated in figure 3 (c) ? Assume the surface crosses the x1 axis at – 1 and x2 axis at 2.

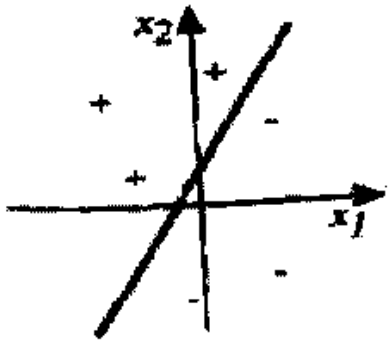


fig 3(c)

5 (CO 2)

4. Answer the following questions :—

(a) Describe Bayes rule. Explain approximations that lead to Naive Bayes classifier. Why Naive Bayes classifier is called "Naive" ? Consider following Dataset and apply Naive Bayes classifier to predict if Bob will Default his loan, for instance :—

Bob :< Homeowner = No , MaritalStatus = Married,
JobExperience = 3>??

Home Owner	Marital Status	Job Experience	Defaulted
Yes	Single	3	No
No	Married	4	No
No	Single	5	No
Yes	Married	4	No
No	Divorced	2	Yes
No	Married	4	No
Yes	Divorced	2	No
No	Married	3	Yes
No	Married	3	No
Yes	Single	2	Yes

5 (CO 3)

(b) What is the significance of minimum description length principle in machine learning ? Derive an expression for MDL. 5 (CO 3)

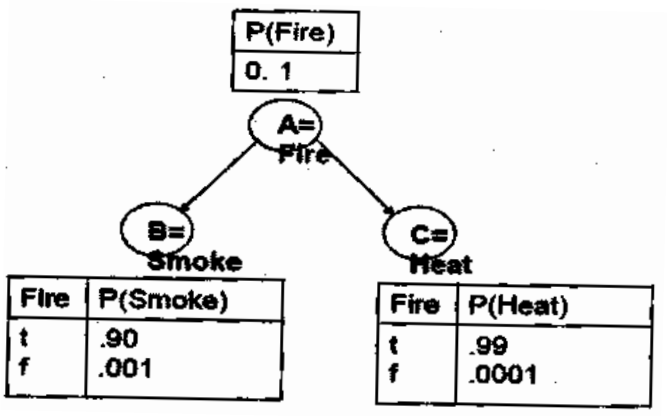
5. Answer the following questions :—

(a) Given the data points :—

X1	2	2	8	5	7	6	1	4
X2	10	5	4	8	5	4	2	9

The task is to cluster these points into three clusters. We initially assign (2 , 10) , (5 , 8) and (1 , 2) as the center of each cluster.
Use K – means algorithm and Euclidean distance as similarity measure. (k = 3). 7 (CO 3)

(b) Consider the following Bayesian network.



Compute P (Smoke = t) ? Where t = true and f – false

Compute P (Fire = t | Smoke = t) ?

3 (CO 3)

6. Answer the following questions :—

(a) Multiple models learned from the data may be combined to improve classification accuracy. Outline the basic features of the following popular ensemble methods :—

(i) Bagging (ii) Boosting and Adaboost 5 (CO 4)

(b) Explain how support vector machine can be used for classification of the data. 5 (CO 4)