

**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.
- (3) Illustrate your answers with neat sketches wherever necessary.

1. Answer the following questions :

- (a) NASA wants to be able to discriminate between Martians (M) and Humans (H) based on the following characteristics: Green $\in \{N, Y\}$, Legs $\in \{2, 3\}$, Height $\in \{S, T\}$, Smelly $\in \{N, Y\}$. Our available training data is as follows.

Green	legs	Height	Smelly	Species??
N	3	S	Y	M
Y	2	T	N	M
Y	3	T	N	M
N	2	S	Y	M
Y	3	T	N	M
N	2	T	Y	H
N	2	S	N	H
N	2	T	N	H
Y	2	S	N	H
N	2	T	Y	H

Greedily learn a decision tree using the ID3 algorithm and draw the decision tree.
8(CO 2,CO 3)

- (b) State true or False, with 1–2 line Justification :
- (i) Over fitting is more likely when the set of training data is small.
 - (ii) Over fitting is more likely when the hypothesis space is small.
 - (iii) The ID3 algorithm is guaranteed to find the optimal decision tree.
 - (iv) K-NN algorithm is easily fooled by irrelevant features.
- 2(CO 2,CO 3)

2. Answer any **Two** questions :—

- (a) "Conjunctions of Boolean literals are PAC Learnable". Justify your answer with a suitable example. 5(CO 1,CO 3)
- (b) Describe Radial Basis function (RBF) network. Solve XNOR classification problem using RBF network. 5(CO 2,CO 3)
- (c) Differentiate between following :—
 - (i) RBF and Multi layer perceptron.
 - (ii) Lazy and Eager Learners. 5(CO 2)

3. Answer any **Two** questions :—

- (a) Derive Gradient Descent rule in a single neuron with activation as unipolar sigmoid function. 5(CO 2,CO 3)
- (b) We take three parameters (features) to differentiate between a **banana** and an **apple** i.e.

Shape : {1 for round; -1 for elliptical}

Texture : {1 for smooth ; -1 for rough}

Weight of fruit : {1 if weight > 100g; -1 weight < 100 g}

The input vector for Apple is {1 1 -1} and Target $t_1 = -1$,

and banana is {-1 1 -1} and Target $t_2 = 1$. The Initial weight vector

$w=[0.5 \ -1 \ -0.5]$ bias weight =0.5, Learning Rate (η) = 1.

Apply perceptron learning Upto one epoch. Use Step (Sign) Activation function.

5(CO 2,CO 3)

- (c) Can a neural network be used to model the following machine learning algorithm ? If so, state the neural network structure and the activation function(s) used. If not, describe why not in one or two sentences.

(i) K-nearest Neighbors.

(ii) Linear Regression.

(iii) Logistic Regression.

5(CO 2,CO 3)

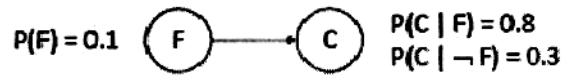
4. Answer any **Two** questions :—

- (a) Describe naïve bayes classifier and apply it on following car theft dataset to Classify the unknown instance $x = \langle \text{Red.SUV,Domestic} \rangle$

Example	Color	Type	Origin	Stolen??
1	Red	Sports	Domestic	Yes
2	Red	Sports	Domestic	No
3	Red	Sports	Domestic	Yes
4	Yellow	Sports	Domestic	No
5	Yellow	Sports	Imported	Yes
6	Yellow	SUV	Imported	No
7	Yellow	SUV	Imported	Yes
8	Yellow	SUV	Domestic	No
9	Red	SUV	Imported	No
10	Red	Sports	Imported	Yes

5(CO 3)

- (b) Describe Brute Force MAP Learning and give its derivation. 5(CO 3)
- (c) Consider the following Bayesian network, where F = having the flu and C = coughing :—



Fig(1)

Write down the joint probability table specified by the Bayesian network for fig(1). 5(CO 3)

5. Answer any **Two** questions :—

- (a) What do you mean by unsupervised Learning ? Perform a hierarchical clustering of the following data using the **average linkage algorithm** and Euclidean distance. Show the distance matrices and the dendrogram.

Sample no.	X	Y
1	4	4
2	8	4
3	15	8
4	24	4
5	24	12

5(CO 2,CO 3)

- (b) Describe Expectation Maximization algorithm. 5(CO2)
- (c) Apply K-means clustering algorithm on following data and identify cluster for each individual. Show the values of the centroids and the distance from the sample to the centroids. Assume K=2 and initial cluster centroids as A(1, 1) and B(2, 1). Show the graphical result of cluster formation.

Sample	X	Y
A	1	1
B	2	1
C	1	2
D	4	3
E	5	4

5(CO2)

6. Answer any **Two** questions :—

- (a) In a support vector machine, suppose we only have four training examples in two dimensions, positive examples at $x_1 = [0, 0]$, $x_2 = [2, 2]$ and negative examples at $x_3 = [h, 1]$, $x_4 = [0, 3]$, where we treat $0 \leq h \leq 3$ as a parameter :—
- (i) How large can $h \geq 0$ be so that the training points are still linearly separable ?
 - (ii) 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 2)
- (b) Describe Hidden Markov model with suitable example. 5(CO 2)
- (c) Write short notes on Bagging and Boosting. 5(CO 1,CO 4)