Total No. of Pages 04 *Roll No. ……*

FIFTH SEMESTER **B. Tech.[CSE]**

END SEMESTER EXAMINATION ***November-2022***

CO327 MACHINE LEARNING

Time: 3:00 Hours ***Max. Marks: 40***

Note: Answer ***ALL*** questions.
Assume suitable missing data, if any.
CO# is course outcome(s) related to the question.

1[a] Determine whether the following tasks belong to supervised learning, unsupervised learning, or reinforcement learning. **[0.5x8=4] [CO1]**

(i) You want to predict whether or not it will be raining at 5pm tomorrow based on previous weather-related database.

(ii) Take a collection of notes written on multiple topics of a subject, and find a way to automatically group these topics into a small number of groups of topics that are somehow "similar".

(iii) You have given a large database of student records which contains students' performance in different subjects, try to learn whether there might be different clusters of such students which needs a separate teaching strategy.

(iv) In stock market, given data on stock prices over the last 25 years, determining the future stock price of a company.

(v) Grouping of pixels in an image with similar colors.

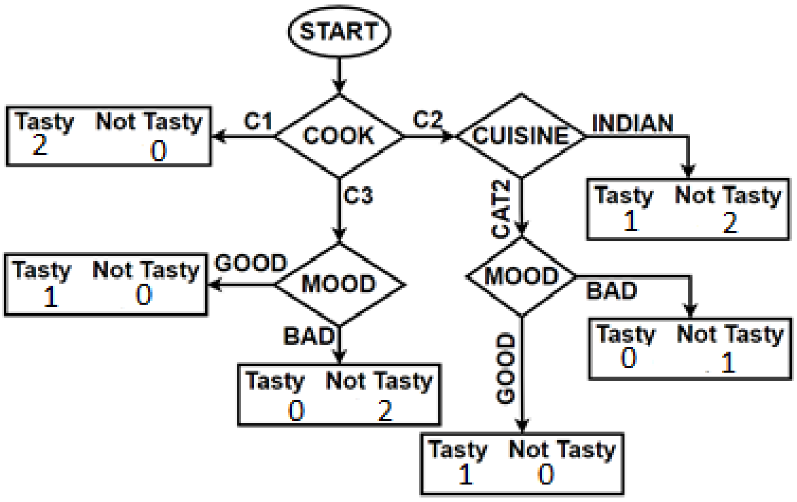
(vi) To train a robot to find a path in a given scenario.

(vii) Inspect a large database of emails that are known to be spam, to discover if there are sub-types of spam.

(viii) Observe the past record and current statistics of two cricket teams, and predicting which team will win in the next match.

[b] The data is collected from a restaurant for food (cuisine type: Indian or continental) made by various cooks (C_1, C_2, C_3). Feedback of customers is taken as tasty or non-tasty along with customers' mood (good or bad). The data is used to build a decision tree, shown in Fig. below. Decision trees are prone to overfitting (explain why? In maximum 2 sentences). Pruning is one way to deal with overfitting. Let the classification error for the given decision tree is 0.31. If we decide to build a decision tree up to level 2 only (consider root node as level 0), the classification error is 0.37. Let the value of the tuning parameter $\lambda = 0.03$. Explore, whether pruning at deepest node is advisable or not? Also, discuss the effect of tuning parameter *i.e.* $\lambda = 0$ and $\lambda = \infty$. [Hint: Cost function is a combination of classification error and complexity of the decision.]

[1+2+1] [CO2]



2. Answer *any TWO* of the followings

[a] In a sports competition of a university, the management want to design a machine learning based model to select team members. The data for the previous selection of athletes is given in Table I.

[4] [CO1, CO2]

Table I

S. No.	1	2	3	4	5	6	7	8	9	10
Speed	2.50	3.75	7.00	4.25	7.25	4.50	5.25	7.50	5.50	4.25
Agility	6.00	8.00	4.25	3.75	5.75	5.00	9.50	8.00	6.75	3.75
Draft	No	No	Yes	No	Yes	No	Yes	Yes	Yes	No

Use k-nearest neighbors (KNN) with k = 3, and 5 to predict whether the player with speed 6.75 and agility 3.0 will be selected or not. Use Euclidean distance.

[b] Support vector machines (SVM) consider a margin between the classes to improve the classification. If a hyper plane $\mathbf{w}^T\mathbf{x} + w_0 = 0$ is decision boundary between two classes, suggest (with proper justification) an objective function with suitable constraints for SVM to get the desired margin. Also, explain how the logistic regression is different from SVM (use maximum 5 sentences/expressions). **[3+1] [CO2]**

[c] An agency collected the data for crop health across the country as given in Table II. It has three attributes: Season = summer (S) or winter (W), Region = plane (P) or hilly (H), Seed type = traditional (T) or genemodified (GM). Predict the heath of crop (Y = heathy, N = not healthy) of hilly region in summer season with tradition seed using naïve Bayes classifier. Show all the steps. **[4] [CO2, CO3]**

Table-II

S. No.	1	2	3	4	5	6	7	8	9	10
Season	S	S	S	W	W	W	W	W	S	S
Region	P	P	P	P	P	H	H	H	H	P
Seed Type	T	T	T	T	GM	GM	GM	T	GM	GM
Healthy	Y	N	Y	N	Y	N	Y	N	N	S

3. Answer *any TWO* of the followings

[a] Find the cluster using single link technique for data shown in Table III. Use Euclidean distance, and draw dendogram. **[4] [CO2]**

Table III

Label	A	B	C	D	E
X	0.22	0.27	0.18	0.35	0.25
Y	0.38	0.19	0.41	0.12	0.35

[b] Find the clusters with the help of k-means clustering algorithm for the data shown in Table IV. Consider A and B as the initial centroids, and use Euclidean distance. Solve for maximum three iterations. **[4] [CO2]**

Table IV

Label	A	B	C	D	E
X	168	179	182	187	170
Y	60	68	76	77	59

[c] Explain the concept of soft assignment in clustering. Discuss the limitations of *k*-means clustering algorithm and how they are addressed in expectation maximization (EM) approach of clustering.

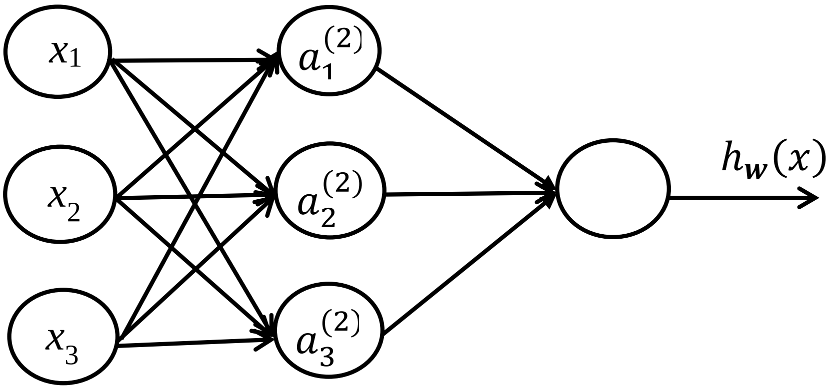
[1+1+2] [CO2]

4. Answer *any TWO* of the followings

[a] Use perceptron learning rule to train the network. The input training vectors are as follows: $X_1 = [1 \;\; -2 \;\; 0 \;\; 1]^T$, $X_2 = [0 \;\; 1.5 \;\; -0.5 \;\; -1]^T$ and $X_3 = [-1 \;\; 1 \;\; 0.5 \;\; -1]^T$. The initial weight vector is $[1 \;\; -1 \;\; 0 \;\; 0.5]^T$, The learning rate is 0.1 and the desired outputs are $d_1 = -1$, $d_2 = -1$ and $d_3 = 1$. Calculate the weight after one complete cycle. The activation function is given by: **[4] [CO2]**

$$s(t) = \begin{cases} 1, & t \geq 0 \\ -1, & otherwise \end{cases}$$

[b] The Fig. below shows a neural network. Compute the output at $a_2^{(2)}$ and $h_w(x)$ by using feed-forward approach. Further, obtain the expression of gradients using backpropagation at $a_2^{(2)}$. Assume the sigmoid activation function. **[4] [CO2]**



[c] A dataset shown in Table V, contains four samples {S_1, S_2, S_3, S_4} having two features {X_1, X_2}. Compute the principal components of dataset. Justify that first principal component is the direction of maximum variance (in Maximum one page of answer sheet). **[4] [CO2]**

Table V

Feature	S_1	S_2	S_3	S_4
X_1	4	8	13	7
X_2	11	4	5	14

5[a] What is the role of exploration and exploitation in reinforcement learning. Write one step for exploration and exploitation for the following cases: **[4] [CO1, CO3]**

(i) Restaurant Selection
(ii) Online Advertisements
(iii) Oil Drilling
(iv) Chess Playing

[b] Table VI shows a dataset and response of a probabilistic classifier F. Classifier F is binary classifier with class levels 0 and 1. F gives the probability of a sample x belonging a class 1 i.e. p(C=1|x). Draw the ROC curve with four points corresponding to thresholds = 0.1, 0.5, 0.7, 0.8. **[4] [CO2]**

Table VI

S. No.	1	2	3	4	5	6	7	8	9	10
Class C (Actual)	0	1	1	0	1	0	0	0	1	0
p(C=1|x)	0.98	0.67	0.58	0.18	0.55	0.86	0.39	0.89	0.82	0.26

----Best of Luck----