Total No. of Pages *Roll No. ……*

FIFTH SEMESTER **B.Tech. (CSE)**

MID SEMESTER EXAMINATION *September-2022*

CO327 MACHINE LEARNING

Time: 1:30 Hours ***Max. Marks: 20***

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

1[a] A tollbooth collects the data of various cars passing through it. The following attributes are recorded: speed of the car, gender of the driver, time of arrival, car registration number, age bracket of the driver (young, middle, old), number of co-passengers, and driving license number. The booth operators want to design a machine learning (ML) model to predict the speed of cars using this model. Identify the features that can be used for ML model design. Also, identify whether it is a classification problem, regression problem, or none. [Give one-line justification (not more) for each selection/answer] **[1+1] [CO1] [L2]**

[b] The probability distribution $f(X)$ of a random variable X is given in Table. I. Compute the mean and variance of X. **[1+1] [CO2] [L3]**

Table. I

X	0	1	2	3
$f(X)$	1/7	3/7	2/7	1/7

2[a] A travel agency wants an automated system to predict travel costs. The agency has the following data available with it.

Table II

S. No.	Distance (in Km)	Travelling Cost (in Rupees)
1	1	2.75
2	2	3.5
3	3	4.25
4	4	5
5	5	5.75

Formulate the above problem as a linear model $h(x) = w_0 + w_1 x$ to predict the travelling cost for a given distance. The parameter w_0 is 2 (optimal). Apply gradient descent algorithm to find optimal parameter w_1. The learning rate for the first epoch is 0.073, and for the second epoch and later, the learning rate is 0.091. Let the initial value of w_1 is 0.5.

[4] [CO1, CO2] [L3]

[b] In logistic regression, binary cross-entropy is used as the cost function for two-class classification. Illustrate (considering one sample) that the cost function will have a single optimum so that the gradient descent algorithm converges to the global optima. **[3] [CO2] [L4]**

3[a] A factory is producing papers. The quality control unit applies two types of testing (durability test and strength test) to assess paper quality. The data for the same is given below:

Table III

S. No.	1	2	3	4	5	6	7	8
Durability	7	6	7	6	3	1	4	3
Strength	7	4	4	5	4	4	3	5
Quality	Good	Bad	Good	Good	Bad	Bad	Bad	Bad

In general, the factory produces 720 good quality papers out of 1000. Use k-nearest neighbor (KNN) with k = 1, and 3 to predict the quality of a new paper (durability = 5, strength = 5). **[2+1] [CO3] [L3]**

[b] Now, suppose (in above question 3[a]), we define some distance-based probabilistic classifier instead of KNN. The likelihood of belonging to a class for a new sample is $1/d$. Here d is the Euclidian distance of a new sample from nearby samples of the same class. If there are multiple neighbouring samples of a class, the overall likelihood is calculated by the union of all likelihoods. Assume a cutoff distance d_{cf}; beyond that, no sample is considered in calculating overall likelihood. [Hint: $p(\text{A}\cup\text{B}) = p(\text{A}) + p(\text{B}) - p(\text{A}\cap\text{B})$]

Consider d_{cf} is the maximum distance of new sample (durability = 5, strength = 5) from other samples in 3[a] for KNN with k = 3. Predict the quality of a new paper (durability = 5, strength = 5) using posterior probability. Also, compare the performance of this probabilistic classifier with KNN {Maximum two sentences}. **[2+1] [CO3] [L3, L5]**

4 A career counselling agency wants an automated system to advise for MS programs. It has previous data (given in Table IV) of students who have succeeded or failed in MS programs. The data contains two attributes of each student: CGPA (High, Medium, Low) and whether or not they have published a good research paper (Yes, No). An ML engineer is hired to develop such a system. He thought of applying a decision-tree algorithm but wanted a new criterion of data division (in subsets). He got an idea for the same, inspired by the F1 score. In the F1 score, he replaced precision and recall with the two classes (succeed, failed) probabilities and named it the G1 score. Apply this newly defined G1 score and develop a full decision tree. [Use the weighted average of G1 scores of subsets to compare with the G1 of the original set (before division)]. **[3] [CO1, CO2] [L3]**

Table IV

S. No.	CGPA	Publication	Result (MS)
1	Low	No	Failed
2	Low	Yes	Succeed
3	Medium	No	Failed
4	Medium	Yes	Succeed
5	High	No	Succeed
6	High	Yes	Succeed

---Best of Luck---