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 $h(x) = w_0 + w_1x$ 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: $p(A \cup B) = p(A) + p(B) - p(A \cap B)$]

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---

Name- Hemant Goyal
Roll no - 2K21/CO/196

Date:
Page No.:

Assignment-I

Ans 1 (a) The features that can be used for ML model design are

- → Speed of car (target variable)
- → gender of the driver
- → time of arrival
- → age bracket of the driver
- → no of copassenger.

- car registration number and driving licence no are excluded because they are unique values so it is not helpful for model design.

- It is a regression problem because our goal is to predict the speed as it involves predicting a continuous numerical value.

(b)

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

Mean- $E(x) = x f(x)$

$$Ex = 0 \times \frac{1}{7} + 1 \times \frac{3}{7} + 2 \times \frac{2}{7} + 3 \times \frac{1}{7}$$

$$= 0 + \frac{3}{7} + \frac{4}{7} + \frac{3}{7} = \frac{10}{7}$$

$$\text{variance} = x^2 f(x) - E(x)^2$$

$$= 0^2 \times \frac{1}{7} + 1^2 \times \frac{3}{7} + 2^2 \times \frac{2}{7} + 3^2 \times \frac{1}{7} - \left(\frac{10}{7}\right)^2$$

$$= 0 + \frac{3}{7} + \frac{8}{7} + \frac{9}{7} - \frac{100}{49}$$

$$= \frac{20}{7} - \frac{100}{49}$$

$$= \frac{140 - 100}{49} = \frac{7}{49} = \frac{1}{7} \text{ Ans}$$

Ans 2 (a) Initial $w_1 = 0.5$ — The parameter $w_0 = 2$ (optimal).

Linear model $h(x) = w_0 + w_1 x$

Calculation for prediction

for $x = 1 \Rightarrow 2 + 0.5 \times 1 = 2.5$
$x = 2 \Rightarrow 2 + 0.5 \times 2 = 3$
$x = 3 \Rightarrow 2 + 0.5 \times 3 = 3.5$
$x = 4 \Rightarrow 2 + 0.5 \times 4 = 4$
$x = 5 \Rightarrow 2 + 0.5 \times 5 = 4.5$

Calculation of error

actual − Predicted

for $x = 1 \Rightarrow 2.75 - 2.5 = 0.25$
$x = 2 \Rightarrow 3.5 - 3 = 0.5$
$x = 3 \Rightarrow 4.25 - 3.5 = 0.75$
$x = 4 \Rightarrow 5.0 - 4.0 = 1$
$x = 5 \Rightarrow 5.75 - 4.5 = 1.25.$

Date:
Page No.:

→ Updating w_1 after firs epoch.

$$w_1 = w_{int} - \underset{rate}{learing} \times \frac{1}{n}\Sigma(error \times x)$$

n = no of data points.

$$w_1 = 0.5 - 0.073 \times \frac{1}{5}(0.25\times1 + 0.5\times2 + 0.75\times3 + 1\times4 + 1.25\times5)$$

$$w_{1(1)} = 0.3 - 0.073 \times 2.95$$

$$w_{1(1)} = 0.3 - 0.21535$$

$$w_{1(1)} = 0.28465$$

So for second epochs.

$$w_{1(2)} = w_{1(1)} - \cancel{0.21535}\ 0.091 \times 2.95$$

$$w_{1(2)} = 0.28465 - 0.26845$$

$$= 0.0162 \quad \underline{Ans}$$

(b) Consider a single data sample with a binary classification problem. We aim to predict whether the sample belongs to class 0 or class 1.

$$P(Y=1/x) = \sigma(wx+b)$$

Binary cross entropy function.

$$J(w,b) = -y * \log(P(Y=1/x)) + (1-y)\log(1-P(y=1/x))$$

y is the actual binary label (0, 1)

$P(Y=1/x)$ is the predicted probability

The binary cross-entropy cost function in logistic regression has a single global optimum because of its convex nature. This property ensures that when using gradient descent for parameter optimization, the algorithm will converge to the global optimum.

Ans3(a) For KNN we will find the distance and Rank for all the papers.

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
distance	2.82	1.41	2.23	1	2.23	4.123	2.23	2
Rank.	7	2	4	1	5	8	6	3
	Good	Bad	Good	Good	Bad	Bad	Bad	Bad.

Date:
Page No.:

[To Predict k = 1 and 3 with (durability = 5 strength = 5)]

for SNo (1)

distance = $\sqrt{(7-5)^2 + (7-5)^2} = \sqrt{4+4} = \sqrt{8}$

for (2) distance = $\sqrt{(6-5)^2 + (6-4)^2} = \sqrt{2}$

for (3) distance = $\sqrt{(7-5)^2 + (5-4)^2} = \sqrt{3}$

for (4) distance = $\sqrt{(6-5)^2 + (5-5)^2} = 1$

for (5) distance = $\sqrt{(3-3)^2 + (5-4)^2} = \sqrt{3}$

for (6) distance = $\sqrt{(5-1)^2 + (3-4)^2} = \sqrt{16+1} = \sqrt{17}$

for (7) distance = $\sqrt{(5-4)^2 + (5-3)^2} = \sqrt{1+4} = \sqrt{3}$

for (8) distance = $\sqrt{(3-3)^2 + (5-5)^2} = \sqrt{4} = 2$

for k = 1 Its predicted it is of Good quality.

for k = 3 we will look for top 3 rank so it predict Bad.

Scanned with CamScanner

(b) So lets calculate the probabilistic classifier Ans 4

so for good classes

- So for SN_0 3 likelihood $= \frac{1}{d} = \frac{1}{\sqrt{5}}$

for Sno 1 likelihood $= \frac{1}{\sqrt{8}}$

for Sno 4 likelihood $= 1$

Overall likelihood

So $P(Good) = \frac{1}{\sqrt{5}} + \frac{1}{\sqrt{8}} + 1$

$= \frac{1}{2.23} + \frac{1}{2.82} + 1 = 0.44 + 0.35 + 1$

$= 1.79$

So $P(Bad) = 1 - P(Good) = 0.79$.

so the posterior probability of a new paper being of good quality is 1.79 and of it being a bad quality is 0.79.

Date:
Page No.:

Ans 4 Given

S.No	CGPA	Publication	Result
1	Low	No	F
2	Low	Yes	S
3	Medium	No	F
4	Medium	Yes	S
5	High	No	S
6	High	Ye	S.

Decision - Tree.

let Succeed = +1
Failed = - 1.

Step - ① Calculate Information Gain of CGPA.
④

① ~~Std~~ Entropy of Entire Data set.

$$S\{+4, -2\} = -\frac{4}{6}\log\frac{4}{6} - \frac{2}{6}\log\frac{2}{6}$$

4 → passed
2 → Failed

$$= -\frac{2}{3}\times\frac{\log 0.66}{\log 2} - \frac{1}{3}\frac{\log 0.33}{\log 2} = 0.4 + 0.52$$
$$= 0.92$$

② Entropy of all attributes.

① Entropy of Low $\{+1, -1\} = -\frac{1}{2}\log\frac{1}{2} - \frac{1}{2}\log\frac{1}{2}$
$= +\log 2 = 1$

② Entropy of Medium $\{+1, -1\} = -\frac{1}{2}\log\frac{1}{2} - \frac{1}{2}\log\frac{1}{2} = 1.$

③ Entropy of High $\{+2, 0\} = -\frac{2}{2}\log\frac{2}{2} - \frac{0}{2}\log\frac{0}{2} = 0$

Information Gain = Entropy (whole Data) $- \frac{2}{6}$ Ent(L)
$- \frac{2}{6}$ Ent(M) $- \frac{2}{6}$ Ent(H)

$$= 0.92 - \frac{2}{6}(1) - \frac{2}{6}(1) = 0.96 - 0.66$$
$$= 0.30.$$

Now calculate Information Gain of Publication.

① Entropy of entire data set.

$$S\{+4,-2\} = 0.92 \quad \text{(As calculate previously)}$$

② Entropy of all attributes:

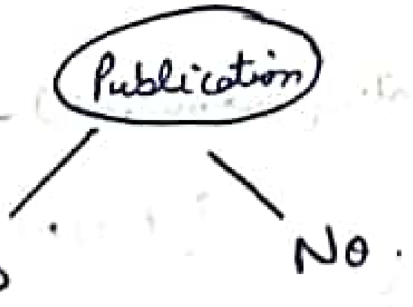
$$\text{Entropy of Yes } \{+3,0\} = -\frac{3}{3}\log\frac{3}{3} - \frac{0}{3}\log\frac{0}{3} = 0$$

$$\text{Entropy of No } \{+1,-2\} = -\frac{1}{3}\log\frac{1}{3} - \frac{2}{3}\log\frac{2}{3} = 0.92$$

$$\text{Information Gain} = \text{Entropy}\{\text{whole Data}\} - \frac{3}{6}\text{Ent(Yes)} - \frac{3}{6}\text{Ent(No)}$$

$$= 0.92 - \frac{1}{2}(0.92) = 0.46$$

As IG of Publication is greater, Hence Publication will become Root Node.



Now will ~~will~~ calculate IG for Yes & No separately.

Yes

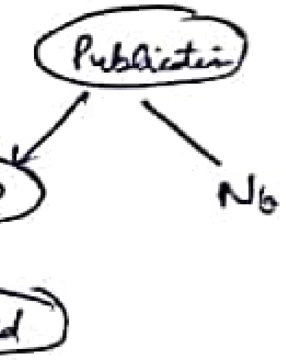
Low	Yes	Succ
Medium	Yes	Succ
High	Yes	Succ

No.

Low	No	Fail.
Med	No	fail
High	No	Succ.

① For Yes, ~~Greater~~ All the results are yes

Hence, Yes leads to success.



② For No, Calculate IG for CGPA,

Entire entropy of Data Set,

$$S\{+1,-2\} = -\frac{1}{3}\log\frac{1}{3} - \frac{2}{3}\log\frac{2}{3} = 0.92$$

⇒ Entropy of all attributes:

① Entropy of Low $\{0,-1\} = -\frac{0}{1}\log\frac{0}{1} - \frac{1}{1}\log\frac{1}{1} = 0$

Med $\{0,-1\} = -\frac{0}{1}\log\frac{0}{1} - \frac{1}{1}\log\frac{1}{1} = 0$

High $\{+1, 0\} = 0.$

Hence Final decision Tree,