VIT-AP UNIVERSITY	Continuous Assessment Test – Fall semester (2024-25) -August 2024	
	Maximum Marks: 50	Duration: 90 Mins
Course Code: CSE3008	Course Title: Introduction to Machine Learning	
Set No:	Exam Type**: Closed Book**	School: SCOPE
Date:	Slot:	Session:

Keeping mobile phone/smart watch, even in 'off' position is treated as exam malpractice

General Instructions if any Open Book/Open Notebook/Closed Book:

1. **"*fx* series" - non Programmable calculator are permitted: YES**
2. **Reference tables permitted : YES (if Yes, Please specify: Logarithm Tables)**

PART – A: Answer any ALL Questions, Each Question Carries 10 Marks (5×10=50 Marks)

1. "Restaurant A" sells burgers with optional flavors: Pepper, Ginger, and Chilly. Every day this week you have tried a burger (A to E) and kept a record of which you liked. Using Hamming distance, show how the 3NN classifier with majority voting would classify **{ pepper: false, ginger: true, chilly: true}**

Sample	**Pepper Pepper**	**Ginger**	**Chilly**	**Liked**
A	TRUE	TRUE	TRUE	FALSE
B	TRUE	FALSE	FALSE	TRUE
C	FALSE	TRUE	TRUE	FALSE
D	FALSE	TRUE	FALSE	TRUE
E	TRUE	FALSE	FALSE	TRUE

Solution:

The training examples contain three attributes, Pepper, Ginger, and Chilly. Each of these attributes takes either True or False as the attribute values. Liked is the target that takes either True or False as the value.

In the k-nearest neighbor's algorithm, first, we calculate the distance between the new example and the training examples. using this distance we find k-nearest neighbors from the training examples.

To calculate the distance the attribute values must be real numbers. But in our case, the dataset set contains the categorical values. Hence we use hamming distance measure to find the distance between the new example and training examples.

Let x1 and x2 be the attribute values of two instances.

Then, in the hamming distance, if the categorical values are the same or matching that is x1 is the same as x2 then the distance is 0, otherwise 1.

For example,

If the value of **x1 is blue** and **x2 is also blue** then the distance between x1 and x2 is **0**.

If the value of **x1 is blue** and **x2 is red** then the distance between x1 and x2 is **1**.

The following table shows the distance between the new example and the training example, calculated using hamming distance.

	Pepper	Ginger	Chilly	Liked	Distance
A	True	True	True	False	1 + 0 + 0 = 1
B	True	False	False	True	1 + 1 + 1 = 3
C	False	True	True	False	0 + 0 + 0 = 0
D	False	True	False	True	0 + 0 + 1 = 1
E	True	False	False	True	1 + 1 + 1 = 3

Next, Based on the distance we find 3 nearest neighbors (3NN), which are marked in the last column.

	Pepper	Ginger	Chilly	Liked	Distance	3NN
A	True	True	True	False	1 + 0 + 0 = 1	2
B	True	False	False	True	1 + 1 + 1 = 3	
C	False	True	True	False	0 + 0 + 0 = 0	1
D	False	True	False	True	0 + 0 + 1 = 1	2
E	True	False	False	True	1 + 1 + 1 = 3	

Finally, majority voting is used to assign the classification label to the new example. In this case, we have, **two False** and **one True** nearest examples. Hence the new example is classified as **FLASE**.

2. Apply ID3 algorithm and determine the root node of the decision tree for the given training data in the table. Predict the class of the following new example: **age<=30, income=medium, student=yes, credit-rating=fair.** Also draw the decision tree post-classification at the root node.

age	income	student	Credit rating	Buys computer
<=30	high	no	fair	no
<=30	high	no	excellent	no

31…40	high	no	fair	yes
>40	medium	no	fair	yes
>40	low	yes	fair	yes
>40	low	yes	excellent	no
31…40	low	yes	excellent	yes
<=30	medium	no	fair	no
<=30	low	yes	fair	yes
>40	medium	yes	fair	yes
<=30	medium	yes	excellent	yes
31…40	medium	no	excellent	yes
31…40	high	yes	fair	yes
>40	medium	no	excellent	no

First, check which attribute provides the highest Information Gain in order to split the training set based on that attribute. We need to calculate the expected information to classify the set and the entropy of each attribute.

The information gain is this mutual information minus the entropy:

The mutual information of the two classes,

Entropy(S)= E(9,5)= $-9/14 \log_2(9/14) - 5/14 \log_2(5/14)=0.94$

Now Consider the Age attribute

For Age, we have three values $age_{<=30}$ (2 yes and 3 no), $age_{31..40}$ (4 yes and 0 no), and $age_{>40}$ (3 yes and 2 no)

Entropy(age) = $5/14 (-2/5 \log_2(2/5)-3/5\log_2(3/5)) + 4/14 (0) + 5/14 (-3/5\log_2(3/5)-2/5\log_2(2/5))$

$= 5/14(0.9709) + 0 + 5/14(0.9709) = 0.6935$

Gain(age) = $0.94 - 0.6935 = 0.2465$

Next, consider Income Attribute

For Income, we have three values $income_{high}$ (2 yes and 2 no), $income_{medium}$ (4 yes and 2 no), and $income_{low}$ (3 yes 1 no)

Entropy(income) = $4/14(-2/4\log_2(2/4)-2/4\log_2(2/4)) + 6/14 (-4/6\log_2(4/6)-2/6\log_2(2/6)) + 4/14 (-3/4\log2(3/4)-1/4\log2(1/4))$

$= 4/14 (1) + 6/14 (0.918) + 4/14 (0.811)$

$= 0.285714 + 0.393428 + 0.231714 = 0.9108$

Gain(income) = $0.94 - 0.9108 = 0.0292$

Next, consider Student Attribute

For Student, we have two values $student_{yes}$ (6 yes and 1 no) and $student_{no}$ (3 yes 4 no)

Entropy(student) = $7/14(-6/7\log_2(6/7)-1/7\log_2(1/7)) + 7/14(-3/7\log_2(3/7)-4/7\log_2(4/7)$

= 7/14(0.5916) + 7/14(0.9852)

= 0.2958 + 0.4926 = 0.7884

Gain (student) = 0.94 – 0.7884 = 0.1516

Finally, consider Credit_Rating Attribute

For Credit_Rating we have two values credit_ratingfair (6 yes and 2 no) and credit_ratingexcellent (3 yes 3 no)

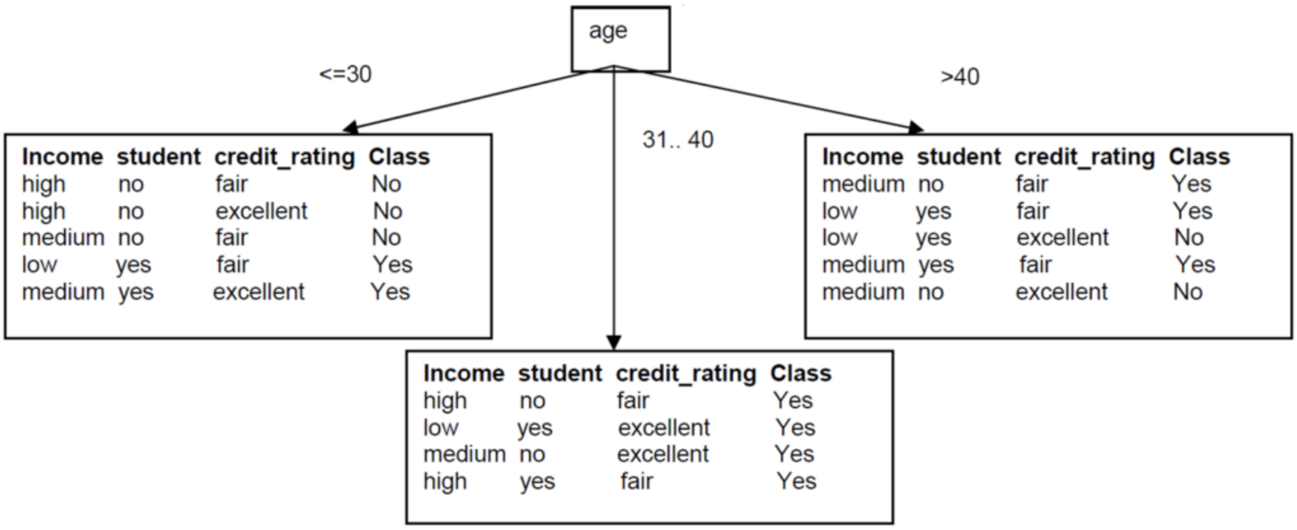
Entropy(credit_rating) = $8/14(-6/8\log_2(6/8)-2/8\log_2(2/8)) + 6/14(-3/6\log_2(3/6)-3/6\log_2(3/6))$

= 8/14(0.8112) + 6/14(1)

= 0.4635 + 0.4285 = 0.8920

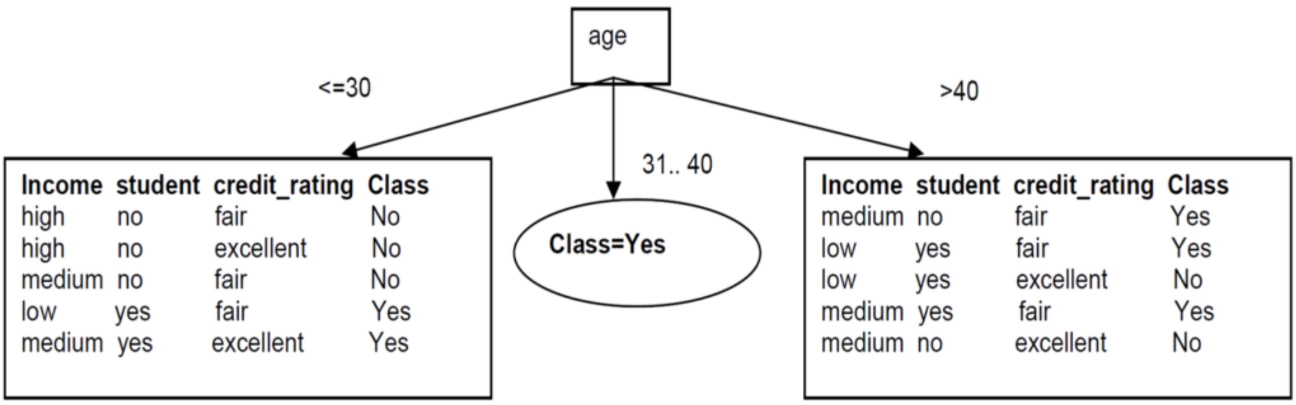
Gain(credit_rating) = 0.94 – 0.8920 = 0.0479

Since Age has the highest Information Gain we start splitting the dataset using the age attribute.



Decision Tree after step 1

Since all records under the branch age31..40 are all of the class, Yes, we can replace the leaf with Class=Yes



Decision Tree after step 1_1

New example: age<=30, income=medium, student=yes, credit-rating=fair

Follow branch(age<=30) Sincs majority is "No" we predict Class=No

Buys_computer = No

3. Implement the AdaBoost algorithm on the following dataset with six data instances. Use 3 Decision Stumps for each of the 3 attributes and "**Job Profile**" as the target attribute.

CGPA	Interactiveness	Communication Skill	Job Profile
>=9	Yes	Good	Yes
<9	No	Moderate	Yes
>=9	No	Moderate	No
<9	No	Good	No
>=9	Yes	Moderate	Yes
>=9	Yes	Moderate	Yes

Solution:

Step 1: initial weight assigned to each item = 1/6

Step2: Iterate for each weak classifier

i. Decision stump for CGPA

a. Train the decision stump H with a random bootstrap sample from the training dataset T.

CGPA	Predicted	Actual	Weight
>=9	Yes	Yes	1/6
<9	No	Yes	1/6
>=9	Yes	No	1/6
<9	No	No	1/6
>=9	Yes	Yes	1/6
>=9	Yes	Yes	1/6

Error = 2/6=0.333

Alpha = 0.347

Z=0.9428

Wt(i+1)=0.1249 for correct classifications

Wt(i+1)=0.2501 for incorrect classifications

Updated weights from the CGPA decision stump:

CGPA	Predicted	Actual	Weight
>=9	Yes	Yes	0.1249
<9	No	Yes	0.2501
>=9	Yes	No	0.2501

<9	No	No	0.1249
>=9	Yes	Yes	0.1249
>=9	Yes	Yes	0.1249

Interactivess:

Interactiveness	Predicted	Actual	Weight
Yes	Yes	Yes	0.1249
No	No	Yes	0.2501
No	No	No	0.2501
No	No	No	0.1249
Yes	Yes	Yes	0.1249
Yes	Yes	Yes	0.1249

Error = 1*0.2501=0.2501

Alpha = 0.5490

Z=0.866

Wt(i+1)(0.1249)=0.0832 for correct classifications

Wt(i+1)(0.2501)=0.1667 for correct classifications

Wt(i+1) (0.2501)=0.5001 for incorrect classifications

Updated weights from the interactiveness decision stump:

Interactiveness	Predicted	Actual	Weight
Yes	Yes	Yes	0.0832
No	No	Yes	0.5001
No	No	No	0.1667
No	No	No	0.0832
Yes	Yes	Yes	0.0832
Yes	Yes	Yes	0.0832

Communication Skill

Communication Skill	Predicted	Actual	Weight
Good	Yes	Yes	0.0832
Moderate	No	Yes	0.5001
Moderate	No	No	0.1667
Good	Yes	No	0.0832
Moderate	No	Yes	0.0832
Moderate	No	Yes	0.0832

Error = 3*0.0832+0.5001=0.7497

Alpha = -0.5485

Step3: Compute the Final Prediction for each instance:

CGPA (0.347)	Interactiveness (0.549)	Communication Skill (-0.5485)	Weighted Avg	Final prediction
Yes	Yes	Yes	0.3475	Y
Yes	No	No	0	N
No	No	No	0.347	Y
No	No	Yes	-0.5485	N
Yes	Yes	No	0.896	Y
Yes	Yes	No	0.896	Y

4. Estimate the conditional probabilities of each attribute {Color, Type, Origin } for the Stolen classes: {Yes, No} using the data given in the table. Using these probabilities estimate the probability values for the new instance – (Color=Yellow, Type=Sports, and Height=Domestic).

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

Solution:

Prior Probabilities:

P(yes)=0.5

P(no)=0.5

Conditional Probabilities:

Color	Yes	No
Red	3/5	2/5
Yellow	2/5	3/5

Type	Yes	No
Sports	4/5	2/5
SUV	1/5	3/5

Origin	Yes	No
Domestic	2/5	3/5
Imported	3/5	2/5

New Instance = (Yellow, Sports, Domestic)

P(Yes|New Instance)=p(yes)*p(Color=yellow|yes)*p(type=sports|yes)*p(origin=domestic|yes)

P(Yes|New Instance)=0.5*2/5*4/5*2/5=0.064

P(No|New Instance)=p(no)*p(Color=yellow|no)*p(type=sports|no)*p(origin=domestic|no)

P(No|New Instance)=0.5*3/5*2/5*3/5=0.072

Since, P(No|New Instance)=0.5*3/5*2/5*3/5=0.072 > P(Yes|New Instance)=0.5*2/5*4/5*2/5=0.064

This new instance can be classified as "**Not Stolen"**

5. Given a dataset with the following points:
   Class 1: (1, 2), (2, 3), (3, 1)
   Class 2: (4, 5), (5, 4), (6, 3)
   If the given dataset is linearly separable, find the equation of the optimal hyperplane that separates the two classes.

5A) Nearest Points from both Class: (2,3) (4,5)

x_1	x_2	Class
2	3	+1
4	5	-1

$N=2$

$x_1 = (2,3)$

$x_2 = (4,5)$

$y_1 = +1$

$y_2 = -1$

$\alpha = (\alpha_1, \alpha_2)$

subject to the Conditions

$\alpha_1 - \alpha_2 = 0$ means $\alpha_1 = \alpha_2$

$\alpha_1 > 0 \cdot \alpha_2 > 0$

$$f(\bar{x}) = \bar{w} \cdot \bar{x} - b$$

$$\phi(\bar{\alpha}) = \sum_{i=1}^{N} \alpha_i - \frac{1}{2} \sum_{i=1}^{N} \sum_{j=1} \alpha_i \alpha_j y_i y_j (\bar{x_i} \cdot \bar{x_j})$$

$$= (\alpha_1 + \alpha_2) = \frac{1}{2}\Big[\alpha_1 \alpha_1 y_1 y_1 (\bar{x_1} \cdot \bar{x_1}) + \alpha_1 \alpha_2 y_1 y_2 (\bar{x_1} \cdot \bar{x_2})$$
$$+ \alpha_2 \alpha_1 y_2 y_1 (\bar{x_2} \cdot \bar{x_1}) + \alpha_2 \alpha_2 y_2 y_2 (\bar{x_2} \cdot \bar{x_2})\Big]$$

$$= (\alpha_1 + \alpha_2) = \frac{1}{2}\left[20\alpha_1^2 - 46\alpha_2\alpha_1 + 41\alpha_2^2\right]$$

$\alpha_1 = \alpha_2$

$$\phi(\bar{\alpha}) = (\alpha_1 + \alpha_1) - \frac{1}{2}\left[20\alpha_1^2 - 46\alpha_2\alpha_1 + \alpha_2^2 41\right]$$

$$\phi(\bar{\alpha}) = 2\alpha_1 - \frac{1}{2} 15\alpha_1^2$$

$$= 2\alpha_1 - 7.5\alpha_1^2$$

For ϕ to be maximum we must have

$$\frac{d\phi}{d\alpha_1} = 2 - 15\alpha_1 = 0$$

$$\alpha_1 = \frac{2}{15}$$

$$\bar{w} = \sum_{i=1}^{N} \alpha_i y_i \bar{x_i} \Rightarrow \alpha_1 y_1 \bar{x}_1 + \alpha_2 y_2 \bar{x}_2$$

$$= \frac{2}{15}(-2, -2)$$