

JAYPEE UNIVERSITY OF INFORMATION TECHNOLOGY, WAKNAGHAT

TEST -3 EXAMINATION- 2023

M. Tech-I Semester (CSE/IT/DS)

COURSE CODE (CREDITS): 22MIWCI132 (3)

MAX. MARKS: 35

COURSE NAME: Artificial Intelligence Techniques

COURSE INSTRUCTORS: Dr. Nancy Singla

MAX. TIME: 2 Hours

Note: (a) All questions are compulsory.

(b) Marks are indicated against each question in square brackets.

(c) The candidate is allowed to make suitable numeric assumptions wherever required for solving problems

Q1. (a) How a utility-based agent differs from a goal-based agent?

[2+3]

(b) Define Tic-Tac-Toe problem and its state representation.

[CO1]

Q2. Consider the following dataset of species having attributes as color, legs, height and smelly.

[10]

[CO2]

S. No.	Color	Legs	Height	Smelly	Species
1	White	3	Short	Yes	M
2	Green	2	Tall	No	M
3	Green	3	Short	Yes	M
4	White	3	Short	Yes	M
5	Green	2	Short	No	H
6	White	2	Tall	No	H
7	White	2	Tall	No	H
8	White	2	Short	Yes	H

Use Naïve Bayes algorithm and show all the precomputations to identify the species with the following attributes.

$X = \{\text{Color}=\text{Green}, \text{Legs}=2, \text{Height}=\text{Tall}, \text{Smelly}=\text{No}\}$

Q3. (a) What are the different attribute selection measures for constructing a decision tree?

[4+4]

[CO5]

(b) Suppose you are using the Decision Tree Learning algorithm to learn a 2-class classification variable, C, and you must decide which attribute to assign to a node in the tree. At this node there are 100 examples; 30 are positive and 70 are

negative. If attribute A is selected, its first child will get 18 positive and 22 negative examples, and its second child will get 12 positive and 48 negative examples.

Use $\log 0.1 = -3.32$, $\log 0.2 = -2.32$, $\log 0.3 = -1.74$, $\log 0.33 = -1.59$, $\log 0.4 = -1.32$, $\log 0.45 = -1.15$, $\log 0.5 = -1.0$, $\log 0.55 = -0.86$, $\log 0.6 = -0.74$, $\log 0.67 = -0.58$, $\log 0.7 = -0.51$, and $\log 0.8 = -0.32$, $\log 0.9 = -0.15$, and $\log 1 = 0$, where all logs are to base 2.

- (i) What is the entropy of C, i.e., $H(C)$, at the node?
- (ii) What is the conditional entropy of choosing attribute A at the node? That is, compute $H(C | A)$.

Q4. What is Overfitting? What are the techniques used to avoid overfitting? [5][CO5]

Q5. (a) Consider the following scenario: [5+2]
[CO4]

In a Covid test of 1000 patients, there were 45 positive tests, of which 30 patients had covid and 15 were falsely tested positive. Of the 955 negative tests there were 5 that were incorrect, these patients had covid but were tested negatively.

Draw the confusion matrix and calculate the accuracy, precision, recall and F1 score from the matrix.

- (b) Suppose you are working on weather prediction, and you would like to predict whether or not it would be raining at 5pm tomorrow. You want to use a learning algorithm for this. Would you treat this as a classification or a regression problem and why?