

Assignment 2(2024)

1. A decision tree can be used to build models for: **(1 Mark)**

- A. Regression problems
- B. Classification problems
- C. Both of the above
- D. None of the above

Ans: C

Explanation: Decision is used for both regression and classification problems.

2. Entropy value of ____ represents that the data sample is pure or homogenous: **(1 Mark)**

- A. 1
- B. 0
- C. 0.5
- D. None of the above.

Ans: B

Explanation: A pure or homogenous data sample is 0.

3. Entropy value of ____ represents that the data sample has a 50-50 split belonging to two categories: **(1 mark)**

- A. 1
- B. 0
- C. 0.5
- D. None of the above

Ans: A

Explanation: Entropy = $-0.5\log_2 0.5 - 0.5\log_2 0.5 = 1$

4. If a decision tree is expressed as a set of logical rules, then: **(1 Mark)**

- A. the internal nodes in a branch are connected by AND and the branches by AND
- B. the internal nodes in a branch are connected by OR and the branches by OR
- C. the internal nodes in a branch are connected by AND and the branches by OR
- D. the internal nodes in a branch are connected by OR and the branches by AND

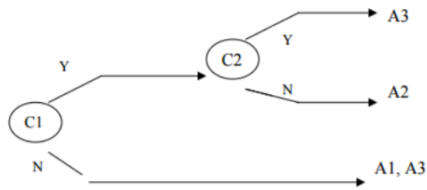
Ans: C

Explanation: definition of decision tree.

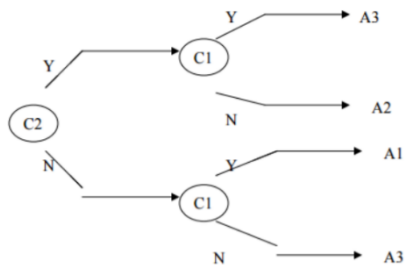
5. The Decision tree corresponding to the following is? **(1 Mark)**

```
if C2 then
  if C1 then A3
  else A2
endif
else A1, A3
endif
```

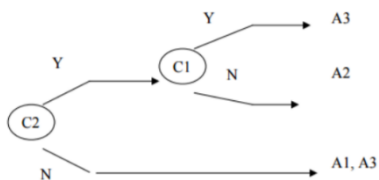
A.



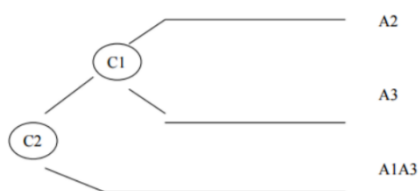
B.



C.



D.



Ans: C

Explanation: option c is the valid DT for the rule.

For questions 6-7, consider the following table depicting whether a customer will buy a flat or not.

GPA	Studied	Passed
Low	F	F
Low	T	T
Medium	F	F
Medium	T	T
High	F	T
High	T	T

6. What is the entropy
Mark)

of the dataset? **(1**

- A. 0.50
- B. 0.92
- C. 1
- D. 0

Ans: B

Explanation: $\text{Entropy}(2,4) = -(2/6)\log(2/6) - (4/6)\log(4/6) = 0.92$

7. Which attribute would information gain choose as the root of the tree? **(2 Marks)**

- A. GPA
- B. Studied
- C. Passed
- D. None of the above

Ans: B

Explanation: From information gain criterion. The **Studied** has the highest information gain.

8. A chemical company has three options: (i) commercial production, (ii) pilot plant and (iii) no production. The cost of constructing a pilot plant is Rs 3 lacs. If a pilot plant is built, chances of high and low yield are 80% and 20% respectively. In the case of high yield from the pilot plant, there is a 75% chance of high yield from the commercial plant. In the case of low yield from the pilot plant, there is only a 10% chance of high yield from the commercial plant. If the company goes for commercial plant directly without constructing a pilot plant, then there are 60% chance of high yield. The company earns Rs 1,20,00,000 in high yield and loses Rs 12,00,000 in low yield. The optimum decision for the company is: **(2 marks)**

- A. Commercial Production.
- B. Pilot plant

C. No Production

D. None of the above.

Ans: A

Explanation: The company should produce commercially. The final estimated cost is Rs 67,20,000

For Commercial Production:

$$\text{Estimated cost} = 0.6 \times 12000000 - 0.4 \times 1200000 = 67,20,000$$

For Pilot Plant:

$$\begin{aligned} \text{Estimated cost} &= 0.8 \times 0.75 \times 12000000 - 0.8 \times 0.25 \times 1200000 + 0.2 \times 0.10 \times 12000000 - 0.8 \times 0.9 \times 1200000 \\ &- 300000 = 60,36,000 \end{aligned}$$