

Н

Score: **74%**

No. of questions: 10

Correct answer: 8

Incorrect answer: 2

Show incorrect attempt only

1 Mark Question 1

In a decision tree, by comparing the impurity across all possible splits in all possible predictors, the next split is chosen. How can we measure the impurity?

Α **ROC**

Entropy, Gini-Index В



C **MAPE**

Correct Answer: B. Entropy, Gini-Index

ROC and MAPE are evaluation metrics and not the criterion to decide splits. Entropy and Gini-Index are used to decide the right split.

Question 2



2 Marks

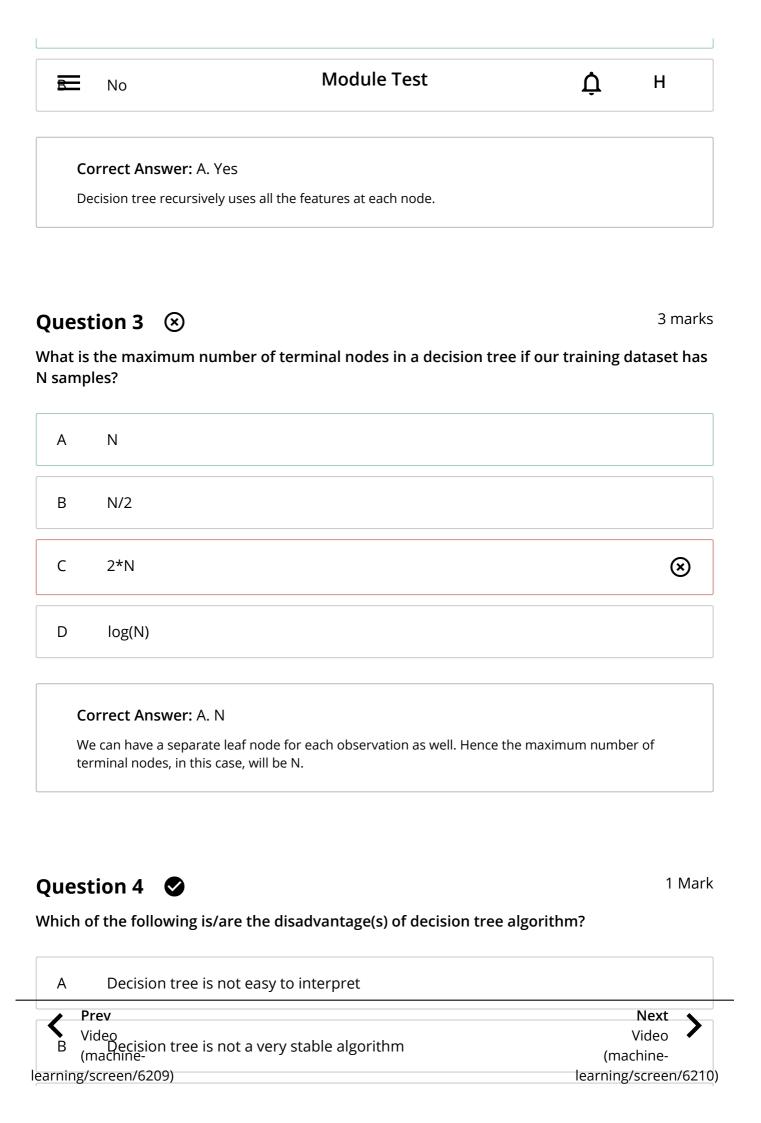
While creating a decision tree, can we reuse a feature to split a node?

Video

Next Video

(machinelearning/screen 210)

(machinelearning/screen/6209)



C =	Decision tree will overfit the data easily if it perfectly memorized it Module Test H
D	Both B and C
	orrect Answer: D. Both B and C ecision tree is easy to interpret. All other are the disadvantages of decision tree.

2 Marks **Question 5**

Which of the following parameter(s) can help to prevent overfitting?

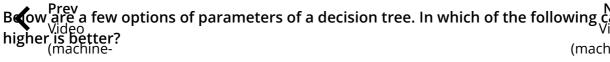
Α Max depth Minimum samples for node split В C Minimum samples for leaf node D Splitting criterion Ε Option A, B, and C

Correct Answer: E. Option A, B, and C

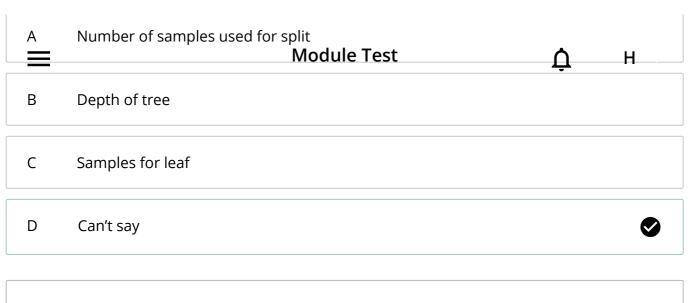
Splitting criterion does not help in preventing overfitting, it just makes sure that we get pure nodes. Tuning other hyperparameters like max_depth, minimum samples for node split, minimum samples for leaf node will help to prevent overfitting.

Question 6

3 marks



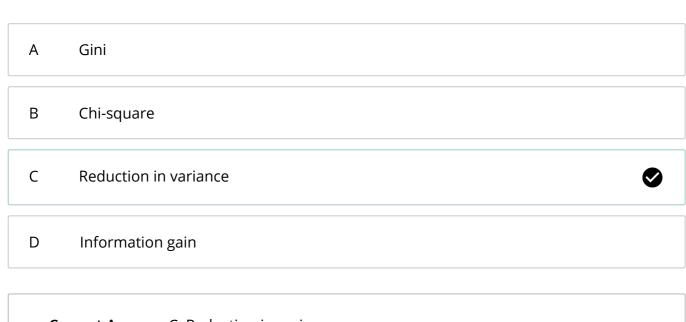




Correct Answer: D. Can't say

For all three options A, B and C, it is not necessary that if you increase the value of parameter the performance will increase. For example, if we have a very high value of depth of tree, the resulting tree may overfit the data, and would not generalize well. On the other hand, if we have a very low value, the tree may underfit the data. So, we can't say for sure that "higher is better".

Let's say you are asked to work on a problem to predict the future sales of a product in a store and decided to use a decision tree model. Which algorithm should be used for splitting?



Correct Answer: C. Reduction in variance

Since the target variable is continuous, it is a regression problem and hence reduction in variance will be used for splitting.



_	tion 8	2 Marks H
А	Decision tree is prone to overfit and accuracy doesn't help to generalize	
В	Information gain is more stable as compared to accuracy	⊗
С	Information gain chooses more impactful features closer to root	
D	All of the above	
	Correct Answer: D. All of the above Il of the given options are correct.	
_	on trees are not affected by multicollinearity in features.	3 marks
А	True	Ø
В	False	
_		

Correct Answer: A. True

True, decision trees are not affected by multicollinearity in features. For example, if there are two 90% correlated features, decision tree would consider only one of them for splitting.



1 Mark

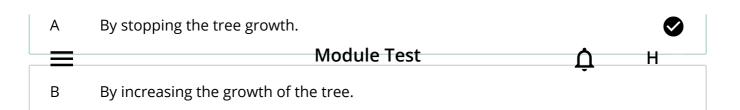
How can we avoid overfitting in a decision tree?

Next Video



(machine-

(machine-



Correct Answer: A. By stopping the tree growth.

Increasing the growth of tree increases the chances of overfitting and hence if we stop the growth, it can help us to avoid overfitting.