Score: **74%**

No. of questions: **10**

Correct answer: **8**

Incorrect answer: **2**

Show incorrect attempt only

Question 1

1 Mark

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?

A ROC

B Entropy, Gini-Index

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?

A Yes

Correct Answer: A. Yes

Decision tree recursively uses all the features at each node.

Question 3

3 marks

What is the maximum number of terminal nodes in a decision tree if our training dataset has N samples?

A N

B N/2

C 2*N

D log(N)

Correct Answer: A. N

We can have a separate leaf node for each observation as well. Hence the maximum number of terminal nodes, in this case, will be N.

Question 4

1 Mark

Which of the following is/are the disadvantage(s) of decision tree algorithm?

A Decision tree is not easy to interpret

B Decision tree is not a very stable algorithm

C Decision tree will overfit the data easily if it perfectly memorized it

D Both B and C ✓

Correct Answer: D. Both B and C

Decision tree is easy to interpret. All other are the disadvantages of decision tree.

Question 5 ✓

2 Marks

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

A Max depth

B Minimum samples for node split

C Minimum samples for leaf node

D Splitting criterion

E 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

Below are a few options of parameters of a decision tree. In which of the following cases higher is better?

A Number of samples used for split

B Depth of tree

C Samples for leaf

D Can’t say ✓

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”.

Question 7 ✓

1 Mark

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?

A Gini

B Chi-square

C Reduction in variance ✓

D Information gain

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.

Question 8

2 Marks

Why do we prefer information gain over accuracy when splitting?

A Decision tree is prone to overfit and accuracy doesn't help to generalize

B Information gain is more stable as compared to accuracy

C Information gain chooses more impactful features closer to root

D All of the above

Correct Answer: D. All of the above

All of the given options are correct.

Question 9

3 marks

Decision trees are not affected by multicollinearity in features.

A True

B 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.

Question 10

1 Mark

How can we avoid overfitting in a decision tree?

A By stopping the tree growth.

B By increasing the growth of the tree.

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.