Bookmarks

- Module 1: The Basics of R and Introduction to the Course
- Entrance Survey
- Module 2: Fundamentals of Probability, Random Variables, Distributions, and Joint Distributions
- Module 3: Gathering and Collecting Data, Ethics, and Kernel Density Estimates
- Module 4: Joint, Marginal, and Conditional Distributions &

Decision Trees - Quiz

Bookmark this page

Question 1

1.0/1.0 point (graded)

True or False: When fitting a decision tree, the decision tree that best fits the data will perform better predictions.

- a. True
- b. False ✔

Submit You have used 1 of 1 attempt

Question 2

1.0/1.0 point (graded)

In the worst case of overfitting, a dataset with n variables can have a tree that has, at most, a depth of _____.

a. $\log_2 n$

b. $\frac{n}{2}$

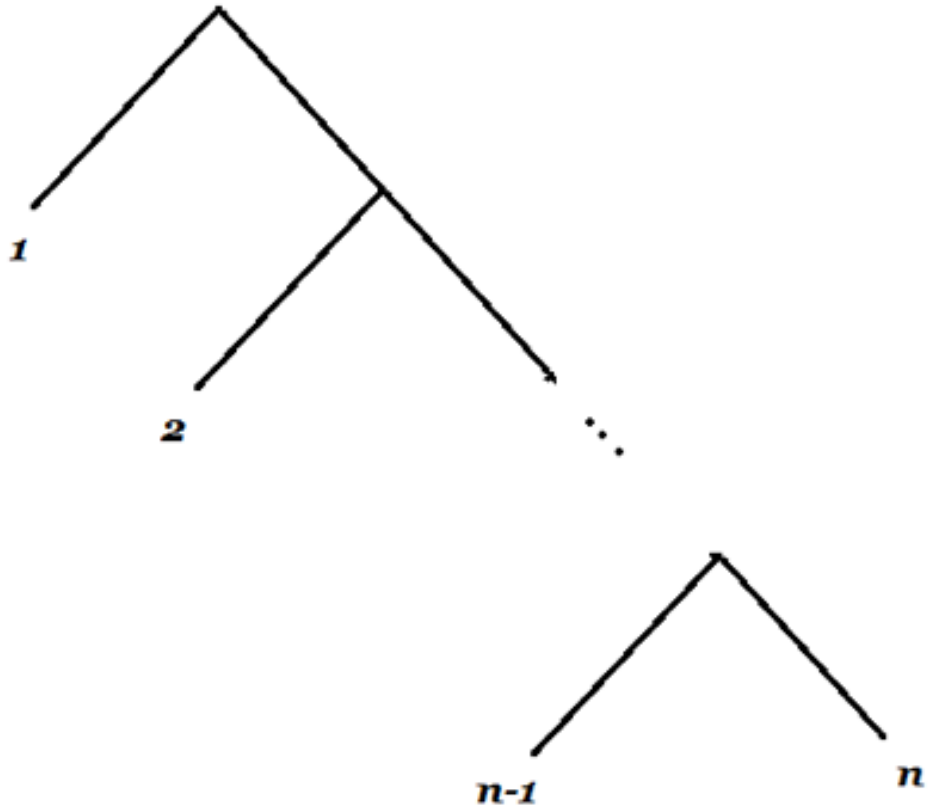
c. 2^n

d. $\frac{n}{4}\log_2(2n)$

e. n ✔

Explanation

A dataset of variables can be represented in a tree in which, in the worst case, each data point has its own leaf. So in the absolute worst case with no other constraints it would have n. An example is below:



Submit You have used 1 of 2 attempts

Question 3

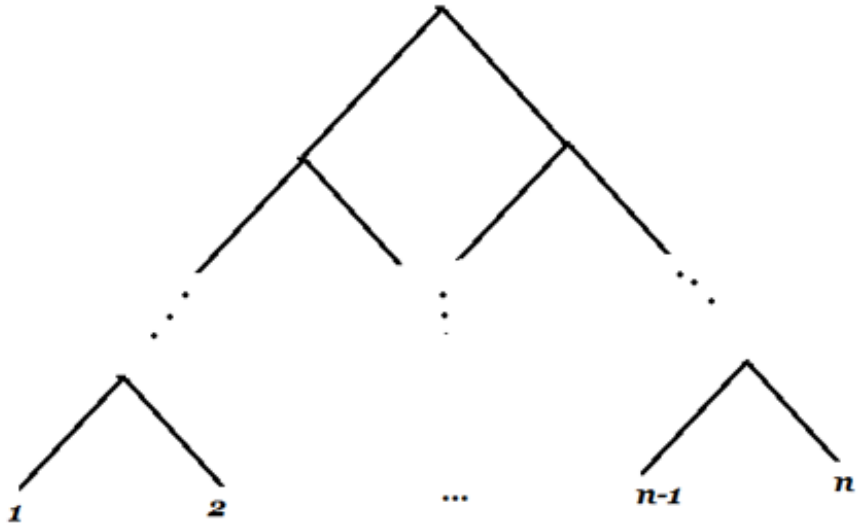
0/1 point (graded)

In the worst case, a dataset with n variables and only binary splits in each branch with equal points in each leaf can have a tree that has at most _____ leafs.

- a. $\log_2 n$
- b. $\frac{n-1}{2}$
- c. 2^n ✖
- d. $\frac{n}{4}\log_2(2n)$
- e. n

Explanation

If a dataset of size n is split equally then it is constantly divided by 2, until every leaf has one point. The question of "how deep can my tree be?" is the same as how many times can I divide n by two, and still get a number greater than 1. This number is then bounded above by $\log_2 n$. The picture below illustrates this example. Although there is still a point per leaf, this illustrates how simple conditions can change the depth in your graph.



Submit You have used 2 of 2 attempts

✖ Incorrect (0/1 point)

Discussion

Topic: Module 11 / Decision Trees - Quiz

Hide Discussion

Add a Post

I agree that the answer is consistent with depth of tree for n variables.

posted 4 days ago by **joycevdev** **Community TA**

I thought variables were the things measured on the data, that provided the criteria for the splits in the tree. Don't we have n DATA VALUES (and not n VARIABLES) to be sorted out by the tree? ;)

posted 3 days ago by **Margaret_Niehaus**

Add a comment

sandipan_dey

less than a minute ago

I thought the question 3 asked the # leafs, but was surprised to find out that it asked the depth of the tree.

Add a comment

− **Collapse discussion**

Decision Trees Quiz - Question 2

question posted 9 days ago by **Margaret_Niehaus**

Do we have to assume that each variable can take exactly two values? Otherwise I don't see that I have enough information to answer the question....

This post is visible to everyone.

+ **Expand discussion**

[Staff] Question 2 on Decision Trees Quiz

question posted 10 days ago by **joycevdev** **Community TA**

Hmm. For question 2, can it really be something other than 2^n? Unless perhaps the question was supposed to be "a dataset containing n items",...

This post is visible to everyone.

+ **Expand discussion**

1