1. Questions 1 to 6 refer to the following common scenario:
Consider the following dataset: 1 point

x1	x2	x3	y
1	1	1	+1
0	1	0	-1
1	0	1	-1
0	0	1	+1

Let us train a decision tree with this data. Let's call this tree T1. What feature will we split on at the root?

- ○ x1
- ○ x2
- ◉ x3

2. Refer to the dataset presented in Question 1 to answer the following. 1 point

Fully train T1 (until each leaf has data points of the same output label). What is the depth of T1?

3

3. Refer to the dataset presented in Question 1 to answer the following. 1 point
What is the training error of T1?

0

4. Refer to the dataset presented in Question 1 to answer the following. 1 point

Now consider a tree T2, which splits on x1 at the root, and splits on x2 in the 1st level, and has leaves at the 2nd level. Note: this is the XOR function on features 1 and 2. What is the depth of T2?

2

5. Refer to the dataset presented in Question 1 to answer the following.
What is the training error of T2?

1 point

0

6. Refer to the dataset presented in Question 1 to answer the following.
Which has smaller depth, T1 or T2?

1 point

- ○ T1
- ◉ T2

7. (True/False) When deciding to split a node, we find the best feature to split on that minimizes classification error.

1 point

- ○ True
- ○ False

8. If you are learning a decision tree, and you are at a node in which all of its data has the same y value, you should

1 point

- ○ find the best feature to split on
- ◉ create a leaf that predicts the y value of all the data
- ○ terminate recursions on all branches and return the current tree
- ○ go back to the PARENT node and select a DIFFERENT feature to split on so that the y values are not all the same at THIS node

9. Consider two datasets D1 and D2, where D2 has the same data points as D1, but has an extra feature for each data point. Let T1 be the decision tree trained with D1, and T2 be the tree trained with D2. Which of the following is true? 1 point

- ○ T2 has better training error than T1
- ○ T2 has better test error than T1
- ◉ Too little information to guarantee anything

10. (True/False) Logistic regression with polynomial degree 1 features will always have equal or lower training error than decision stumps (depth 1 decision trees). 1 point

- ○ True
- ◉ False

11. (True/False) Decision trees (with depth > 1) are always linear classifiers. 1 point

- ○ True
- ◉ False