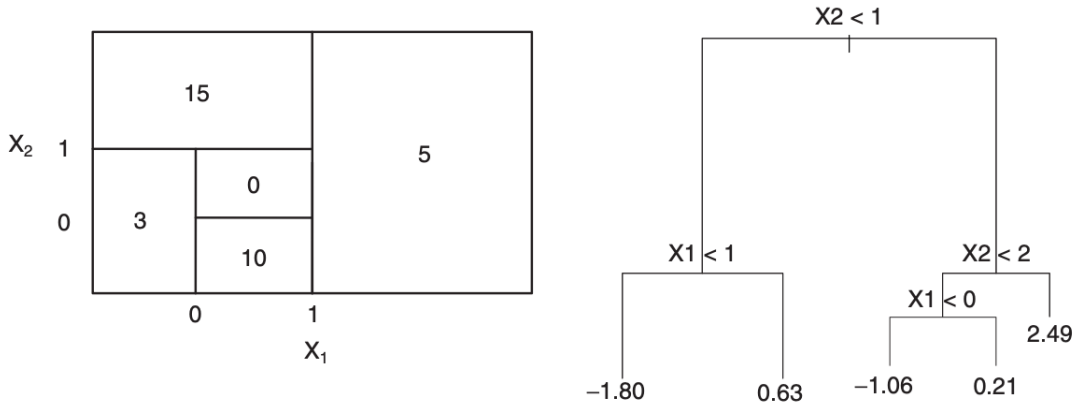
Tree-based methods

Machine Learning 2021-2022 - UMONS
Souhaib Ben Taieb

1

1. Sketch the tree corresponding to the partition of the predictor space illustrated in the left-hand panel of Figure 1. The numbers inside the boxes indicate the mean of Y within each region.

2. Create a diagram similar to the left-hand panel of Figure 1, using the tree illustrated in the right-hand panel of the same figure. You should divide up the predictor space into the correct regions, and indicate the mean for each region.



(This question is from ISLR, Section 8.4, exercise 4).

2

Suppose we produce ten bootstrapped samples from a data set containing red and green classes. We then apply a classification tree to each bootstrapped sample and, for a specific value of X, produce 10 estimates of $P(\text{Class is Red} \mid X = x)$: 0.1, 0.15, 0.2, 0.2, 0.55, 0.6, 0.6, 0.65, 0.7, and 0.75.

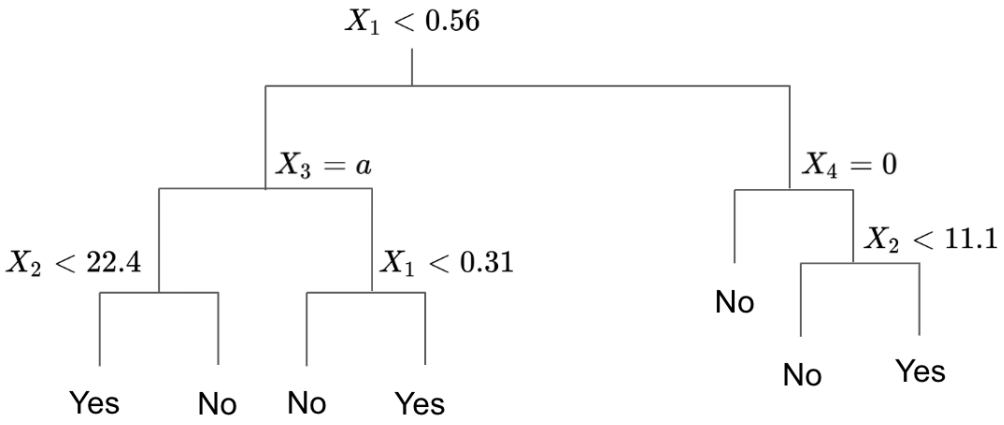
There are two common ways to combine these results together into a single class prediction. One is the majority vote approach. The second approach is to classify based on the average probability. In this example, what is the final classification under each of these two approaches?

(This question is from ISLR, Section 8.4, exercise 5).

3

Given the decision tree of Figure [3], how would the following observations be classified ?

X_1	X_2	X_3	X_4	Y
0.48	18.1	a	1	
0.64	32.5	a	0	
0.12	26.5	b	0	
0.69	6.7	c	1	
0.43	18.6	c	0	
0.84	16.5	a	1	
0.33	28.5	a	1	
0.92	6.3	c	1	
0.96	12.1	b	0	
0.16	13.1	b	1	



4

Build a classification tree using the information gain for the following dataset:

Size	Orbit	Temperature	Habitable
Big	Far	205	No
Big	Near	205	No
Big	Near	260	Yes
Big	Near	380	Yes
Small	Far	205	No
Small	Far	260	Yes
Small	Near	260	Yes
Small	Near	380	No
Small	Near	380	No