

Decision Trees Practice Problems Solutions
--

Consider the following dataset:

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

1. What feature will we split on at the root of our decision tree, and what will our information gain be from splitting on that feature using the Gini impurity measure?

The first thing we should do is look at the input data and see what makes sense. If we were to split the data on x2 or x3, we would get an information gain of 0 (try this out and make sure it makes sense to you).

Let's try splitting the data on x1, meaning our left group would be the data points where x1 = 0, and the right group would be the data points where x1 = 1. Now we have to plug this information into our formula.

$I(D_p)$ , the impurity of our total dataset, is 0.5. This is because the y labels of all four samples are balanced with two +1s and two -1s.

$I(D_{left})$ , the impurity of our dataset with the three datapoints where x1 = 0, isn't obvious, so let's calculate it. The formula for Gini Impurity would result in  $1 - [(\frac{1}{3})^2 + (\frac{2}{3})^2] = \frac{4}{9}$ . This is because there is one +1 in this left group and two -1s in the right group, making three total datapoints.

$I(D_{right})$ , the impurity of the right dataset, is simply 0. Why? There is one data point in this sample, so it's perfectly sorted with only one class.

Now putting everything together into our formula we have

$$IG(D_p, f) = I(D_p) - \frac{N_{left}}{N_p} I(D_{left}) - \frac{N_{right}}{N_p} I(D_{right})$$

which when we substitute our values in we get

$$IG(D_p, f) = \frac{1}{2} - \frac{3}{4} * \frac{4}{9} - \frac{1}{4} * 0 = \boxed{\frac{1}{6}}$$

2. Build a decision tree using the dataset. What is the depth of the tree?

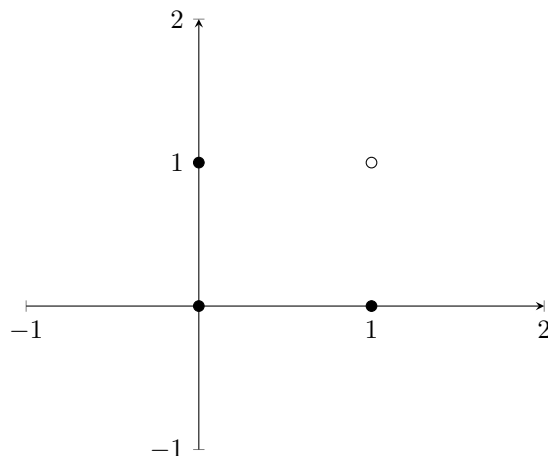
Since we know that the first split is going to be the one we just calculated, we'll have to calculate the remaining splits on the three data points that are still not pure yet. Try this one on your own. With the first layer as layer 0, you should get a depth of  $\boxed{3}$ .

3. What will the decision tree classify a data point with the features  $x1 = 0$ ,  $x2 = 0$ , and  $x3 = 0$  as ( $y = -1$  or  $y = +1$ )?

Running this through the decision tree you just made, you should find that the tree classifies the data point as  $\boxed{-1}$ .

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

It's often helpful to visualize your data, even if it seems simple. Let the  $\bullet = +1$  and  $\circ = -1$ .



4. What will the information gain be after the first split in the above data set with the Gini impurity measure? ~~With entropy as the impurity measure?~~

Again, we plug in our values into the information gain formula

$$IG(D_p, f) = I(D_p) - \frac{N_{left}}{N_p} I(D_{left}) - \frac{N_{right}}{N_p} I(D_{right})$$

and then substitute values according to what we have in our data

$$IG(D_p, f) = \frac{5}{8} - \frac{2}{4} * 0 - \frac{2}{4} * \frac{1}{2} = \boxed{\frac{1}{8}}$$

5. What is the depth of the final decision tree?

We just need to do one more split on the group where there is still one +1 and one -1, so we would have a depth of  $\boxed{2}$ .