Week 9 - Tree-Based Methods

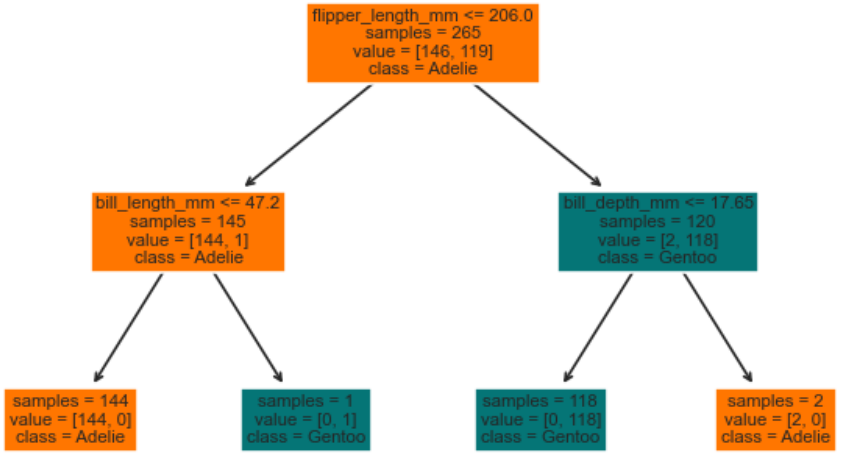
Exercises

Question 1.

Using the Figure below, classify the following penguins (in the table) as either a Adelie or Gentoo penguin.

species	bill_length_mm	bill_depth_mm	flipper_length_mm	body_mass_g
?	49.1	14.8	220.0	5150.0
?	37.7	19.8	198.0	3500.0

Figure 6: Multiple Features Penguins Tree (max_depth = 3)

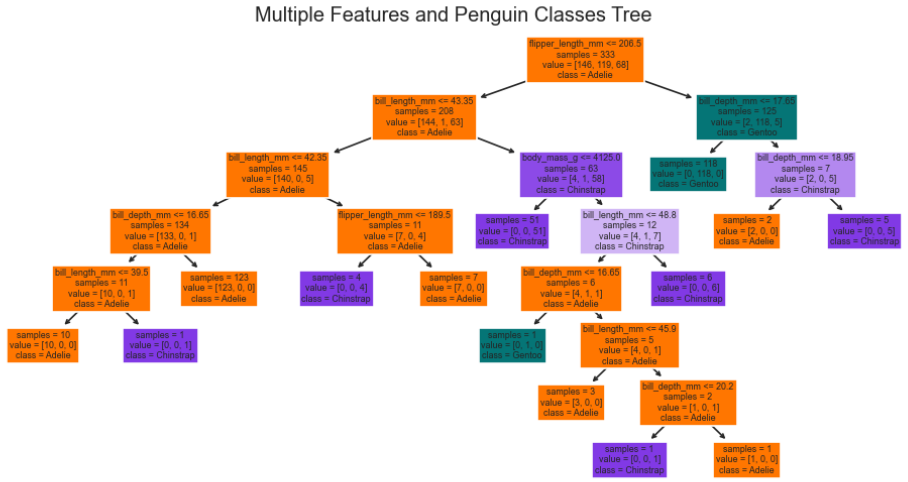


▶ Click here for answer

Question 2.

Using the Figure below, classify the following penguins as either a Adelie, Gentoo, or Adelie penguin

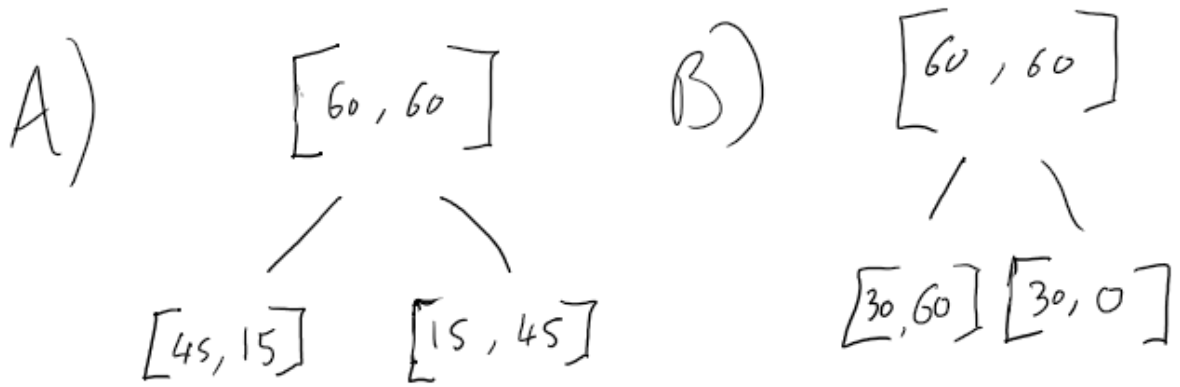
species	bill_length_mm	bill_depth_mm	flipper_length_mm	body_mass_g
?	35.2	15.9	186.0	3050.0
?	51.3	18.2	197.0	3750.0



▶ Click here for answer

Question 3.

Demonstrate why Entropy or Gini impurity is better than classification error for identifying which of the following is a better splitting scenario:



▶ Click here for Classification Error

▶ Click here for Geni Impurity

▶ Click here for Entropy

Question 4.

Although generally Gini impurity is lower in both child nodes compared to the parent node, demonstrate using a 1D dataset with ordered classes A, B, A, A, A that this is not *always* the case.

▶ Click here for answer

Question 5.

If a decision tree is *overfitting* to the training set, would it be a good idea to try decreasing `max_depth`?

▶ Click here for answer

Question 6.

If a decision tree is *underfitting* the training set, would it be a good idea to scale the input features?

▶ Click here for answer

Question 7.

Assume we have three base classifiers in a majority voting ensemble and $C_j(j \in 0, 1)$. Each classifier predicts the following:

$$C_1(x) \rightarrow 0, C_2(x) \rightarrow 0, C_3(x) \rightarrow 1$$

What is the predicted class of the majority voting ensemble if...

a. ...no weights are assigned?

▶ Click here for answer

b. ...C_1 and C_2 have a weight of 0.2, and C_3 has a weight of 0.6?

▶ Click here for answer

c. ...the classifiers have weights as in b, but instead predict
$C_1(x) \rightarrow [0.9, 0.1], C_2(x) \rightarrow [0.8, 0.2], C_3(x) \rightarrow [0.4, 0.6]$

▶ Click here for answer

Question 8.

If you trained five different models on the same training data, and they all achieve 95% precision, would combining these classifier lead to better results? Explain your reasoning.

▶ Click here for answer

Question 9.

Why might out-of-bag evaluation slightly improve training performance when tuning hyperparameters than cross-validation?

▶ Click here for answer

Question 10.

Why are "Extra-Trees" more random than regular "Random Forests"? Why would you want to use "Extra-Trees"? Do you think "Extra-Trees" would be faster or slower to train?

▶ Click here for answer

```
[NbConvertApp] Converting notebook Trees_Exercises.ipynb to html
[NbConvertApp] Writing 498033 bytes to Trees_Exercises.html
```