HW3

Problem 1.1 – (a) How many levels are there in the decision tree?

The level/degree of the tree is "3" (Level0, Level1, Level2) and the Root node is "Setosa" and the left of root node is Setosa and the right to root is Versicolor.

Problem 1.1 – (b) What is the default class label associated with each vertex?

Level 0, Vertex 1: Default class label is Setosa

Level 1, Vertex 1: Default class label is Setosa

Level 1, Vertex 2: Default class label is Versicolor

Level 2, Vertex 1: Default class label is Versicolor

Level 2, Vertex 2: Default class label is Virginica

Problem 1.1 – (c) Starting from the root note, what is the name of the first attribute used for a decision, and what are the split points? Your answer should be of the form

Level 0, split on attribute: Petal.Length

Split points: < 2.5 left subtree, >= 2.5 right subtree

Level 1, split on attribute: Petal.Width

Split points: < 1.8 left subtree, >= 1.8 right subtree

Problem 1.1 – (d)

(i) At each vertex, what do the three numbers in the middle line signify?

Answer

The three numbers in the middle line signify the distribution of observations in the data, they are Setosa, Versicolor, Virginica.

That means, It shows how many Observations are connected to each class in that specific node.

For instance: There are three classes Setosa, Versicolor and Virginica. The distributions are shown as: .63.37.00

The above distributions tells observations that are associated with the node. The 63% of them are labelled as Setosa and the remaining 37% are labelled as Versicolor.

(ii) At each vertex, what does the last line signify?

Answer: The final line indicates how much of the total data is connected to that node before moving on to the next level. If it is stated as 31%, for instance, that means the node has 31% of the data set's total observations. Another illustration is when it says 20% at the leaf node, it signifies that 20% of the data set's total observations are present at that leaf node.