Tiffany Wong A20442087
CS422 - Homework #4 10/03/2021

Problem 1.1: Decision Tree on the Iris Dataset

a. *How many levels are there in the decision tree?*

There are three levels in the decision tree.

b. *What is the default class label associated with each vertex?*

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
Level 3, Vertex 1: Default class label is versicolor
Level 3, Vertex 2: Default class label is virginica

c. *Starting from the root note, what is the name of the first attribute used for a decision, and what are the split points?*

Level 1, split on attribute: Petal.Length
Split points: < 2.5 left subtree, >= 2.5 right subtree
Level 2, split on attribute: Petal.Width
Split points: < 1.8 left subtree, >= 1.8 right subtree
Level 3, split on attribute: Petal.Length
Split points: < 5.0 left subtree, >= 5.0 right subtree

d. *Each vertex has three lines.*

i. *At each vertex, what do the three numbers in the middle line signify?*

The three numbers in the middle line signify the percentage of each flower type at *each* vertex. So for example, on the first level's left subtree, where there is a split on the attribute is petal.length < 2.5, 100% is setosa flowers, 0% is vericolor and virginica flowers.

ii. *At each vertex, what does the last line signify?*

The last line indicates the percentage of the entire dataset size that is at the current vertex/node. So at all the end nodes, the percentages should sum up to 100%, which it does in this example: 33% + 32% + 4% + 31% = 100%.