Sohaib Syed
CS422
9/26/21
Dr. Gurbani

<div align="center">Homework 4</div>

Problem 1.1: Decision Tree on the Iris Dataset

a) There are 3 levels

b)

Level 1, Vertex 1: Default class label is setosa

Level 2, Vertex 1: Default class label is setosa

Level 2, Vertex 2: Default class label is versicolor

Level 3, Vertex 1: Default class label is versicolor

Level 3, Vertex 2: Default class label is virginica

c)

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

d) Interpretation

i) The numbers in the middle signify the split of data for the class. For example, level 3, vertex 1has 0% setosa, 91% versicolor, and 9% virginica.

ii) The last line signifies the total percentage of the data set that is being used for that classification. For example, after the first split on Petal.length, 33% of the whole data is Petal.length <2.5, and 67% is >= than 2.5.