

CS422-01: Homework #3

Gurbani

Sujad Patel

October 3, 2022

Problem 1.1: Decision Tree on the Iris Dataset

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

There are 3 levels (0, 1, 2).

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

Level **0**, Vertex **1**: Default class label is <Setosa>
Level **1**, Vertex **3**: Default class label is <Versicolor>
Level **2**, Vertex **2**: Default class label is <Setosa>
Level **2**, Vertex **6**: Default class label is <Versicolor>
Level **2**, Vertex **7**: Default class label is <Verginica>

(c) Starting from the root node, what is the name of the first attribute used for a decision, and what are the split points?

Level **0**, split on attribute: <Petal.Length>
Split points: < 2.45, >= 2.45
Level **1**, split on attribute: <Petal.Width>
Split points: < 1.75, >= 1.75

(d) Each vertex has three lines.

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

The three numbers signify the *probabilities* of the data belonging to specific class. The format is "Setosa" or "Versicolor" or "Verginica". For example, the left most leaf node has "1.00" in "Setosa" field indicating that it has 100% probability of being classified as "Setosa."

d.(II) At each vertex, what does the last line signify?

The last line signifies the percent of remaining data, out of the whole data set, after the split. After the first split, the model classified that 33% belongs to Setosa. After a second split on the remaining data, 67%, it classified that 36% are Versicolor with 91% probability and the other 31% are Verginica with 98% probability.