CS 422 - Data Mining - Homework 3

Imaduddin Sheikh - `isheikh@hawk.iit.edu`

Due Date: **2/24/2022 11:59:59 PM**

Problem 1.1: Decision Tree on the Iris Dataset

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

There are 3 levels.

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

Level 0, Vertex 1: Default class label is setosa

Level 1, Vertex 2: Default class label is versicolor

(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 0, split on attribute: Petal.Length

Split points: < 2.5 left subtree, ≥ 2.5 right subtree

Level 1, split on attribute: Petal.Length

Split points: < 4.8 left subtree, ≥ 4.8 right subtree

(d) Each vertex has three lines

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

The second line in each vertex indicates the information gain that is calculated at each and every stage of the decision tree.

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

The last line represents the probability of that vertex belonging to these given class labels.