Data Mining – Homework_ 3

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

The decision tree has three level. Starting from the setosa (1), it is divided into two groups: the ones with a Petal.Length (2.1) less than 2.5 cms and the ones with a Petal.Lenght (2.2) equal or higher to 2.5 cm. Then, for the versicolor petals equal or higher than 2.5 cms, it is divided again into two groups taking into account if the Petal.Width is lower (2.2.1), or equal or higher (2.2.2) than 1.8 cms.

(b) What is the default class label associated with each vertex? Your output should look like the following: Level N, Vertex 1: Default class label is Level N, Vertex 2: Default class label is ...

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>(fifty fifty)
Level 3, Vertex 1: Default class label is <versicolor>(91/100)
Level 3, Vertex 2: Default class label is <virginica> (98/100)

(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 N, split on attribute: . Split points: < X.X left subtree, >= X.X right subtree …

Level 1, split on attribute: Petal.Length (2.5 cms)
Split points: <2.5 cms left subtree, >=2.5 cms right subtree

Level 2, split on attribute: Petal.Width (1.8 cms)
Split points: <1.8 cms left subtree, >=1.8 cms right subtree

(d) Each vertex has three lines.

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

The numbers in the middle line mean the predicted probability for every group classification.

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

The last line contains the value of the likelyhood of pertaining to that group in the classification made.

Author: Julen Ferro Bañales