

Question 6: (Performance)

The iris_rounded dataset had a tree size of 61 and an average classification rate of 0.902. Although the dataset had only 150 examples, it was able to have to a reasonable tree size (little less than half the number of examples) demonstrating that the decision tree algorithm was able to reduce the entropy of the original example set enough such that it did not take a larger tree to represent the problem (it was able to find patterns via picking good attributes to split upon). The average classification rate was pretty high, but not 1.00, demonstrating both not-overfitting the examples and accurate splitting and classification of the training set.

Question 7: (Applications)

The iris_rounded dataset and similar datasets can be used in conjunction with GUI for a ecology-classification app where the user can enter in details about the flower or other plant-life found in the wild and the back-end decision tree can display possible classifications of that flower / plant-life. One step further would be if the attribute values (sepal length/width, pedal length/width) were collected using an image recognition software so the user can simply take a picture of the flower / plant-life and the image recognition software would extract the attribute values to be passed into the decision tree.