CS 686-02 - Lab 3 Spring, 2018

Lab 3 - Decision Tree Implementation

The goal of this assignment is to change an implementation of the Decision Tree algorithm as a class in our object-oriented framework. You will test this implementation by executing your algorithm against the Zoo Animals dataset.

Background

We discussed an n-ary Decision Tree classifier during lecture. Specifically, it chooses the split along lines with the highest information gain as measured by two metrics: entropy and gini impurity. Your task is to implement the Decision Tree classifier using both metrics and test it against the Zoo Animals dataset.

Requirements

Perform the following steps:

1. Get the Zoo Animals dataset from the UCI Machine Learning Repository > Zoo Data Set. Alternatively, get this data from Kaggle > Zoo Animal Classification.
2. Create your own decsion_tree class in decision_tree.py, which has classifier.py as its superlcass. Implement three functions in svm_basic.py:
    a. __init__(self) — i.e. the constructor
    b. fit(self, X, Y)
    c. predict(self, X)

    This class must give the user the option of choosing EITHER entropy and gini impurity as the metric to be used. Use entropy as the default. This option must be set through the constructor using the “criterion” parameter, such as is shown in the following:

    ```
    classifier = decision_tree(criterion='gini')
    ```

Submission

Submit your source code or link to your github repository on Canvas.

Grading

Your grade for this assignment will be as follows:

- 100% = Implementation works correctly.
- 75% = Implementation works but contains minor errors (eg. incorrect implementation of metric).
- 50% = Implementation works but contains major errors (eg. missing portions of implementation).
- 0% = Implementation not attempted or not submitted on time.