CS 6375
Assignment 2-Part 2: Decision Trees

Names of students in your group:
ADRITA DUTTA:axd172930
CHIRAG SHAHI:cxs180005

Number of free late days used: **0**
Note: You are allowed a total of 4 free late days for the entire semester. You can use at most 2 for each assignment. After that, there will be a penalty of 10% for each late day.

Please list clearly all the sources/references that you have used in this assignment.
References:
Chapter 3- Tom Mitchell

Programming Part (70 points)

1)DataSet:
https://archive.ics.uci.edu/ml/machine-learning-databases/iris/iris.data

Accuracy before pruning:0.97
Accuracy after pruning(Given method):0.67

Accuracy before pruning:0.97
Accuracy after pruning(Pruning all nodes 1 level above leaf node):0.73

Output code with our method of pruning:

```
Chirags-MacBook-Pro:desktop chiragshahi$ python3 ./driver.py
Is PetalW >= 1.0?
--> True:
  Is PetalW >= 1.8?
  --> True:
    Is PetalL >= 4.9?
    --> True:
      Predict {'Iris-virginica': 43}
Iris-virginica 43
    --> False:
      Is SepalW >= 3.2?
      --> True:
        Predict {'Iris-versicolor': 1}
Iris-versicolor 1
      --> False:
        Predict {'Iris-virginica': 2}
Iris-virginica 2
Decision Node Depth      3
Decision Node Id       13
Decision Node Depth    2
Decision Node Id     6
  --> False:
    Is PetalL >= 5.0?
    --> True:
      Is PetalW >= 1.6?
      --> True:
        Is PetalL >= 5.8?
        --> True:
          Predict {'Iris-virginica': 1}
Iris-virginica 1
        --> False:
          Predict {'Iris-versicolor': 2}
Iris-versicolor 2
Decision Node Depth        4
Decision Node Id         26
      --> False:
        Predict {'Iris-virginica': 3}
Iris-virginica 3
Decision Node Depth      3
Decision Node Id       12
    --> False:
      Is PetalW >= 1.7?
      --> True:
        Predict {'Iris-virginica': 1}
Iris-virginica 1
      --> False:
        Predict {'Iris-versicolor': 47}
Iris-versicolor 47
Decision Node Depth      3
Decision Node Id       11
Decision Node Depth    2
Decision Node Id     5
```

```
Decision Node Depth  1
Decision Node Id  2
--> False:
  Predict {'Iris-setosa': 50}
Iris-setosa 50
Decision Node Depth0
Decision Node Id0
********** Leaf nodes ****************
id = 14 depth =3
id = 28 depth =4
id = 27 depth =4
id = 54 depth =5
id = 53 depth =5
id = 25 depth =4
id = 24 depth =4
id = 23 depth =4
id = 1 depth =1
********** Non-leaf nodes ****************
id = 0 depth =0
id = 2 depth =1
id = 6 depth =2
id = 13 depth =3
id = 5 depth =2
id = 12 depth =3
id = 26 depth =4
id = 11 depth =3
**************Tree before pruning*******
Is PetalW >= 1.0?
--> True:
  Is PetalL >= 4.8?
  --> True:
    Is PetalL >= 5.1?
    --> True:
      Predict {'Iris-virginica': 30}
Iris-virginica 30
    --> False:
      Is SepalL >= 6.7?
      --> True:
        Predict {'Iris-versicolor': 3}
Iris-versicolor 3
      --> False:
        Is SepalW >= 3.2?
        --> True:
          Predict {'Iris-versicolor': 1}
Iris-versicolor 1
        --> False:
          Is PetalW >= 1.8?
          --> True:
            Predict {'Iris-virginica': 6}
Iris-virginica 6
          --> False:
            Is PetalL >= 5.0?
```

```
Iris-virginica 1
            --> False:
              Predict {'Iris-versicolor': 1}
Iris-versicolor 1
Decision Node Depth              6
Decision Node Id              111
Decision Node Depth          5
Decision Node Id          55
Decision Node Depth        4
Decision Node Id        27
Decision Node Depth      3
Decision Node Id      13
Decision Node Depth    2
Decision Node Id      6
  --> False:
    Is PetalW >= 1.7?
    --> True:
      Predict {'Iris-virginica': 1}
Iris-virginica 1
    --> False:
      Predict {'Iris-versicolor': 37}
Iris-versicolor 37
Decision Node Depth      2
Decision Node Id      5
Decision Node Depth  1
Decision Node Id  2
--> False:
  Predict {'Iris-setosa': 40}
Iris-setosa 40
Decision Node Depth0
Decision Node Id0
Accuracy on test = 0.97
**************Tree after pruning*******
Is PetalW >= 1.0?
--> True:
  Is PetalW >= 1.8?
  --> True:
    Predict {'Iris-versicolor': 1, 'Iris-virginica': 45}
Iris-versicolor 1
Iris-virginica 45
  --> False:
    Predict {'Iris-versicolor': 49, 'Iris-virginica': 5}
Iris-versicolor 49
Iris-virginica 5
Decision Node Depth  1
Decision Node Id  2
--> False:
  Predict {'Iris-setosa': 50}
Iris-setosa 50
Decision Node Depth0
Decision Node Id0
Accuracy on test = 0.73
```

2)DataSet:
https://archive.ics.uci.edu/ml/machine-learning-databases/car/car.data

Changes made in driver.py while trying this:
header = ['a', 'b', 'c', 'd', 'e','f','Class']
df = pd.read_csv('https://archive.ics.uci.edu/ml/machine-learning-databases/car/car.data',
header=None, names=['a','b','c','d','e','f','Class'])

Accuracy before pruning:0.97
Accuracy after pruning (Given method):0.97

Accuracy before pruning:0.98
Accuracy after pruning(Our method):0.77

3)DataSet:
Changes made in driver.py while trying this:
header = ['top-left', 'top-middle', 'top-right', 'middle-left', 'middle-middle', 'middle-right', 'bottom-left',
'bottom-middle', bottom-right', 'Class']
df = pd.read_csv('https://archive.ics.uci.edu/ml/machine-learning-databases/tic-tac-toe/tic-tac-
toe.data',
header=None, names= ['top-left', 'top-middle', 'top-right', 'middle-left', 'middle-middle', 'middle-right',
'bottom-left', 'bottom-middle', 'bottom-right', 'Class'])

Accuracy before pruning:0.95
Accuracy after pruning (Given method):0.96

Accuracy before pruning:0.98
Accuracy after pruning(Our method):0.41

Pruning method used by us: Pruning all nodes that are one level above leaf node.

Result summary:
Dataset1:
In case of pruning (using given method)- Accuracy decreases

In case of pruning (using our chosen method of pruning all nodes 1 level above leaf)-

Since in this case we pruned all nodes 1 level above leaf nodes irrespective of whether that is increasing or decreasing the accuracy, thus in this case also accuracy decreases, but it was observed that the rate of decrease in accuracy is less compared to the other case.

Dataset2:
In case of pruning (using given method)- Accuracy remains constant

In case of pruning (using our chosen method of pruning all nodes 1 level above leaf)-
Since in this case we pruned all nodes 1 level above leaf nodes irrespective of whether that is increasing or decreasing the accuracy, thus in this case accuracy decreases.

Dataset3:
In case of pruning (using given method)- Accuracy increases slightly

In case of pruning (using our chosen method of pruning all nodes 1 level above leaf)-
Since in this case we pruned all nodes 1 level above leaf nodes irrespective of whether that is increasing or decreasing the accuracy, thus in this case accuracy decreases.In this case accuracy decreases drastically.