CPS 844 Lab 4

03/02/2022
Section 1
Ryan Soliven
500840952

#1

1) (10 points) Load the data (Y is the class labels of X)

Code

```
X = np.load('Xdata.npy')
Y = np.load('Ydata.npy')
```

#2

2) (15 points) Split the training and test data as follows:
80% of the data for training and 20% for testing.
Preserve the percentage of samples for each class using the argument 'stratify'.
Use the argument 'random' so that the data splitting is the same every time your code is run.

Code

```
dataTrain, dataTest, classTrain, classTest = train_test_split(X, Y, test_size=0.2, stratify=Y, random_state=3)
```

#3

3) (50 points) Test the fit of different decision tree depths
Instruction 1: Use the range function to create different depths options, ranging from 1 to 50, for the decision trees
Instruction 2: As you iterate through the different tree depth options, please:
create a new decision tree using the 'max_depth' argument
train your tree
apply your tree to predict the 'training' and then the 'test' labels
compute the training accuracy
compute the test accuracy
save the training & testing accuracies and tree depth, so that you can use them in the next steps

Code

```
trainAccuracy = []
testAccuracy = []
treeDepth = []
for i in range(1, 51):
    clf = tree.DecisionTreeClassifier(criterion = 'entropy', max_depth=i)
    clf = clf.fit(dataTrain, classTrain)
    predA = clf.predict(dataTrain)
    trainAccuracy.append(accuracy_score(classTrain, predA)*100)
    predB = clf.predict(dataTest)
    testAccuracy.append(accuracy_score(classTest, predB)*100)
    treeDepth.append(i)
```

#4

4) (10 points) Plot of training and test accuracies vs the tree depths

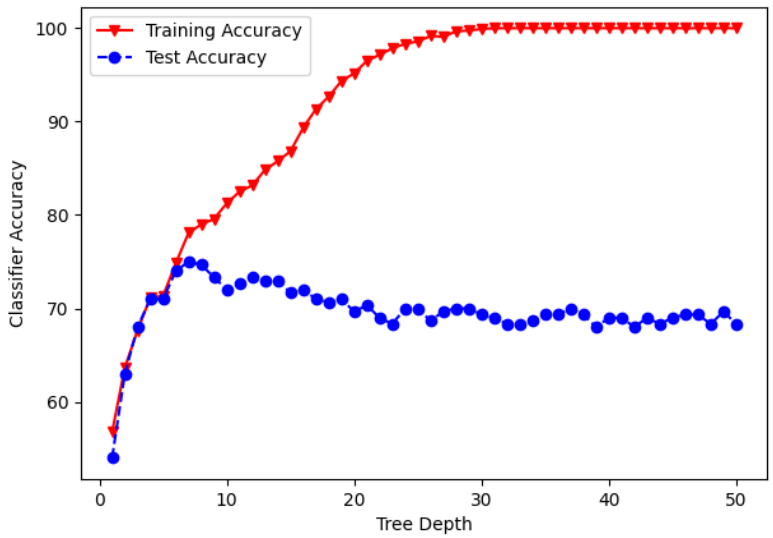
Code

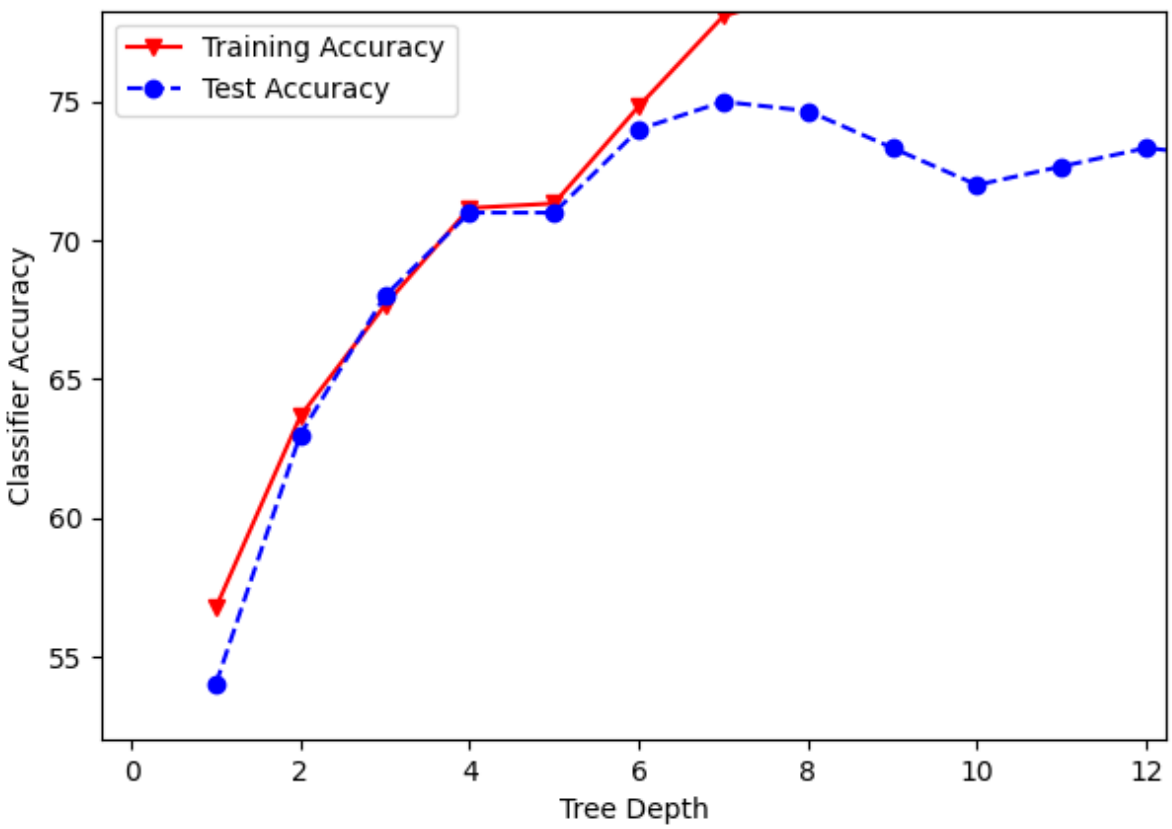
```
plt.plot(treeDepth,trainAccuracy,'rv-',treeDepth,testAccuracy,'bo--')
plt.legend(['Training Accuracy','Test Accuracy'])
plt.xlabel('Tree Depth')
plt.ylabel('Classifier Accuracy')
plt.show()
```

#5

5) (15 points) Fill out the following blank:

Model overfitting happens when the tree depth is greater than **7**, approximately.




Training Accuracy
Test Accuracy
Classifier Accuracy
Tree Depth
75
70
65
60
55
0
2
4
6
8
10
12