

Part 1: Bagging

```
In [1]: import numpy as np
import multiprocessing
import Process
from array import *
from sklearn.datasets import make_classification
from sklearn.ensemble import BaggingClassifier
from sklearn.tree import DecisionTreeClassifier
from sklearn.tree import plot_tree
from sklearn import tree
import graphviz
from IPython.display import Image
import matplotlib.pyplot as plt

In [2]: #Get the data from figure 3.36 to a 2D array for simplicity
A = C
TrainingAttr = [[0, 0, 0],
               [0, 0, 1],
               [0, 1, 0],
               [0, 1, 1],
               [1, 0, 1],
               [1, 0, 0],
               [1, 1, 0],
               [1, 1, 1],
               [1, 0, 1],
               [1, 1, 0],
               [1, 1, 0],
               [1, 1, 1]]

TrainingClass = [1, 1, 1, -1, 1, 1, -1, 1, -1, -1, -1]

# A B C
ValidationAttr = [[0, 0, 0],
                 [0, 1, 1],
                 [1, 1, 0],
                 [1, 0, 1],
                 [1, 0, 0],
                 [1, 0, 0]]

ValidationClass = [1, 1, 1, -1, 1, 1]

classifiers = ['+', '+', '-']

attr = ['A', 'B', 'C']

#-----
# function to generate the Bagging Ensemble model
def baggingModel():
    # make the model
    clf = BaggingClassifier(base_estimator=None, n_estimators=10, random_state=0)
    clf.fit(TrainingAttr, TrainingClass)

    #get the decision trees that are part of this ensemble
    decisionTrees = get_attr(clf, 'estimators_')

    #get the data that was used in each tree
    decisionTreeData = get_attr(clf, 'estimators_samples_')

    #print the decision trees along with the classes
    print("Bagging Ensemble Individual Decision Trees:")
    itr = 0
    for x in decisionTrees:
        #print the data that went into each data set
        print("Decision Tree Number: " + str(itr) + "\n")
        print("A, B, C [Class]")
        for i in range(10):
            #format attributes into printable set
            dset = ", ".join(str(m) for m in TrainingAttr[decisionTreeData[itr][i]])

            #print with formatting
            print("[% + dset + %]" % (classifiers[TrainingClass[decisionTreeData[itr][i]] + 1]))

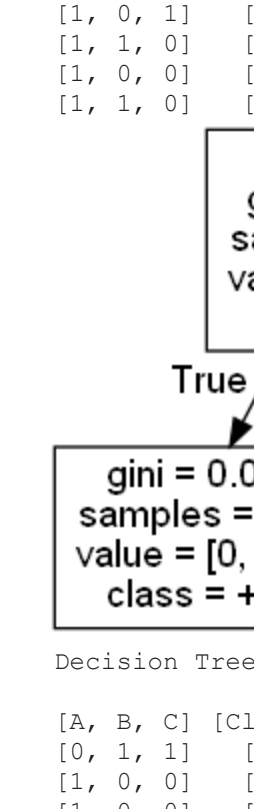
        #display the decision tree
        d_data = tree.export_graphviz(x, feature_names=attr, class_names=['-', '+'])
        graph = graphviz.Source(d_data, format="png")
        graph.render('Tree', view=False)
        graph.display(filename="Tree.png")
        itr += 1

    #return variables for predictions
    return clf, decisionTrees, decisionTreeData

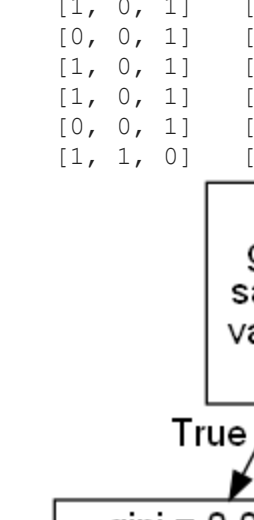
clf, decisionTrees, decisionTreeData = baggingModel()
```

Bagging Ensemble Individual Decision Trees:

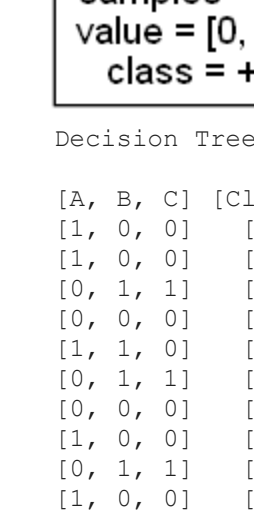
Decision Tree Number: 0



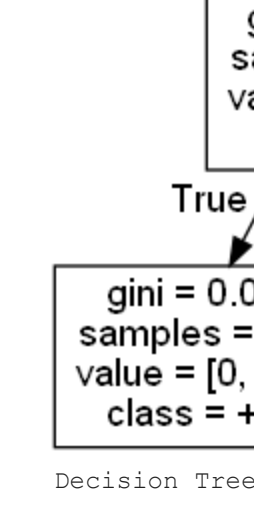
Decision Tree Number: 1



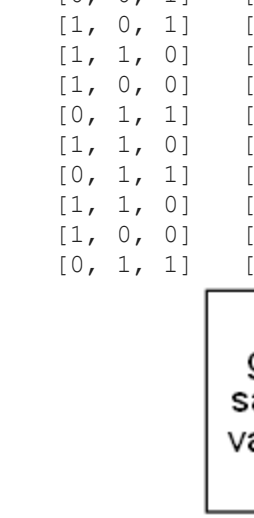
Decision Tree Number: 2



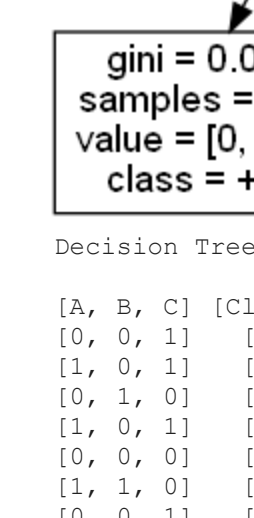
Decision Tree Number: 3



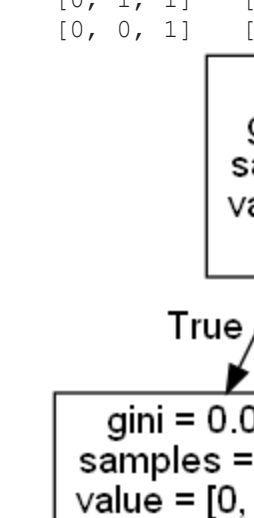
Decision Tree Number: 4



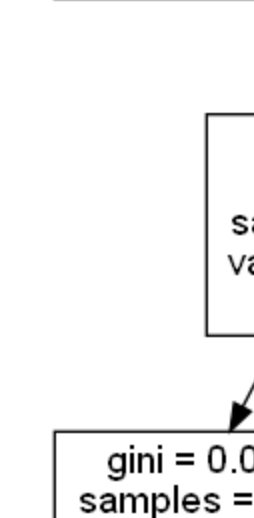
Decision Tree Number: 5



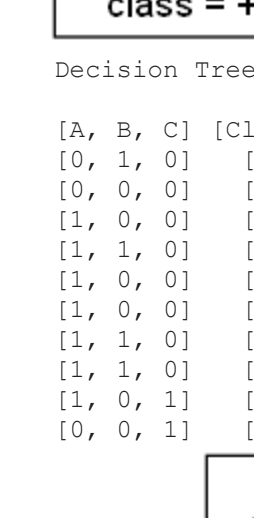
Decision Tree Number: 6



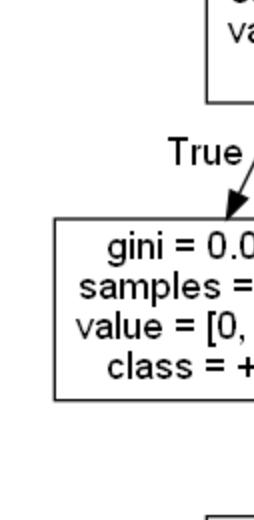
Decision Tree Number: 7



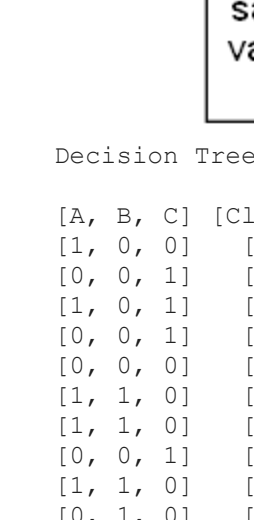
Decision Tree Number: 8



Decision Tree Number: 9



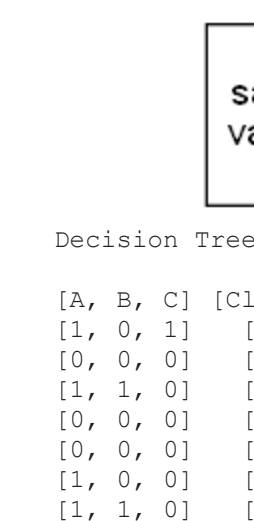
Decision Tree Number: 10



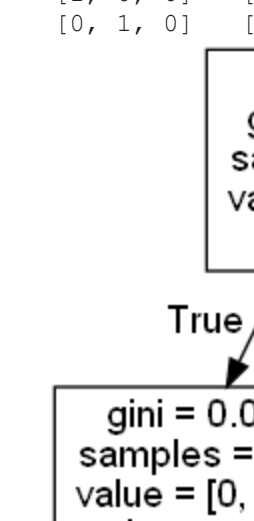
Decision Tree Number: 11



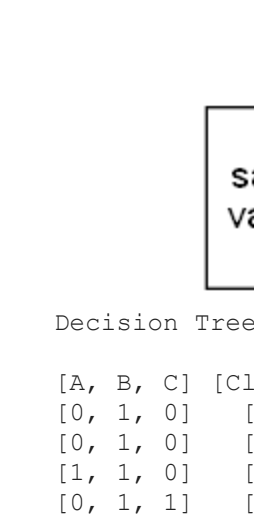
Decision Tree Number: 12



Decision Tree Number: 13



Decision Tree Number: 14



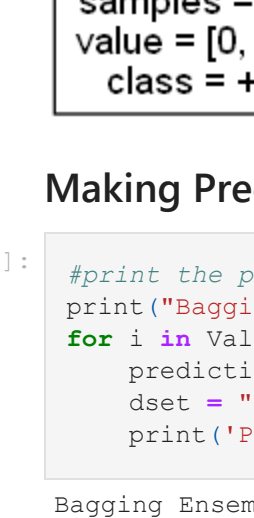
Decision Tree Number: 15



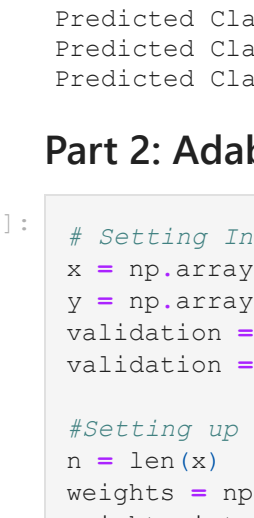
Decision Tree Number: 16



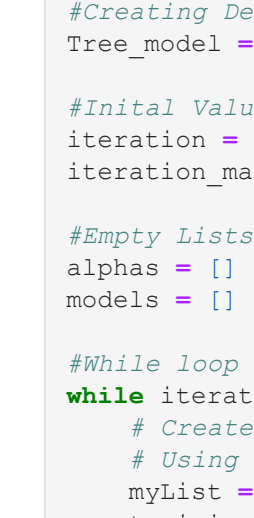
Decision Tree Number: 17



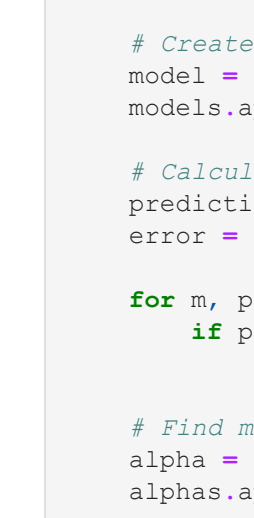
Decision Tree Number: 18



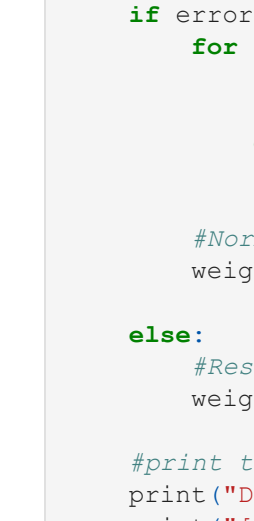
Decision Tree Number: 19



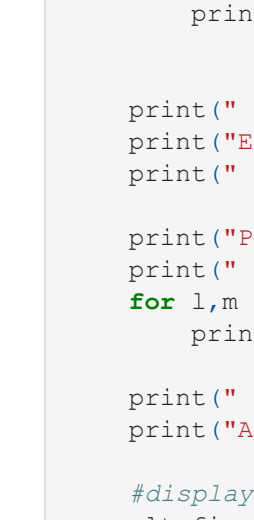
Decision Tree Number: 20



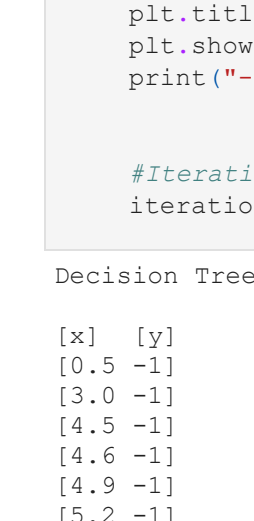
Decision Tree Number: 21



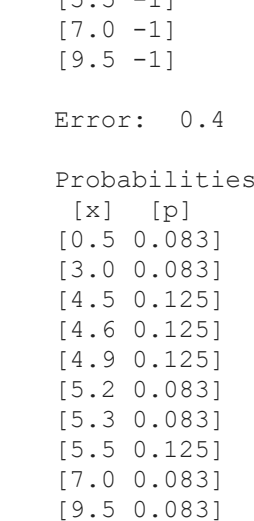
Decision Tree Number: 22



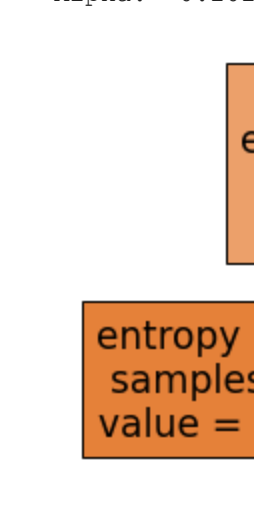
Decision Tree Number: 23



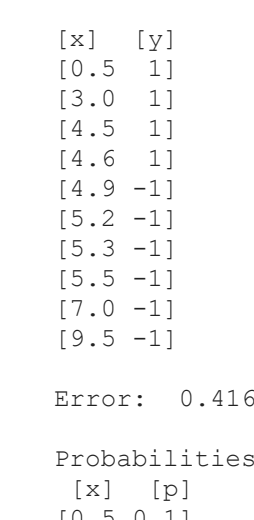
Decision Tree Number: 24



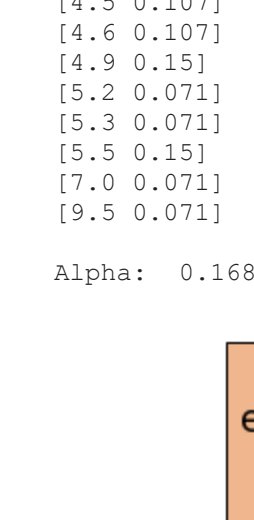
Decision Tree Number: 25



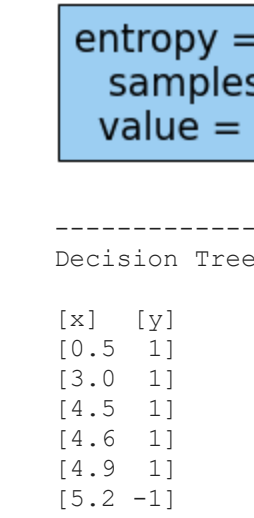
Decision Tree Number: 26



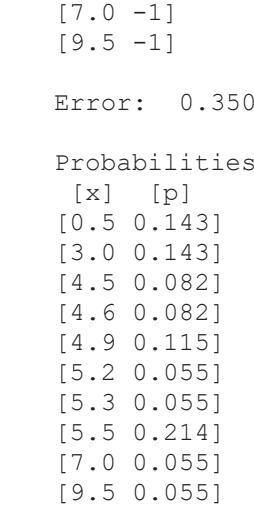
Decision Tree Number: 27



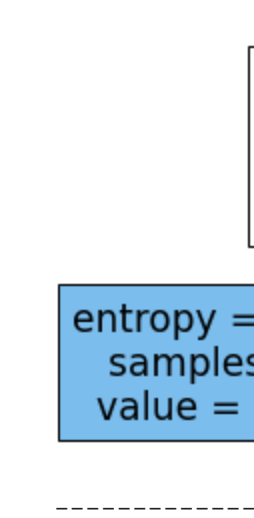
Decision Tree Number: 28



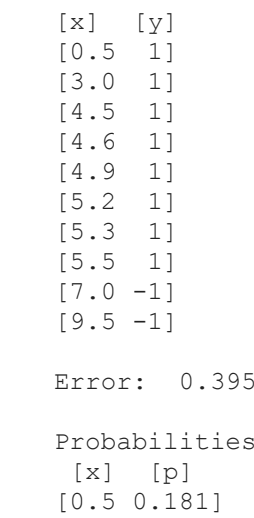
Decision Tree Number: 29



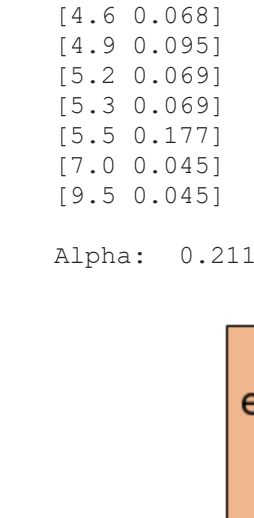
Decision Tree Number: 30



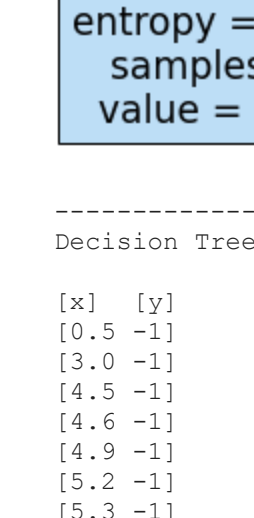
Decision Tree Number: 31



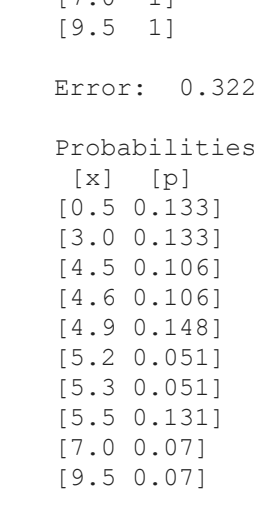
Decision Tree Number: 32



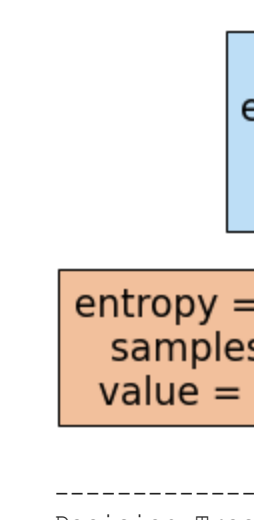
Decision Tree Number: 33



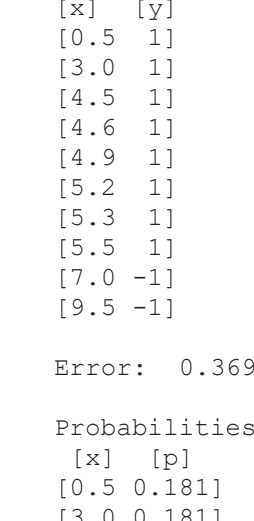
Decision Tree Number: 34



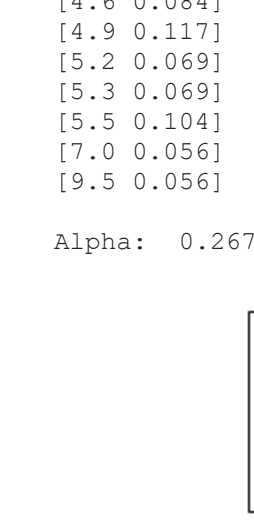
Decision Tree Number: 35



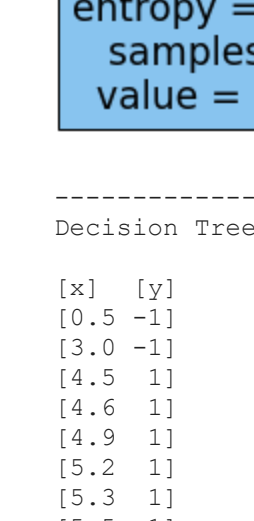
Decision Tree Number: 36



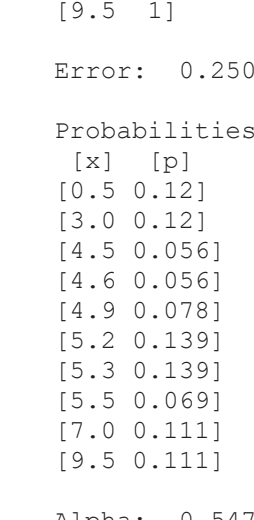
Decision Tree Number: 37



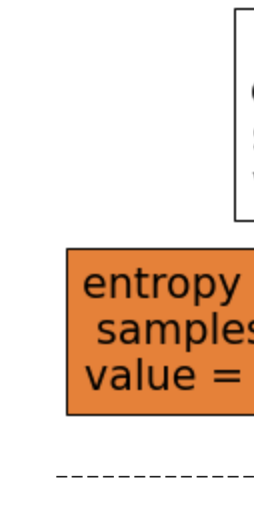
Decision Tree Number: 38



Decision Tree Number: 39



Decision Tree Number: 40



Decision Tree Number: 41

Decision Tree Number: 42

Decision Tree Number: 43

Decision Tree Number: 44

Decision Tree Number: 45

Decision Tree Number: 46

Decision Tree Number: 47

Decision Tree Number: 48

Decision Tree Number: 8

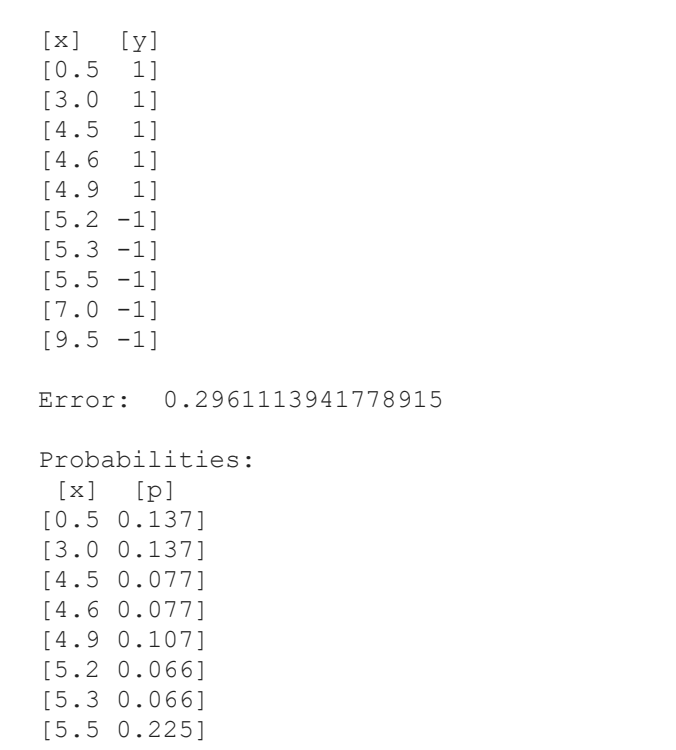
```
[x] [y]
[0.5 -1]
[3.0 -1]
[4.5 -1]
[4.6 -1]
[4.9 -1]
[5.2 -1]
[5.3 -1]
[5.5 -1]
[7.0 -1]
[9.5 -1]
```

Error: 0.2590934837899975

Probabilities:

```
[x] [p]
[0.5 0.081]
[3.0 0.081]
[4.5 0.108]
[4.6 0.108]
[4.9 0.151]
[5.2 0.094]
[5.3 0.094]
[5.5 0.134]
[7.0 0.075]
[9.5 0.075]
```

Alpha: 0.5253427602632438



Decision Tree Number: 9

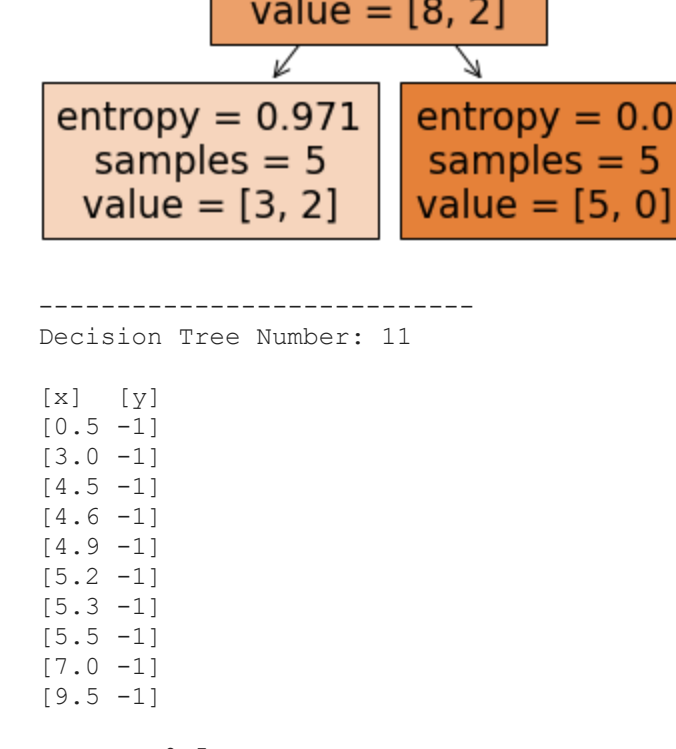
```
[x] [y]
[0.5 1]
[3.0 1]
[4.5 1]
[4.6 1]
[4.9 1]
[5.2 -1]
[5.3 -1]
[5.5 -1]
[7.0 -1]
[9.5 -1]
```

Error: 0.2961113941778915

Probabilities:

```
[x] [p]
[0.5 0.137]
[3.0 0.137]
[4.5 0.077]
[4.6 0.077]
[4.9 0.107]
[5.2 0.066]
[5.3 0.066]
[5.5 0.225]
[7.0 0.053]
[9.5 0.053]
```

Alpha: 0.4329421990304272



Decision Tree Number: 10

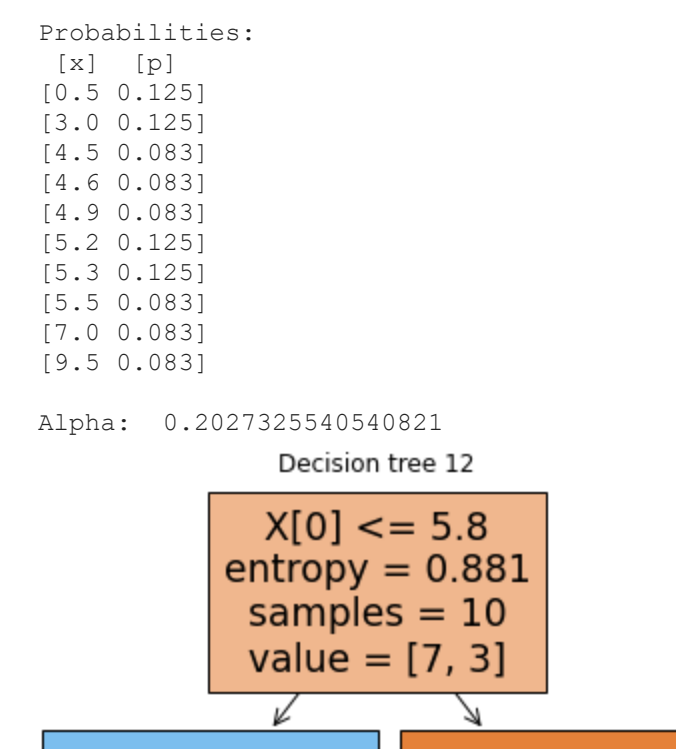
```
[x] [y]
[0.5 -1]
[3.0 -1]
[4.5 -1]
[4.6 -1]
[4.9 -1]
[5.2 -1]
[5.3 -1]
[5.5 -1]
[7.0 -1]
[9.5 -1]
```

Error: 0.4857965103707136

Probabilities:

```
[x] [p]
[0.5 0.133]
[3.0 0.133]
[4.5 0.079]
[4.6 0.079]
[4.9 0.11]
[5.2 0.065]
[5.3 0.065]
[5.5 0.232]
[7.0 0.052]
[9.5 0.052]
```

Alpha: 0.028414624025551817



Decision Tree Number: 11

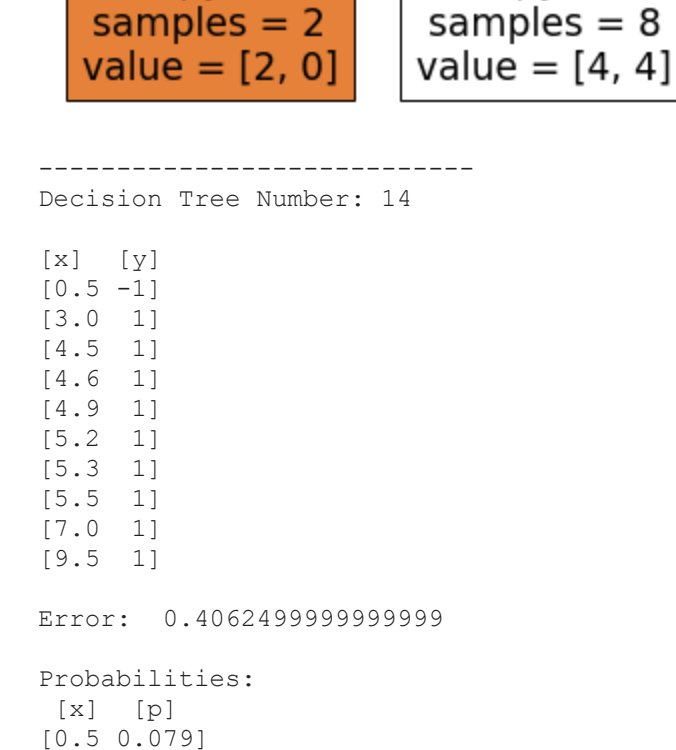
```
[x] [y]
[0.5 -1]
[3.0 -1]
[4.5 -1]
[4.6 -1]
[4.9 -1]
[5.2 -1]
[5.3 -1]
[5.5 -1]
[7.0 -1]
[9.5 -1]
```

Error: 0.5

Probabilities:

```
[x] [p]
[0.5 0.1]
[3.0 0.1]
[4.5 0.1]
[4.6 0.1]
[4.9 0.1]
[5.2 0.1]
[5.3 0.1]
[5.5 0.1]
[7.0 0.1]
[9.5 0.1]
```

Alpha: 0.0



Decision Tree Number: 12

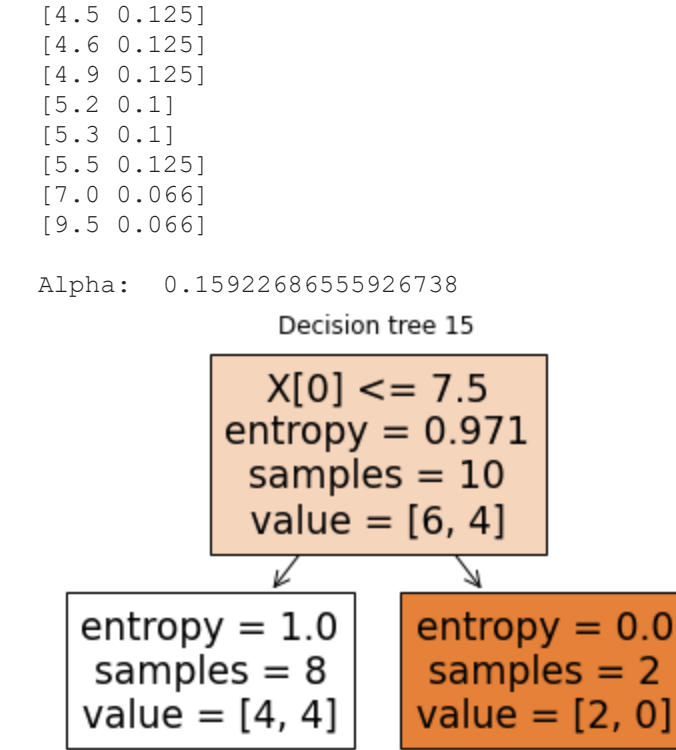
```
[x] [y]
[0.5 1]
[3.0 1]
[4.5 1]
[4.6 1]
[4.9 1]
[5.2 1]
[5.3 1]
[5.5 1]
[7.0 -1]
[9.5 -1]
```

Error: 0.4

Probabilities:

```
[x] [p]
[0.5 0.125]
[3.0 0.125]
[4.5 0.083]
[4.6 0.083]
[4.9 0.083]
[5.2 0.125]
[5.3 0.125]
[5.5 0.083]
[7.0 0.083]
[9.5 0.083]
```

Alpha: 0.2027325540540821



Decision Tree Number: 13

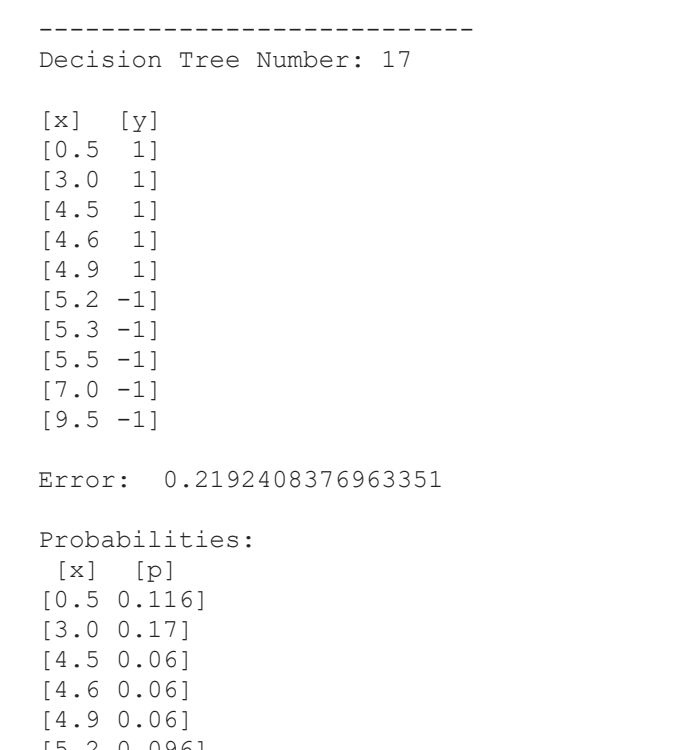
```
[x] [y]
[0.5 -1]
[3.0 -1]
[4.5 -1]
[4.6 -1]
[4.9 -1]
[5.2 -1]
[5.3 -1]
[5.5 -1]
[7.0 -1]
[9.5 -1]
```

Error: 0.3333333333333333

Probabilities:

```
[x] [p]
[0.5 0.094]
[3.0 0.094]
[4.5 0.125]
[4.6 0.125]
[4.9 0.125]
[5.2 0.094]
[5.3 0.094]
[5.5 0.125]
[7.0 0.062]
[9.5 0.062]
```

Alpha: 0.34657359027997275



Decision Tree Number: 14

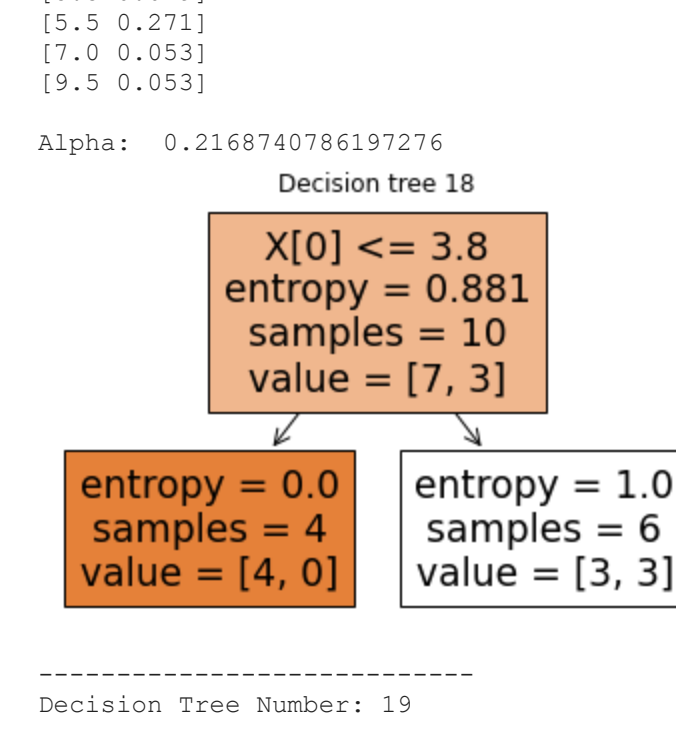
```
[x] [y]
[0.5 -1]
[3.0 1]
[4.5 1]
[4.6 1]
[4.9 1]
[5.2 1]
[5.3 1]
[5.5 1]
[7.0 1]
[9.5 1]
```

Error: 0.4062499999999999

Probabilities:

```
[x] [p]
[0.5 0.079]
[3.0 0.115]
[4.5 0.105]
[4.6 0.105]
[4.9 0.105]
[5.2 0.115]
[5.3 0.115]
[5.5 0.105]
[7.0 0.077]
[9.5 0.077]
```

Alpha: 0.18974481085245207



Decision Tree Number: 15

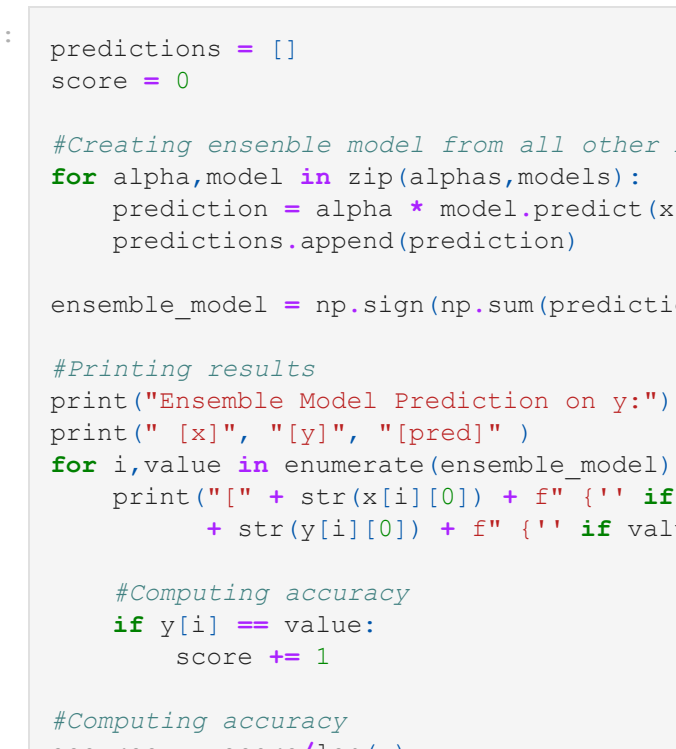
```
[x] [y]
[0.5 -1]
[3.0 -1]
[4.5 -1]
[4.6 -1]
[4.9 -1]
[5.2 -1]
[5.3 -1]
[5.5 -1]
[7.0 -1]
[9.5 -1]
```

Error: 0.4210526315789473

Probabilities:

```
[x] [p]
[0.5 0.068]
[3.0 0.1]
[4.5 0.125]
[4.6 0.125]
[4.9 0.125]
[5.2 0.1]
[5.3 0.1]
[5.5 0.125]
[7.0 0.066]
[9.5 0.066]
```

Alpha: 0.15922686555926738



Decision Tree Number: 16

```
[x] [y]
[0.5 -1]
[3.0 -1]
[4.5 1]
[4.6 1]
[4.9 1]
[5.2 1]
[5.3 1]
[5.5 1]
[7.0 1]
[9.5 1]
```

Error: 0.3321678321678322

Probabilities:

```
[x] [p]
[0.5 0.051]
[3.0 0.075]
[4.5 0.094]
[4.6 0.094]
[4.9 0.094]
[5.2 0.15]
[5.3 0.15]
[5.5 0.094]
[7.0 0.1]
[9.5 0.1]
```

Alpha: 0.3491982682230444



Decision Tree Number: 17

```
[x] [y]
[0.5 1]
[3.0 1]
[4.5 1]
[4.6 1]
[4.9 1]
[5.2 -1]
[5.3 -1]
[5.5 -1]
[7.0 -1]
[9.5 -1]
```

Error: 0.2192408376963351

Probabilities:

```
[x] [p]
[0.5 0.116]
[3.0 0.17]
[4.5 0.06]
[4.6 0.06]
[4.9 0.06]
[5.2 0.096]
[5.3 0.096]
[5.5 0.213]
[7.0 0.064]
[9.5 0.064]
```

Alpha: 0.6350479451364248



Decision Tree Number: 18

```
[x] [y]
[0.5 -1]
[3.0 -1]
[4.5 -1]
[4.6 -1]
[4.9 -1]
[5.2 -1]
[5.3 -1]
[5.5 -1]
[7.0 -1]
[9.5 -1]
```

Error: 0.3932316623087414

Probabilities:

```
[x] [p]
[0.5 0.096]
[3.0 0.14]
[4.5 0.076]
[4.6 0.076]
[4.9 0.076]
[5.2 0.079]
[5.3 0.079]
[5.5 0.271]
[7.0 0.053]
[9.5 0.053]
```

Alpha: 0.2168740786197276



Decision Tree Number: 19

```
[x] [y]
[0.5 1]
[3.0 1]
[4.5 1]
[4.6 1]
[4.9 1]
[5.2 1]
[5.3 1]
[5.5 1]
[7.0 -1]
[9.5 -1]
```

Error: 0.3944568614998887

Probabilities:

```
[x] [p]
[0.5 0.122]
[3.0 0.178]
[4.5 0.063]
[4.6 0.063]
[4.9 0.063]
[5.2 0.1]
[5.3 0.1]
[5.5 0.224]
[7.0 0.044]
[9.5 0.044]
```

Alpha: 0.21430801019251342



Using final ensemble model to predict X

```
In [6]:
predictions = []
score = 0

#Creating ensemble model from all other models
for alpha,model in zip(alphas,models):
    prediction = alpha * model.predict(x)
    predictions.append(prediction)

ensemble_model = np.sign(np.sum(predictions,axis=0))

#Printing Results
print("Ensemble Model Prediction on y:")
print(" [x] ", "[y]" )
for i,value in enumerate(ensemble_model):
    print("[" + str(x[i][0]) + " + i" ["' if y[i] < 0 else ' ']'
    + str(y[i][0]) + " + i" ["' if value < 0 else ' ']' + str(value) + "])")

#Computing accuracy
if y[i] == value:
    score += 1

#Computing accuracy
accuracy = score/len(y)

#Printing Accuracy
print("Accuracy: ", accuracy)
```

Ensemble Model Prediction on y:

```
[x] [y] [pred]
[0.5 -1 1.0]
[3.0 -1 1.0]
[4.5 1 1.0]
[4.6 1 1.0]
[4.9 1 1.0]
[5.2 -1 1.0]
[5.3 -1 1.0]
[5.5 1 1.0]
[7.0 -1 -1.0]
[9.5 -1 -1.0]
```

Accuracy: 0.6

Using final ensemble model to predict validation set

```
In [6]:
predictions = []

#Creating ensemble model from all other models
for alpha,model in zip(alphas,models):
    prediction = alpha * model.predict(validation)
    predictions.append(prediction)

ensemble_model = np.sign(np.sum(predictions,axis=0))

#Printing Results
print(" [x] ", "[y]" )
for i,value in enumerate(zip(validation, ensemble_model)):
    print("[" + str(i) + " + i" ["' if m[i] < 0 else ' ']' + str(m[i]) + "])")

[x] [y]
[0 1.0]
[1 1.0]
[2 1.0]
[3 1.0]
[4 1.0]
[5 1.0]
[6 -1.0]
[7 -1.0]
[8 -1.0]
[9 -1.0]
```