Homework 2

a) To implement Information Gain

Get index from the user that is for which attribute we want to get information gain and get the split value of that attribute from the dataset.

I have define entropy function to get entropy for whole data set and weighted entropy for partitioned dataset.

Information gain is difference between Entropy and weighted entropy.

b) To implement Gini Index

Get index from the user that is for which attribute we want to get gini index and get the split value of that attribute from the dataset.

c) To implement CART

Get index from the user that is for which attribute we want to get CART and get the split value of that attribute from the dataset.

Solutions for all the above function:

```
C:\Python27\python.exe C:/Users/Dell/PycharmProjects/ICSI531/Homework_2/Kalpita_HW2.py
100
Enter Column from (0 to 9)4
('Given Index', 4)
('Given Value', 3)
---------------
('Information Gain', 0.0013747650883804141)
('Gini Index', 0.41919999999999996)
('Classification and Regression tree value is', 0.03999999999999994)
```

d) Bestsplit

By executing above function for all the attributes and comparing all the value of measure with each attribute we get the bestsplit.

e) Loadfile()

To load train data set and test dataset.

f) Decision Tree

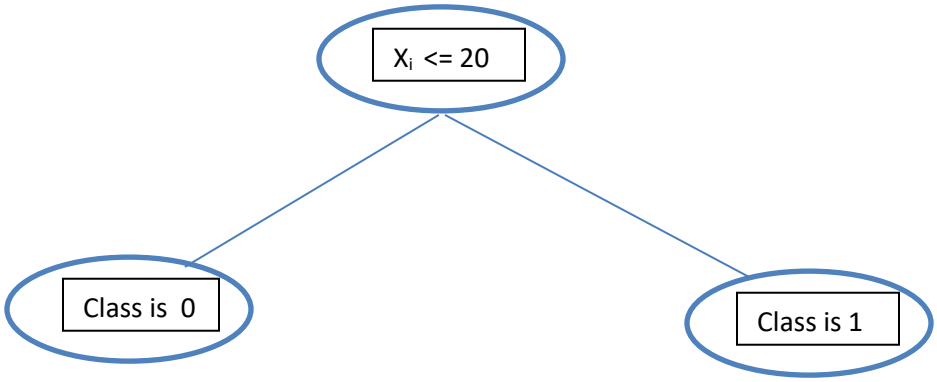
a. **Information gain**

('Best Index', 1, 'Best Split Value', 20.0)

('Classified using IG', [1, 1, 1, 1, 0, 1, 1, 1, 1, 1])

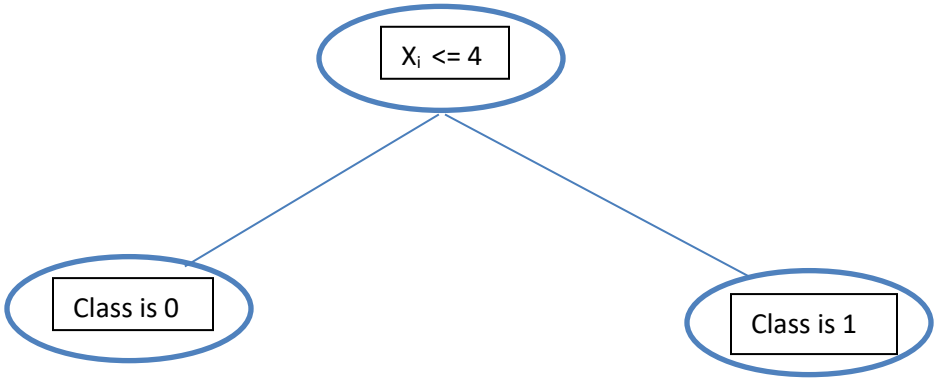
('Predicted class values using Information gain', [1, 1, 1, 1, 0, 1, 1, 1, 1, 1])

Homework 2



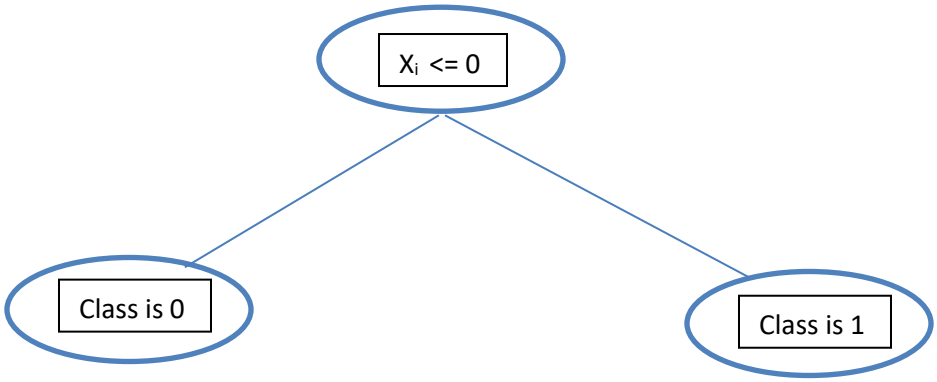
b. **Gini Index**

('Best Index', 5, 'Best Split Value', 4.0)
('Classified using GINI', [0, 0, 0, 0, 1, 1, 0, 0, 0, 0])
('Predicted class values using Gini Index', [0, 0, 0, 0, 1, 1, 0, 0, 0, 0])



c. **CART**

('Best Index', 2, 'Best Split Value', 0.0)
('Classified using CART', [0, 0, 0, 0, 1, 1, 0, 1, 0, 0])
('Predicted class values using CART', [0, 0, 0, 0, 1, 1, 0, 1, 0, 0])



g) Classification error:

Actual class column in test dataset: [0,0,0,0,1,1,0,1,0,0]

a. **Information gain:**

Homework 2

Predicted class with information gain is: [1, 1, 1, 1, 0, 1, 1, 1, 1, 1]
Two values are correctly classified
Zero 0 is correctly classified
Two 1's are correctly classified

b. **Gini Index:**

Predicted class with gini index is: [0, 0, 0, 0, 1, 1, 0, 0, 0, 0]
Nine values are correctly classified
seven 0's is correctly classified
Two 1's are correctly classified

c. **CART:**

Predicted class with CART is: [0, 0, 0, 0, 1, 1, 0, 1, 0, 0]
All values are correctly classified
Seven 0 is correctly classified
Three 1's are correctly classified