

# CSCI630 – Foundations of Artificial Intelligence

## Homework 3P

Author – Archit Satish Joshi (aj6082)

---

### 1. Decision Tree till level 2

(Note – The script dtree.py does not create a tree. It simply calculates information gained which will help us decide which attribute needs to be selected for deciding)

For attributes labelled between A1 .... A8 and labels A & B -

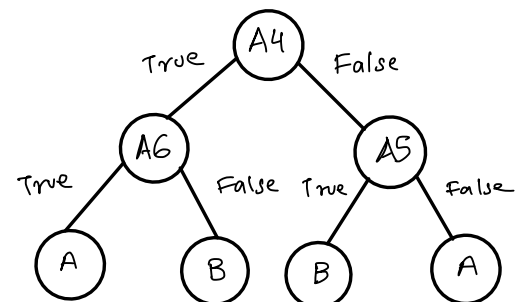
- First split would be done on attribute A4 which has a gain of 0.2634311261883375
- Second split for the True node from above step should be done on A6 which has a gain of 0.17767948750724227. This becomes left child of root node.  
For True value we would get result A as A\_true count is maximum. For False we would get B.
- Second split for the False node from (a) should be done on A5 which has a gain of 0.1045025776631836. This becomes the right child of root node.  
For True value we would get result B as B\_true count is maximum. For False we would get A.

The decision tree will look like follows:

(a) Program output

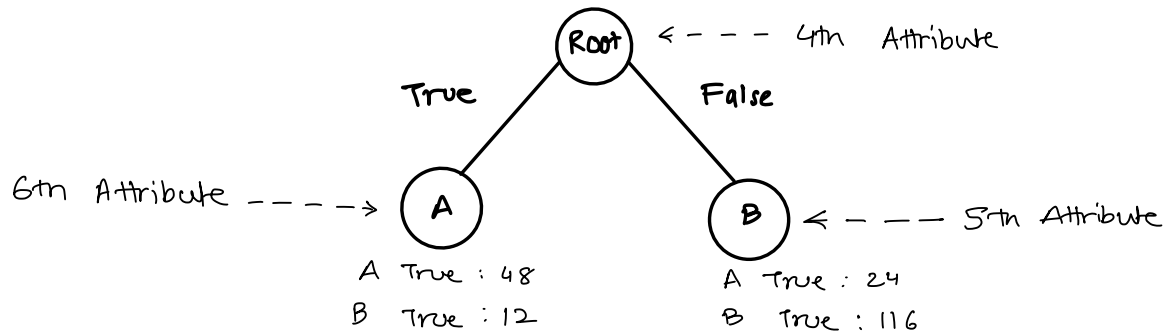
```
"D:\Projects\CSCI630 - Foundations of Artificial Intelligence\Homework_3\  
Attribute A4 to get to Level 1  
Gain using attribute A4 " 0.2634311261883375  
A True : 65  
B True : 102  
  
Attribute A6 for True node split  
Gain using attribute A6 : 0.17767948750724227  
A True : 45  
B True : 11  
  
Attribute A5 for False node split  
Gain using attribute A5 : 0.1045025776631836  
A True : 20  
B True : 91  
  
Process finished with exit code 0
```

(b) Decision Tree



## 2. Adaboost

For adaboost purposes let's consider our tree to look like this with stumps –



The first split is done on A4 which becomes our stump which is the root.

True split for root node was done on attribute A6 and False split for root node was done on attribute A5.

Node A will be the node of True split as A True count is greater for the split.  
(A True count = 48, B True count = 12 i.e., incorrectly classified A instances = 12)

Node B will be the node of False split as B False count is greater for the split.  
(A False count = 24, B False count = 116 i.e., incorrectly classified B instances = 24)

a. What would the error rate of the first stump be?

$N = 200$

Initial weight =  $1/N = 0.005$

$$\begin{aligned}\text{Error} &= (\text{Incorrectly classified A instances} * \text{Initial Weight}) + \\ &\quad (\text{Incorrectly classified B instances} * \text{Initial Weight}) \\ &= (12 * 0.005) + (24 * 0.005) \\ &= 0.18\end{aligned}$$

b. What would be the hypothesis weight of the first stump?

$$\text{Hypothesis Weight} = \ln \left( \frac{1 - \text{Error}}{\text{Error}} \right) = \ln \left( \frac{1 - 0.18}{0.18} \right) = 1.52$$

c. What would be the initial weights of each example?

$$N = 200$$

$$\text{Initial weight} = 1/N = 0.005$$

d. What would be the new weights of each example that was correctly classified, and each that was incorrectly classified?

$$\text{Incorrectly classified weight} = \text{Initial weight} = 0.005$$

$$\begin{aligned}\text{Correctly classified weight} &= \text{Initial weight} * ( \text{Error} / ( \text{Error} - 1 ) ) \\ &= 0.005 * ( 0.18 / ( 1 - 0.18 ) ) \\ &= 0.0010975\end{aligned}$$