Decision Tree Classifier

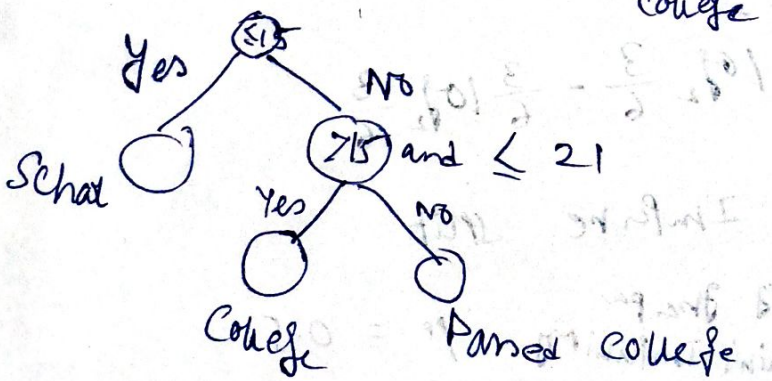
- ID3
- cart

Purity split:
1. Entropy
2. Gini Index

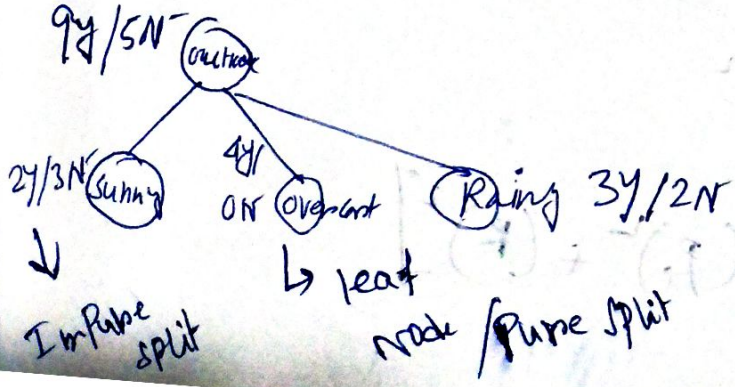
3. Information gain → feature

D. T Tree split

```
age = 14
if age <= 15:
    print ("school")
elif (age > 15 and age <= 21):
    print ("college")
else:
    print ("Passed college")
```



Tennis ville Play or not



① Purity → Pure split!
- Entropy
- Gini Index

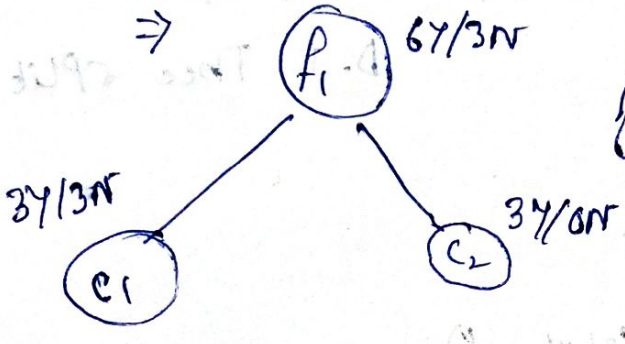
② Information Gain → How the features are selected.

Entropy

Gini Impurity

$$H(S) \Rightarrow -P_+ \log_2 P_+ - P_- \log_2 P_-$$

$$\Rightarrow$$



$$G.I = 1 - \sum_{i=1}^{n} (P_i)^2$$

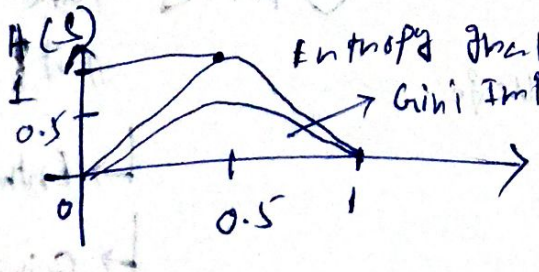
$$H(S) = -\frac{3}{3}\log_2\frac{3}{3} - \frac{0}{3}\log_2\frac{0}{3}$$

$$= -\log_2 1 = 0 \Rightarrow \text{Pure split.}$$

$$H(S) \Rightarrow -\frac{3}{6}\log_2\frac{3}{6} - \frac{3}{6}\log_2\frac{3}{6}$$

$H(C_1)$

$$= 1 \Rightarrow \text{Impure split}$$



$p_+ = 0.5$

$p_- = 1 - 0.5$

$= 0.5$

$$G.I = 1 - \sum_{i=1}^{n} (P)^2$$

$$= 1 - \sum_{i=1}^{n} \left[(P_+)^2 + (P_-)^2 \right]$$

3Y/3N

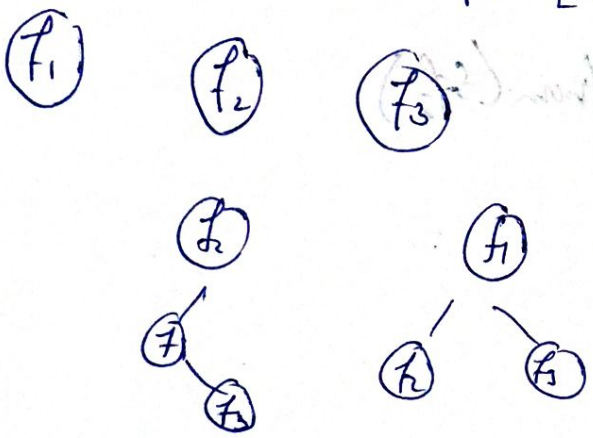
$$1 - \left[\left(\frac{1}{2}\right)^2 + \left(\frac{1}{2}\right)^2\right]$$

$$= 1 - \frac{1}{2} = 0.5$$

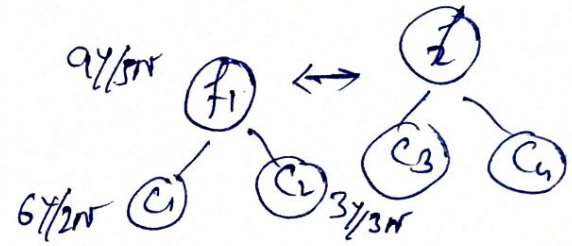
3Y/0N

$$= 1 - \left[\left(\frac{3}{3}\right)^2 + 0\right]$$

$$= 1 - [1+0] = 0$$



Information gain :-



Gain (S, f1) =>

Entropy of Root Node

$$H(S) - \sum_{V \in Val} \frac{|S_V|}{|S|} H(S_V)$$

$$H(S) = -P_+ \log_2 P_+ - P_- \log_2 P_-$$

$$= -\frac{9}{14} \log_2 \frac{9}{14} - \frac{5}{14} \log_2 \frac{5}{14}$$

$$\Rightarrow 0.94$$

$$H(C_2) = -\frac{3}{6} \log_2 \frac{3}{6} - \frac{3}{6} \log_2 \frac{3}{6}$$

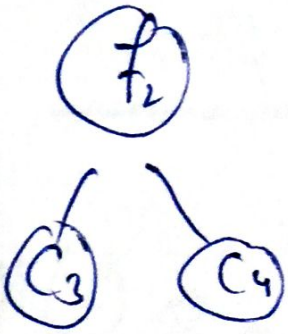
$$= 1$$

$$H(C_1) = -\frac{6}{8} \log_2 \frac{6}{8} - \frac{2}{8} \log_2 \frac{2}{8}$$

$$= 0.81$$

$$Gain(S, f_2) = 0.94 - \left[\frac{8}{14} * 0.81 + \frac{6}{14} * 1\right]$$

$$\Rightarrow 0.049$$



$0.052 > 0.049$

$Gain(S, f_2) = 0.052$

then we will split from (S, f_2)