Discision Tree

Description:

→ A Flowchart for making decisions. It starts with question at the root and branches into answers at each node, leading to a final desecision at the leaves. Each decision is based on data, helping the algorithm learn patterns and make predictions or classifications, making it a powerful too like sorting and predicting outcomes.

There are two types of decision Tree:

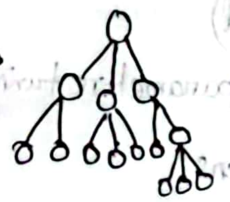
1. Decision Tree Classifier [Classification]
2. Decision Tree Regressor [Regression]

Decision Tree Classifier: (Two types)

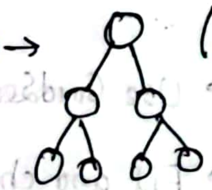
1. ID3 [Iterative Dichotomiser 3]
2. CART [Classification and Regression Tree]

Scikit learn use this

ID3 technique →



CART Technique →

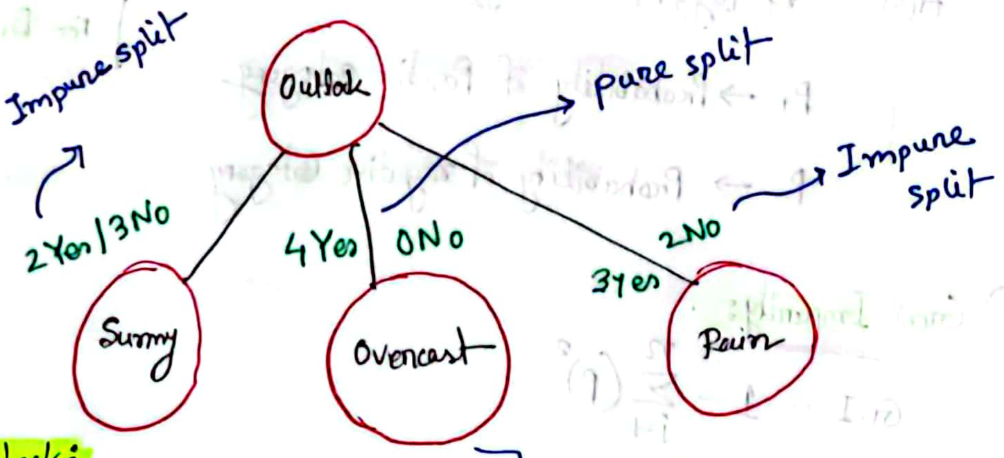


(only binary)

Let's take an example of a dataset of predicting Play tennis or not.

Day	Outlook	Temp	Humidity	Wind	Play Tennis
1	Sunny	Hot	High	Weak	No
2	Sunny	Hot	High	Strong	No
3	Overcast	Hot	High	Weak	Yes
4	Rain	Mild	High	Weak	Yes
5	Rain	Cool	Normal	Weak	Yes
6	Rain	Cool	Normal	Strong	No
7	Overcast	Cool	Normal	Strong	Yes
.	.	.	.	.	.
.	.	.	.	.	.
.	.	.	.	.	.
14	Rain	Mild	High	Strong	No

Take one independent feature say → "Outlook" and compare with 'target feature'



① Purity check:

pure split → Only yes/ only no

impure split → Some Yes and Some no combination (Need further splitting)

To check purities we use two techniques:

1) Entropy 2) Gini Impurity.

Using these techniques to find pure or impure split and decide to decide further split

② What Feature you need to select to start the split:

For this, we use "information Gain" technique

This helps to understand which independent feature has to be chosen to start with.

Technique explanation of purity check:

① Entropy:

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

[For Binary classification]

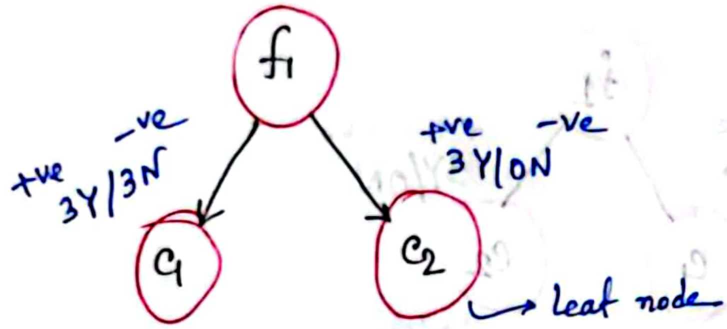
$P_+ \rightarrow$ Probability of Positive category

$P_- \rightarrow$ Probability of negative Category

② Gini Impurity:

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

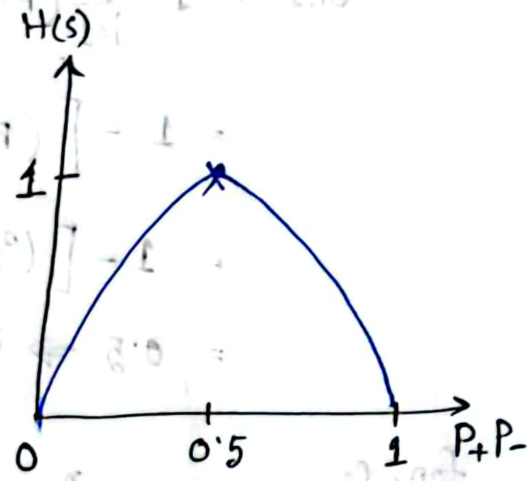
Let's take an example for Entropy calculation:



$$H(C_1) = -P_+ \log_2 P_+ - P_- \log_2 P_-$$

$$= -\tfrac{3}{6}\log_2\left(\tfrac{3}{6}\right) - \left(\tfrac{3}{6}\right)\log_2\left(\tfrac{3}{6}\right)$$

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



$$H(C_2) = -P_+ \log_2 P_+ - P_- \log_2 P_-$$

$$= -\tfrac{3}{3}\log_2\left(\tfrac{3}{3}\right) - \left(\tfrac{0}{3}\right)\log_2\left(\tfrac{0}{3}\right)$$

$$= 0 \Rightarrow \text{Pure split}$$

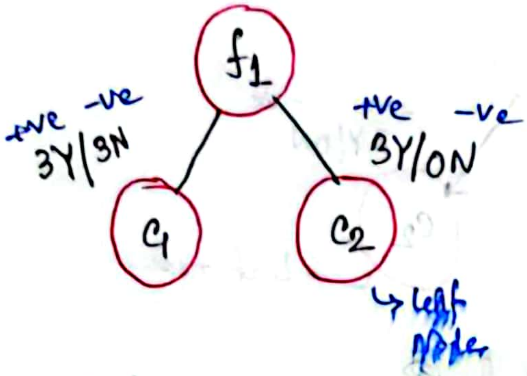
That was for Binary classification (Yes/no)

P → Probability
f → feature
c → category

For "Multi Class Classification" (C_1 C_2 C_3 : Yes/No/Maybe)

$$\text{Entropy} \rightarrow H(S) = -P_{C_1}\log_2 P_{C_1} - P_{C_2}\log_2 P_{C_2} - P_{C_3}\log_2 P_{C_3}$$

Now let's take an example for Gini Impurity Calculation:



for, C1,

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

$$= 1 - \left[(P_+)^2 + (P_-)^2 \right]$$

$$= 1 - \left[(3/6)^2 + (3/6)^2 \right]$$

$$= 0.5 \Rightarrow \text{impure split}$$

for, C_2,

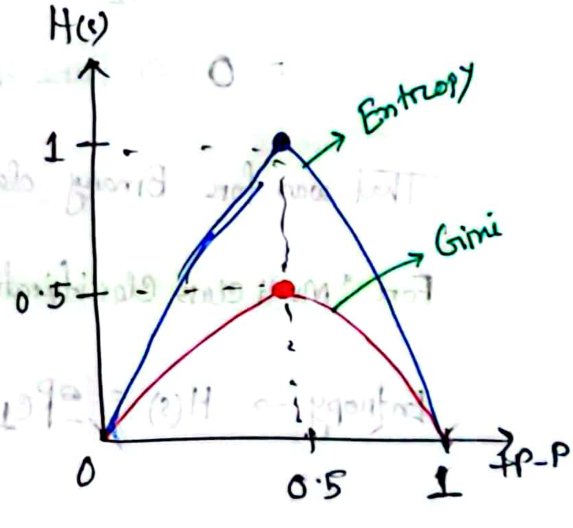
$$G.I = 1 - \sum_{i=1}^{n} (p)^2$$

$$= 1 - \left[(P_+)^2 + (P_-)^2 \right]$$

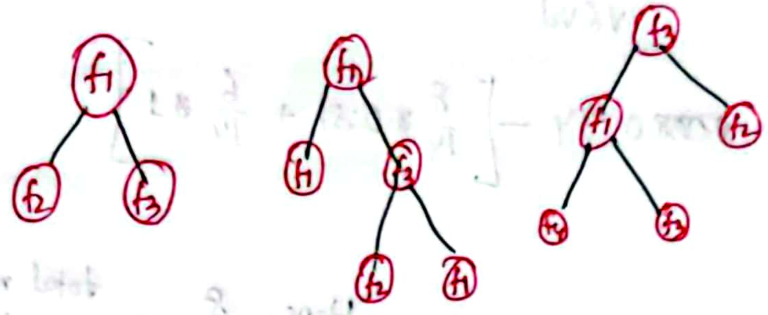
$$= 1 - \left[(3/3)^2 + (0/3)^2 \right]$$

$$= 1 - 1$$

$$= 0 \Rightarrow \text{Pure split}$$



Explanation of "Information Gain":



which feature should ~~so~~ be selected at first, then after splitting which ~~one~~ new feature should be selected, that is decided by information gain.

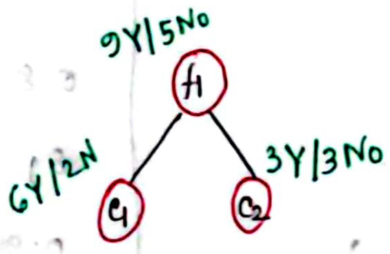
Information Gain:

$$Gain(S, f_1) = H(S) - \sum_{v \in val} \frac{|S_v|}{|S|} H(S_v)$$

$H_S \rightarrow$ Entropy of root node

$S \rightarrow$ Sample

$V \rightarrow$ value



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

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

$$= 0.94$$

using the same calculation formula, $H(S_v) = H(C_1), H(C_2)$

$$\Rightarrow H(C_1) = 0.81$$

$$H(C_2) = 1$$

$$Gain(S, f_1) = H(S) - \sum_{v \in Val} \frac{|S_v|}{|S|} H(S_v)$$

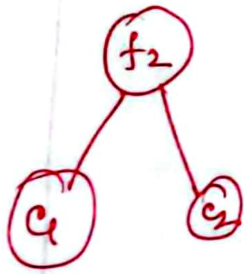
$$= 0.94 - \left[\frac{8}{14} * 0.81 + \frac{6}{14} * 1 \right]$$

$\therefore Gain(S, f_1) = 0.049$

Here, $\frac{8}{14} \rightarrow \frac{\text{total num of Y and N in } c_1}{\text{total No of Y and N in } f_1}$

Then suppose we take another feature and measure the Information gain

Here, $\frac{6}{14} = \frac{\text{total No of Y and N in } c_2}{\text{total no of Y and N in } f_1}$



$0.8 = H(c_1)$

$1 = H(c_2)$

$0.94 = H(S) = H(f_1)$

$Gain(S, f_2) \rightarrow$ Gain found $\rightarrow 0.051$

The greater the Gain value, that feature will be used by us.

So, we will use "feature f2".

Entropy vs Gini Impurity: (Which to use When)

Whenever dataset small (1000,2000 records) → Use Entropy

Whenever dataset large (1M, 100000, more) → Use Gini Impurity