I)- Decision Tree:

Dataset - Problem Statement

Day	Outlook	Temperature	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
8	Sunny	Mild	High	Weak	No
9	Sunny	Cool	Normal	Weak	Yes
10	Rain	Mild	Normal	Weak	Yes
11	Sunny	Mild	Normal	Strong	Yes
12	Overcast	Mild	High	Strong	Yes
13	Overcast	Hot	Normal	Weak	Yes
14	Rain	Mild	High	Strong	No

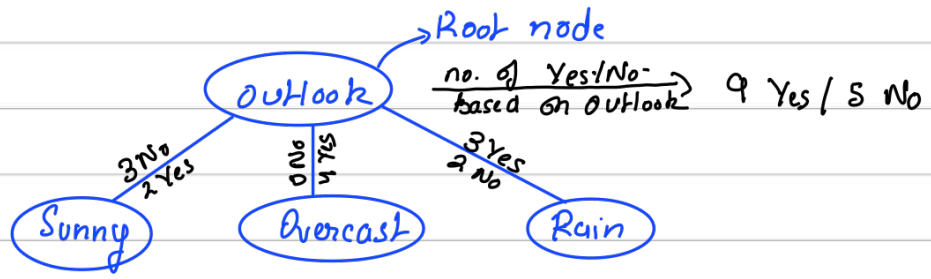
asit@gmail.com

We have to predict based on the 4 i/p features Temp, Humidity, Wind, Outlook weather a person will go out to play Tennis or not

Using decision tree to solve the following problem:

① Step 1 :- Decision Tree will select the root node

Let's consider



② Step 2: Purity Split Check.

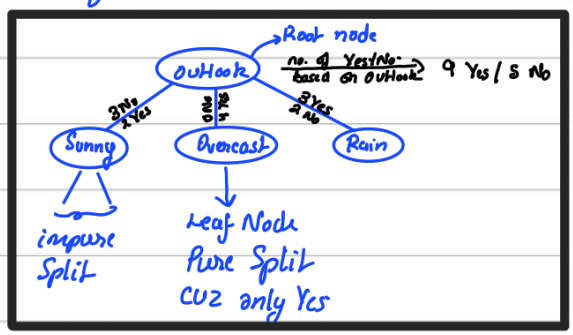
a → Pure Split: Either entirely Yes or entirely No

b → Impure Split: further divided to other features

Purity Split Check is done using techniques

- → Entropy
- → Gini Impurity

} measure of Impurity



③ Step 3: Information Gain: What feature you need to select to start the split (i.e choose as Root node)?

PRIORITY CHECK

Entropy ← → **Gini Impurity**

$$\boxed{H(S) = -P_{(+)} \log_2(P_{(+)}) - P_{-} \log_2(P_{(-)})}$$

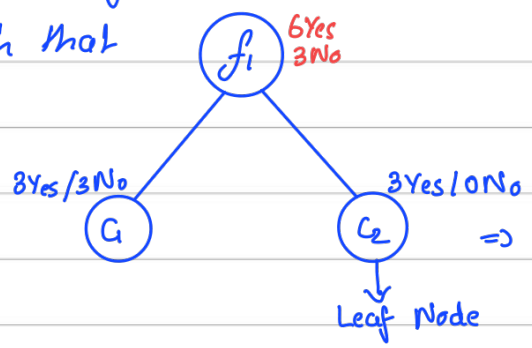
, $P_{(+)}$ → Probability of +ve Category

$P_{(-)}$ → " " +ve "

$$\boxed{G.I = 1 - \sum_1^n (P)^2}$$

1.1)- Entropy:

ex: Consider f_1 as root node and use f_1 for doing 1st split such that



⇒ C_2 considered as leaf node/Pure split cuz only has Yes o/p.

Now, checking if C_1 is either pure split or impure split (cuz we know already the C_2 is a pure split) using Entropy :-

$$P_{(+)} = 3/6 = 1/2 \quad \& \quad P_{(-)} = 3/6 = 1/2$$

$$\text{Entropy}(C_1) = H(S) = -P_{(+)} \log_2 P_{(+)} - P_{(-)} \log_2 P_{(-)}$$

$$= -\frac{1}{2} \log_2\left(\frac{1}{2}\right) - \frac{1}{2} \log_2\left(\frac{1}{2}\right)$$

$$= -\frac{1}{2} \times (-1) - \frac{1}{2}(-1)$$

$$= \frac{1}{2} + \frac{1}{2}$$

$$\boxed{H(C_1) = 1}$$ Entropy of Category 1

$H(C_1) = 1$ Entropy indicates that we are having completely Impure state cuz equal no. of Yes & No's exists

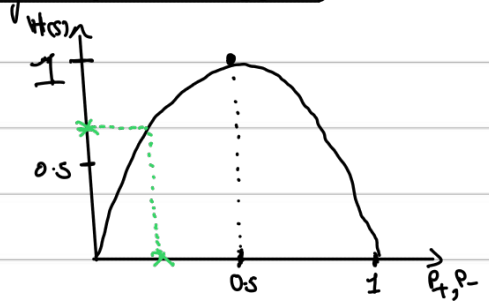
$$\text{Entropy}(C_2) = H(C_2) = -\frac{3}{3} \log_2(3/3) - \frac{0}{3} \log_2 0/3$$

$$= -\frac{3}{3}(0) - 0$$

$$\boxed{H(C_2) = 0}$$

$H_{C_2} = 0$ Entropy indicate that we are having a pure split

Plotting $H(S)$ Vs $P_{(-)}$ $P_{(+)}$:



1.2)- Gini Impurity:

$$GI = 1 - \sum_i^n (P)^2 = 1 - [(P_+)^2 + (P_-)^2]$$

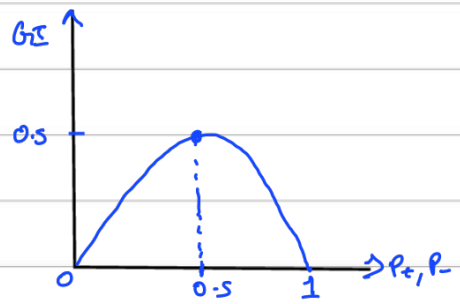
$$GI(C_1) = 1 - [(1/2)^2 + (1/2)^2] = 0.5$$

$$GI(C_2) = 1 - [(3/3) + (0/3)] = 1 - 1 = 0$$

$\boxed{GI = 0.5}$ means impure split

$\boxed{GI = 0}$ means pure split

GI Vs (P_+, P_-)



1.3)- Entropy and Gini-Impurity in cas of Multiclass Classification:

for 3 o/p categories

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

$GI = 1 - [(P_{C_1})^2 + (P_{C_2})^2 + (P_{C_3})^2]$

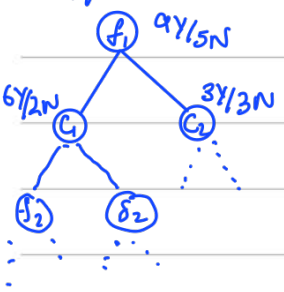
1.4)- Information Gain:

Information gain helps in finding out which feature to select as a root node (i.e which feature to select to start split)

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

Considering 3 features f_1, f_2, f_3 o/p

Suppose,



Entropy $H(S) = -\frac{9}{14}\log_2\left(\frac{9}{14}\right) - \frac{5}{14}\log_2\left(\frac{5}{14}\right) \approx \underline{0.94}$

Entropy $H(C_1) = -\frac{6}{8}\log_2\left(\frac{6}{8}\right) - \frac{2}{8}\log_2\left(\frac{2}{8}\right) \approx \underline{0.81}$

Entropy $H(C_2) = \underline{\underline{1}}$ (impure split)

using above 3 eq Gain for f_1 is :-

$$Gain(S, f_1) = H(S) - \sum \frac{|S_v|}{S} H(S_v)$$

$$= 0.94 - \left[\frac{6+2}{9+5} \times 0.81 + \frac{3+3}{9+5} \times 1\right]$$

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

$$\underline{Gain(S, f_1) = 0.044}$$

lly suppose

$$\underline{Gain(S, f_2) = 0.051}$$

So, as

$$Gain(S, f_2) > Gain(S, f_1)$$

$\therefore$, we will be splitting using feature f_2

Qy When to use Entropy and when to use GI?

Solⁿ → Small dataset : Entropy (log formula)

large dataset : Gini Impurity (simple math)

Qy What if feature (Root node) is continous and Numerical?

Solⁿ →

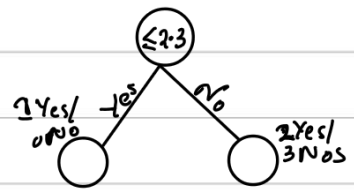
f_1	O/p
2.3	Yes
3.6	Y
4	N
5.2	N
6.7	Y
7.8	N

Step 1 :- Sort the feature f_1

Step 2 :- Threshold is set to 1st (lowest) value

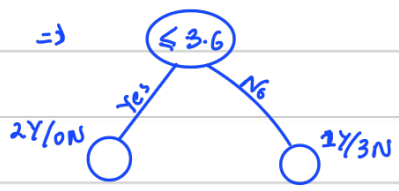
⇒ Threshold = 2.3

∴ Split will look like

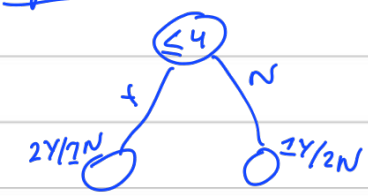


Step 3 :- Now Threshold is set 2 2nd lowest value

⇒



Step 3 : Threshold = 4 . . . . .

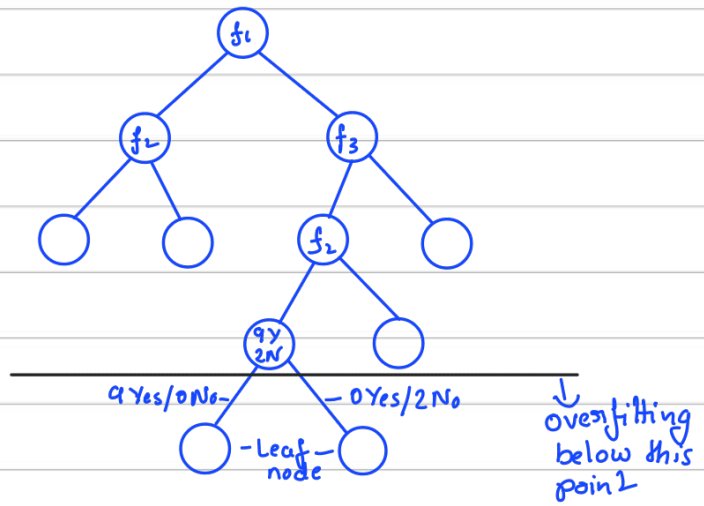


internall all the Entropy, GI and Gain is calculated and whichever have Higher gain will be selected.

Disadvantage:- as Dataset ↑es, Time Complexity ↑es.
(☆ can be fixed through Post-Pruning & Prepruning in DT)

1.4)- Post-Pruning & Pre-Pruning: (Reduce Overfitting)

In decision tree whenever we are given a training data we perform a decision tree split and this split keeps going on unless and we get the leaf node. Like:-



But, whenever we split down the decision tree down to every Leaf node the training accuracy tends to be Overfitting:
(Train Acc ↑es
Test Acc ↓es)

1.4.1)- Post-Pruning:

Step 1:- Construct entire Decision Tree to the complete leaf node
Step 2:- We start Pruning (cutting) the Decision tree after max depth reached (max_depth = 3)
3:- Suitable for smaller dataset.

1.4.2)- Pre-Pruning:

① Hyperparameter tuning to select the best parameters.

∴ all these parameters like

'criterion'	'min_weight_fraction_leaf'
'splitter'	'max_features'
'max_depth'	'max_leaf_nodes'
'min_samples_split'	'class_weight'
'min_samples_leaf'	'ccp_alpha'

all the parameters a picked and set ahead of constructing the decision tree.

II)- Decision Tree Regressor:

- 2.1- Variance
- 2.2- Variance Reduction

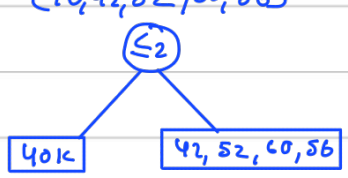
given dataset :-

Exp	Career Gap	Salary (o/p)
2	Yes	40k
2.5	Yes	42k
3	No	52k
4	No	60k
4.5	Yes	56k

$\bar{y} = 50k$

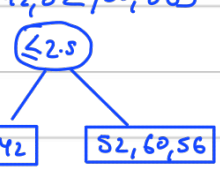
Split 1 : Threshold = 2

[40, 42, 52, 60, 56]



Split 2 :- Threshold = 2.5

[40, 42, 52, 60, 56]



. . . . . .

note :- Our final aim is to calculate Variance Reduction.

$$\boxed{\text{Variance on Error} = \frac{1}{n}\sum_{i}^{n}(y_i - \bar{y}_i)^2}$$ (mean Squared Error)

For Split 1 :-

① Variance of Root :- $Var(root) = \frac{1}{5}\left[(40-50)^2 + (42-50)^2 + (52-50)^2 + (60-50)^2 + (58-50)^2\right]$

$= \frac{1}{5}(100 + 64 + 4 + 100 + 36) = \frac{304}{5} = 60.8$

② $Var(\text{left child}) = \frac{1}{1}(40-50)^2 = 100$

③ $Var(\text{Right Child}) = \frac{1}{4}\left[(42-50)^2 + (52-50)^2 + (60-50)^2 + (58-50)^2\right] = \frac{204}{4} = 51$

④ Calculating Variance Reduction for Split :-

$$\boxed{\text{Variance Reduction} = Var(\text{Root}) - \sum w_i \, Var(\text{child})}$$

, $w_i \rightarrow$ weight of i = $\frac{\text{child elements}}{\text{total elements}}$

•) For split 1

Var Reduction $= 60.8 - \left[\frac{1}{5}\times(100) + \frac{4}{5}(51)\right] = 20 + \frac{204}{5} = 60.8 = 0$

lly

•) For Split 2 :

Variance Reduction $= 0.004$

As, Variance Red^n(Split 1) < Variance Red^n (Split 2)

∴, Split 2 will be selected for decision tree building