Decision Tree

1) Decision Tree classifier [classification]

2) Decision Tree Regressor [Regression]

* Decision Tree classifier *

1) ID3 [Iterative Dichotomiser 3]
2) CART [classification & Regressor Tree]

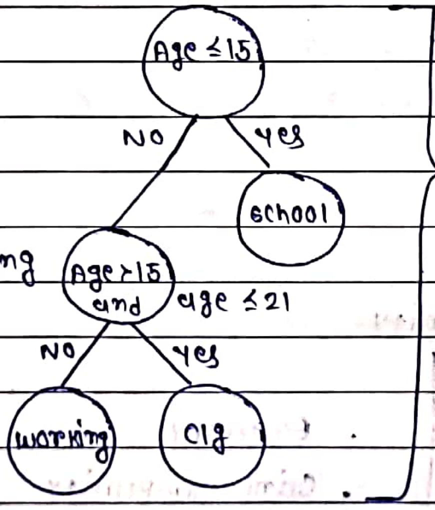
→ ID3 Have child more than two In tree
→ CART Have Binary child.

Example :-

age=14 → school

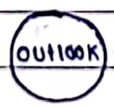
age=25 → working

age=20 → clg



decision tree classifier [multinested If else]

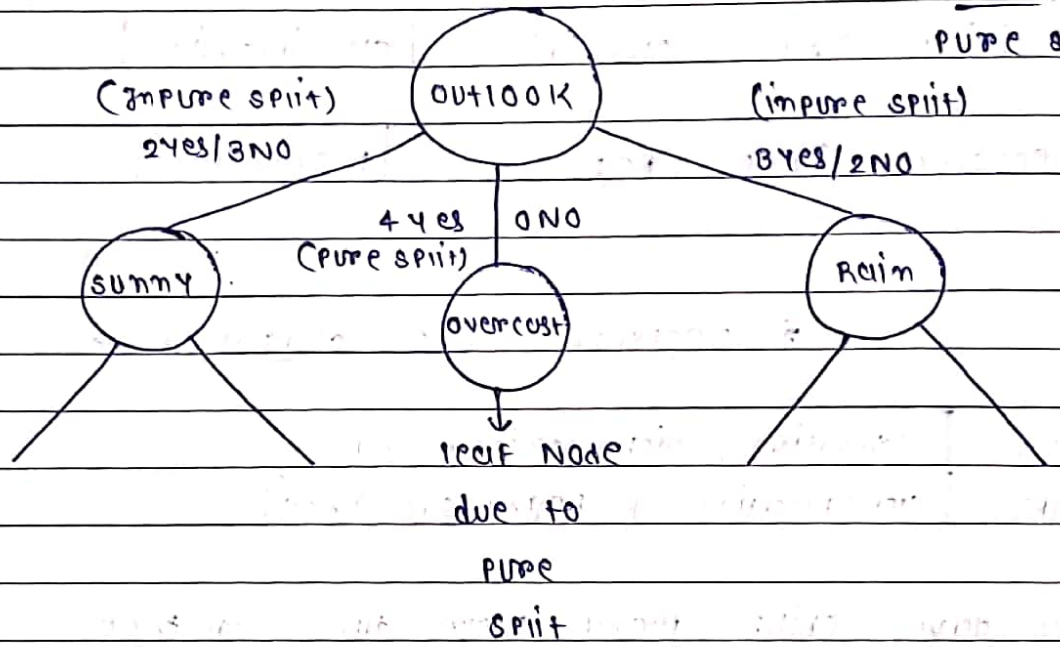
→ Example → predict play tennis or not.



→ we select Outlook from given example feature

→ Initial step we will select any one feature as Root of the tree.

1> Purity check:-
pure split or impure



→ For checking purity
- Entropy
- Gini impurity

} measure of purity.

→ what feature you need to select as root?:-

we use "information gain"

a> "Entropy"

b> "gini impurity"

a) Entropy

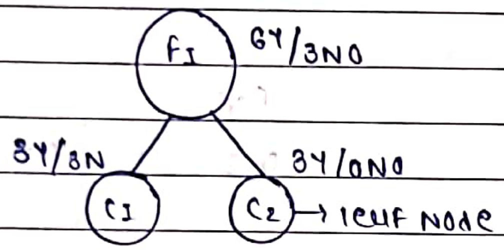
$$\boxed{H(S) = -P_+ \log_2 P_+ - P_- \log_2 P_-}$$

b) gini impurity

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

P_+ - probability of positive category

P_- = probability of negative category



Entropy of $C_1 = -P_+ \log_2 P_+ - P_- \log_2 P_-$

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

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

$= 1 \rightarrow$ Impure split [50% yes & 50% no].

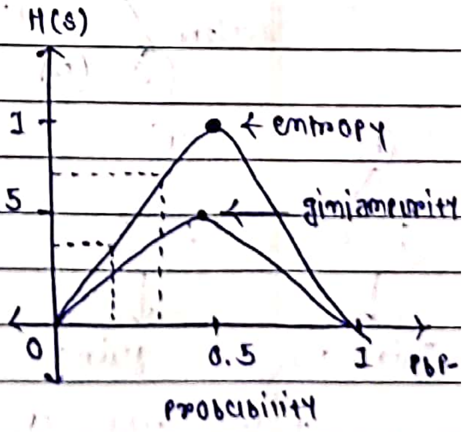
Entropy of $C_2 = -\frac{3}{3}\log_2\left(\frac{3}{6}\right) - \frac{0}{6}\log_2\left(\frac{0}{6}\right)$

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

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

$= -1 \log_2(1)$

$= 0 \rightarrow$ Pure split



MilAn

Page No.

Date

* for multiclass category

Ex → yes / No / maybe

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

b) gini impurity

(C_1)

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

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

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

$= \frac{1}{2} = 0.5$ → impure

(C_2)

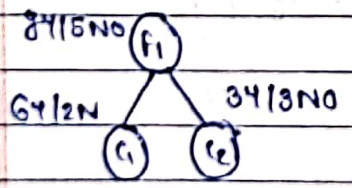
$G.I = 1 - \left[\left(\frac{3}{3}\right)^2 + \left(\frac{0}{3}\right)^2 \right]$

$= 1 - 1$

$= 0$ → pure split.

→ information gain

$F_1 \quad F_2 \quad F_3 \qquad O/P$



Entropy of Root

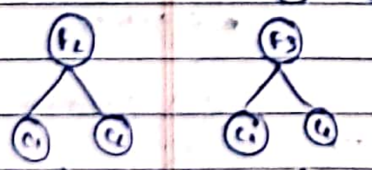
$Gain(S_1, F_1) = H(S) - \sum_{v \in val} \frac{|S_v|}{|S|} H(S_v)$

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

$\boxed{H(S) = 0.94}$

$H(C_1) = 0.81$

$H(C_2) = 1$



gain = 0.051 $\qquad$ gain = 0.048

$gain(S_1, F_1) = 0.94 - \left[\frac{8}{14} * 0.81 + \frac{6}{14} * 1 \right]$

$\boxed{gain(S_1, F_1) = 0.049} < gain(S_1, F_2)$

* we use F_2 as root node.

* information gain is more when we split using F...

Scanned with OKEN Scanner

Decision Tree for Numeric Feature

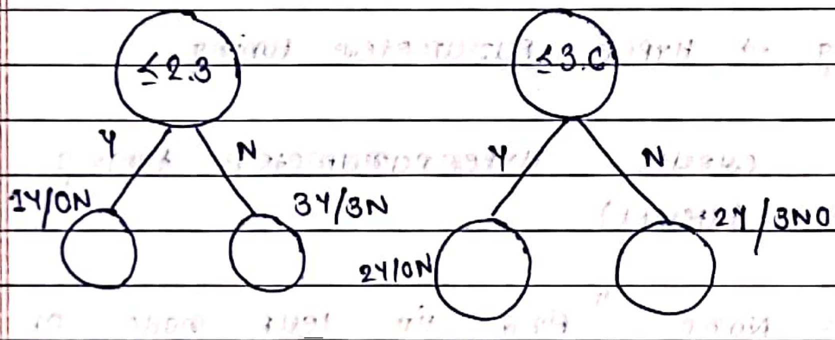
→ dataset

f_1	O/P
2.3	Y
3.6	Y
4	N
5.2	N
6.7	Y
7.8	N
9.0	Y

1) sorting the feature.

* Thresold value = 2.3 [first feature]

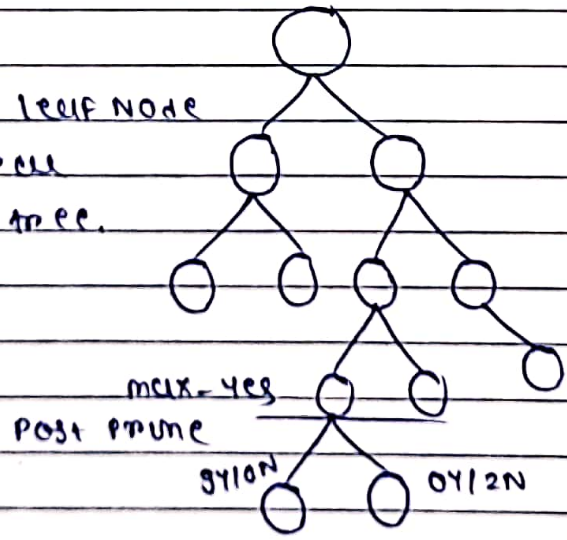
Next Thresold value = 3.6 [Next feature]



→ disadvantage is time complexity.

→ Decision tree post pruning and pre pruning.

→ we find leaf node in general decision tree.



overfitting

TRAINING ACC↑↑
TEST DATA ACC↓↓

MilAn

Page No.

Date

* Two method are use for Reduce overfitting.

1) pre pruning
2) post pruning

1) post pruning → 1) construct decision tree
2) prune the tree (cut the tree)

Hyper-parameter → max-depth

2) pre-pruning → Hyper parameter tuning

First create Hyperparameter tuning
(For small dataset)

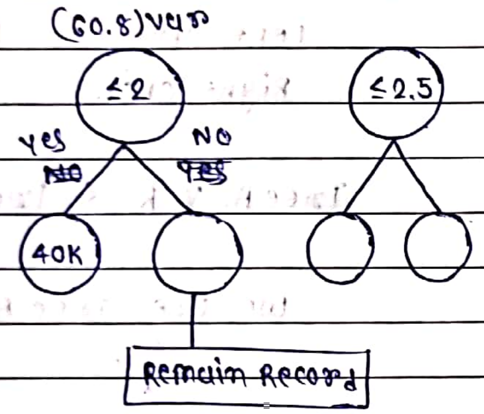
→ max_depth = None "Find till leaf node of decision tree"

* Decision Tree Regressor

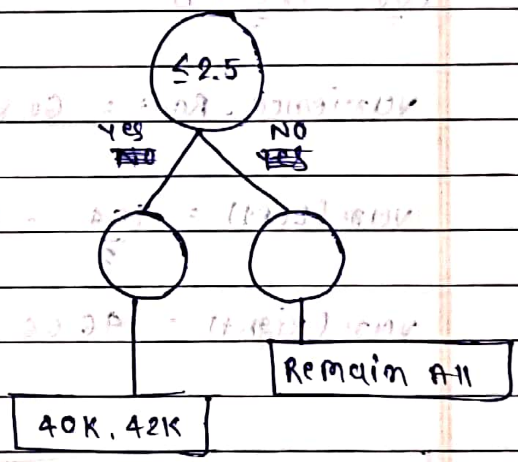
Dataset

Exp	career gap	Salary
		O/P
2	yes	40K
2.5	yes	42K
3	No	52K
4	No	60K
4.5	yes	50K

Avg = 50K = $\hat{y}$



Tree - A



Tree-B

1) we use ~~"variation~~ "(variance) Reduction"

$$\text{varience} = \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y})^2 \quad [\text{mean squared error}]$$

TREE-A

$$\text{varience (Root)} = \frac{1}{5}\left[(40-50)^2 + (42-50)^2 + (52-50)^2 + (60-50)^2 + (56-50)^2\right]$$

$$= 60.8$$

varience (Left) = 100

varience (Right) = 51

MilAn

Page No.

Date

Variemce - Reduction = Var(Root) − $\sum$ wi Var(child)

For Tree A (L) wi = $\frac{1}{5}$ wi(R) = $\frac{4}{5}$ cuz of Elements.

$$= 60.8 - \left[\frac{1}{5} * 100 + \frac{4}{5} * 51\right]$$

$\boxed{\text{Variemce-Reduction} = 0}$

For Tree-B

Variemce - Root = 60.8

Var(Left) = $\frac{164}{2}$ = 82

Var(Right) = 46.66

$\boxed{\text{Variemce-Reduction} = 0.004}$

Variemce-Reduction of Left Split is Less than Right Split.

Tree A - V_R < TreeB - V_R

we use Tree B