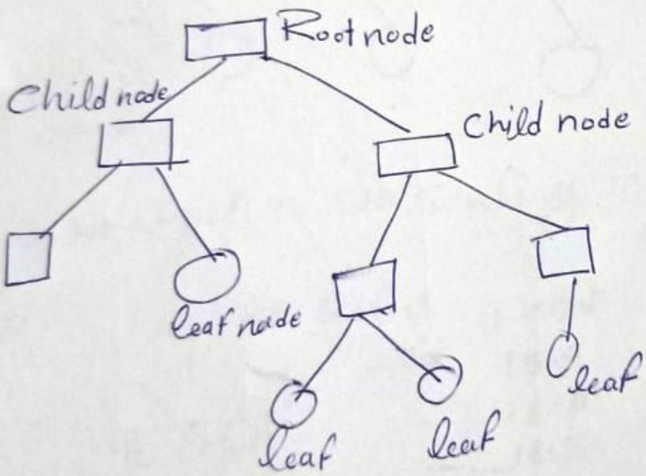
ML Day-7 Decision tree

- Random forest
- Xgboost
- Gradient boosting
- Ada Boost

Ensemble Tech.
↓
Bagging
Boosting
Staking

Decision tree (supervise algo)

Decision tree Classifire
Decision tree regression



Algo for solving Decision tree
1) ID3 → (entropy and info gain)
2) CART → (Classification and Regression Tree) which we are Build on the bases of Gini ~~impurity~~ Indexing)
3) C4.5 → it is Probability

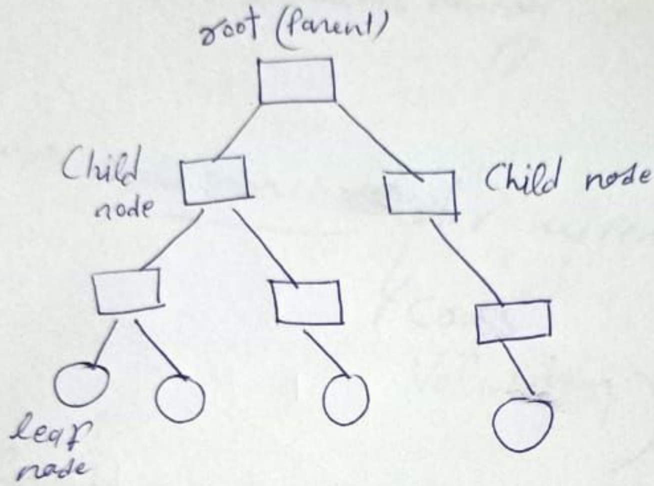
→ for the regression ~~inste~~ instead of Gini and entropy we use Standard Deviation (MSE, RMSE) then Decide Rootnode.

Decision tree → classification
Regression

Data set

weight	hight	(target) obese/No obese
50	150	0
60	160	NO
70	170	0
80	175	NO
90	180	0

weight	height	Target BMI
50	150	21
60	167	22
70	170	23
80	175	20
90	180	18

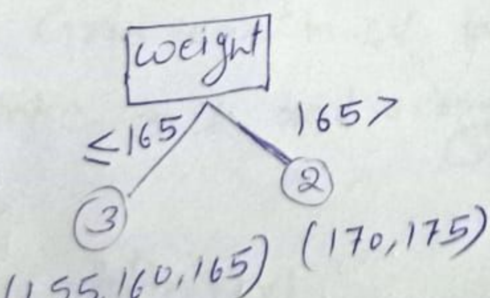


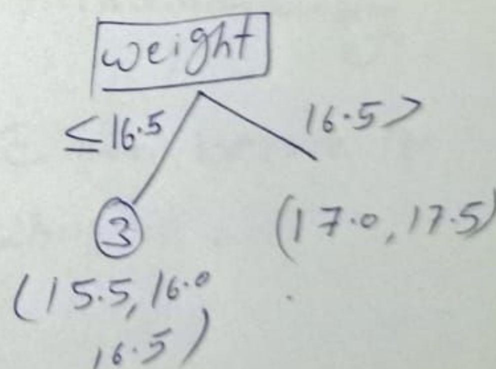
Which column or feature i take. Decide by

1) entropy → info gain
2) Gini $\boxed{\text{indexing / impurity}}$
3) std / MSE
   ↳ Decision tree Regressor

→ Need not To Do Scaling in Decision Tree

weight	after scaling → weight
155	15.5
160	16.0
165	16.5
170	17.0
175	17.5





→ outlier ~~doesn't~~ does not empact on decision tree

→ sklearn.tree.DecisionTreeClassifier

hyperparameter → it is responsible for creating model or model structure.

1) Criterian = 'gini' → (gini or entropy) ~ 2

2) max-depth = None → we can take (2, 3, 4, 5, 6 - - -)
max-depth = {2, 3, 4, 5, 6} ~ 5

3) min_sample_leaf = {5, 10, 15, 20, 25} → 5

4) min_sample_split = {10, 15, 20, 25, 30} ~ 5

Now do hyper Parameter tuning   No. of ittration

$5 \times 5 \times 5 = 125 \times 2 = 250$

for exp

1. use Grid search cv (for Hyper Patrameter Tuning & Cross Validation)

2) Random search.cv

means Picking the value for HyperParameter tuning.

→ Grid search.cv take more time becoze it take each and every Combination

Cross validation. (cv)   10 Data Point

CV = 5 $\frac{10}{5} = 2$ Test

8 training

Test   Training

| x x | x x x x x x x x | — check Accuracy

x | x x | x x x x x x → check Accuracy

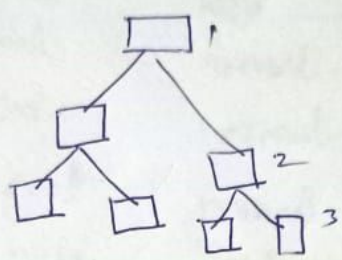
× × [× ×] × × × × × × — then check Accuracy

like that do with the whole data

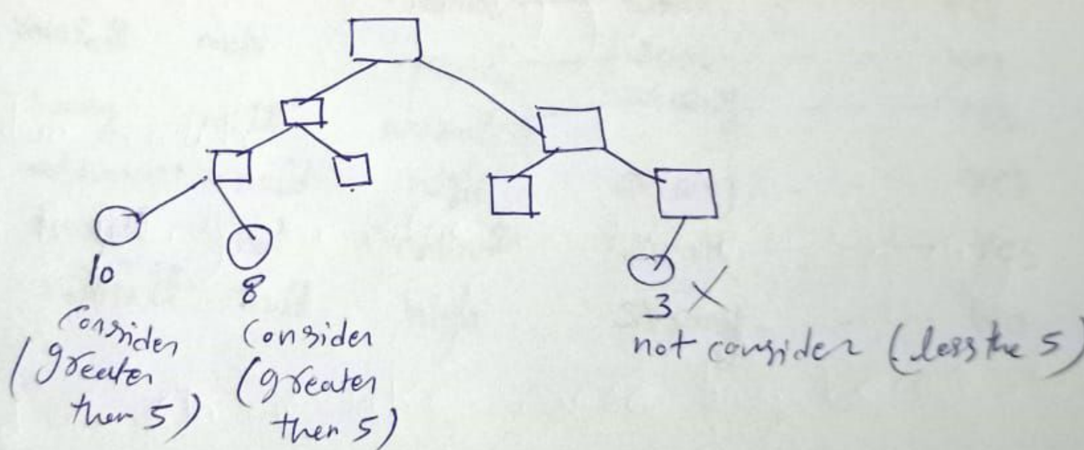
→ we take Max Accuracy or avg Accuracy

Decision tree

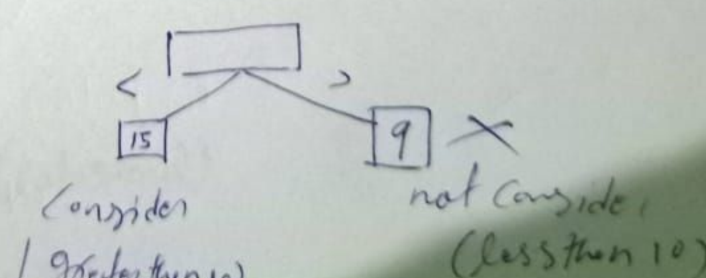
① Max_depth = 3



② min_Sample_leaf = 5



③ min_Sample_Split = 10



4) Splitter : ['best', 'random']

ML Day-8 Decision-Tree

Decision tree classifier, Decision Tree regressor

F_1 Day	F_2 outlook	F_3 temp	F_4 humidity	F_5 Wind	output Decision
1	sunny	hot	high	weak	No
2	Sunny	hot	high	strong	No
3	overcast	hot	high	weak	yes
4	rainfall	mild	high	weak	Yes
5	rainfall	cool	normal	weak	yes
6	rainfall	cool	normal	strong	No
7	overcast	cool	normal	strong	yes
8	sunny	mild	high	weak	No
9	sunny	cool	normal	weak	yes
10	rainfall	mild	normal	weak	yes
11	sunny	mild	normal	strong	yes
12	~~ou~~ overcast	mild	high	strong	yes
13	overcast	hot	normal	weak	yes
14	rainfall	mild	high	strong	No

take

f_1 (cat, Num) O/P (cat, Num)

DT Classifier

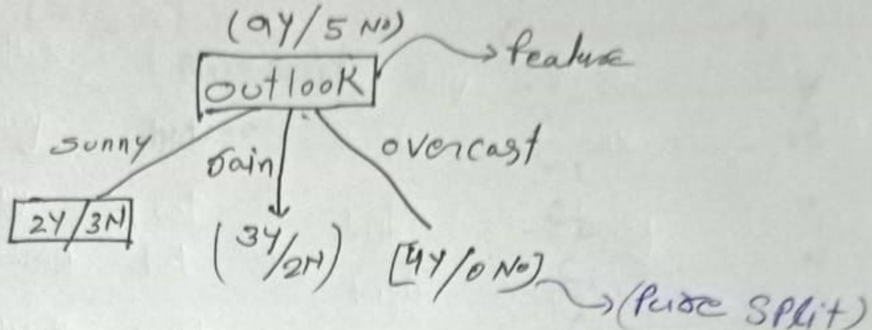
→ f_1 (cat, Num) O/P (categorical)

→ f_1 (cat, Num) O/P (Num)

DT Regressor

① ID_3 (Iterative Decotomiser) (entropy)

② CART (Classification and regression trees (gini impurity)



Check Purity → ① entropy

② gini-coff, gini impurity.

Entropy $\overset{H(S)}{\Rightarrow}$ $- \sum_{i=1}^{n} P_i \times \log_2(P_i)$

we have 2 class

Y/N

Binary

$- P_Y \log_2(P_Y) - P_N \log_2(P_N)$

if class 3

C_1, C_2, C_3

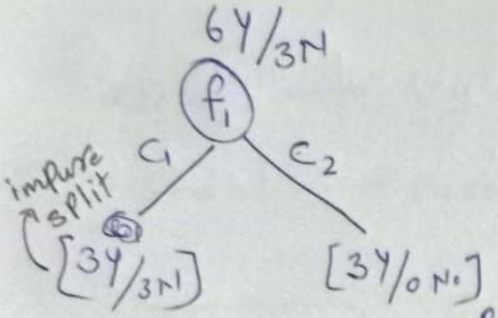
$$\boxed{-P_{C_1} \log(P_{C_1}) - P_{C_2} \log(P_{C_2}) - P_{C_3} \log(P_{C_3})}$$

for Multi Class

Gini-coff ⇒

$$1 - \sum_{i=1}^{n} P_i^2$$

$$1 - [P_Y^2 + P_N^2]$$



Pure split

50% Y

50% N

H(S) = 1

H(S) = 0

f_1	O/P
C_1	Y
C_2	Y
C_1	Y
C_2	Y
C_1	Y
C_1	N
C_2	Y
C_1	N
C_1	N

Entropy

$$H(S) = -\sum_{i=1}^{n} P_i \log_2(P_i)$$

for C_1

$$= -P_Y \log_2(P_Y) - P_N \log_2(P_N)$$

$$= -\frac{3}{6}\log_2\left(\frac{3}{6}\right) - \frac{3}{6}\log_2\left(\frac{3}{6}\right) \Rightarrow -\frac{1}{2}\log_2\left(\frac{1}{2}\right) - \frac{1}{2}\log_2\left(\frac{1}{2}\right)$$

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

$$= -\frac{1}{2}[0-1] - \frac{1}{2}[0-1]$$

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

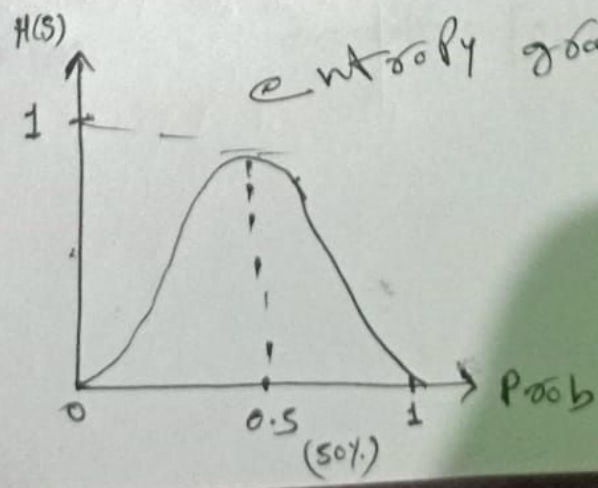
$$\boxed{H(S) = 1}$$

for C_2

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

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

entropy graph w.R.T Probability



$H(S) = 1 \longrightarrow$ Very impure split

$H(S) = 0 \longrightarrow$ Pure split

$\boxed{2\text{Yes}/3\text{NO}}$

$$H(S) = -\frac{2}{5}\log_2\left(\frac{2}{5}\right) - \frac{3}{5}\log_2\left(\frac{3}{5}\right)$$

$\boxed{H(S) = 0.97}$

② Gini-Coff or Gini-impurity

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

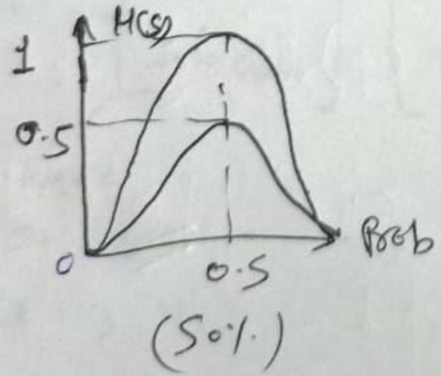
① $3Y/3N \longrightarrow$ Entropy $\rightarrow H(S) = 1$ (Very impure split)

② $3Y/0N \longrightarrow 0$

③ $2Y/3N$

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

$$= 1 - \left[\left(\frac{1}{2}\right)^2 + \left(\frac{1}{2}\right)^2\right] \Rightarrow 1 - \left[\frac{1}{4} + \frac{1}{4}\right] \Rightarrow 1 - \left[\frac{2}{4}\right] \Rightarrow 1 - \frac{1}{2} = \frac{1}{2} = 0.5$$



entropy = [0, 1]

gini-coff = [0, 0.5]

[4Y/8N] ⇒ Gini-Coff value

[8Y/2N] ⇒ " "

$$= 1 - \left[\left(\frac{4}{12}\right)^2 + \left(\frac{8}{12}\right)^2\right]$$

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

$G.I = \underline{\underline{0.444}}$

$$\Rightarrow = 1 - \left[\left(\frac{8}{10}\right)^2 + \left(\frac{2}{10}\right)^2\right]$$

$G.I = 1 - 0.68$

$G.I = \underline{\underline{0.32}}$

Features 1, feature 2, feature 3

↳ Purity

ID3 → [entropy, Info Gain] [gini impurity, Info Gain] CART ← fast

it can have more than 2 split like 3, 4

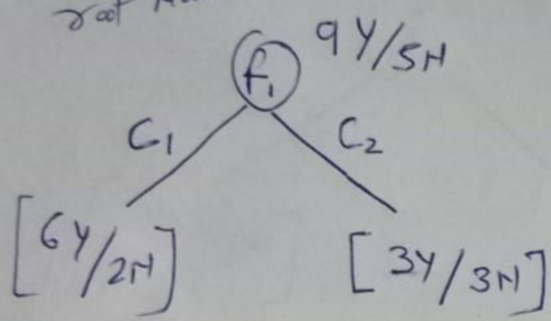
CART → This for Binary Approach

Information Gain

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

root Node

→ Value



$H(S) \Rightarrow$ root feature entropy

$$= -P_Y \log_2(P_Y) - P_N \log_2(P_N)$$

$$= -\left(\frac{9}{14}\right) \log\left(\frac{9}{14}\right) - \frac{5}{14} \log\left(\frac{5}{14}\right)$$

$$= -(0.64) \log(0.64) - (0.35) \log(0.35)$$

(f_1) root feature $H(S) \approx 0.94$

$6Y/2N \Rightarrow -\frac{6}{8} \log_2\left(\frac{6}{8}\right) - \frac{2}{8} \log\left(\frac{2}{8}\right) = 0.81$

$3Y/3N \Rightarrow H(S) = 1$

→ This Approach called ID3 Approach → No of Split

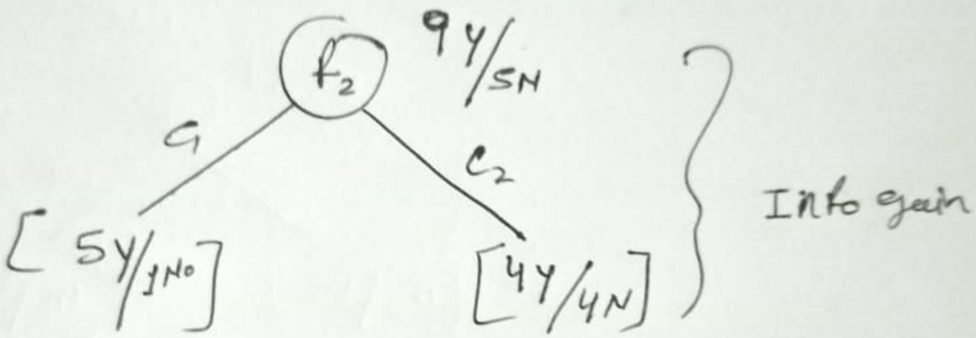
$$\text{Info Gain}(S, f_1) = 0.94 - \left[\frac{8}{14} \times 0.81 + \frac{6}{14} \times 1\right]$$

$$= 0.94 - [0.462 + 0.42]$$

$$\boxed{f_1 = 0.058}$$

Feature 2

$f_2 =$



$$\text{Gain} = 0.94 - \left[\left(\frac{6}{14}\right) \times 0.65 + \left(\frac{8}{14}\right) \times 1\right]$$

$$\text{Gain}(f_2) = 0.094$$

$$\boxed{\text{Info gain}(f_2) > \text{Info gain } f_1}$$

So we select f_2

it is providing more information