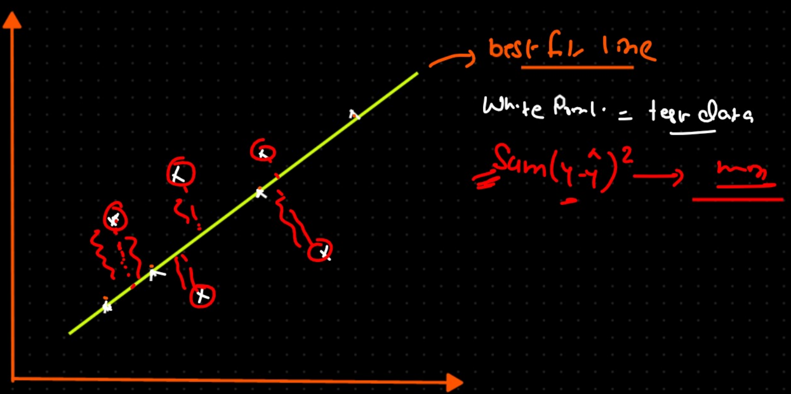
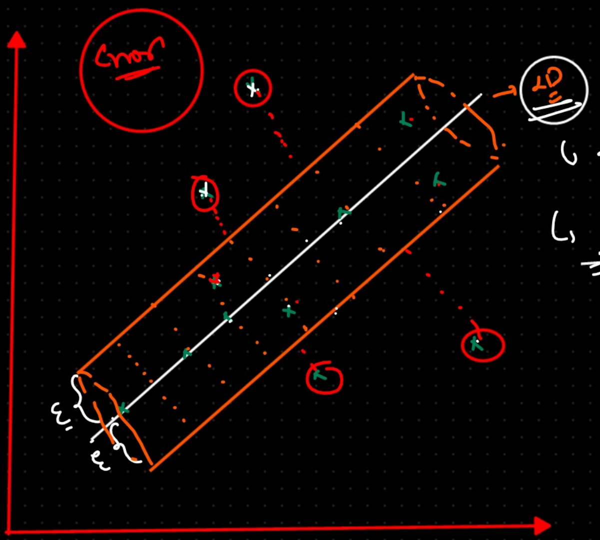
1. SVR.
2. Python implement.
3. DT
4. hyperparameter

SVR →

Linear reg.

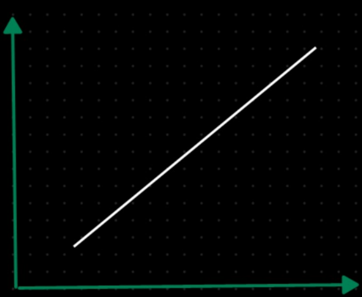


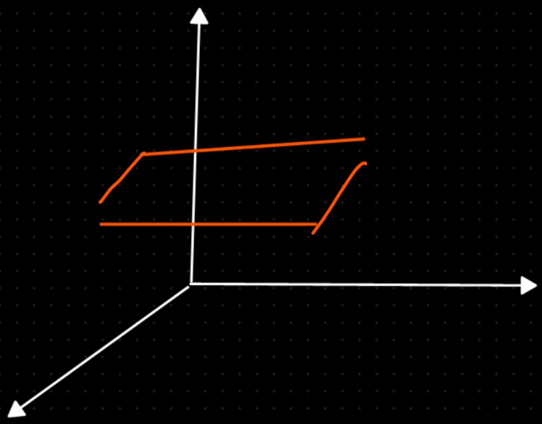


Regression ?

↳ train → traindata

↳ test data





SVR

Constraint for SVM(SVC) ⇒ $\left\{ \hat{y}_i (w^T x + b) \geq 1 \right\}$

$\underset{(w,b)}{\text{minimise}} = \dfrac{\|w\|}{2}$

SVC → hard classifier
soft classifier

Cost function ⇒

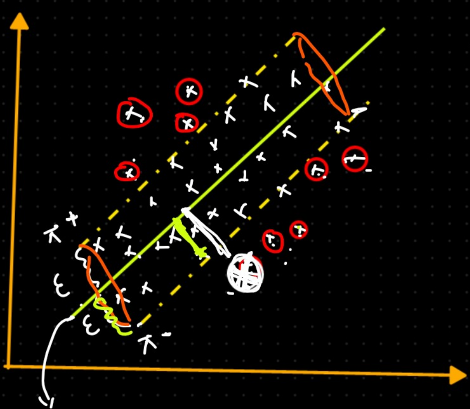
$$\boxed{\min_{(w,b)} \frac{\|w\|}{2} + C_i \sum_{i=1} \xi_i}$$

w i.e ω

ETA

hinge loss.

Support Vector Regression



hyperplane = ε

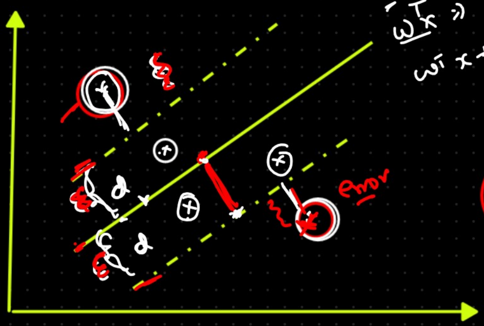
$w^T x + c$

⇒ { Reduce the error
minimize the error }

= Constraint

$$\boxed{|Y_i - (w^T x + c)| \leq \varepsilon}$$

$$Y_i - (w^T x + c) \leq \varepsilon$$



$w^T x \Rightarrow$ origin

$w^T x + b \Rightarrow$

$$\boxed{w_1 x_1 + w_2 x_2 + b = 0}$$

$$\begin{bmatrix} w_1 \\ w_2 \end{bmatrix} \cdot \begin{bmatrix} x_1 & x_2 \end{bmatrix}$$

$$\text{Dist}(y_i, w^T x + c) \Rightarrow \{ w_1 x_1 + w_2 x_2 \}$$

$$\leq \varepsilon(d)$$

Cost fun

minimize

$$\left\{ \min_{(w,b)} \frac{\|w\|}{2} \right\} + C \sum_{i=1}^{n} |\varepsilon_i|$$

Decision tree $\Rightarrow$ Tree which take the Decision based on a condition

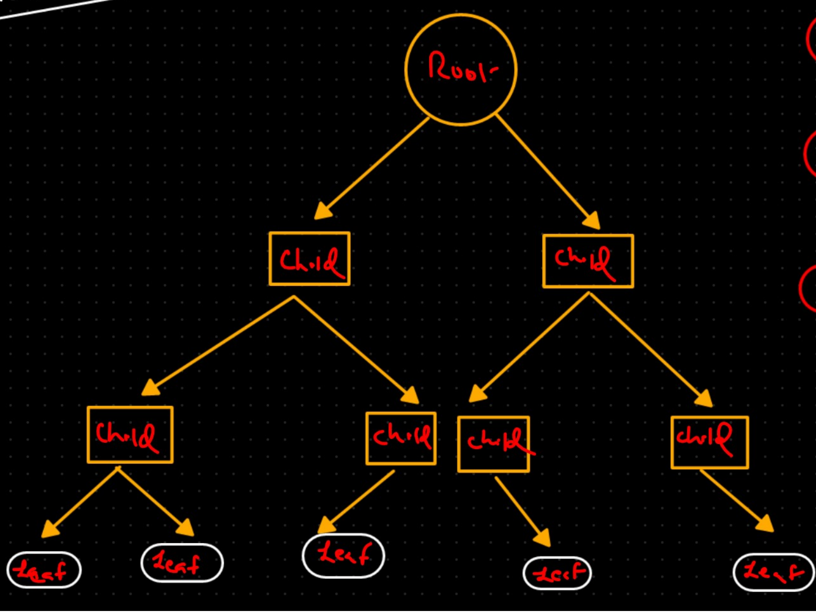
It is based on a condition

Sunny = 27

- if age ≤ 15 ;
  Print ("School")
- if age ≥ 15 & age ≤ 25 ;
  Print ("College")
- if age ≥ 25 ;
  Print ("working")

Tree

Decision tree



Diff $\Rightarrow$ Node

1. Root
2. Child
3. Leaf

Decision tree algo → ① ID3 algo } iterative Dichotomiser
② CART } classification & Reg tree

C4.5 → updated version of ID3

ID3

① Itreative Dichotomiser

② Entropy ⇒ information gain

③ it will split the Data in multiclass

④ it only work with the classification

C4.5

regression

CART

① classification & Req. tree

② Gini impurity (Gini Coef, Gini index)

③ it will split the data into Binary class

④ it works for the Reg & class.

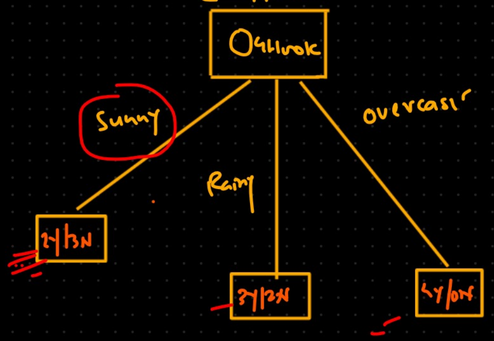
Decision tree → Regression, classification

ID3

CART

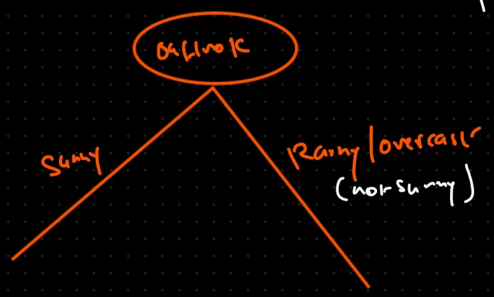
Outlook ⇒ 3

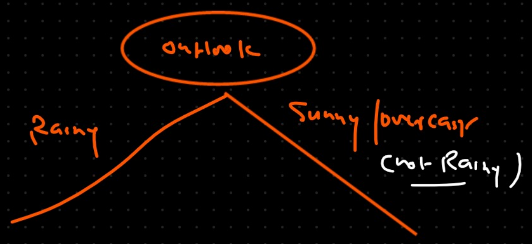
[9Y | 5N]



Outlook	Temperature	Humidity	Windy	PlayTennis
Sunny	Hot	High	False	No
Sunny	Hot	High	True	No
Overcast	Hot	High	False	Yes
Rainy	Mild	High	False	Yes
Rainy	Cool	Normal	False	Yes
Rainy	Cool	Normal	True	No
Overcast	Cool	Normal	True	Yes
Sunny	Mild	High	False	No
Sunny	Cool	Normal	False	Yes
Rainy	Mild	Normal	False	Yes
Sunny	Mild	Normal	True	Yes
Overcast	Mild	High	True	Yes
Overcast	Hot	Normal	False	Yes
Rainy	Mild	High	True	No

CART ⇒ Binary split





Split ⇒ (1) ID3 ⇒ entropy → information gain

(2) CART ⇒ Gini impurity

Entropy ⇒ measure of Randomness

DT:- (we have to find out best Root Node which is having a minimum entropy → Randomness

Entropy → Range [0-1]

{Entropy = 0} ⇒ No Randomness

thermodynamics { if near to zero = less Randomness
if near to one = more Randomness }

$$\boxed{\text{Entropy} = -\sum_{i=1}^{N} P_i \log(P_i)}$$

2 class ⇒ Y | N

$$\text{Entropy} = -P_y \log_2(P_y) - P_N \log_2(P_N)$$

f=3

$$-P_{c_1}\log_2(P_{c_1}) - P_{c_2}\log_2(P_{c_2}) - P_{c_3}\log_2(P_{c_3})$$

N ⇒ classes

Dummy data

$(c_1|c_2)$ feature — target $(Y|N)$

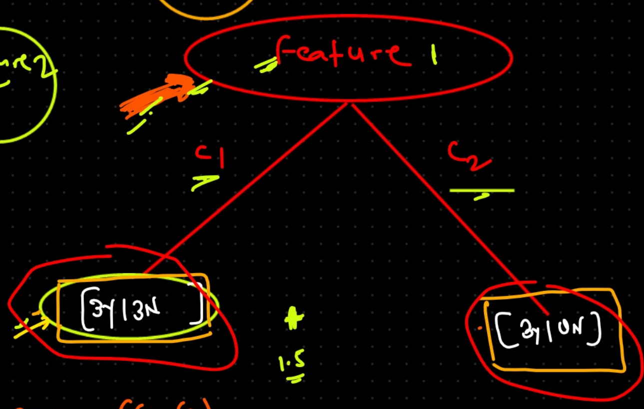
feature	target
c_1	Y
c_2	Y
c_1	Y
c_2	Y
c_1	Y
c_1	N
c_2	Y
c_1	N
c_1	N

Entropy

[6Y|3N]

feature 2

0.9



feature (c_1, c_2)

$$P(Y) = \frac{3}{6} = \frac{1}{2}$$
$$P(N) = \frac{3}{6} = \frac{1}{2}$$

$$\log(m/n) = \log m - \log n$$

$$\log_n n = 1$$

$$\Rightarrow -P_Y \log_2(P_Y) - P_N \log_2 P(N)$$
$$\Rightarrow -\frac{3}{6}\log_2\left(\frac{3}{6}\right) - \frac{3}{6}\log_2\left(\frac{3}{6}\right)$$
$$= -\frac{1}{2}\log_2\left(\frac{1}{2}\right) - \frac{1}{2}\log_2\left(\frac{1}{2}\right)$$
$$= -\frac{1}{2}\left[\log_2(1) - \log_2(2)\right] - \frac{1}{2}\left[\log_2(1) - \log_2(2)\right]$$
$$= -\frac{1}{2}[0-1] - \frac{1}{2}[0-1]$$
$$= \frac{1}{2} + \frac{1}{2} = 1$$

3Y|3N ⇒ 50|50%

Range = [0-1]

1 = high entropy

(3Y|0N) ⇒ 0

high random

3Y|0N

$$= -\frac{3}{3}\log_2\left(\frac{3}{3}\right) - \frac{0}{3}\log_2\left(\frac{0}{3}\right)$$
$$= -1(0) - 0$$
$$= 0 - 0 = 0$$

Y|N

6 3Y|3N ⇒ Y|N ⇒ 1

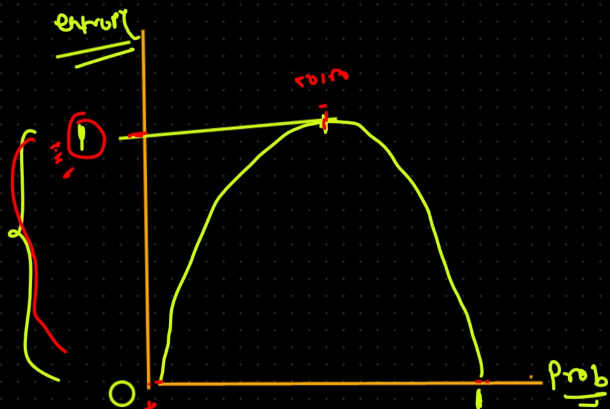
4Y|2N ⇒ yes

(Randomness)

< 1

5Y|1N <<1

6Y|0N = 0



%

(2y/3y) = ?

Entropy

{ Gini
i.g.
ch.s
D+R } ⇒ Practical

Num

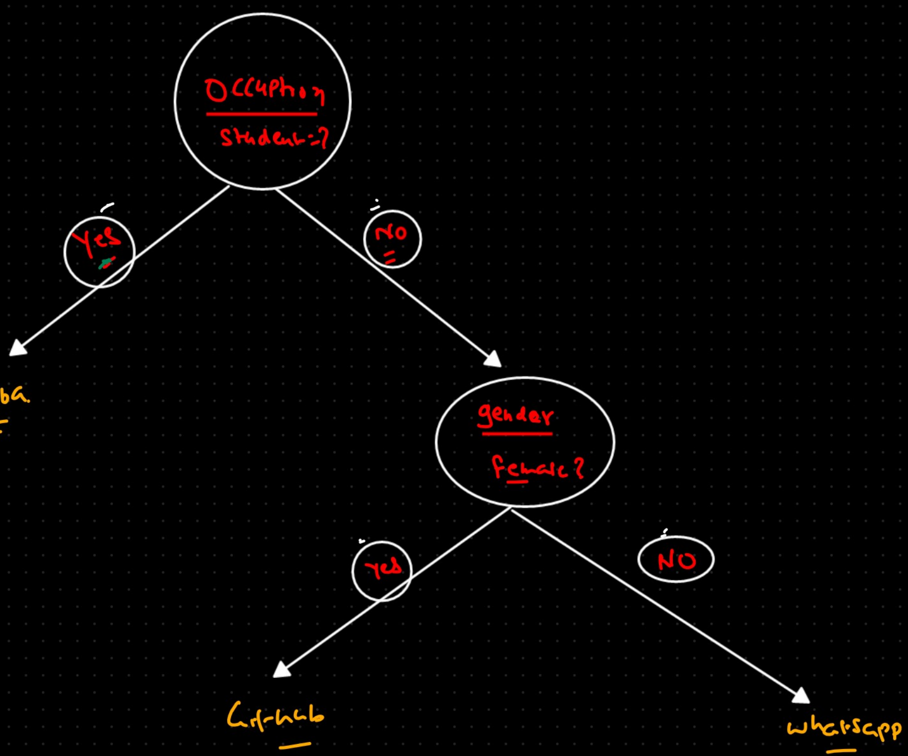
Weight

Cat. Cat.

Gender	Occupation	Suggestion - app
F	Student	Pubg
F	Programmer	github
M	Prog.	whatsapp
F	Prog.	Github
M	Stu	Pubg
M	Stu.	Pubg

```
if occupation == "Student"
    Print("Pubg")
else if gender == "female"
    Print("Git")
else Print("whatsapp")
```

Tree



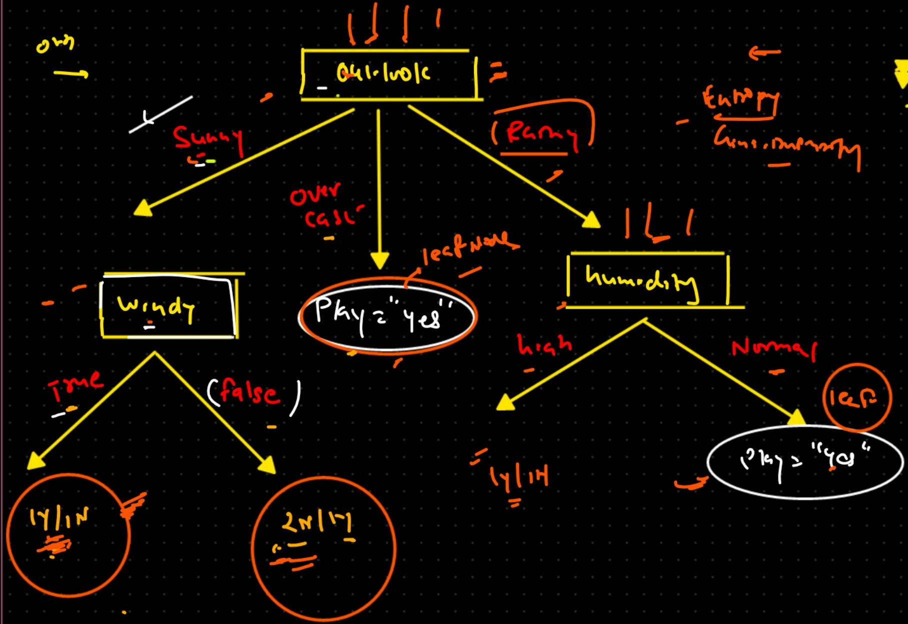
Question: No. of Features?

{ how will decide that which one will be my Root Node? }

ID3 ⇒ entropy → info. gain
CART ⇒ Gini impurity

Weighted Sum

Rights



Outlook	Temperature	Humidity	Windy	PlayTennis
Sunny	Hot	High	False	No
Sunny	Hot	High	True	No
Overcast	Hot	High	False	Yes
Rainy	Mild	High	False	Yes
Rainy	Cool	Normal	False	Yes
Rainy	Cool	Normal	True	No
Overcast	Cool	Normal	True	Yes
Sunny	Mild	High	False	No
Sunny	Cool	Normal	False	Yes
Rainy	Mild	Normal	False	Yes
Sunny	Mild	Normal	True	Yes
Overcast	Mild	High	True	Yes
Overcast	Hot	Normal	False	Yes
Rainy	Mild	High	True	No

DT => { DTR → target => Num
DTC → target = Cat }

DT
DTR

feature	target
gender	salary → Num

classification

feature	target
gender	obes / not-obes

DTC

Cat/Num

feature
target

Gender → Cat

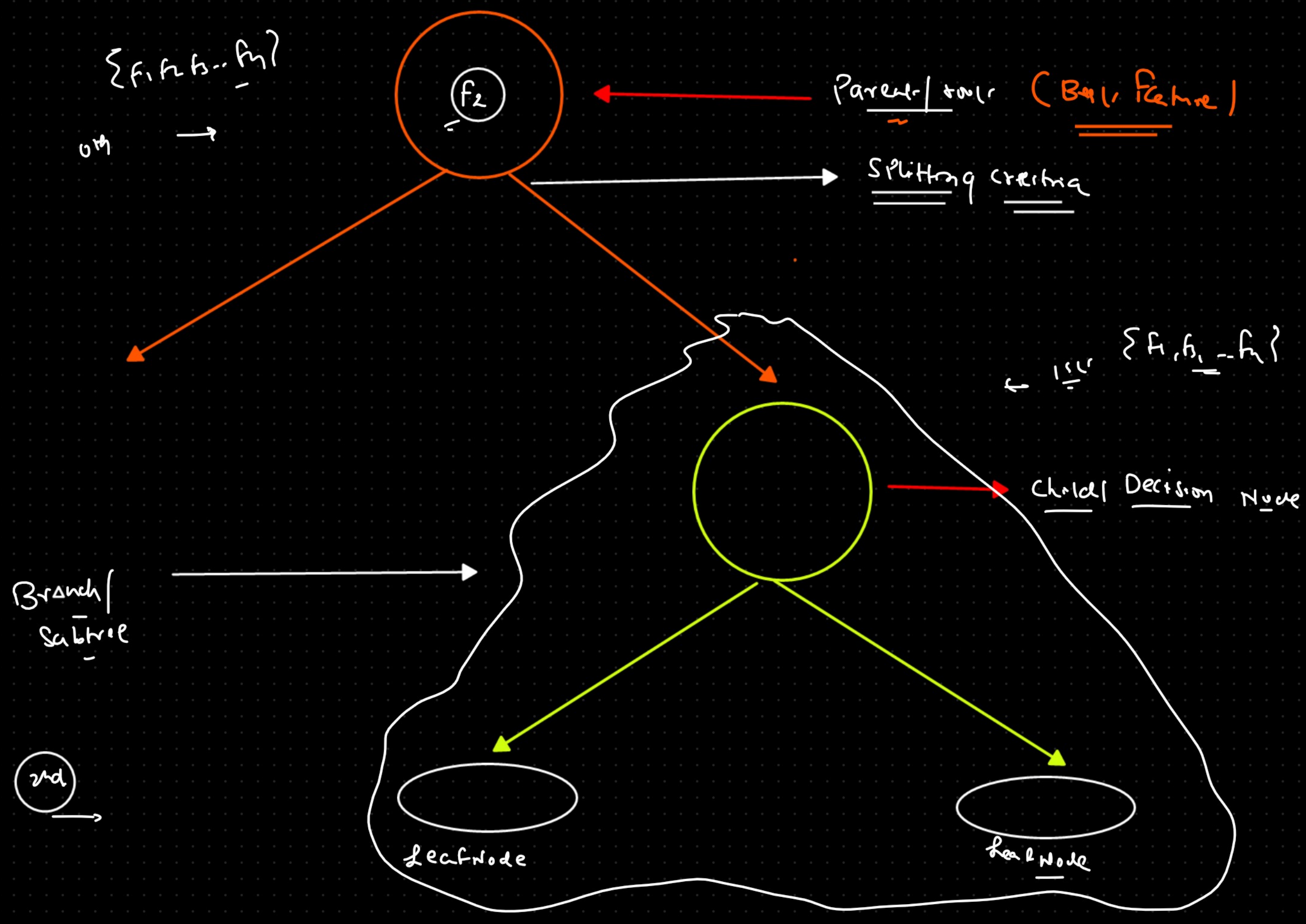
Num
split

class

DT => classification

Reg.

Independent	target	
Categorical	Categorical	DTC
Numeric	Categorical	
Categorical	Num	DTR
Num	Num	



1. reg/classification (target)

   independent ($f_1 \ldots f_n$) => root — best feature
   - 1. entropy
   - 2. Gini-index

2. we split the data based on best feature into the subtree

3. again we perform the same step

   recursively generate a new tree

Advantage

1. It is intuitive easy to understand
2. minimal Data Prep. is req. (with less preprocessing (raw) DT)
3.

Disadvantage

1. overfitting
2. prones to error for imbalance data

linreg, log, svm ⇒ {$y = mx + c$, $ax + by + c = 0$}

DT ⇒ {Pre Pruning, Post-Pruning} ⇒ overfitting / optimize

Hyperparameter

{DT} ⇒ {Non-Parametric} {hyperparameter}

{$f_1 f_2 f_3 \dots f_n$} ⇒ best feature, best split criteria

Entropy :- Layman, En. is nothing but the of Disorder. You can say it is a measure of Purity / impurity

Phy / chem / eng.

Solid.

ICE < water < GAS / vapour

$$\text{Entropy} = -\sum_{i=1}^{N} p_i \log_2 p_i$$

2 class (yes | No)

$$E = -P_{yes}\log_2(P_y) - P_{No}\log_2(P_N)$$

more than 2 class

3 class

$$E = -P_{c_1}\log_2(P_{c_1}) - P_{c_2}\log_2(P_{c_2}) - P_{c_3}\log_2(P_{c_3})$$

imp

$\log_2$ & $\log_e$

{ P = Probability } ⇐ imp

$$P(A) = \frac{\text{The number of ways event can occur}}{\text{The total number of Possible Outcomes}}$$

3 ⇒ fruit

⇒ 2 orange ⇒ 2/3

⇒ 1 banana ⇒ 1/3

Outlook	Temperature	Humidity	Windy	PlayTennis
Sunny	Hot	High	False	No
Sunny	Hot	High	True	No
Overcast	Hot	High	False	Yes
Rainy	Mild	High	False	Yes
Rainy	Cool	Normal	False	Yes
Rainy	Cool	Normal	True	No
Overcast	Cool	Normal	True	Yes
Sunny	Mild	High	False	No
Sunny	Cool	Normal	False	Yes
Rainy	Mild	Normal	False	Yes
Sunny	Mild	Normal	True	Yes
Overcast	Mild	High	True	Yes
Overcast	Hot	Normal	False	Yes
Rainy	Mild	High	True	No

— target
14 (9Y | 5N)

⇒ (9Y | 5N) = 14

Humidity

high: { (3Y | 4N) } = 7

Normal: { (6Y | 1N) } 7

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

$$E_L = -\frac{3}{7}\log\left(\frac{3}{7}\right) - \frac{4}{7}\log\left(\frac{4}{7}\right)$$

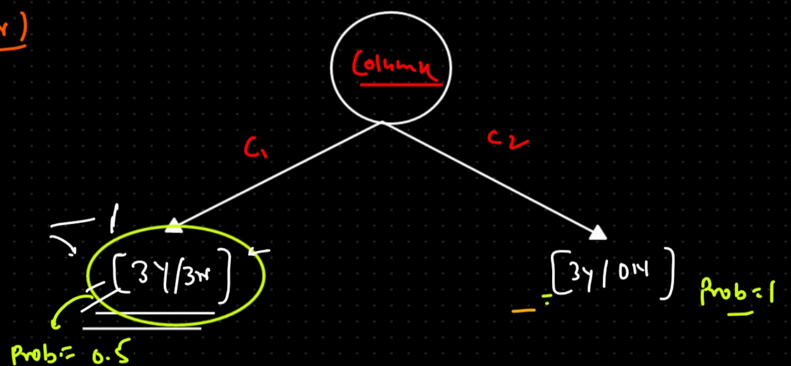
$$E_R = -\frac{6}{7}\log\left(\frac{6}{7}\right) - \frac{1}{7}\log\left(\frac{1}{7}\right)$$

Dummy

Entropy ⇒ Imp

C1/C2 — Column — O/P (Y/N)

Column	O/P
C1	Y
C2	Y
C1	Y
C2	Y
C1	Y
C1	N
C2	Y
C1	N
C1	N



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

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

{ Binary ⇒ E = 0 ⇒ P = 1
E = 1 = P = 0.5 }

Y/N

Random, uncertainty / impure ↑↑↑↑

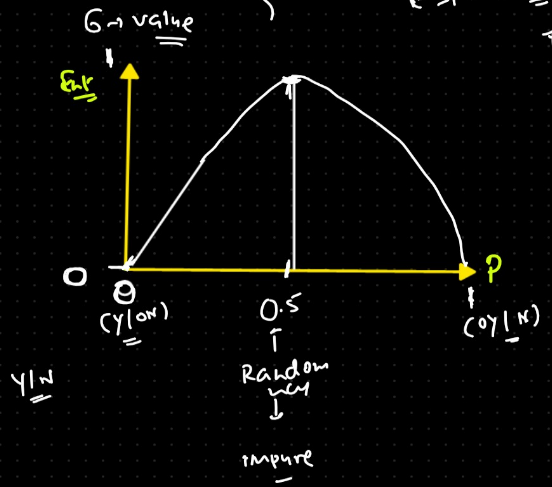
⇒ Entropy high ① 3Y / 3N ⇒ P = 0.5 (P)

② 4Y / 2N ⇒

③ 5Y / 1N ⇒

④ 6Y / 0N ⇒ = 1 (P_2)

entropy ↓↓↓↓ impurity



6Y / 0N

$$\left\{ P = \frac{6}{6} = 1, \quad E = 0 \right\}$$

Gini - coeff ⇒

$$\boxed{\text{Gini impurity} = 1 - \sum_{i=1}^{N} (P_i)^2}$$

2-class (Y|N)

$$1-\left[P_y^2+P_n^2\right]$$

n-number of class more than 2-class

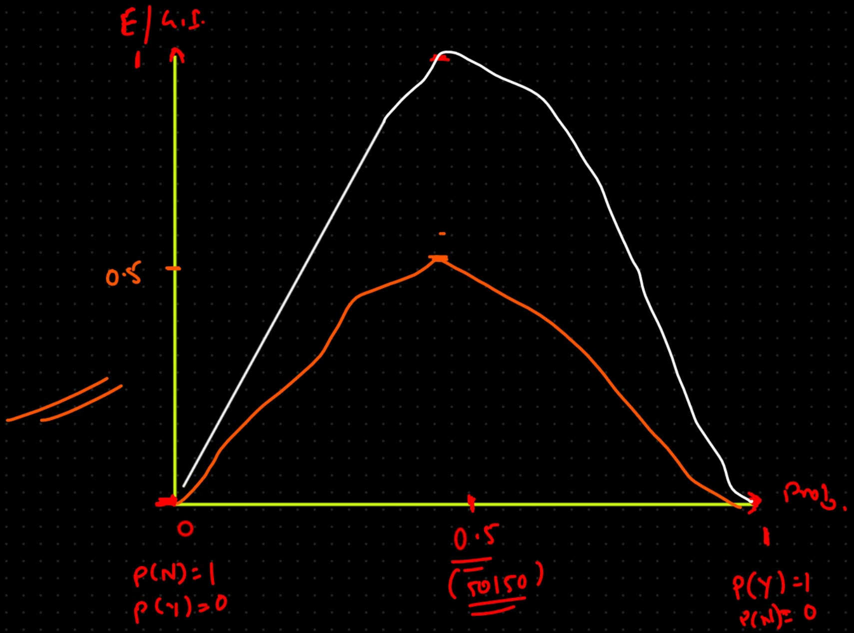
$$\left(1-\left[P_{c_1}^2+P_{c_2}^2+P_{c_3}^2\right]\right)$$

Gini-imp ⇒ ① 3y/3N ⇒ $1-\left[\left(\frac{3}{6}\right)^2+\left(\frac{3}{6}\right)^2\right] \Rightarrow 1-\left[\frac{9}{36}+\frac{9}{36}\right] \Rightarrow 0.5$

Gini-imp ② 3y/0N ⇒ $1-\left[\left(\frac{3}{3}\right)^2+\left(\frac{0}{3}\right)^2\right] \Rightarrow 1-\left[1+0\right] \Rightarrow 0$

Binary ⇒ highest ⇒ P ⇒ 0.5 ⇒ GI = 0.5] E = 1
lowest P = 1 ⇒ GI = 0] E = 0

{ P(N)=0
P(y) = 1-0
= 1 }



Entropy	Gini impurity
① { $-\sum P_i \log(P_i)$ }	{ $1-\sum (P_i)^2$ }
log	square

Research — entropy

Entropy — {more storage}

Some split

Entropy →

Assignment

→ Gini-impurity vs Entropy?

② { log ⇒ Slower compare to square }

99% ⇒ { Entropy ⇒ Slower
big data ⇒ entropy slow Gini-imp ↑ faster }

② [0-1] binary

multiclass

E > 1

[0-0.5]

Gini-imp > 1

③ ID3 ⇒ Entropy

③ {CART} Gini

Info.gain

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

0.7 0.3

$(0.7)^2 \quad (0.3)^2$

$0.7 \times 0.7 \quad 0.3 \times 0.3$

$0.49 \quad 0.09$ } more sensitive

root

child

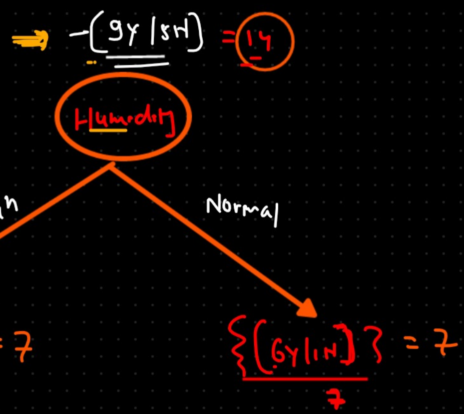
$$\{\text{Info gain}\} = E(\text{Parent}) - (\text{weighted avg.}) * E(\text{Child})$$

$$I.G. = H(S) - \sum_{v \in val} \frac{|S_v|}{|S|} H(S_v)$$

$$I.G. = E(P) - [(\text{weight avg.})] \times E(C)$$

$E(P)$ ↓ Parent

$E(C)$ ↓ Child (L, R)



Parent: $E(H) = -\frac{9}{14}\log\left(\frac{9}{14}\right) - \frac{5}{14}\log\left(\frac{5}{14}\right) \Rightarrow \simeq 0.9402$

L Child: $E_L = -\frac{3}{7}\log\left(\frac{3}{7}\right) - \frac{4}{7}\log\left(\frac{4}{7}\right)$

Right: $E_R = -\frac{6}{7}\log\left(\frac{6}{7}\right) - \frac{1}{7}\log\left(\frac{1}{7}\right)$

$$E(L) = -\frac{3}{7}(\log(3) - \log(7)) - \frac{4}{7}(\log(4) - \log(7)) \Rightarrow 0.68$$

$$E(R) = -\frac{6}{7}(\log_2(6) - \log(7)) - \frac{1}{7}(\log(1) - \log(7)) \Rightarrow 0.41$$

$\log(4) = 0.60$, $\log_2(3) = 0.47$, $\log_2(7) = 0.84$, $\log_2(6) = 0.77$, $\log_2(1) = 0$

h.w

$$L = -\frac{3}{7}(0.47 - 0.84) - \frac{4}{7}[0.60 - 0.84]$$

$$R = -\frac{6}{7}[0.77 - 0.84] - \frac{1}{7}[0 - 0.84]$$

Final info. gain =

$E(R)$ ↓

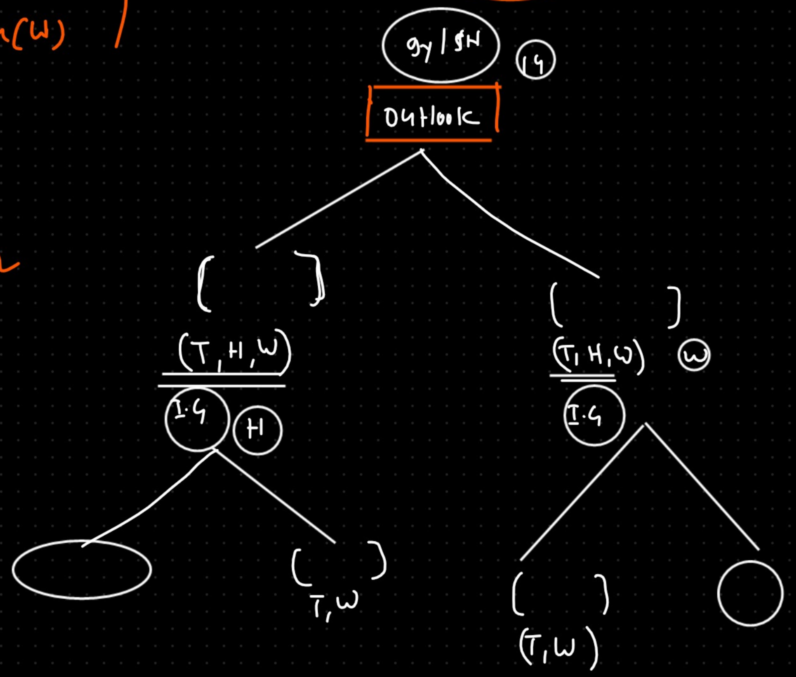
$$0.94 - \left[\frac{7}{14} \times 0.68 + \frac{7}{14} \times 0.41\right]$$

$$= 0.94 - [\quad]$$

$$I.G.(H) = \quad =$$

$$\left\{ \begin{matrix} IG(O) \\ IG(T) \\ IG(W) \end{matrix} \right\} = \text{Highest} \Rightarrow \text{Root Node}$$

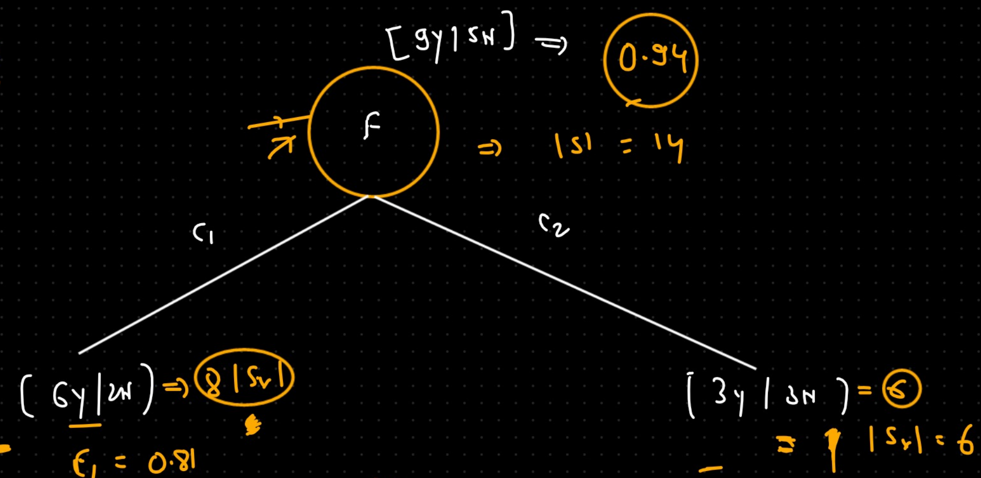
T, H, O, W
↓
highest



?

How to calculate

I.G.



Summation

$$IG = H(S) - \left\{ \sum_{v \in val} \frac{|S_v|}{|S|} * H(S_v) \right\}$$

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

$$I.G = 0.049$$

Outlook	Temperature	Humidity	Windy	PlayTennis
Sunny	Hot	High	False	No
Sunny	Hot	High	True	No
Overcast	Hot	High	False	Yes
Rainy	Mild	High	False	Yes
Rainy	Cool	Normal	False	Yes
Rainy	Cool	Normal	True	No
Overcast	Cool	Normal	True	Yes
Sunny	Mild	High	False	No
Sunny	Cool	Normal	False	Yes
Rainy	Mild	Normal	False	Yes
Sunny	Mild	Normal	True	Yes
Overcast	Mild	High	True	Yes
Overcast	Hot	Normal	False	Yes
Rainy	Mild	High	True	No

⇒ manullay

Entropy, Gini index

Entr → I.G.

features ⇒ outlook = 0.01
Tem = 0.5
Hum = 0.62
Windy = 0.012
I.G.

(9y|5N)

H

{O | T | W} → T

{O | T | W} ⇒ W

O = 0.01
T = 0.5
W = 0.12

(O W) ⇒ info → W

O = 0.01
W = 0.12

O

Now to create a DT = ?

Assignment

Doc

sunny.savita@ineuron.ai

E, I.G. = ? ⇒ Create

1 week

DTR

Hyperparameter ?

(Post, Pre)

I/V

N

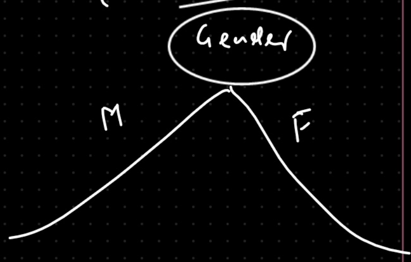
How to approach this type of Data ?

1st Senerio

already sed

Classification

{entropy / Gini-impurity}



(Gender)	Heart Disease
M	Y
F	N
M	Y
F	Y
M	N

Senerio-2

Classification

Optimize

Weight	Heart Disease
220	Y
180	Y
225	Y
190	N
155	Y

=> {Gini-imp / Entropy} => I.G

1. Sort your Data
2. take avg. of adjustant value
3. W.r.t every avg. value i have to do a split

[entropy / Gini-impurity]

1. Sort

2.

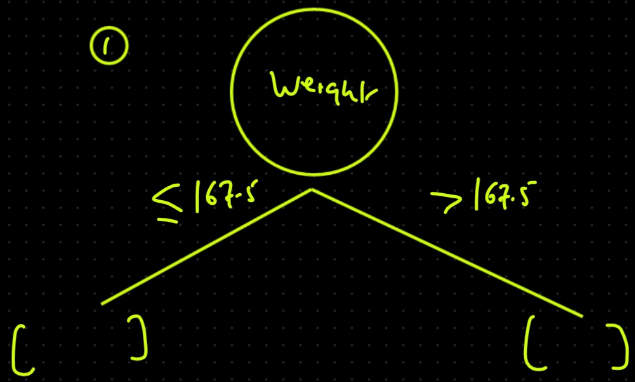
	Weight	H.D.
167.5	155	N
185	180	Y
	190	N
205	220	Y
222.5	225	Y

best split

② average adjustant ←

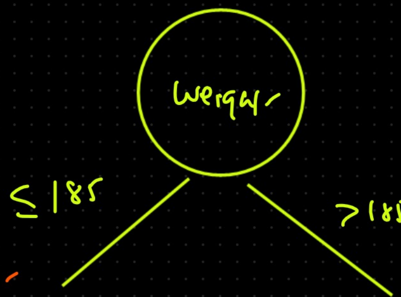
③ i will take every avg. and will do the split

①



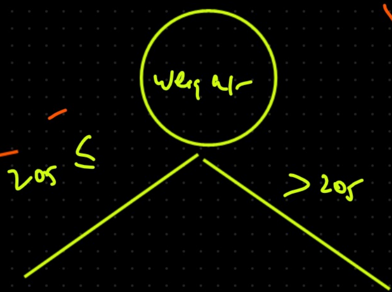
⇒ E/IG = 0.7

②



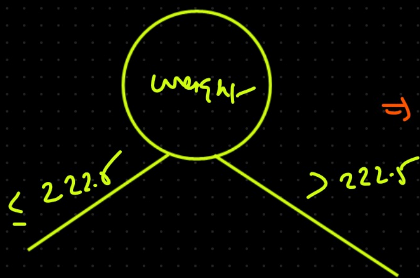
⇒ E/I.G. ⇒ 0.85

③



⇒ E/IG = 0.91

④



⇒ E/IG ⇒ 0.021

Pruning, DTR, Practical, H.T.

↓

Ensemble learning

RF ↑ bagging,

AB, GB, XB ↑ boosting,

Stacking