Decision Tree

DT Regression

DT classification.

① DT classifier -

Two Technique

① ID3
② CART

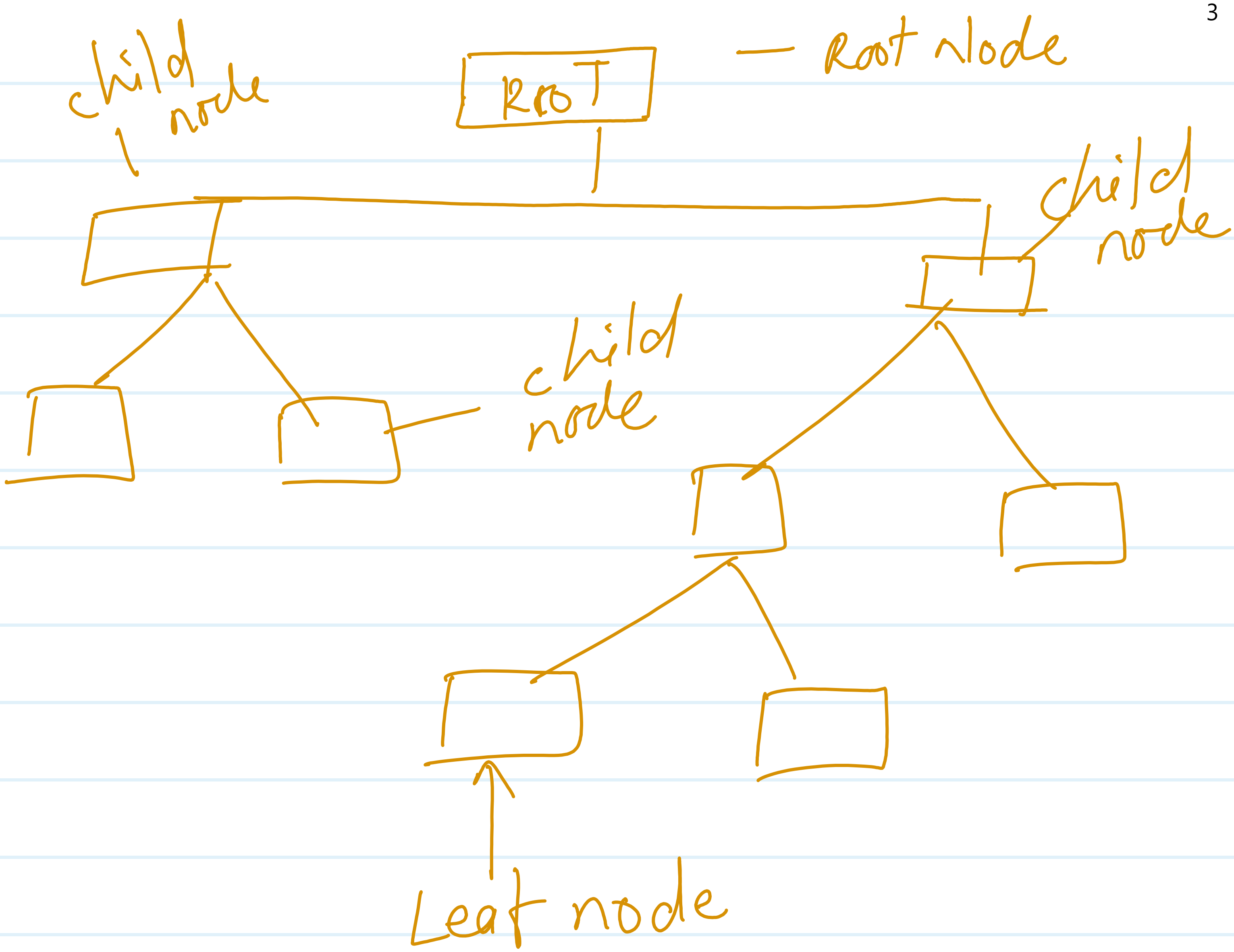
Entropy and Gini Index -
We use it to split dataset

Information Gain

X weight	X heigh	Y o/p	obese/No.
			obese.
60	160	ob	
70	170	No	
80	180	ob	
90	190	No	
100	200	No	

II DT Regression

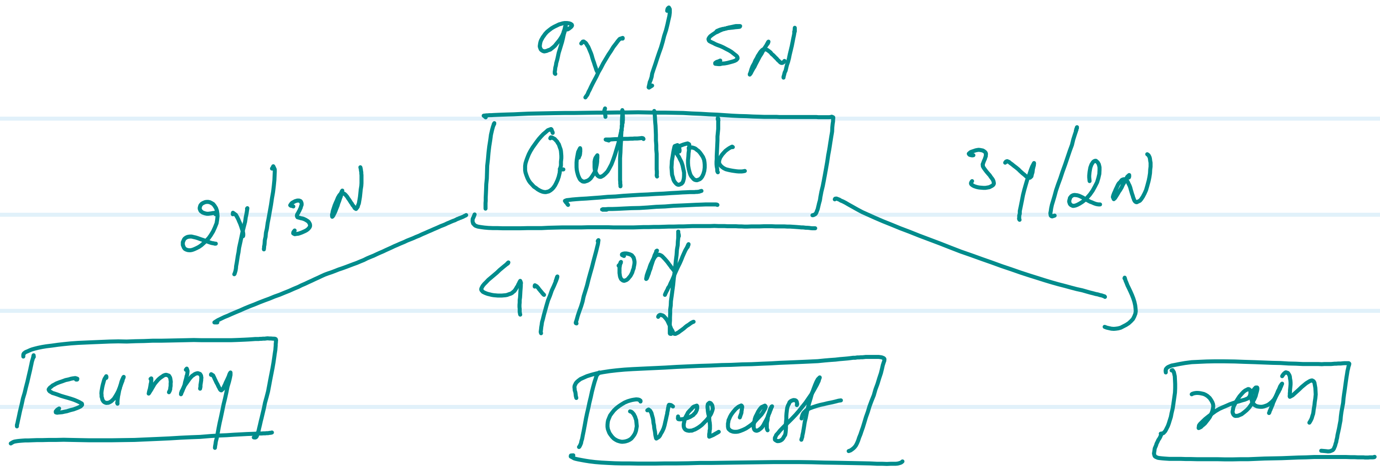
weigh	heigh	o/p BMI
60	160	21
70	170	22
80	180	20
85	190	23
90	195	24


ROOT
Root Node
child node
child node
child node
Leaf node

DT classifier

outlook	temp	humidity	wind	play
Sunny	H	H	W	N
sunny	H	H	S	N
overcast	H	H	W	Y
rain	M	H	W	Y
rain	C	N	W	Y
rain	C	N	S	N
overcast	C	N	W	Y
sunny	M	H	W	N
sunny	C	N	W	Y
rain	M	N	W	Y
sunny	M	N	S	Y
overcast	M	H	S	Y
overcast	C	N	W	Y
rain	M	H	S	N

1. Feature can be numeric and categorical
2. o/p can be numeric and categorical.



For each feature we have to do the same process like we did above.

* Entropy and gini impurity will decide which feature is do complete classification.

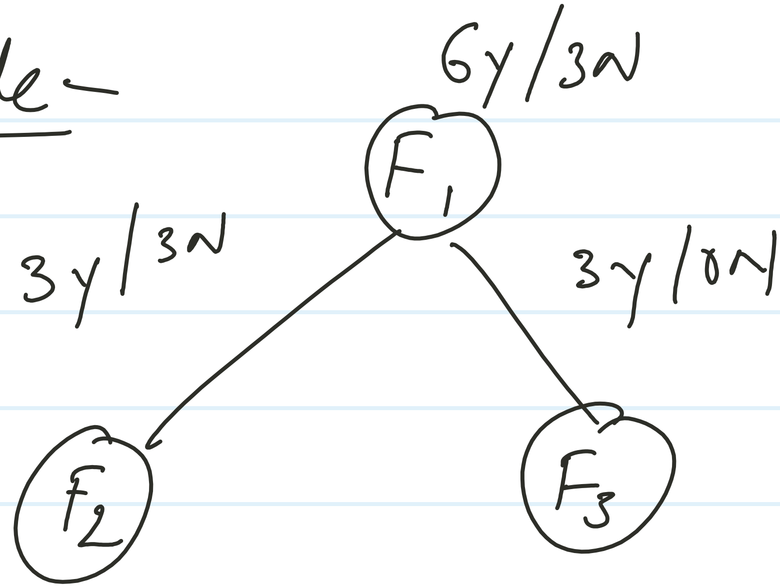
(a) Entropy (H)

Binary class $H = -P_{yes}\log_2(P_{yes}) - P_{No}\log_e(P_{no})$

multiclass

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

Example –



$$C_1 = -\frac{3}{6}\log\left(\frac{3}{6}\right) - \frac{3}{6}\log\frac{3}{6}$$

$$= 1$$ impure split

$$C_2 = -\frac{3}{3}\log\frac{3}{3} - \frac{0}{3}\log\frac{0}{3}$$

$$= 0$$ pure split

Imp For the pure split of data/feature

entropy value $= 0$

for impure split

entropy value $= 1$

(b) Gini Impurities – G.I.

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

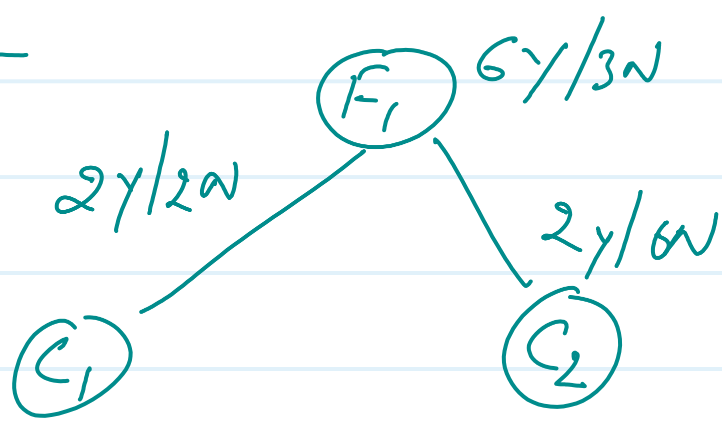
binary class

$$G.I. = 1 - \sum_{i=1}^{n} \left[(PC_1)^2 + (PC_2)^2 \right]$$

multi class

$$G.I. = 1 - \sum_{i=1}^{n} \left[(PC_1)^2 + (PC_2)^2 + \cdots\cdots \right]$$

for Eg –



$$C_1 = 1 - \left[\left(\frac{2}{4}\right)^2 + \left(\frac{2}{4}\right)^2 \right]$$

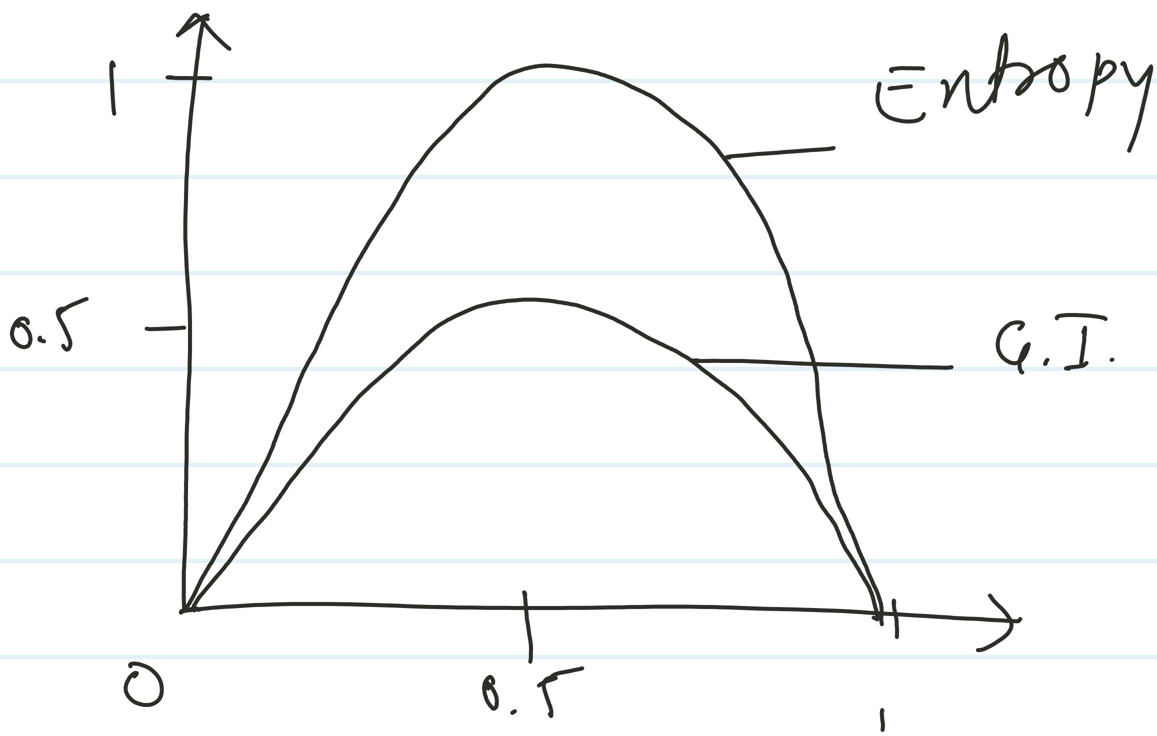
$$= 0.5$$

$$C_2 = 1-\left[\left(\frac{2}{2}\right)^2+\left(\frac{0}{2}\right)^2\right]$$

$$= 0$$

$\Rightarrow$ Range of Entropy is = 0 to 1

$\Rightarrow$ Range of G.I is = 0 to 0.5

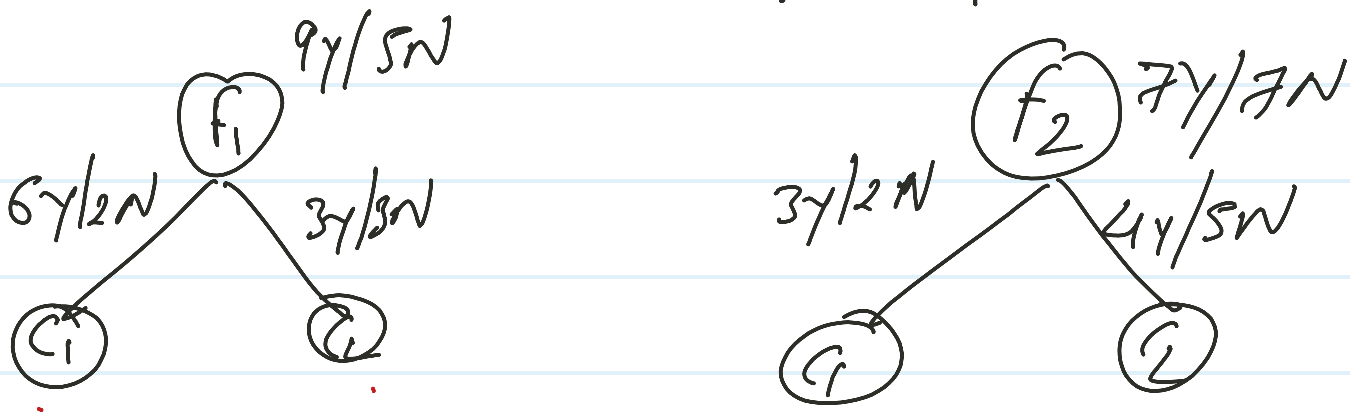


(c) Information Gain –

Will tell us among the all feature which is best for root node.

$$gain(S, f) = H(S) - \sum \frac{|S_v|}{|S|} H(S_v)$$

<u>for eg:-</u> we have two feature f_1 and f_2



<u>F_1</u>

$$H(S) = -\frac{9}{14}\log\frac{9}{14} - \frac{5}{14}\log\frac{5}{14}$$

$$H = 0.94$$

$$C_1 = 0.81$$

$$C_2 = 1$$

$$gain(S, f_1) = 0.94 - \left[\frac{8}{14} \times 0.81 + \frac{6}{14} \times 1\right]$$
$$= 0.049$$

f_2

$H = 0$

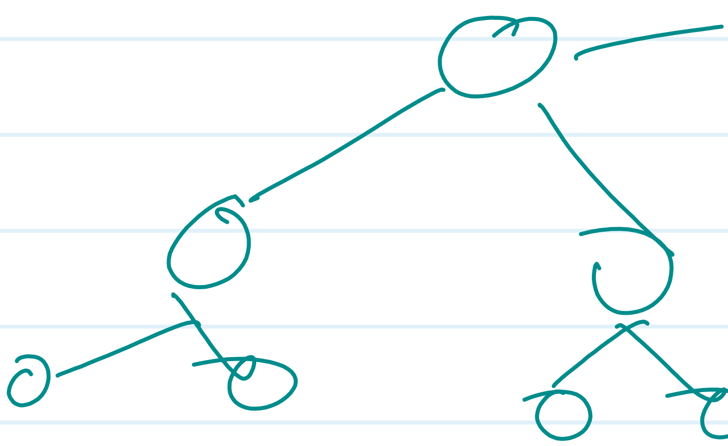
$C_1 = 0.29$

$C_2 = 0.014$

$$gain(S, f) = 0.014$$

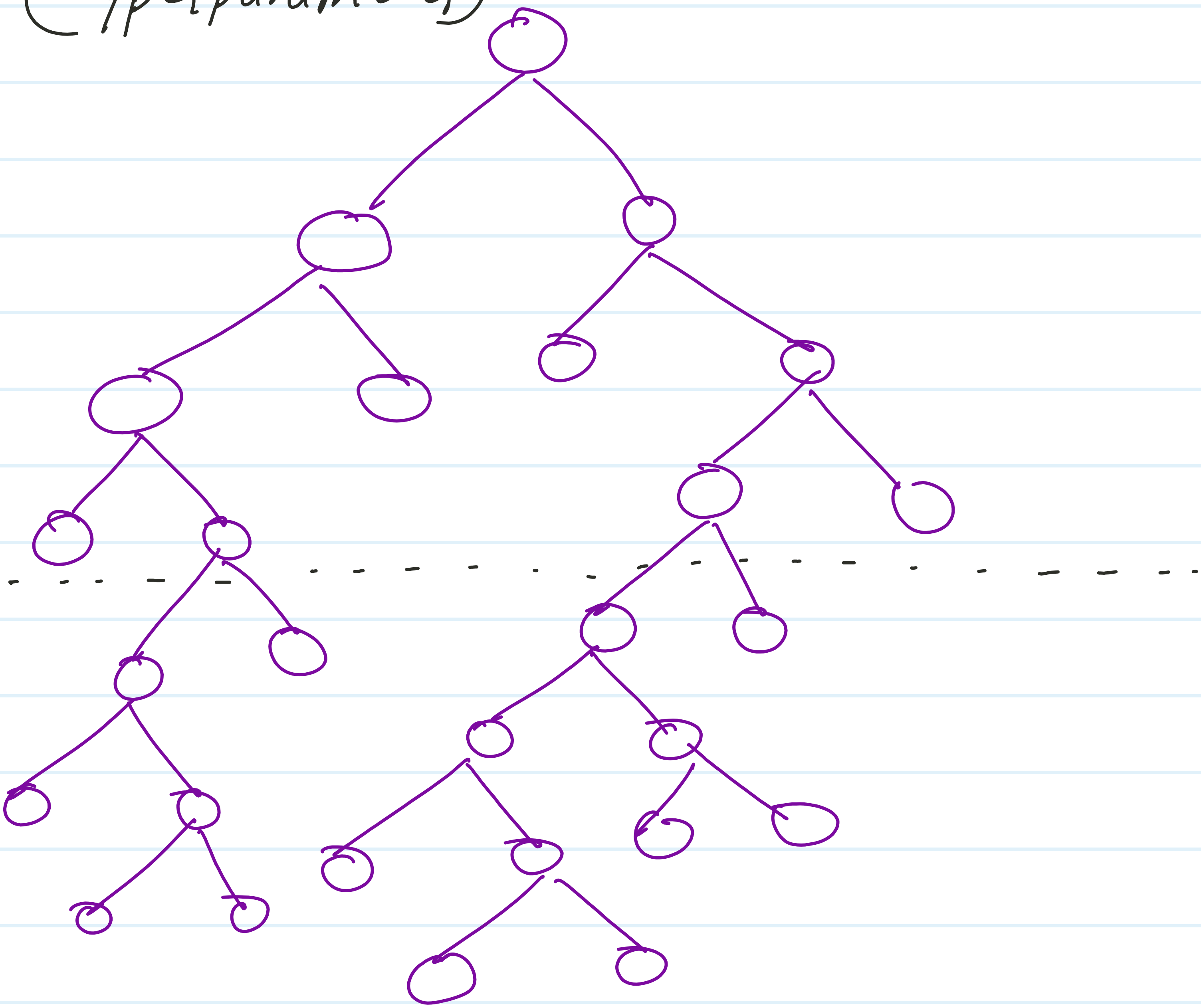
Information gain

f_1	f_2	f_3
0.049	0.014	0.009



Pre Pruning and Post pruning

(Hyperparameters)



1. Bigger dataset — pre-pruning
2. Small dataset — post-pruning.