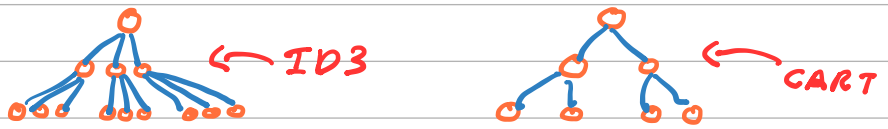
07 - Decision Tree

1) Decision Tree Classifier [classification]
2) Decision Tree Regressor [Regression]

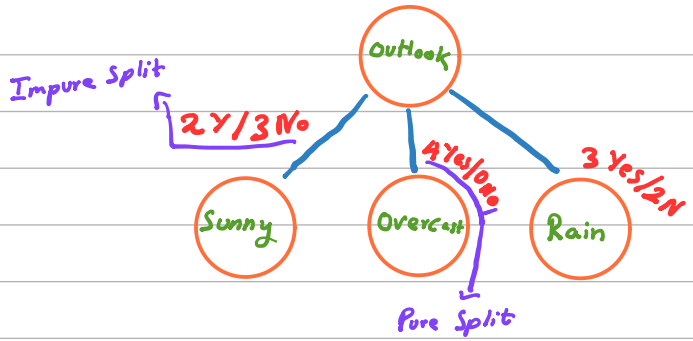
Decision Tree Clasifier

1) ID3 [Iterative Dichotomiser 3]
2) CART [Classification And Regression Tree]



ID3 When the child splits are more than binary (more than 2)

CART When the child splits are only binary (only 2)



(1) Purity Check: Pure Split & Impure Split

Pure Split - In which it cannot be further classified
Impure Split - In which it can be further classified

Entropy
Gini Impurity
} Check Measure of purity

② What feature you need to select to start the split ← Information Gain

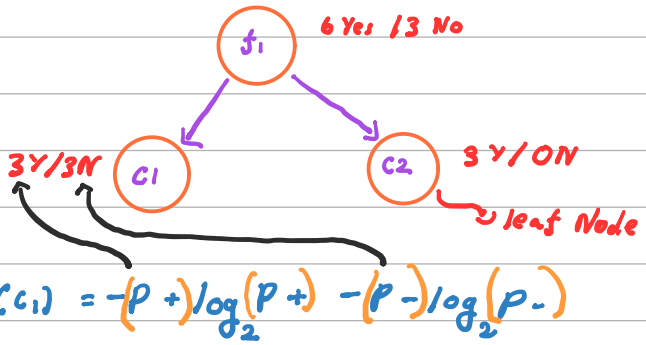
Binary classification

① Entropy

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

P_+ = Probability of +ve category

P_- = Probability of -ve category



$$H(C_1) = -(P_+)\log_2(P_+) - (P_-)\log_2(P_-)$$

$$= -\tfrac{3}{6}\log_2(\tfrac{3}{6}) - (\tfrac{3}{6})\log_2(\tfrac{3}{6})$$

$$= -\tfrac{1}{2}\log_2(\tfrac{1}{2}) - (\tfrac{1}{2})\log_2(\tfrac{1}{2})$$

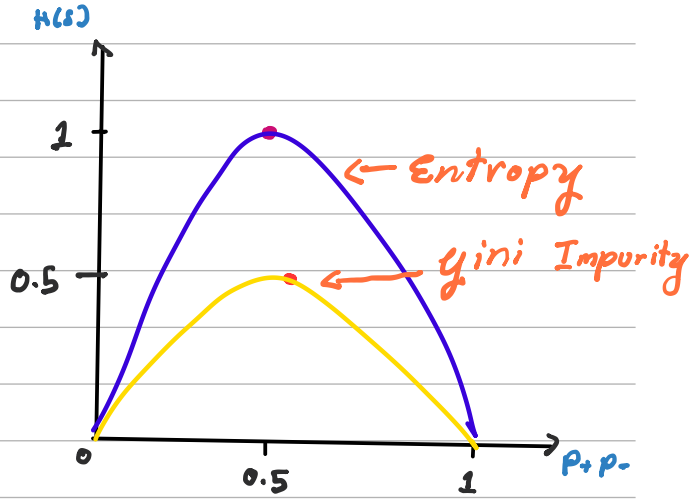
$$= 1 \Rightarrow \text{Impure split}$$

$$H(C_2) = -\tfrac{3}{3}\log_2(\tfrac{3}{3}) - \tfrac{0}{3}\log_2(\tfrac{0}{3})$$

$$= 0 \Rightarrow \text{Pure Split}$$

② Gini Impurity

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

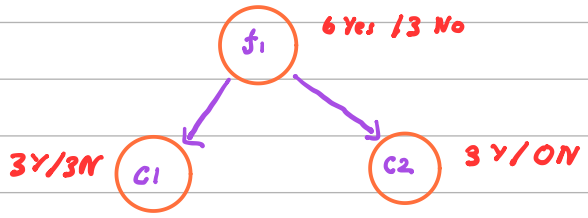


Multi classification

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

② Gini Impurity

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



$C_1 = 1 - [(P+1)^2 + (P-)^2]$

$= 1 - [(3/6)^2 + (3/6)^2]$

$= 1 - [1/4 + 1/4]$

$= 1/2 = 0.5 \Rightarrow$ Impure Split

$C_2 = 1 - [(3/3)^2 + (0/3)^2]$

$= 1 - 1$

$= 0 =$ Pure Split

In Entropy the graph of Impure split forms 1 where as in Gini Impurity the Impure split forms 0.5

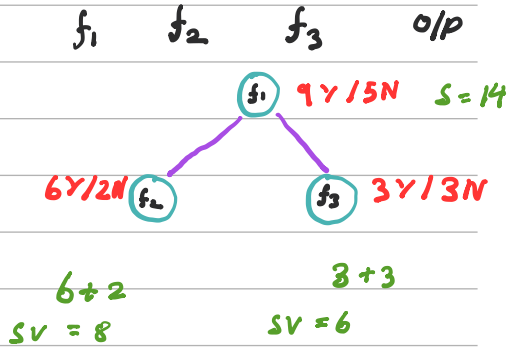
② What feature you need to select to start the split

↓

Information Gain

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

Entropy of the Root node



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

$$= -\frac{9}{14} \log_2 \frac{9}{14} - \frac{5}{14} \log_2 \frac{5}{14}$$

$$H(S) = 0.94$$

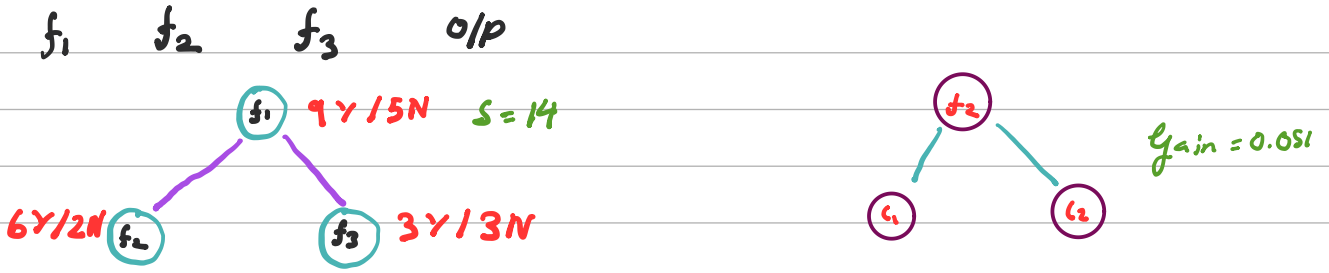
$$H(c_1) = -\frac{6}{8} \log_2 \left(\frac{6}{8}\right) - \frac{2}{8} \log_2 \left(\frac{2}{8}\right)$$

$$= 0.81$$

$$H(c_2) = \underline{1}$$

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

$$Gain(S, f_1) = 0.049$$



We calculate gain of each idividual features and find the max gain and use it.

$$Gain(f_2) > Gain(f_1)$$

Entropy vs Gini Impurity

When the dataset is small → Entropy (1000's - 10,000)

When the dataset is huge → Gini Impurity (1M - 100k)

Decision Tree Post Pruning And Pre Pruning

Due to decision Tree Overfitting occurs due to this we use two different methods to control overfitting

① Post Pruning:

First we make decission tree then later we prune the Tree.

We use `max_depth` to decide the level of splits.

hyperparameter

2) Pre Pruning = hyperparmeter Tuning

The tree is split at the same time of construction of Tree

It will set up the parameters while constructing tree

for large datasets

Pre-Pruning

Post Pruning

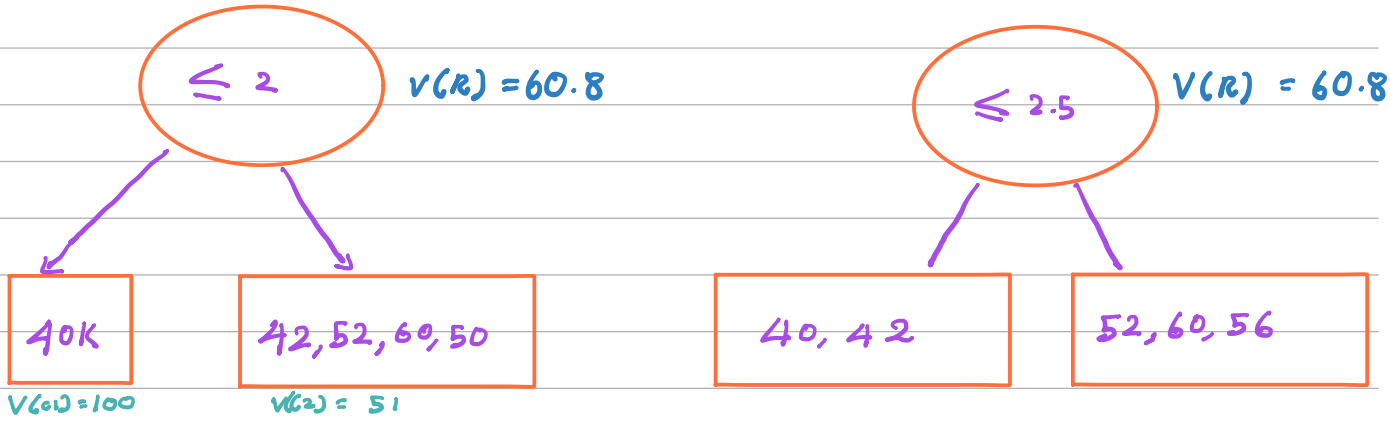
for small dataset

23 Decision Tree Regression

* Output is always Continuous Values

Exp	gap	salary
2	Yes	40K
2.5	Yes	42K
3	No	52K
4	No	60K
4.5	Yes	56K

avg = 50k



In decision Tree Regressor we use Variance Reduction to know which split to use.

$$\text{Variance} = \frac{1}{n}\sum_{i=1}^{n}(y_i - \bar{y})^2 \Rightarrow \text{MSE}$$

$$\text{Variance (Root)} = \frac{1}{5}(40-50)^2 + (42-50)^2 + (52-50)^2 + (60-50)^2 + (56-50)^2$$

$$= \frac{1}{5}\left[100 + 64 + 4 + 100 + 36\right] = 60.8$$

Variance (child 1) $= \frac{1}{1}\left[(40-50)^2\right]$

$= 100$

Variance (child 1) $= \frac{1}{2}\left[100+64\right]$

$= \frac{164}{2} = 82$

Variance (child 2) $= \frac{1}{4}\left[164+4+100+36\right]$

$= 51$

Variance (child 2) $= \frac{1}{3}\left[4+100+36\right]$

$= 46.66$

Variance Reduction:

$= Var(ROOT) - \Sigma wi\,(child)$

how many child nodes are there after split

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

$= 60.8 - 20 - 40.8$

$= 0$

$= 60.8 - \left[\frac{2}{8}(82) + \frac{3}{5}(46.66)\right]$

$= 0.004$

$= 0.004$