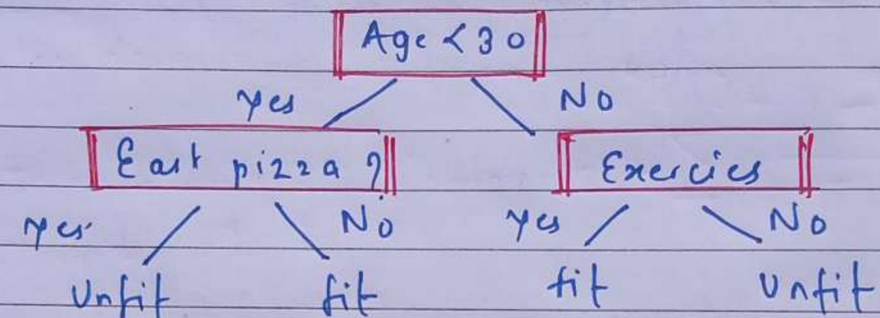
Desion Tree.

About decision Tree.

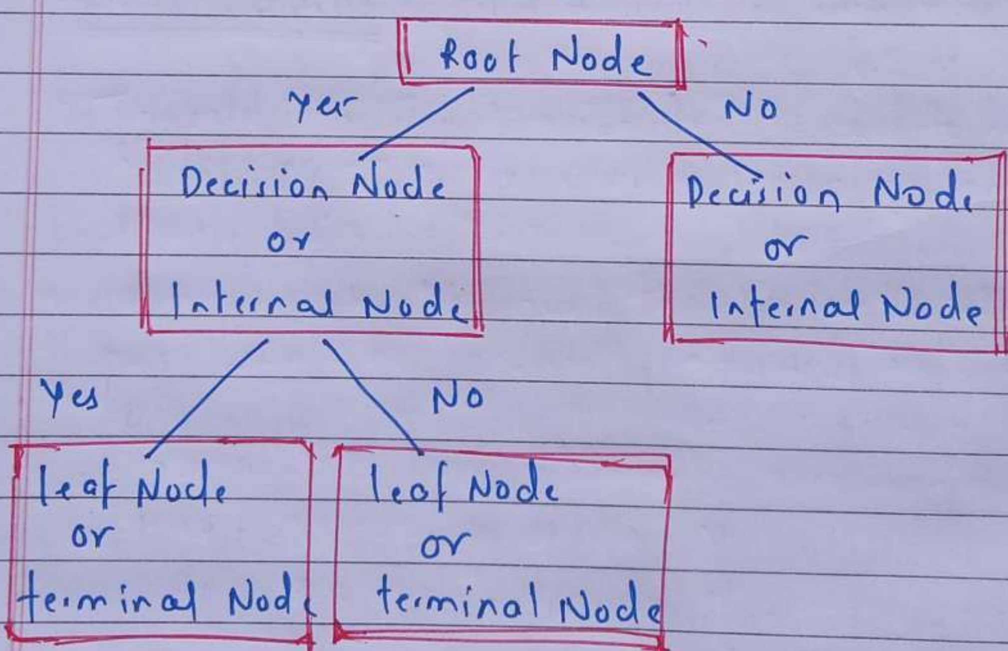
1) it is supervised ML Model
2) Used both classification & Regression.
3) Build Decision Nodes at each step.
4) Basis of Tree based model.

Let understand with Example.

is person fit or Not?



Structure & terminology of DT :-



Advantages :-

1). Can be used for both classification & Regression
2). Easy to interprete
3) No need for Normalization or scaling
4) Not sensetive to outliers.

Disadvantages:

1.) Overfitting issue
2) Small changes in the data alter the tree structure causing instability.
3) training time is relatively high.

Some concepts in DT :-

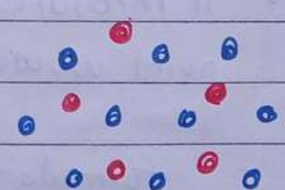
a) Entropy
b) Information Gain
c) Gini Impurity

If Entropy : High	entropy : low
Information gain : low	Information gain : High
Gini Impurity : High	Gini Emparity : low

{ Entropy and gini impurity are inversly proportional to each other }

Entropy :- In ML Entropy is the quantative measures of randomness of information being processed.

Low entropy



High Entropy.

- A high value of entropy means that randomness in the system is high and thus making accurate prediction is tough.
- A low value of Entropy means that randomness in the system is low and thus making accurate prediction is easier.

Entropy : $\sum_{i=1}^{c} - p_i \log_2 p_i$ $\quad$ c = number of classes, p_i : Probability of ith class

Date: / /
Page

- Information Gain :- once we find entropy to find which Feature to be selected as root Node or internal Node we use information Gain:
- it is measure of how much information a feature provide about class low entropy leads to increased Information Gain and high entropy lead low information gain
- information gain computes the difference between entropy before split and average entropy after split of the dataset based on given value

$$\text{Info gain}(T, f) = \text{Entropy}(T) - \sum_{v \in F} \frac{|T_v|}{T} \cdot \text{Entropy}(T)$$

Target ↗ ↑ features

- Gini Impurity :- it is measure of impurity at Node
- The split made in decision tree is said to be pure if all the data point are accurately seperated into different class.
- it measure the likelihood that randomly selected data point would be incorrectly classified by specific Node.

formula : $1 - (P_Y^2 + P_N^2)$

P_Y = probability of class Y
P_N = probability of class N.

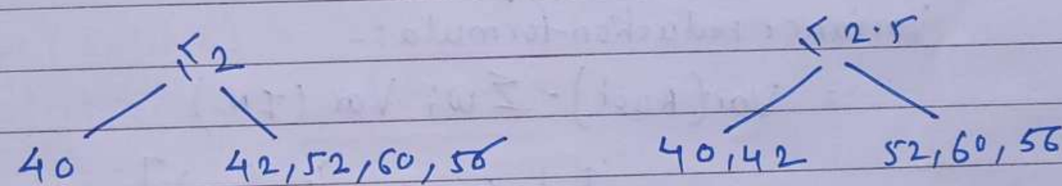
Decision Tree for regression.

Let understand with Example

Exp	Gap	Salary (K)
2	Yes	40
2.5	Yes	42
3	No	52
4	No	60
4.5	Yes	56
		$\bar{y} = 50$ ← Average.

Let take experience at root node
{Note:- since exp is continuous data DT arrange it in ascending order}

- Now for comparison we will take two node example



- Now to decide which split is suitable we used one Concept called "variance reduction."

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

where $\bar{y}$ = Average output.

- Now we have to calculate variance at each node.
- 1st we will calculate variance of root.

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

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

$$\boxed{= 60.8}$$

Now we will calculate variance at each internal node or decision node

$$Variance(IN1) = \frac{1}{1}\left[(40-50)^2\right]$$

$$= 100$$

$$variance(IN2) = \frac{1}{4}\left[(42-50)^2 + (52-50)^2 + (60-50)^2 + (50-50)^2\right]$$

$$= \frac{1}{4}\left[(-8)^2 + (2)^2 + (10)^2 + (6)^2\right]$$

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

$$= 51$$

Variance reduction formula:-

$$= Var(Root) - \sum w_i \, Var(IN)$$

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

$$= 60.8 - 20 - 40.8$$

$$= 0$$

Same we will calculate for second condition ≤ 2.5 whoever have large variance reduction we will finalize it for splitting.

$$var(IN1) = \frac{1}{2}\left[(40-50)^2 + (42-50)^2\right]$$

$$= \frac{1}{2}\left[100 + 64\right]$$

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

$$Var\ IN2 = \frac{1}{2}\left[(52-50)^2 + (60-50)^2 + (56-50)^2\right]$$

$$= \frac{1}{3}\left[\cancel{8}4 + 100 + 36\right]$$

$$= \frac{140}{3} = 46.66$$

Variance reduction for next split i.e ≤ 2.5

$$= Var(Root) - \sum w_i\ Var(IN)$$

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

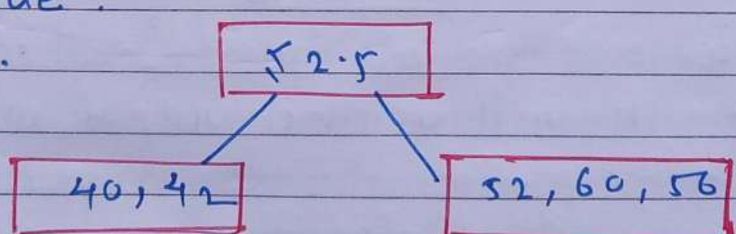
$$= 0.304$$

var(split 2) > var(split 1) thats why we will select second split.

How to calculate o/p for the test data

- whichever leaf node your test data reached to take avg of all the numbers present in the same leaf Node.

for Ex.



if test data reached to 1st leaf then

$$\frac{40+42}{2} = 41$$

and it reached to second leaf then

$$\frac{52+60+56}{3} = 56$$