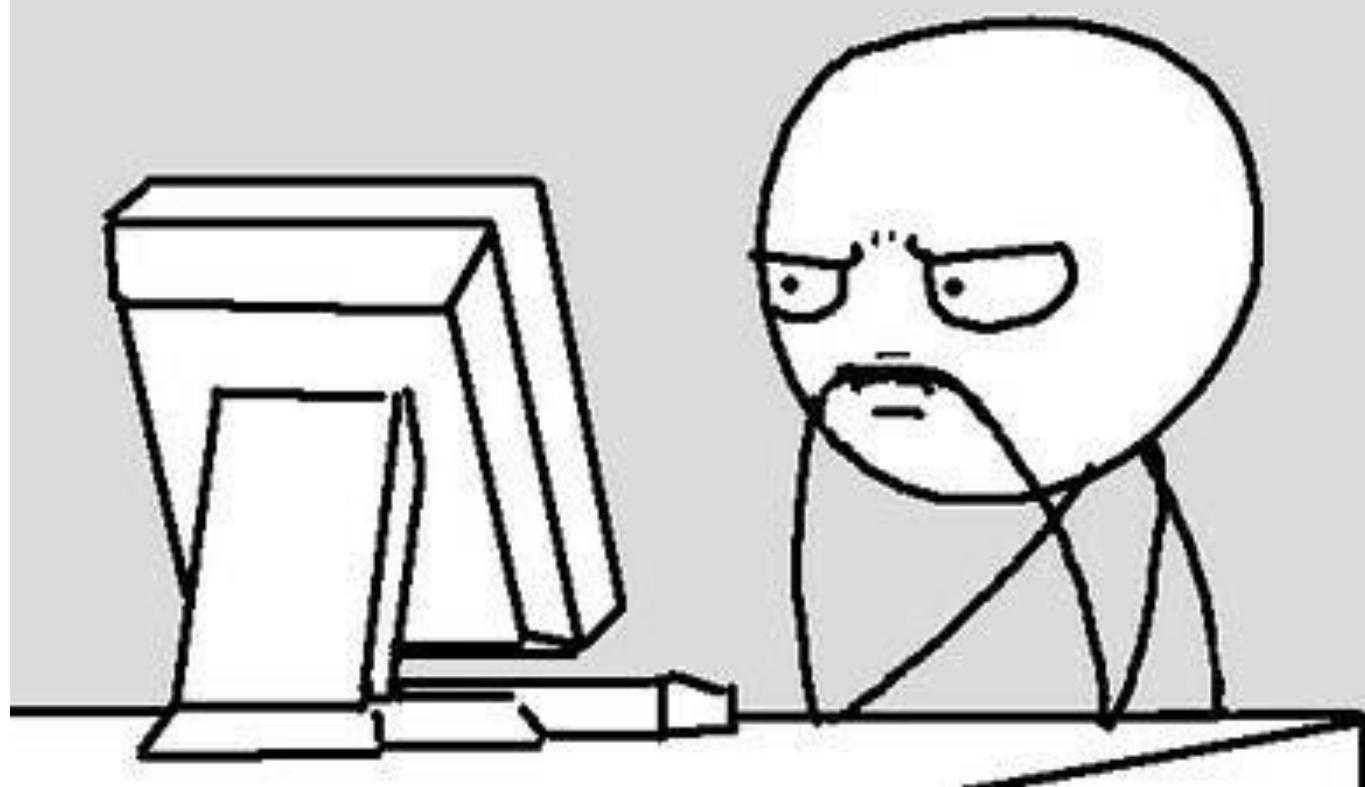
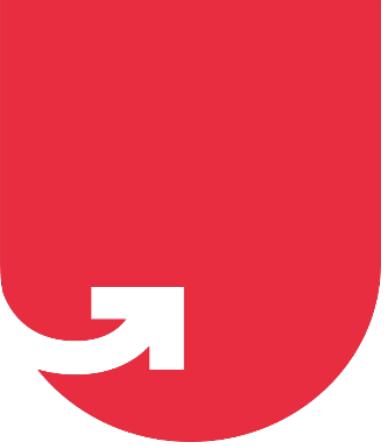


LET'S WAIT





#LifeKoKaroLift

Data Science Certification Program

Course : Machine Learning

Lecture On : Tree Models

Instructor : Shivam Garg



AGENDA

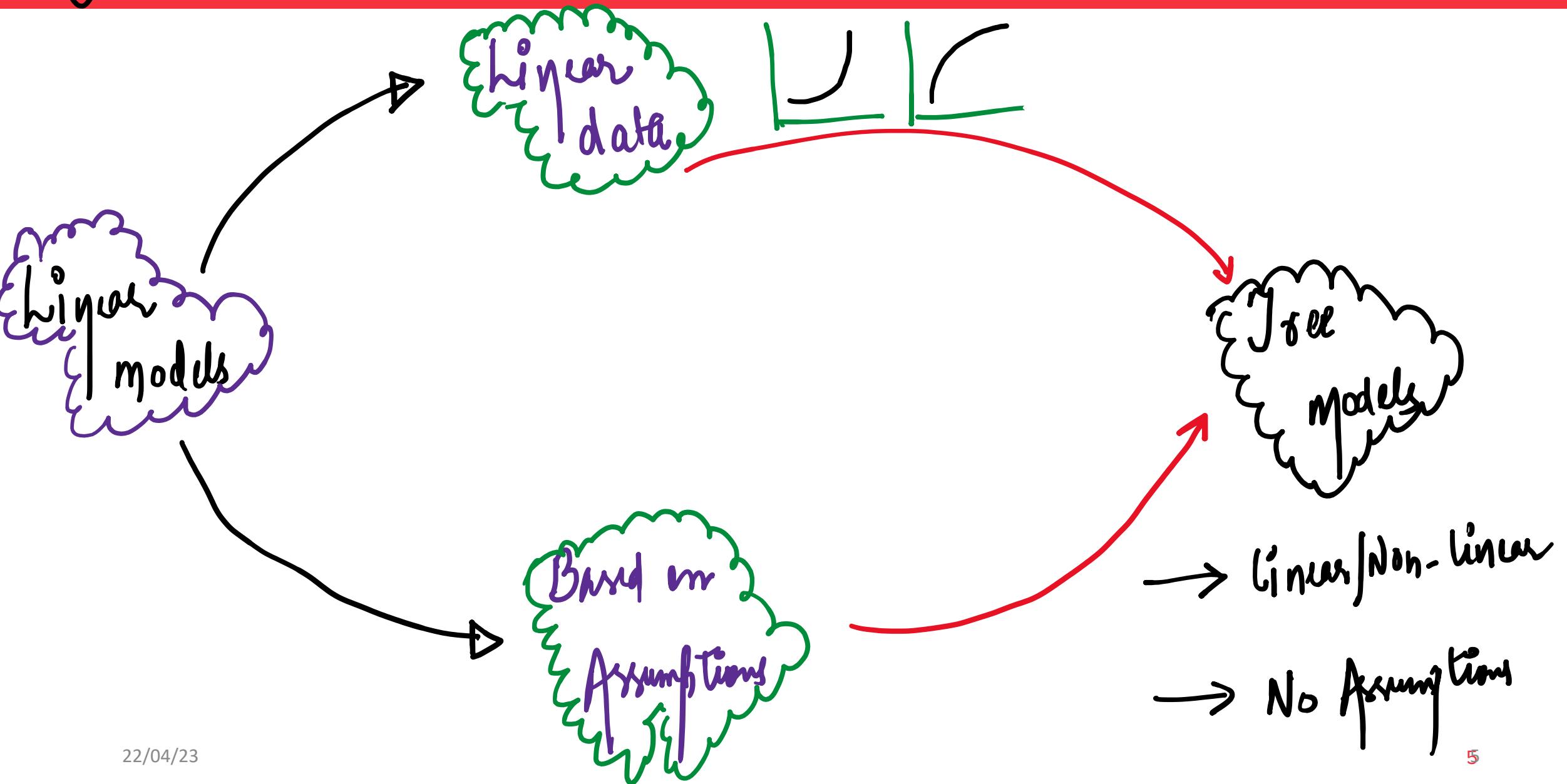
- Why Tree Models?
 - Decision Trees
 - Python Demo
 - Doubts
- 16 mins

few hyperparameters

Classification
or
Regression

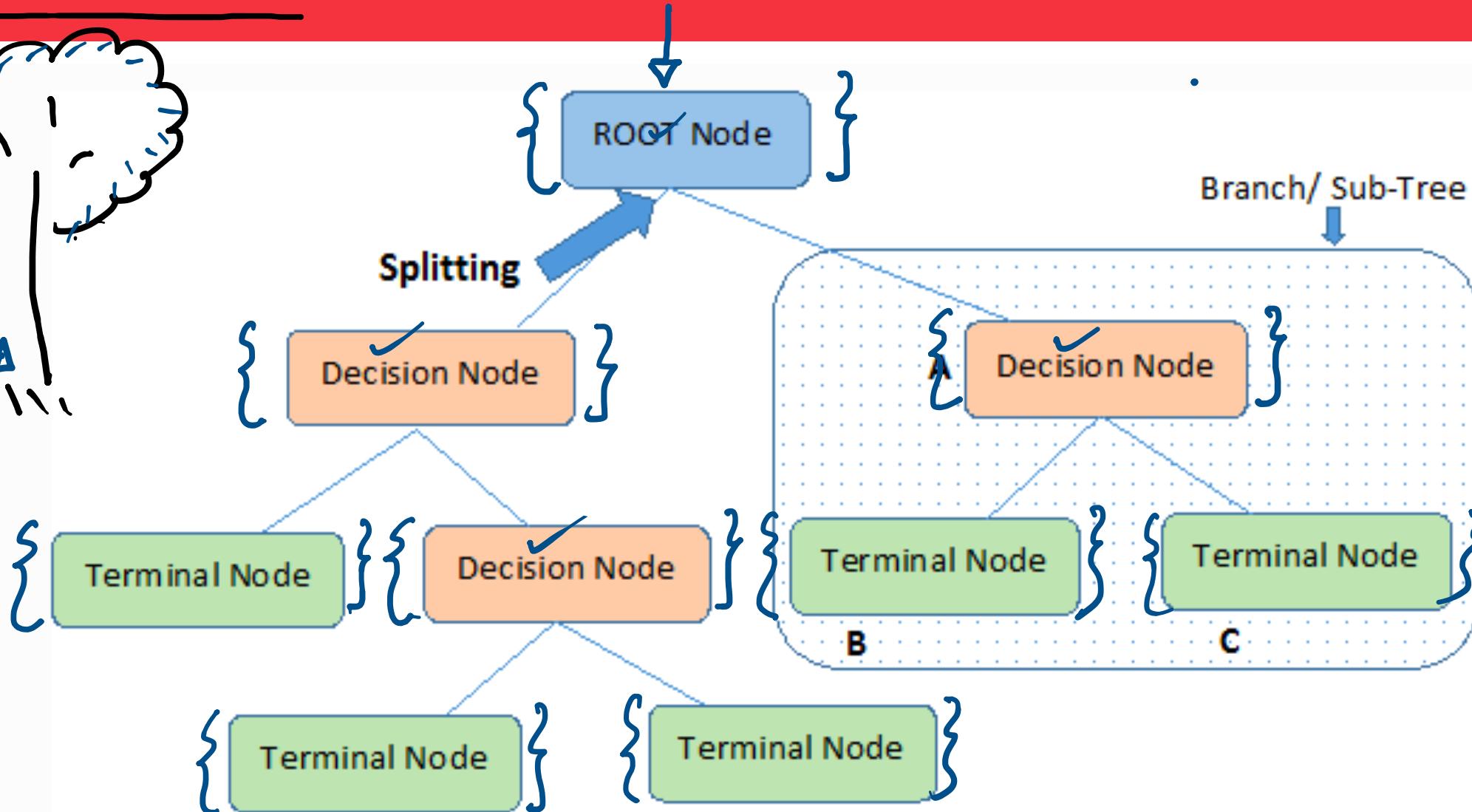
Architecture, Read/Interpret &
Confound

{ Why Tree Models?:



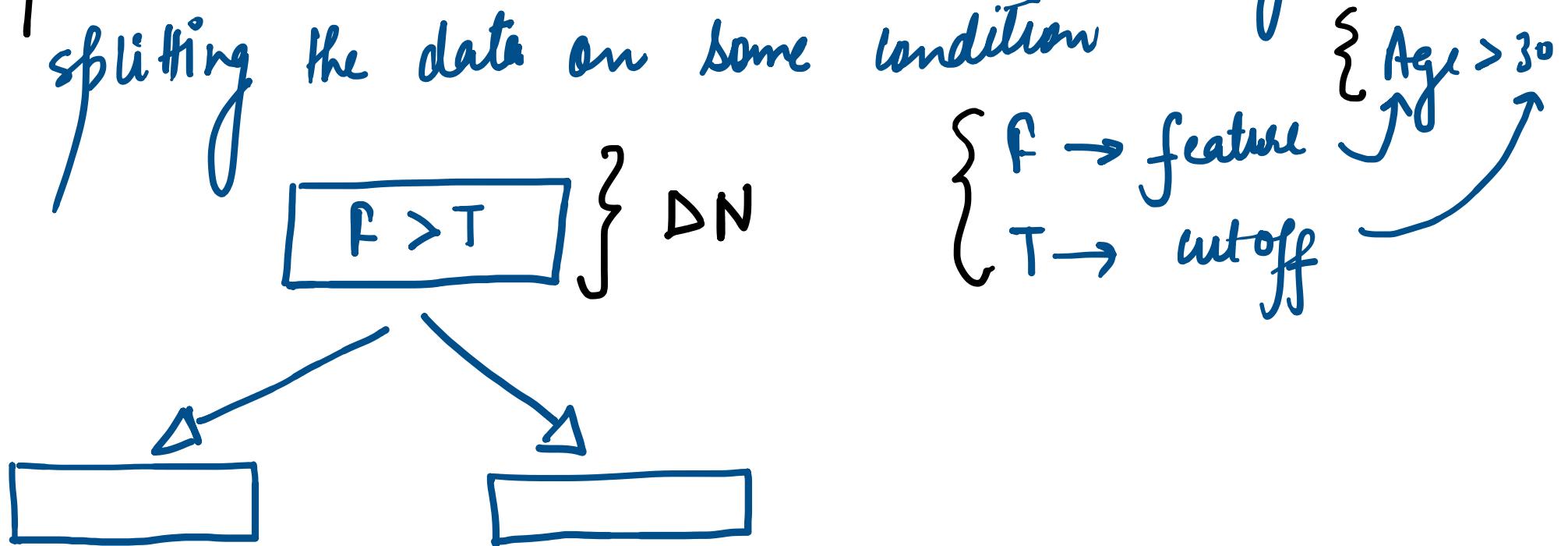
Decision Tree: looks like a tree used to take a decision

upGrad



Note:- A is parent node of B and C.

→ Decision Node :- These are nodes where tree is taking some decision by splitting the data on some condition

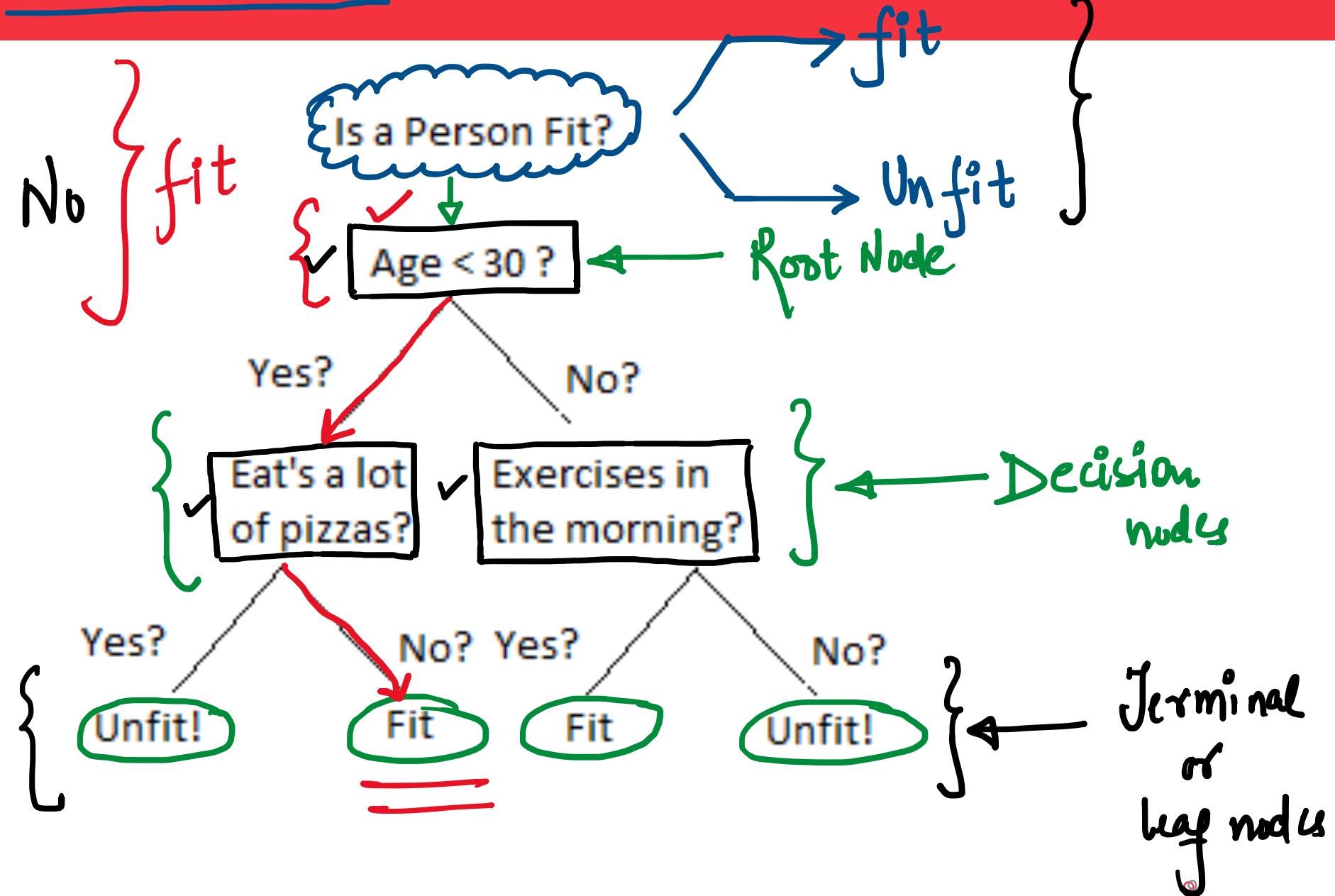


→ Root Node :- Also, a decision node, it is very decision node.

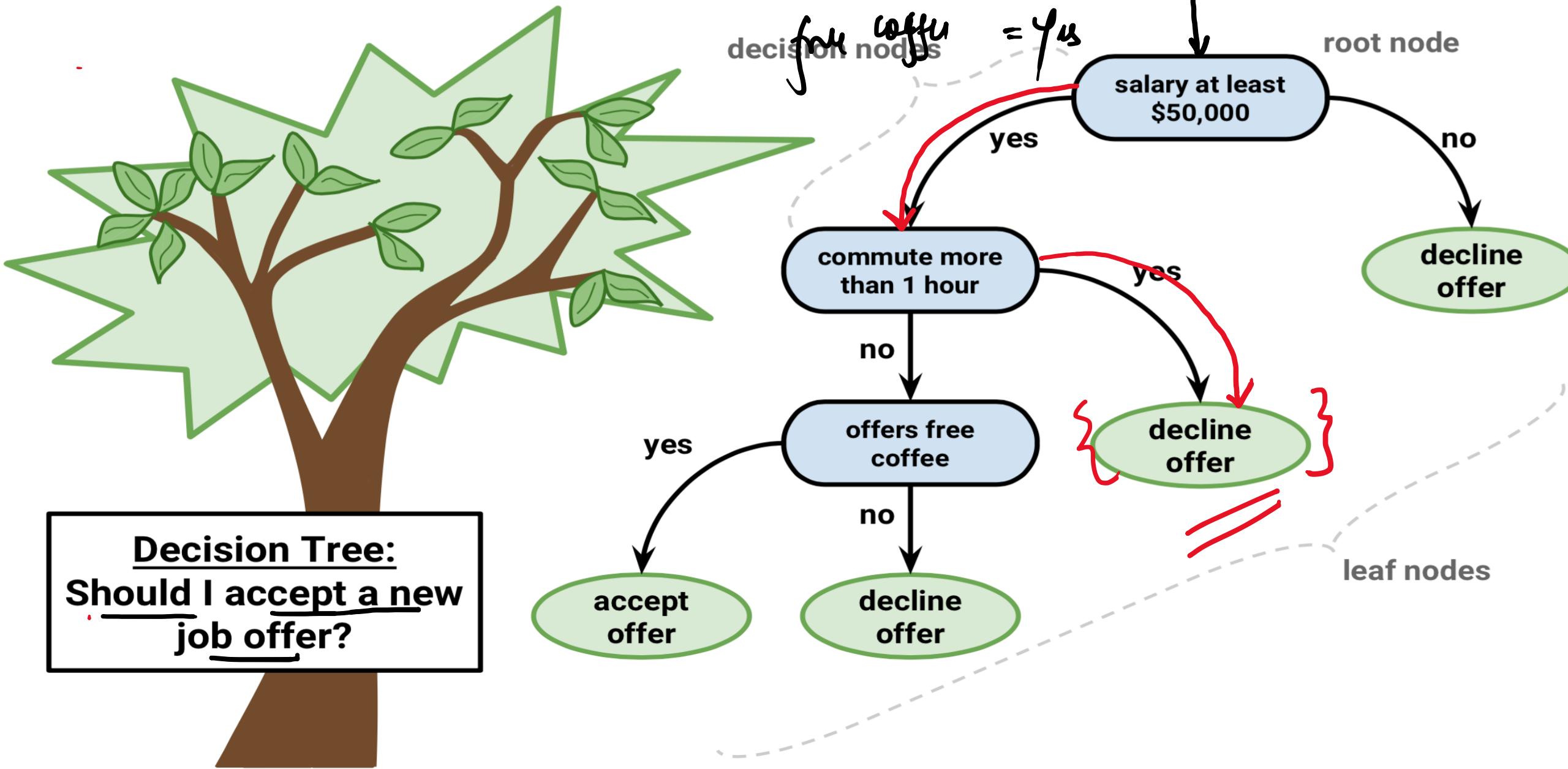
→ Terminal node :- A node where tree terminates. Also,
contains the o/p from the tree. (leaf nodes)

Interpretation of Decision Tree:

✓ Age = 25
 ✓ Eat lot of pizzas = No
 ✓ Exercise = Yes

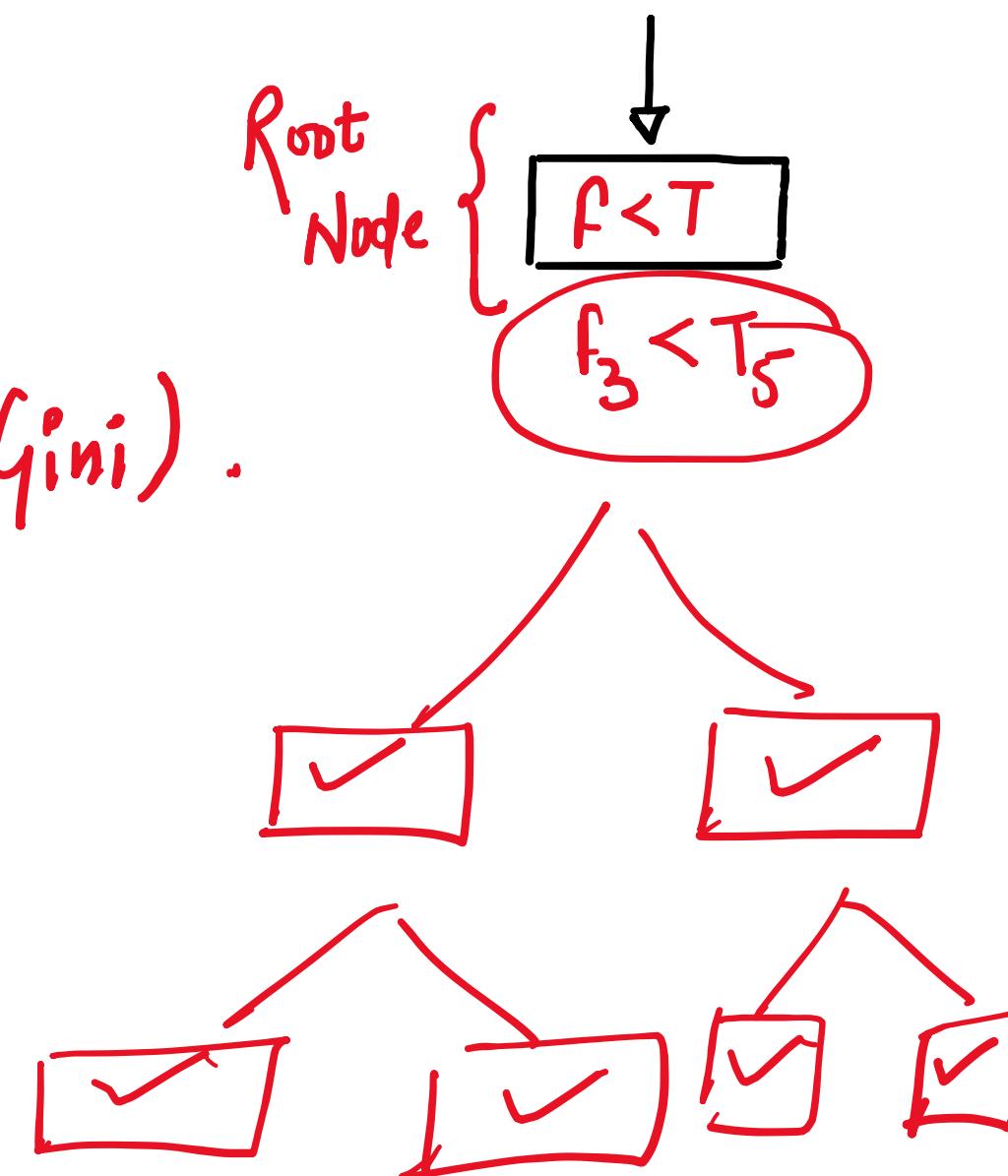


Decision Tree Classification:

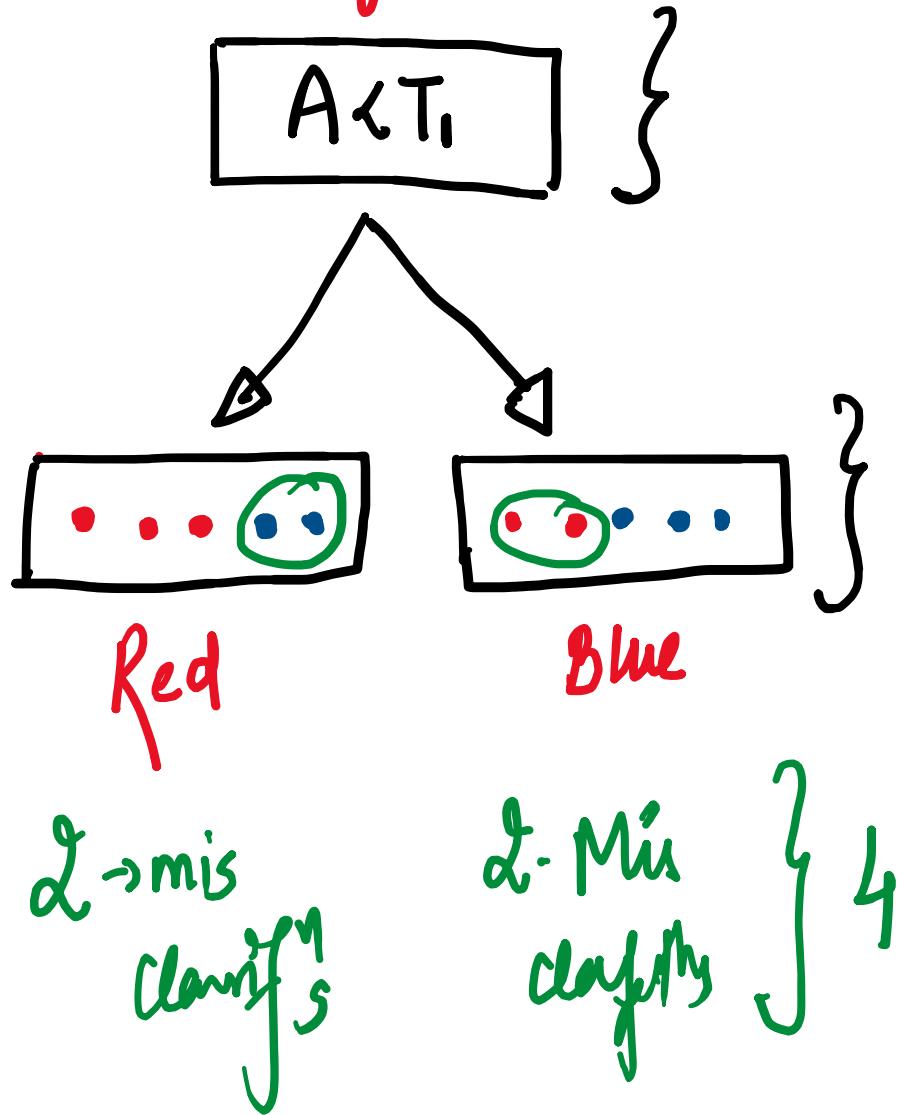


→ Decision Tree Algorithm :-

- * feature & threshold (cutoff) is being the concept of gini (min Gini).
- { * Stop at some stopping criterias



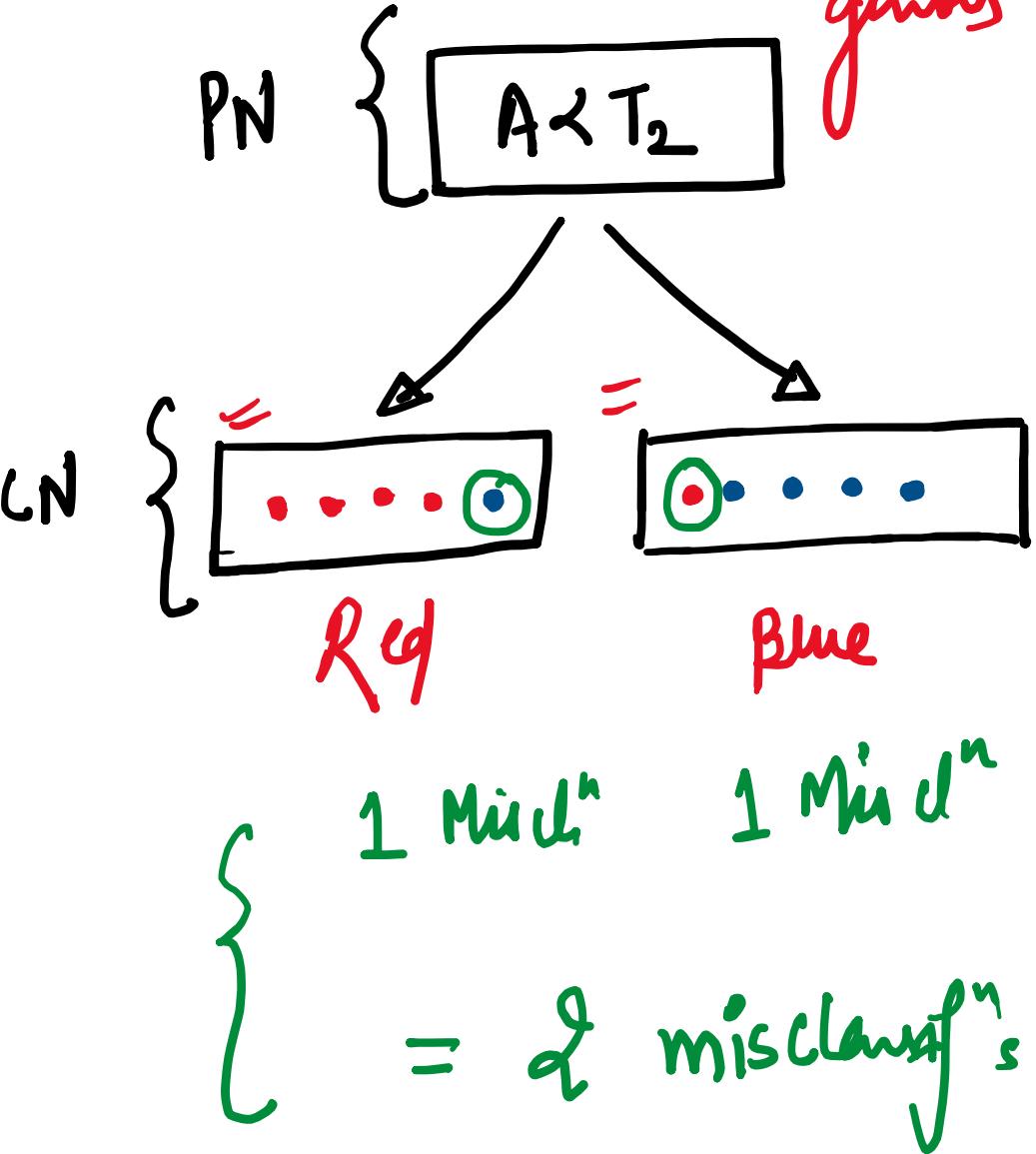
less homogeneous



10 data points

5 B
5 R

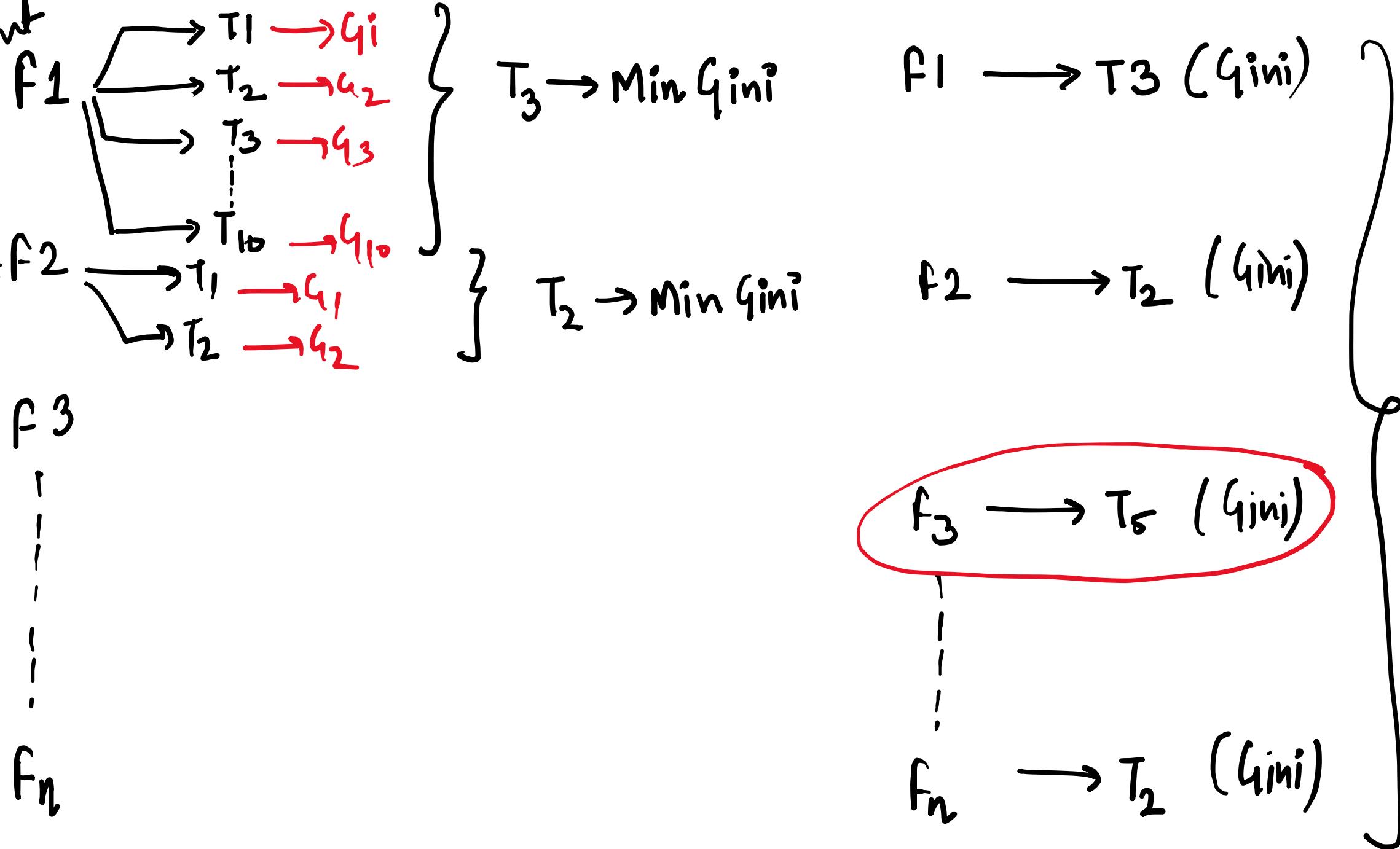
better More homo
genous



Gini : It is the statistical measure of this misclassification

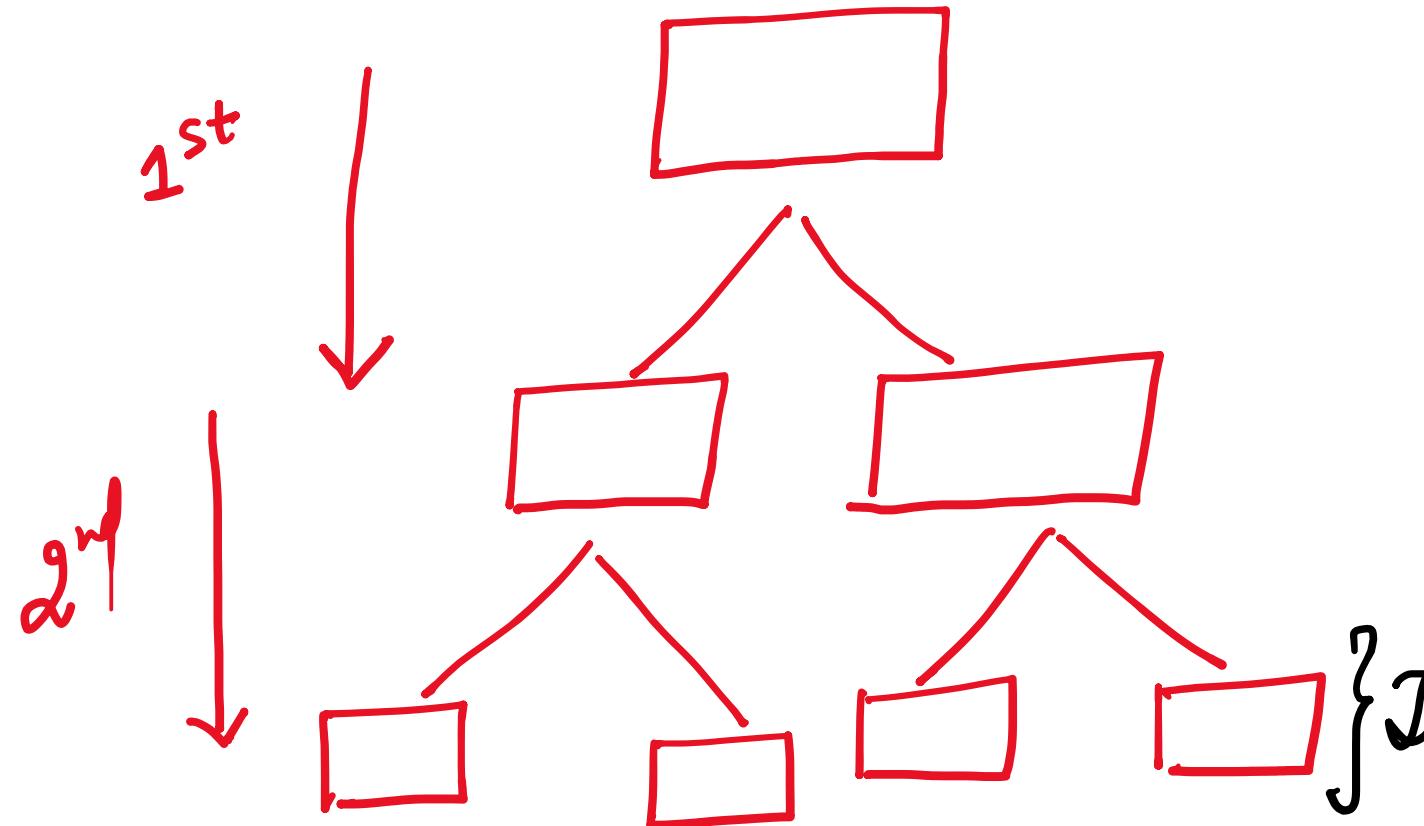
Misclassifⁿ ↑
↓ Misclassifⁿ

→ categorical
* $f = c_1$
* $f = c_2$
* $f = c_3$
→ continuous Age
10% 21 → T_1
20% 23 → T_2
30% 27 → T_3
100% 63 → T_n



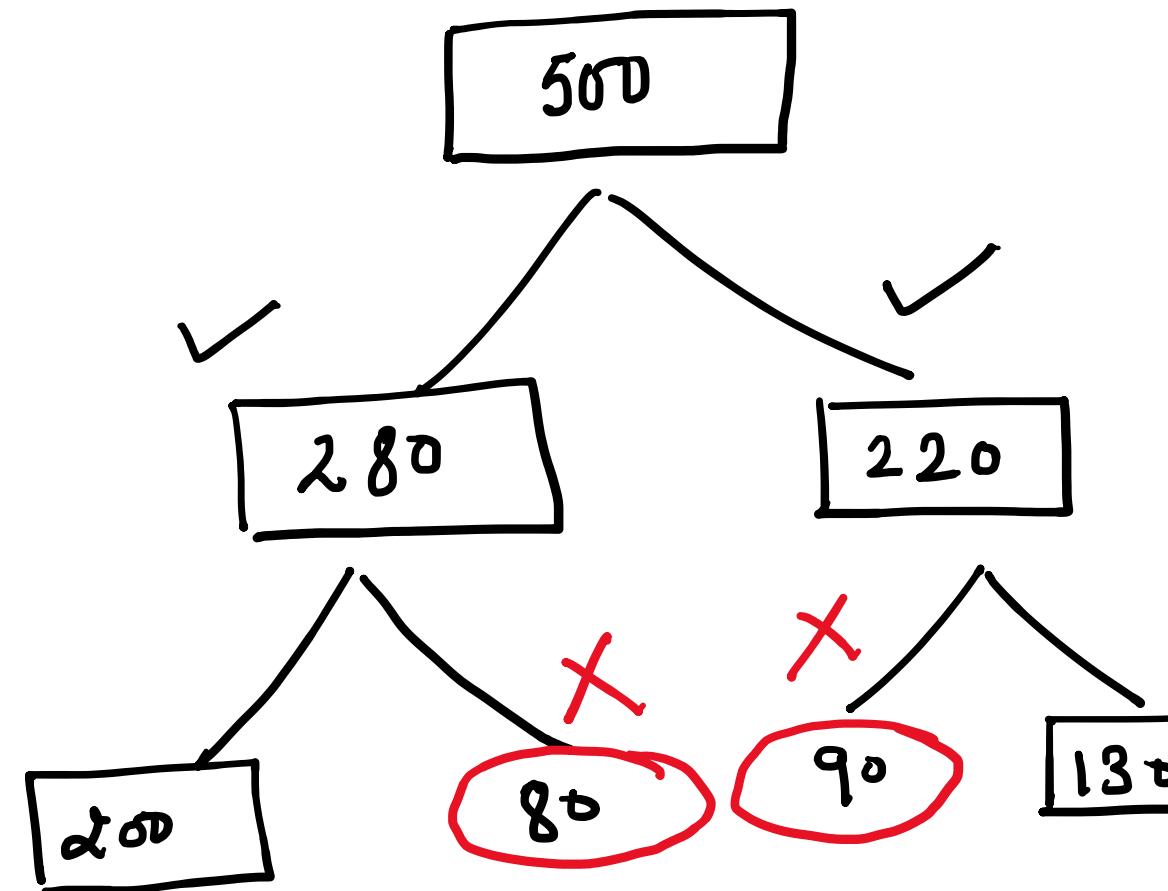
→ Hyperparameters :-

① max_depth :- 2



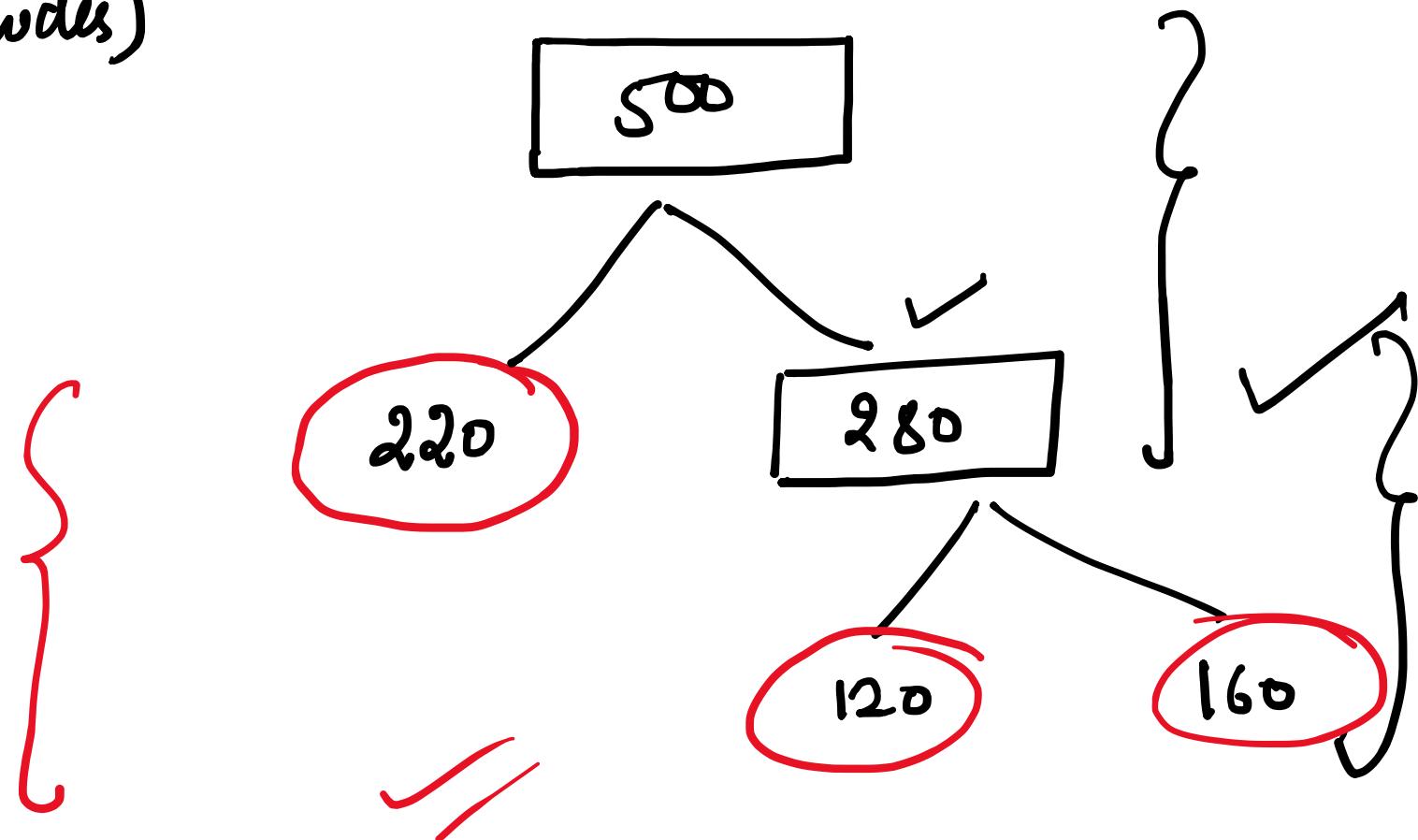
② min_sample_split :- Minimum no. of samples required to make the split.

$$\text{min-sample split} = 100$$



③ min_sample_leaf: Minimum no. of sample required at leaf
after split (child nodes)

min_sample_leaf = 100



Recall that the Gini index is calculated as follows:

- $$G = \sum_{i=1}^k p_i(1 - p_i) = 1 - \sum_{i=1}^k p_i^2$$

where p_i is the probability of finding a point with the label i , and k is the number of classes.