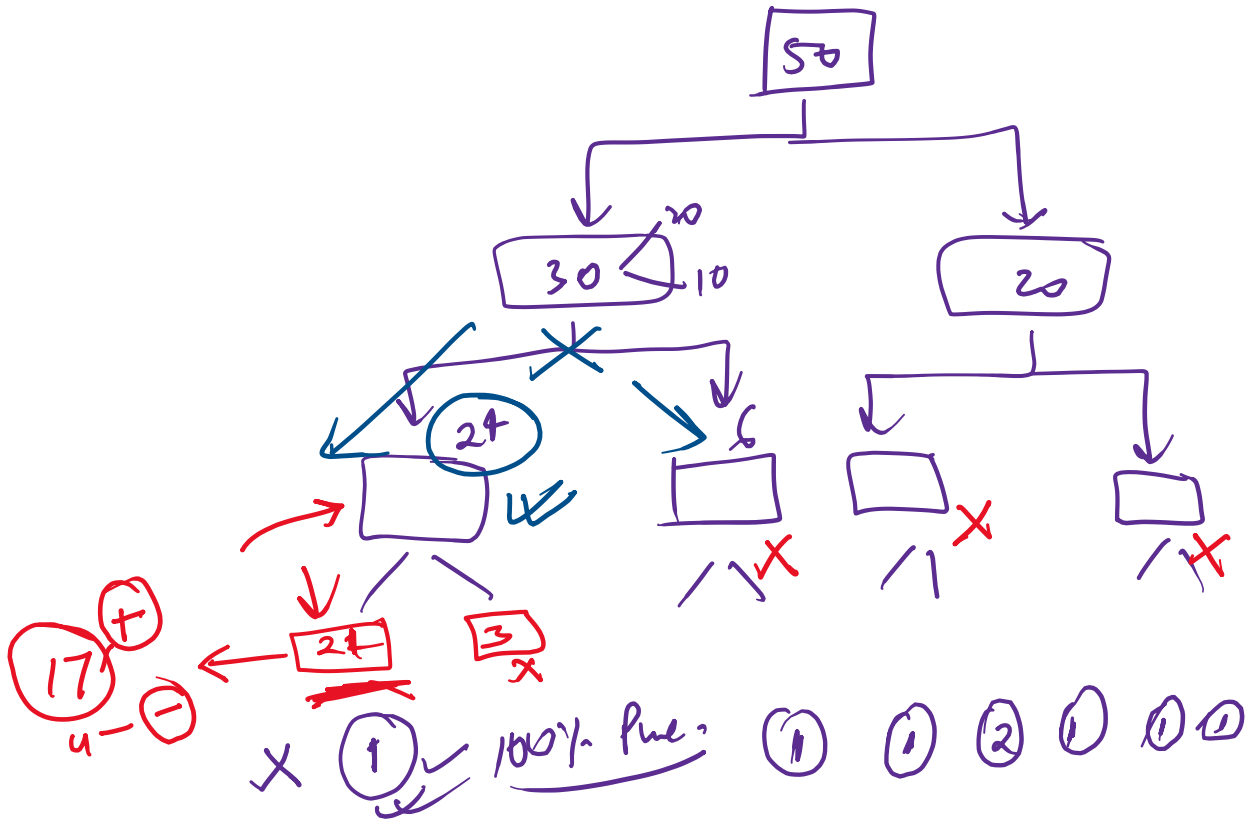
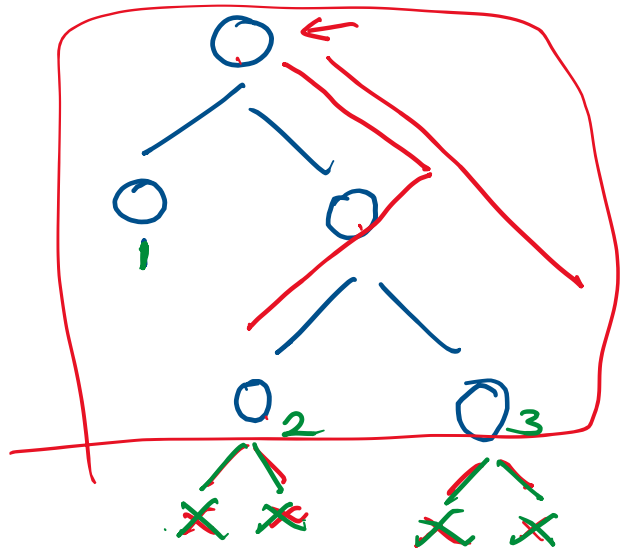


## setting constraint on Tree



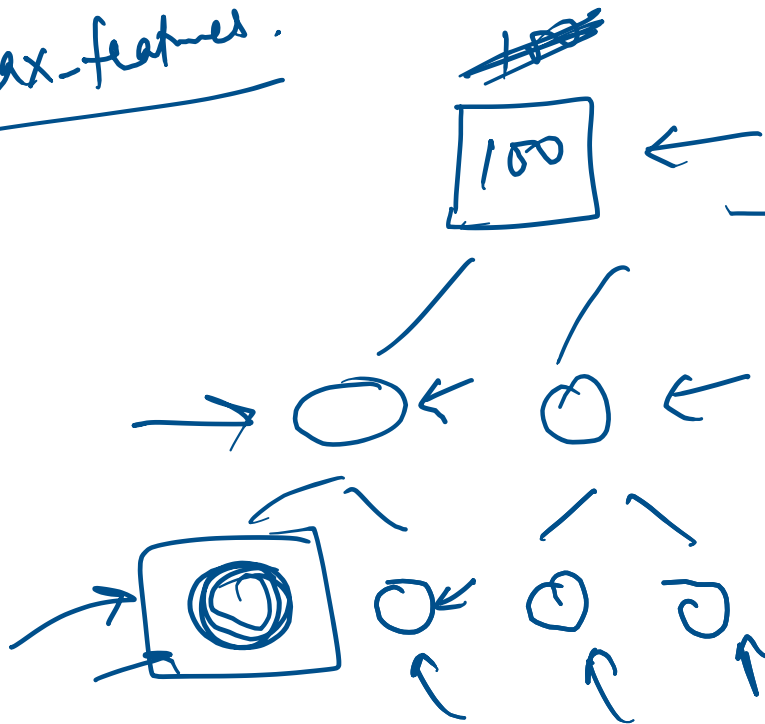
- ① Min<sup>m</sup> samples for a node to split.  $\Rightarrow$  20
- ② Min samples for a terminal node.  $\Rightarrow$  10
- ③ Max<sup>m</sup> depth = 3



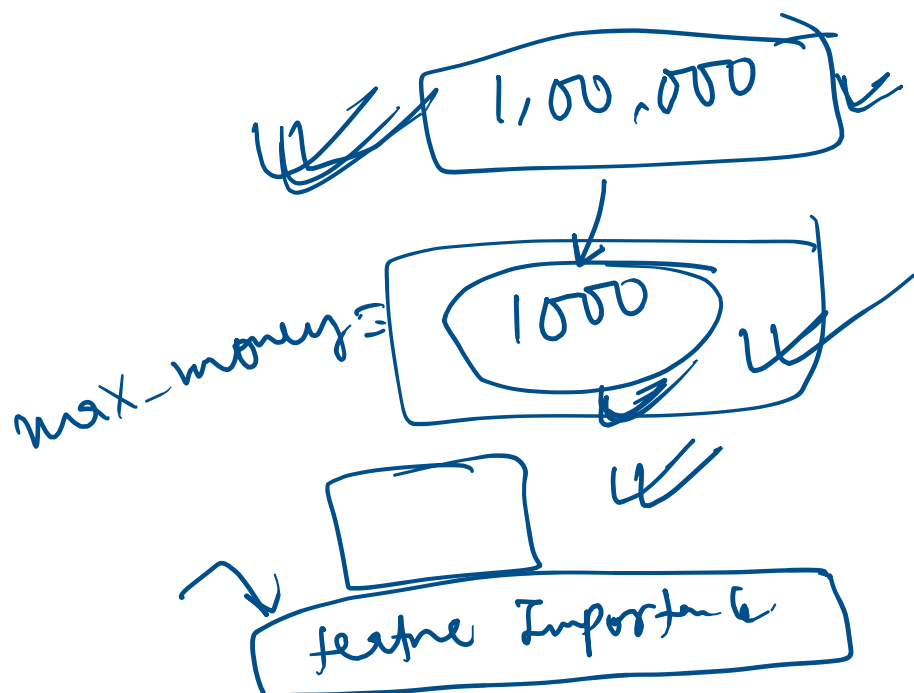
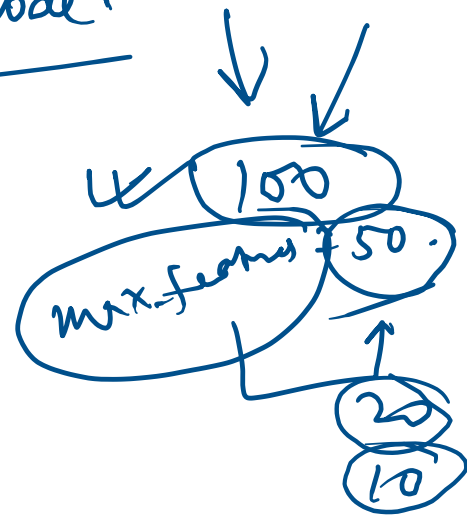
④  $\text{Max}^m$  no. of terminal nodes. = 5

⑤  $\text{Max}^m$  features to consider for split.  
=  $\sqrt{n}$ .

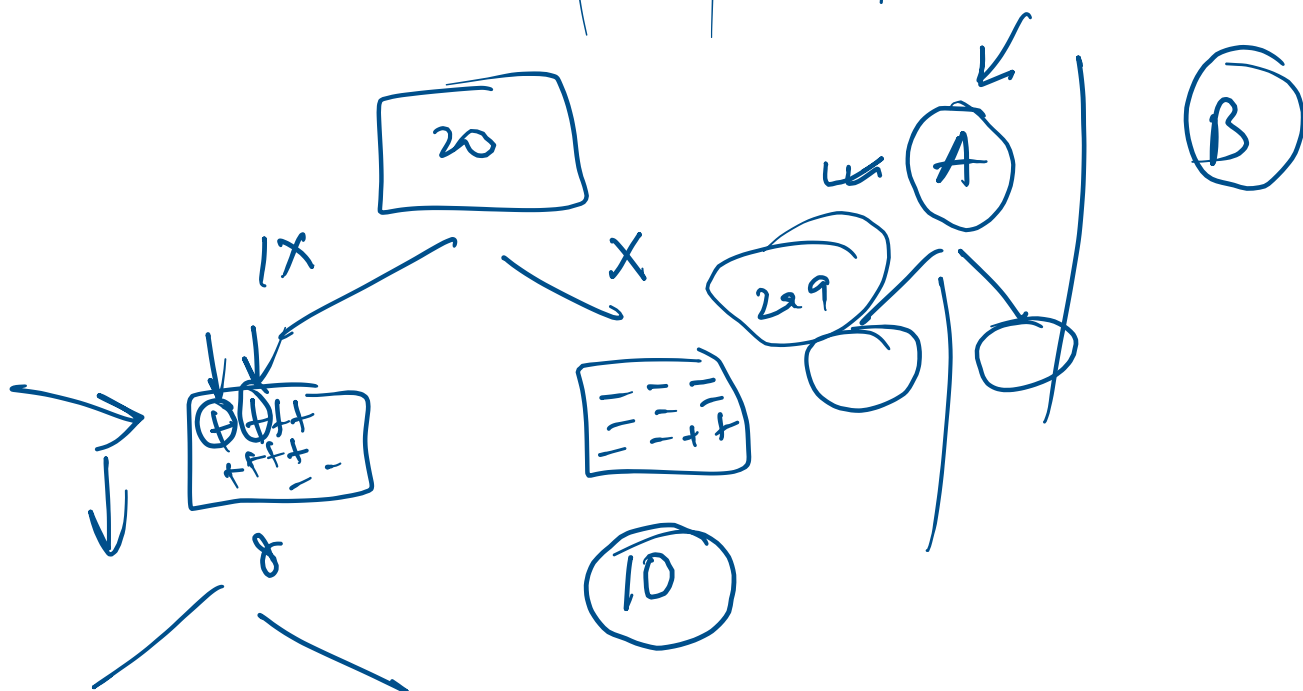
max-features.



Root node.



↓	↓	↓				
Age	Class	Height				
✓	↓	✓				



Real life  $\rightarrow$  2 3 5

$\xleftrightarrow{30}$  50

$n \rightarrow 10$   
 $\hookrightarrow 20 \checkmark$   
 $\hookrightarrow 30 \checkmark$   
 $\sqrt{n}$   
 log n ✓

Recall ✓

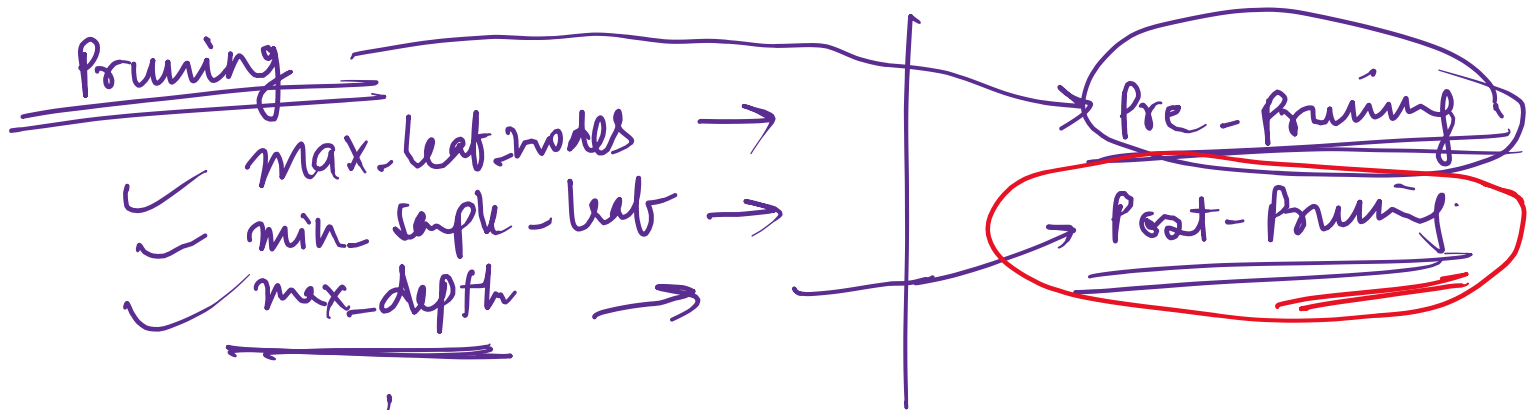
Precision

F1 score

$\hookrightarrow$  overfit  $\leftarrow$   
 $\hookrightarrow$  underfit  $\leftarrow$

Accur ↑  
AUC of ROC

Misc ↓



## Bias & variance (Tradeoff)

Bias : diff bet<sup>n</sup> avg. Prediction of our model & the correct value which it's trying to predict.

Bias ↑

Model is paying little attention to data.  
(oversimplify)

variance : variability of the prediction.  
(spread of our dataset)

variance ↑

pays a lot of attention to training data.  
score diff (Tr. ↑ Te. ↓)

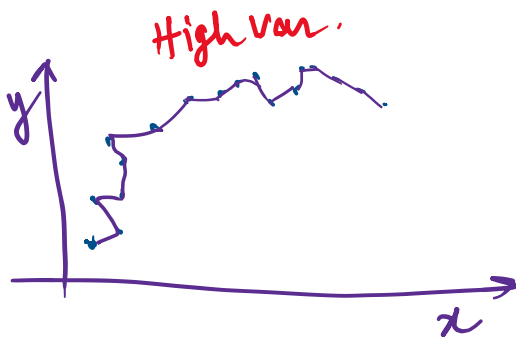
Mathematically

$$\text{Err} = \text{Bias}^2 + \text{Variance} + (\text{Irr. Error})$$

### underfitting

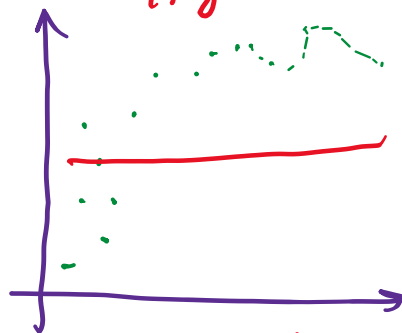
\* Model is unable to capture underlying pattern of data.

Bias  $\uparrow$  var  $\downarrow$



overfitting

High Bias



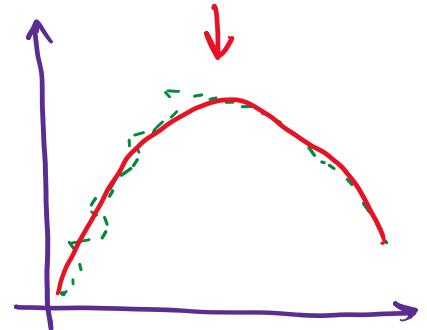
underfitting

### overfitting

\* when our model captures the noise along with the data.

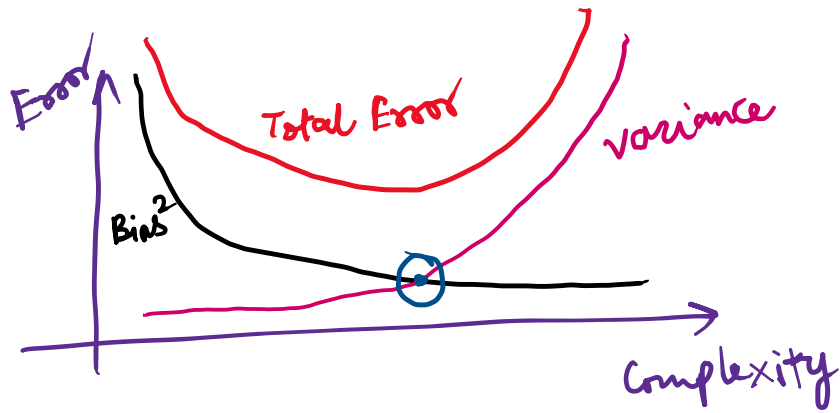
Bias  $\downarrow$  var  $\uparrow$

low var., low Bias



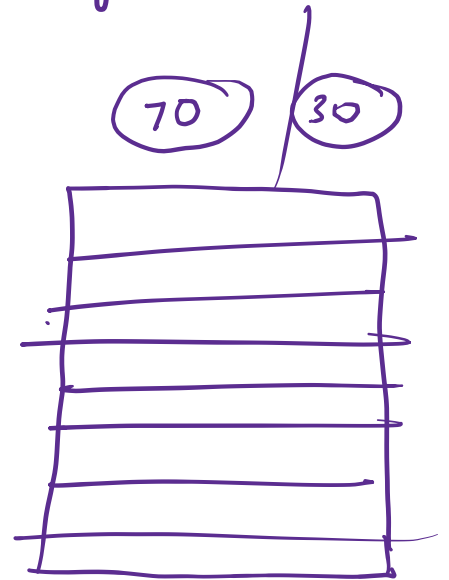
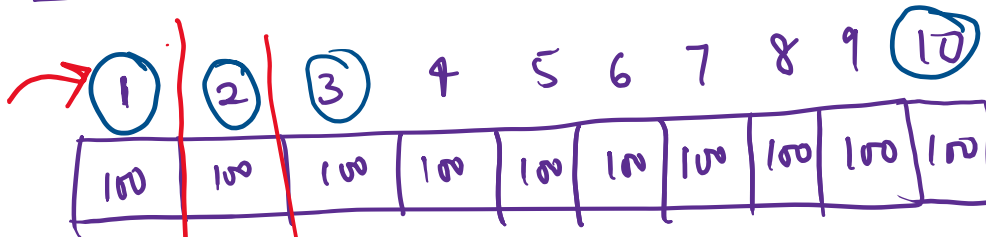
Balance  
(Bias - Var.)

# Bias Variance Tradeoff.



## cross validation

### ✓ (K-fold cross validation)



✓ 9

	Test	Trn
①	1	All the
②	2	1, 3, 4, 5...10
3	3	1, 2, 4, 5...10
4		

10 ←

1st Test ⇒ Training

	9 Tr	200
200	200	200
200	200	200
200	200	200

200 200 200

K = 10

5

←

out of  $k$  subsets

- one of the subset → test set
- $(k-1)$  → Training Set.

Iteration =  $K$   
Trails

$$\frac{\text{operation}}{\text{Trails}} = K$$
 Error estimation is averaged over all Trails.

Result  $\rightarrow$  (Reduces Bias) /  
reduce variance!

[illegible]