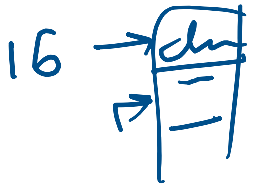
Ensemble Model.

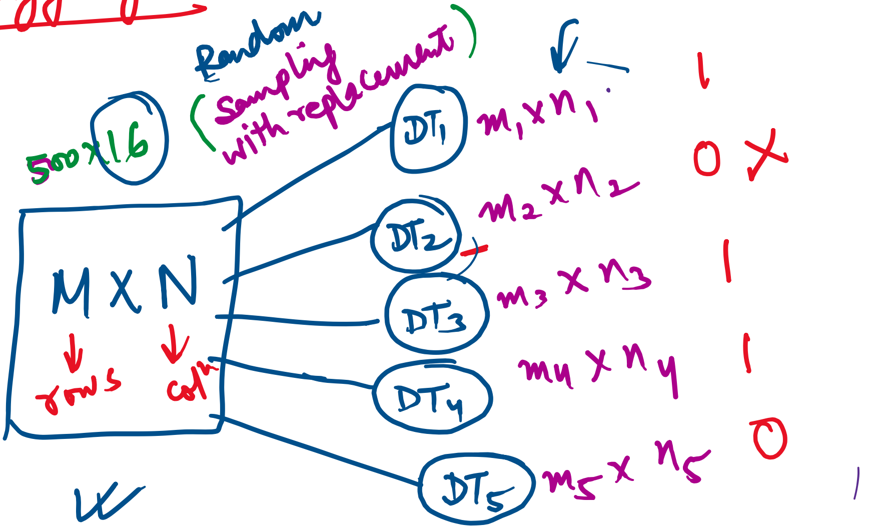
D.T. (Collection/Aggregation)

Bagging (bootstrap aggregation). (Parallel)

Boosting (Sequentially).



① Bagging. (Random Forest).



$m_i \leq M$

$n_i \leq N$

✓ $m_i \Rightarrow \left(\frac{2}{3}\right) M$ ✓

✓ $n_i \Rightarrow \sqrt{N}$

Random forest.

→ forest is made of Trees & R.F. is made of D.T.

→ Ensemble model, Trained using Bagging technique.

→ <u>Adv.</u> – Combination of learning models.

DT : Searches for most imp. feature while splitting a node.

RF : Searches for most imp. feature among a random subset of feature.

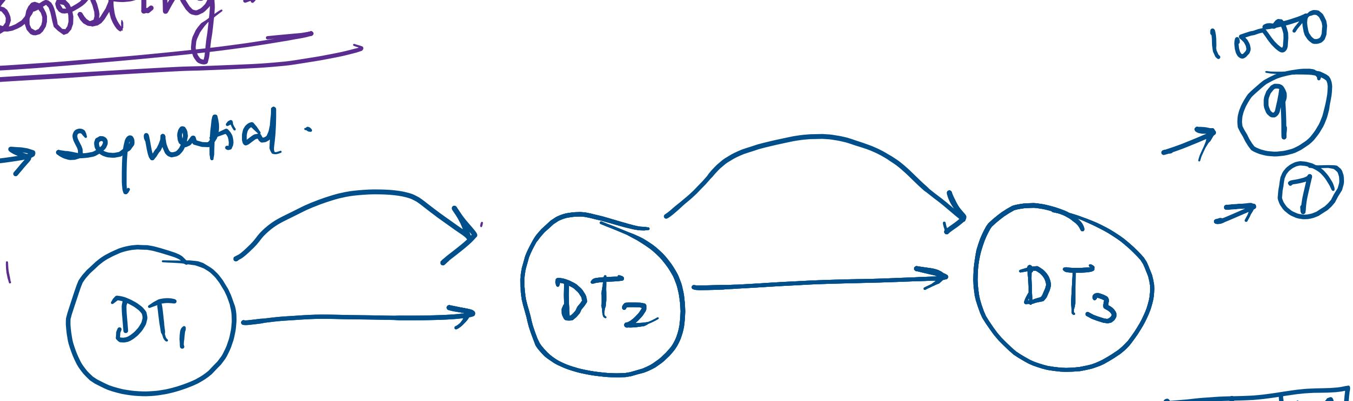
⇓

(High diversity.)

Most imp. features of DT is not included (by chance) in DT_3, then also the split will happen. (based on best available feature).

② Boosting.

→ sequential.



weak learners :

Strong learners :

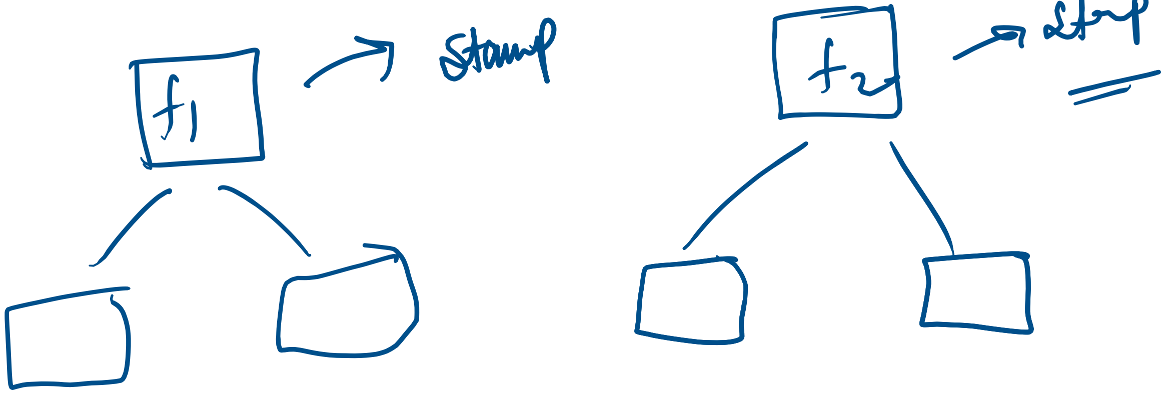
			y	y_Pr
			0	0
			1	0
			1	1
			0	1
			0	0
			1	1

Step 1 : Initialize equal weights.

	x_1	x_2	x_3	x_4	x_5	y	o/p	weights	
1							yes	1/10	→ $\frac{1}{10}e^{-x}$
2							no	1/10	$\frac{1}{10}e^{x}$
3							yes	|	
4							no	|	
5							yes	|	
6							yes		
7							no		
8							yes	1/10	
10									

$$\frac{1}{n} = \boxed{\frac{1}{10}} \checkmark$$

Step 2 : create Base learners in Sequential fashion..



Lesser Entropy : High info gain.

9 1

e_t = sum of all the sample weights of that error row $= \frac{1}{10}$

Step 3 : Performance of stamp.

$$= \frac{1}{2} \log_e \left(\frac{1 - e_t}{e_t} \right)$$

$$= \boxed{x} \longleftarrow$$

Step 4 : ↑ weights of incorrect classification (weak learners)

↓ weight of correct classified record (strong learners)

New sample weight (of incorrect classified row) $= \text{old weight} \times e^{\text{Perf. st.}}$

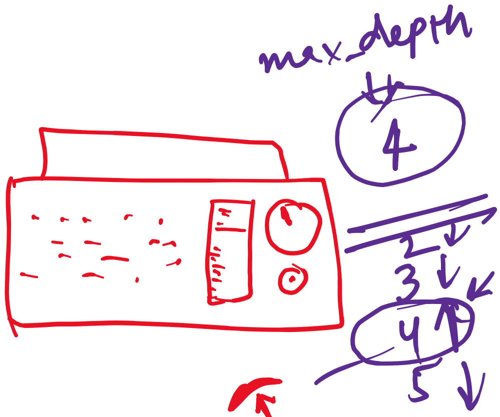
$$= \frac{1}{10} \times \boxed{e^{x}}$$

New sample wt. (strong lear) $= \text{old wt.} \times e^{-\text{Perf. sn}}$

$$= \frac{1}{10} \times e^{-x}$$

(Hyperparameter Tuning) in RF.

★ Increasing the Predictive Power.



① n_estimators

→ More Trees

✓ ↑ Perf.

✓ ↓ Computational Power

✓ ↓ speed.

② max_features :

③ min_sample_leaf :

④ Max_depth.

② Model's Speed.

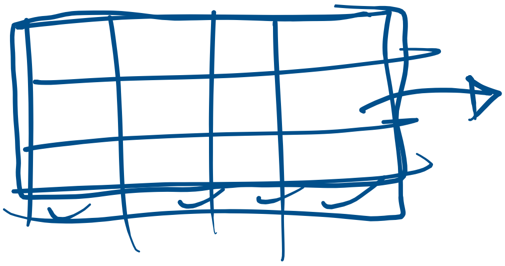
① n_jobs :

1 → one Processor

-1 → no limit.

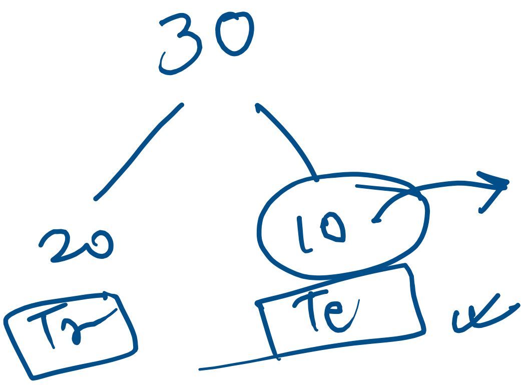
(2) oob_score

out of bag :-

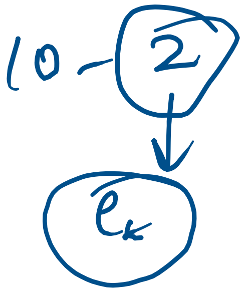


~~60%~~

70%



oob_error_rate.



Grid Search CV

→ member of model_selection.

→ Fits to the hyperparameters & fit the estimator on

max_depth = 3, 4, 5, 8, 10

= 10, 12, 15, 18

= 3, 8, 9

EOF >>