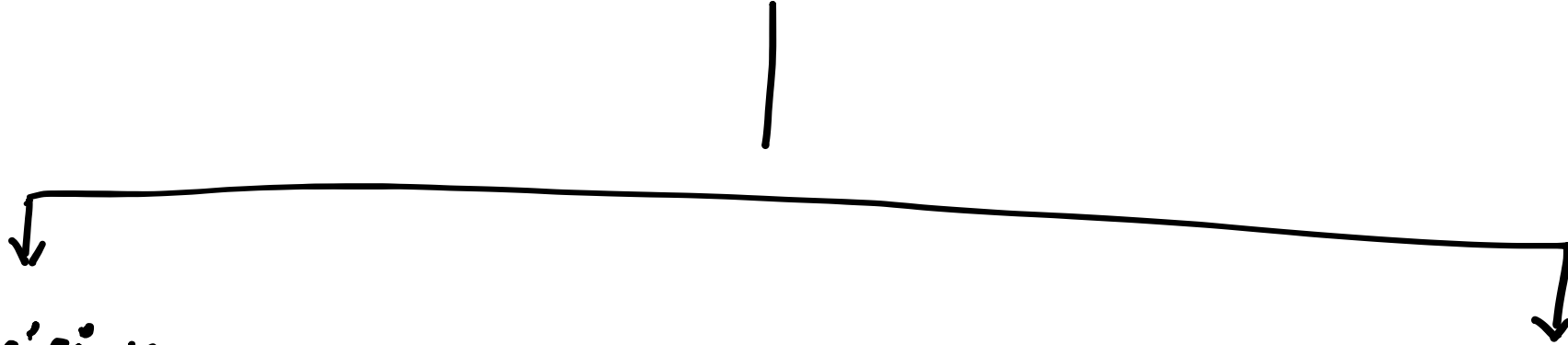


TREE MODELS



Decision
Tree

* Highly Unstable

Ensemble

* We combine the o/p
of various D.T to
get the final o/p.

Ensemble

n trees

Bagging

- * Parallel Approach
- * Independent from each other

n trees

Boosting

- * Sequential Approach
- * Trees are dependent upon each other

* Random forest Algorithm :- (Bagging Algorithm)

↓
Collection of trees

→ All those trees are constructed on randomly
selected data.

65 Random Splits with replacement

70

Data
(100) ✓

(30)

(35)

S1

S2

S3

S4

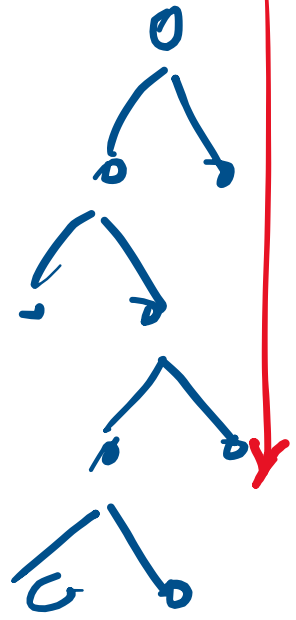
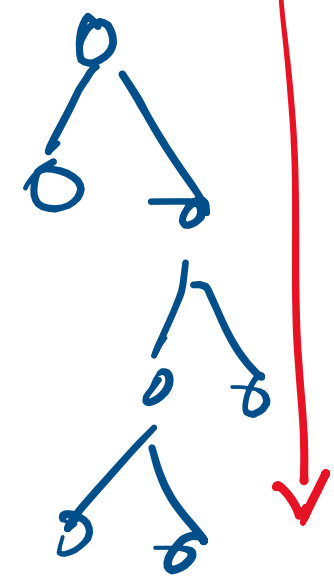
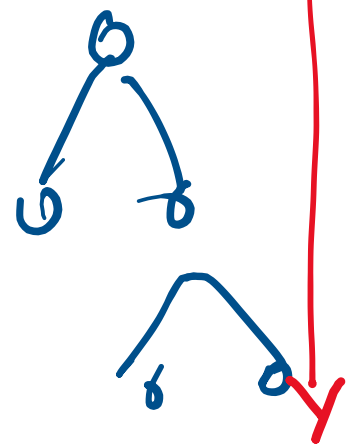
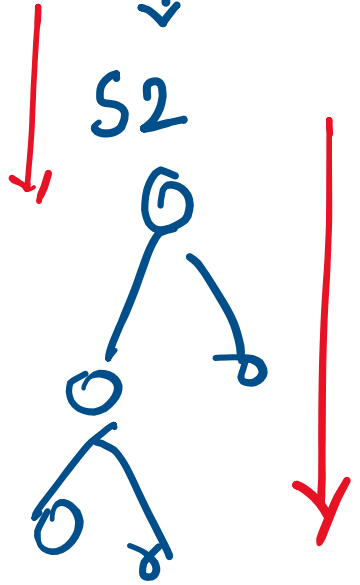
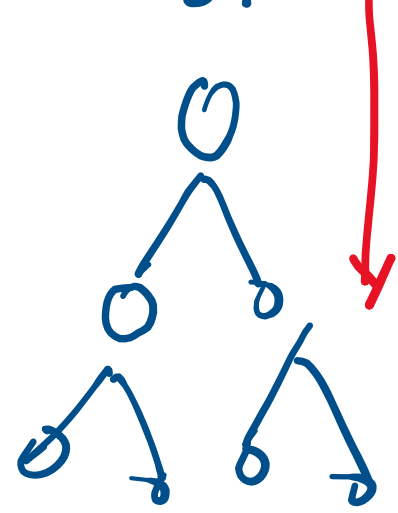
S_n

(100) \swarrow 70-1 ✓
 \searrow 30-0 ✓

n subsets

↓
n trees

final o/p



Voting / Mean

↓
final prediction

Counting

$T_1 \rightarrow 1$

$T_2 \rightarrow 1$

5 trees } \rightarrow ① $T_3 \rightarrow 0$

predict

$T_4 \rightarrow 1$

$T_5 \rightarrow 0$

$0 \rightarrow 2$

$1 \rightarrow 3$

$T_1 \rightarrow 20000$

$T_2 \rightarrow 30000$

$T_3 \rightarrow 25000$

$T_4 \rightarrow 27000$

$T_5 \rightarrow 23000$

Average

25000

final o/p

Bagging \rightarrow No Need of test data

\swarrow OOB

X {out of bag}

\searrow In industry (we keep the test data)

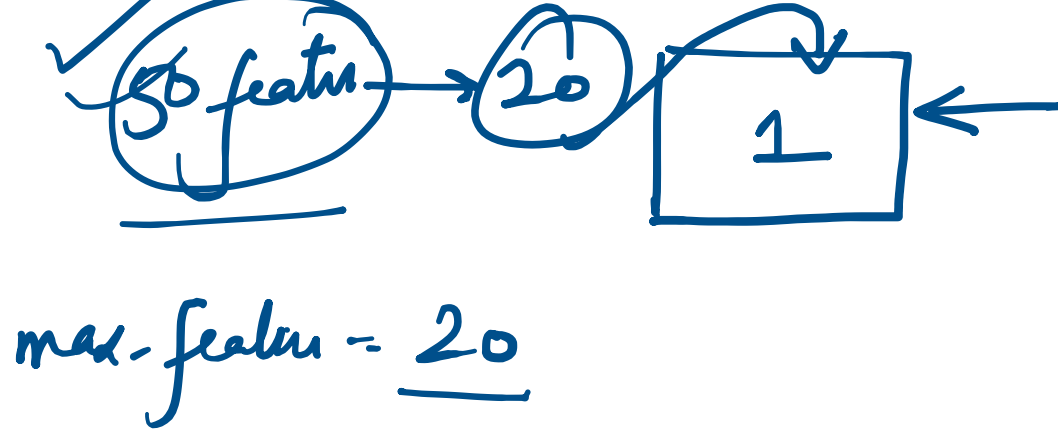
* Working of Random forest

* ✓ OOB {interview}

	<u>D11</u>	<u>D12</u> ✓	<u>D13</u>	<u>D14</u>	<u>D15</u>	Pred	Actual
①	✓	X	✓	✓	X	1	1
2	X	✓	X	✓	✓	0	0
3	✓	✓	X	X	X	0	1
4	✓	✓	X	✓	X	1	1
5	X	X	X	X	✓	0	1
6	✓	✓	✓	✓	X	1	1
7	X	✓	✓	✓	X	0	0
8	✓	X	✓	✓	✓	0	0
9	✓	✓	X	✓	✓	0	1
10	✓	X	✓	X	✓	1	0

→ Hyper parameters :-

- ① max-depth
- ② min-samples-split
- ③ min-samples-leaf
- ④ max-features
- ⑤ n-estimators } ✓



1 tree → 5 nodes → 50
250

100 tree 20 x 100
2500