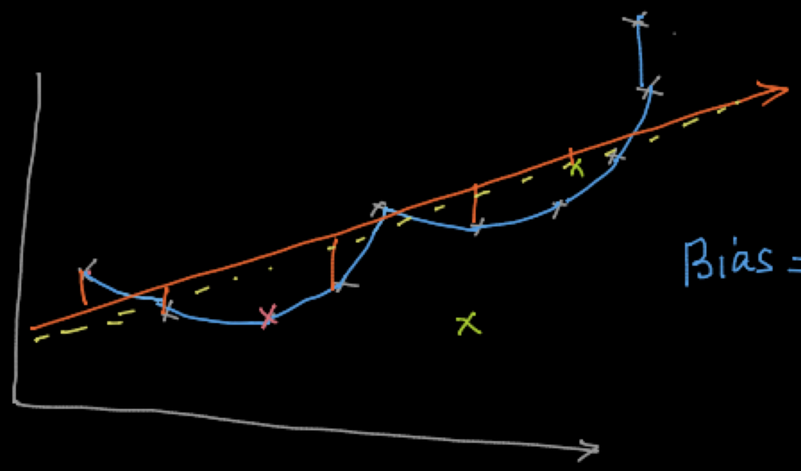
← Ensemble Techniques →.

Training Error → Bias



Bias = 0.

Testing Error → Variance. ?

→

x_1 x_2 x_3 y

80%

1
2
3
.
.
.
n

80%
Randomly Selection.

Overfitting.

Low Bias, Low Variance

Ideal outcome-

↓ 98%. ↑ 97-%. → Target.

Ensemble → Group / collection

— Multiple Models.
— Take the result from all the models.

Base Models → M_1, M_2, M_3

Combining multiple Models → Random Forest →

Homogenous models:
1. Bagging → Random Forest — Techniques.
2. Boosting → Ada Boost — / XG Boost / LGBM / Cat Boost

Heterogenous:
3. Stacking.
4. Cascading.

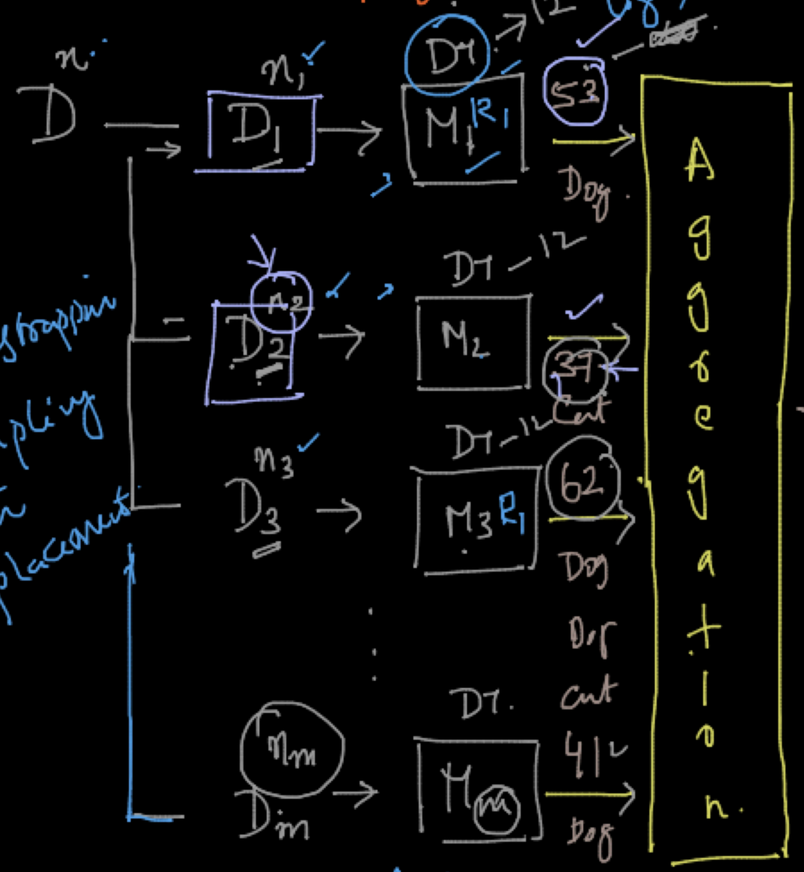
UNSEEN Data.

→ (1300. 3Bad 12 km)

Bagging → Algorithm (fit & fs)

↳ Bootstrapped ← ↳ Aggregation

Row sampling.



Bootstrapping - Sampling with replacement

→ Homogeneous

x_1	x_2	x_3	y
1			
2			
3			
.			
.			
$n = 1000$			

Regression (Average / Median) → 1 o/p.

Classification (Majority Vote).

Aggregation → Reduces Variance

Base Models → Low Bias & High Variance.

DT → Build the tree - Conditionable to a depth 30

Low Bias & Low Variance Model

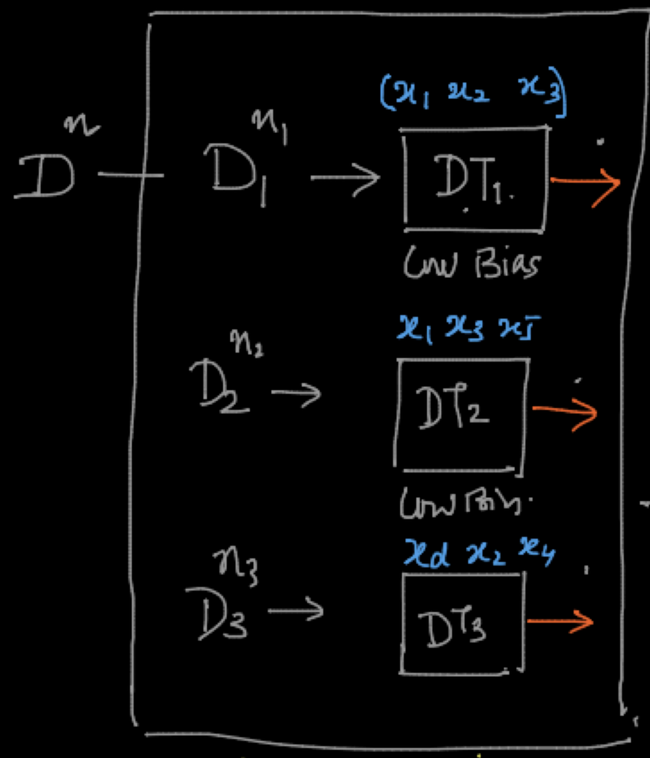
D^n → Ensemble Model → Y = 1 o/p.

→ Low Bias & Low Variance

← RandomForest Algorithm → model = RandomForest( )
→ { How many n. depth.

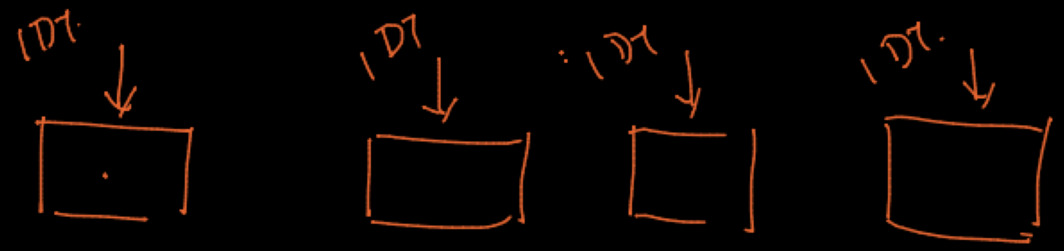
Parallelism → Decision Trees. ✓ →

→ Column / Feature Sampling. 3.



x_1	x_2	x_3	...	x_d	y
1					
2					
3					
⋮					
n					

All Base Models are independently built ✓



⇒ 4 DT ⇒ 4 Trees.

$$\begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 5 & 7 & 8 \end{bmatrix} \qquad \begin{bmatrix} 1 & 2 & 3 \\ 6 & 7 & 8 \\ 8 & 9 & 10 \end{bmatrix}$$

matrix matrices.

Accuracy
Prec
Rec
} Metrics. (measurements)

Boosting → 'Reduces Bias'
- Additive Model.
- Sequential Model

$y = f(x)$.
↳ Model.

	x_1	x_2	x_3	y	$\hat{y}$	ε_0 ($y - \hat{y}$)
1				–	–	
2						
3						
⋮						
n						

Stage 0: Predict y.

M_0: $y = f_0(x)$; $\varepsilon_0 = y - f_0(x)$.

$\{x_i, y_i\}_{i=1}^{n}$ → Predict 'y'

Stage 1:

$\{x_i, \varepsilon_0\}_{i=1}^{n}$ ⇒

Dep.

x_1	x_2	x_3	ε_0	$\hat{\varepsilon}_0$	Residuals from Stage 1: $\varepsilon_1 = \varepsilon_0 - \hat{\varepsilon}_0$
.	.	.	–	–	

$\varepsilon_0 = f_1(x)$.

$y - f_0(x) = f_1(x)$

$(y) = f_0(x) + f_1(x)$

$\varepsilon_0 = y - f_0(x) + f_1(x)$

Stage 2.

$y = f_0(x) + f_1(x) + f_2(x)$

x_1	x_2	x_3	ε_1	$\hat{\varepsilon}_1$	ε_2

→ $\varepsilon_1 = f_2(x)$.

Stage 3:

x_1	x_2	x_3	ε_2	$\hat{\varepsilon}_2$	ε_3

M_0 $\boxed{y = f_0(x)} \Rightarrow \varepsilon_0 = y - f_0(x)$.

$y = 60$ L.

$M_0 \Rightarrow \hat{y} = 48$ L

High Bias, Low Variance

↓

→ [ ] → Low Variance, Low Bias

$M_1 \rightarrow \boxed{\varepsilon_0 = f_1(x)} \Rightarrow \varepsilon_1 = \varepsilon_0 - f_1(x)$

M_2 $\boxed{\varepsilon_1 = f_2(x)}$

DT → Decision Stump

→ [ ] [ ] Depth = 1

High Bias

→ [ ] Low Bias

$$y = f_0(x) + f_1(x) + f_2(x) + \cdots f_m(x)$$

$$y = \alpha_0 f_0(x) + \alpha_1 f_1(x) + \alpha_2 f_2(x) + \cdots \alpha_m f_m(x)$$

→ Additive / Sequential Model

Bias: ↓

Boosting → High Bias + Low Variance Models.

Reduces Bias. ← → Decision Stumps (only one split).

$$y = [f_0(x) + f_1(x) + f_2(x) + f_3(x) \dots]$$

$M_0 = f_0(x)$

y	$\hat{y}$	$\varepsilon_0 = y - \hat{y}$	$\hat{\varepsilon}_0$ ($f_1(x)$)	ε_1	$\hat{\varepsilon}_1$ ($f_2(x)$)	ε_2	$\hat{\varepsilon}_2$ ($f_3(x)$)
50	35	15	5	10	2	8	3

$y =$ 50 L → 49.9 Lakh

45 ↑

42 ↑ ⇒

40 ↑

35

$y =$

Bias ↓ Variance ↓

2 →

$(\hat{y}\ \hat{y}) \Rightarrow$ Ensembling → $M_1\ M_2\ M_3$

$M_1 \rightarrow$

{ Low Bias ←. Low Variance ← } ↓

Bagging → Reduce Variance → High Variance, Low Bias.

Boosting → Reduces Bias → High Bias, Low Variance ↓

Ada Boost
— Adaptive Boosting