Boosting ⇒ Ada boosting
Gradient boosting
XGB Boosting

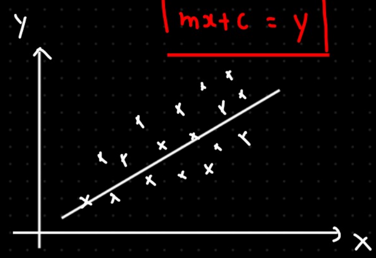
Classification & Regression

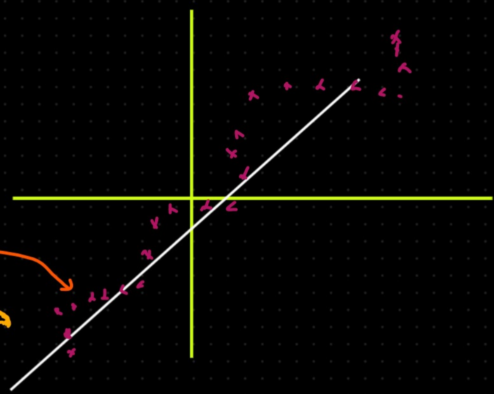
DT

$$\boxed{\text{Gradient boosting} = f_0(x) + f_1(x) + f_2(x) + f_3(x) + \dots + f_n(x)}$$

X | Y

Agenda ⇒ $f(x) = y$
⇓
linear reg
$\boxed{mx + c = y}$





= Polynomial
⇓

Additive
⇓
fn

$\{ f(x) = x + \sin(x) \}$

Final answer

Adaboosting v/s Gradient Boosting

$\alpha_1 \alpha_2 \alpha_3 \dots \alpha_n$

① Ada boosting ⇒ $\alpha_1 M_1 + \alpha_2 M_2 + \alpha_3 M_3 + \dots + \alpha_n M_n$

Gradient boosting ⇒ $\text{Base\_model} + \alpha M_1 + \alpha M_2 + \alpha M_3 + \dots \alpha M_n$

Regression → avg value

α = Learning rate
{α = hyperparameter}

② Ada - Boosting => max_Depth = 1 => stump

= Gradient Descent => { min_leaf_node = 8, max_leaf_node = 32 } 8-32

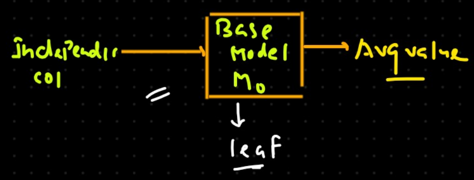
$GD = M_0 + M_1 + M_2 + M_3 \ldots$

M_0 = Base Model = avg value of output column = leaf

$\frac{3+4+8+6+3}{5} = 4.8$

Act · Pred · Psudo res

IQ	CGPA	Salary	Pred1	Res1 (Act-Pred1)	
90	5	3	4.8	-1.8	
100	7	4	4.8	-0.8	
110	6	8	4.8	3.2	
120	9	6	4.8	1.2	
80	5	3	4.8	-1.8	



Avg. val

Boosting => Ada boosting => Random sampling => assigned weight

next sample → misclassified D.P. ← weight

DT = error

error

Performance

Loss

M_0 + M_1

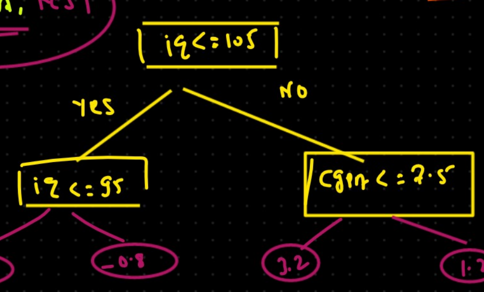
Base Model (Avg) · DT

{ Residual }

{ IQ, CGPA, Res1 }

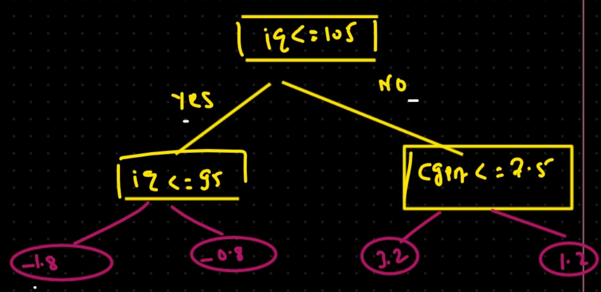
Score => training score => how well it is fitting my training

DT (M_1) => IQ, CGPA, Res1 =>



How well it is fitting my TD.

Iq	CGPA	Salary	Pred1	Res1 (Act-Pred1)	Pred2
90	5	3	4.8	-1.8	-1.8
100	7	4	4.8	-0.8	-0.8
110	6	8	4.8	3.2	3.2
120	9	6	4.8	1.2	1.2
80	5	3	4.8	-1.8	-1.8



getting overfitted

$M_0 + \alpha M_1$

Prediction $= M_0 + M_1$

$= 4.8 + (-1.8)$

Prediction $= 3$

$= 4.8 + (-0.8)$

$= 4$

$= 4.8 + 3.2 = 8$

$= 4.8 + 1.2 = 6$

$= 4.8 - 1.8 = 3$

fp	Act
3	3
4	4
8	8
6	6
3	3

$\Rightarrow M_0 + \alpha M_1$ → Learning Rate

$\alpha = 0.1$

	$(M_0 + \alpha M_1)$	$(M_0 + M_1)$
3	4.62	3
4	4.72	4
6	5.12	6
8	4.92	8
3	4.62	3

$\Rightarrow 4.8 + (0.1 \times (-1.8))$

$= 4.8 - 0.18$

$= 4.62$

$\Rightarrow 4.8 + (0.1 \times (-0.8)) = 4.8 - 0.08 = 4.72$

$\Rightarrow 4.8 + (0.1 \times 3.2) = 5.12$

$= 4.8 + (0.1 \times 1.2) = 4.8 + 0.12 = 4.92$

$= 4.8 + (0.1 \times (-1.8)) = 4.62$

Initially $\Rightarrow (M_0 + M_1)$ Overfitting

$\{ M_0 + \alpha M_1 + \alpha M_2 + \alpha M_3 \Rightarrow$ well fitted?

M_0 ⇒ base model ⇒ leaf ⇒ avg value

$M_0 \rightarrow M_1 \rightarrow M_2$

M_2 ⇒ iq, cgpa, Res_2

$$Res\ 2 = Actual - [M_0 + \alpha M_1]$$

⇒ train my model

iq	cgpa	Salary (Lpa) (Act)	$Pred_1(M_0)$ ($Pred_1$, BM)	Res_1 (Act-$Pred_1$)	$Pred_2$ (DT1, M_1)	$Resd_2$ = (Act - $[M_0 + \alpha M_1]$)
90	8	3	4.8	-1.8	-1.8	-1.62
100	7	4	4.8	-0.8	-0.8	-0.72
110	6	8	4.8	3.2	3.2	2.88
120	9	6	4.8	1.2	1.2	1.08
80	5	3	4.8	-1.8	-1.8	-1.62

over-fitting

✗ final_Pred (without alpha)

$M_0 + M_1$

$$\begin{bmatrix} 3 \\ 4 \\ 8 \\ 6 \\ 3 \end{bmatrix}$$

⇒ Slowly and gradually we reaching toward the target col

– with alpha

$[M_0 + \alpha M_1]$ α ⇒ LR (HP)

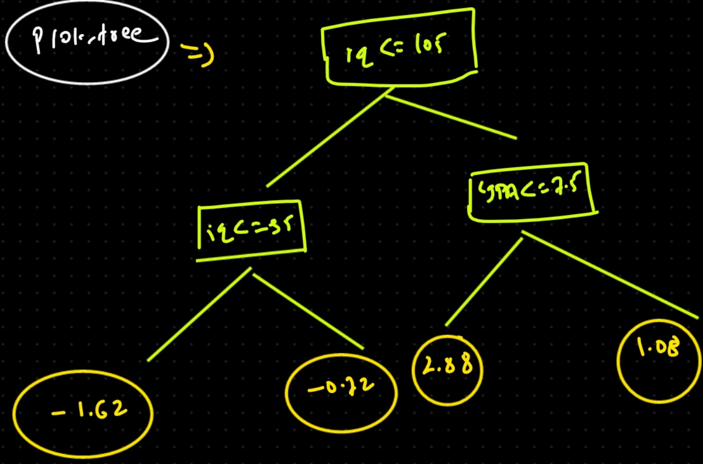
$$\begin{bmatrix} 4.62 \\ 4.72 \\ 5.12 \\ 4.92 \\ 4.62 \end{bmatrix}$$

M_2 ⇒ DT_2 ⇒ {iq, cgpa, Res 2}

IQ	CGPA	Res2	Pred3
90	8	-1.62	-1.62
100	7	-0.72	-0.72
110	6	2.88	2.88
120	9	1.08	1.08
80	5	-1.62	-1.62

$M_2 \Rightarrow DT_2$

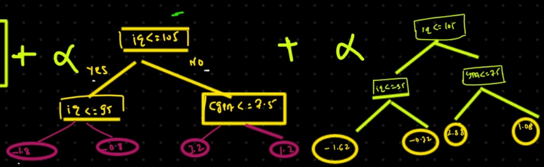
plot_tree $\Rightarrow$



$$Pred = M_0 + M_1 + M_2$$

$$Pred = M_0 + \alpha M_1 + \alpha M_2$$

{60, 5}

$= \boxed{\text{Base model}} + \alpha$ 

$$= 4.8 + 0.1 \times (-1.8) + 0.1 \times (-1.62)$$

$$= 4.8 - 0.18 - 0.162$$

$$= 4.8 - 0.342$$

$$= 4.458$$

$\{60, 5\} = 4.5$ LPA

Resd 3 $\quad M_0 + \alpha M_2 + \alpha M_2 \xrightarrow{} F \quad =$ Act - F

DT_3