x_1	x_2	x_3	y	R_B	R_1	R_2

$R_B = y - \hat{y}$ – from Base Model
$R_1 \to \hat{y}$ of DT_1 | $R_2 \to \hat{y}$ of DT_2
$DT \to \{x, R\}$ | $DT_1 = \{x, R_B\}$

3 # GRADIENT BOOSTING

$\{x_i, y\}$

1.→ A Base Model is trained to give an Constant value as a output

↳ either avg value or a value which minimizes the Error (Reg = MSE, class = Log Loss).

$$\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y})^2 \quad \Big| \quad -\left(y\log\hat{y} + (1-y)\log(1-\hat{y})\right)$$

$$\hat{y} = \frac{1}{1+e^{-(mx+b)}}$$

↳ R_B is calculated (Residuals from Base Model)

2.→ Construct a DT → $\{x_i, R_B\}$ ... $\{x_i, R_n\}$
└ output → Residual

R_1 → output of DT_1 | R_2 → output of DT_2 ...

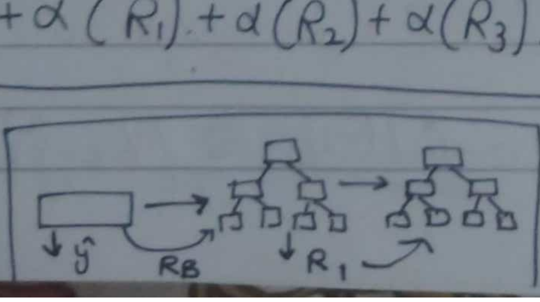
3.→ $\hat{y}$ → output from Base Model → minimize the Error
R_B → Residuals from Base Model $(y - \hat{y})$
R_1 → output of DT_1 $(\{x_i, R_B\})$
R_2 → output of DT_2 $(\{x_i, R_1\})$

↳ every DT learns from Residual of the previous DT

Final answer = $\hat{y} + \alpha(R_1) + \alpha(R_2)$ ....
└ learning Rate – Scalling to avoid overfitting

$\boxed{\hat{y} + \alpha(R_1) + \alpha(R_2) + \alpha(R_3) \ldots}$

★ [R_B is calculated, to fit input as to the 1^{st} DT] ★



* → How split happens → Base split → Har feat Ka hota hai – IG Compare → Choose Best → Level 1 split
Repeat — $\boxed{\text{All other feat} - \text{Base split feat}}$ ← IMP

$$F(x) = h_0(x) + \alpha_1 h_1(x) + \alpha_2 h_2(x) + \dots + \alpha_n (h_n x).$$

y — Output

└ Final ~~output~~

$$\rightarrow \boxed{\sum_{i=1}^{n} \alpha_i h_i(x)}$$

[The Residual value keeps getting smaller]

— × — × — × —

IMP → Pseudo [I/p → $\{x_i, y_i\}$, $L(y,\hat{y})$, No. of Trees]

(1) Initiate Base Model with Constant value.

$$F_0(x) = \underset{\gamma}{\arg\min}\left(\sum_{i=1}^{n} L(y,\gamma)\right) \quad [\gamma = \hat{y}]$$

└ Constant value → The Base Model will give → R_B calc → R_B = Input for DT_1 → R_1 → R_1 Input for DT_2 → R_3 → Final ans = $\hat{y} + \alpha R_1 + \alpha R_2 + \alpha R_3$.

— o — o — o — o — o — o —

└ For Reg → MSE and Class → log loss [Should be diff]

→ Let's say For Reg → $\frac{1}{n}\sum_{i=1}^{n}(y-\hat{y})^2$ (For $F_0(x)$ → loss should be min].

How → for y value → (3 rec)

$$\rightarrow \frac{1}{3}(50-\hat{y})^2 + \frac{1}{3}(70-\hat{y})^2 + \frac{1}{3}(60-\hat{y})^2$$

⇓ First order der

$$\frac{2}{3}(50-\hat{y})(-1) + \frac{2}{3}(70-\hat{y})(-1) + \frac{2}{3}(60-\hat{y})(-1)$$

$\rightarrow \frac{2}{3}\left[3\hat{y} - 180\right] + 2\left[\hat{y} - 60\right] = \boxed{\hat{y} = 60}$

The Model will give 60 as the output and then R_B is calc.
Base

$\boxed{2}$ Iterate $M = 1$ to M (No. of Tress).

$\boxed{2.1} \rightarrow R_B$ = Compute pseudo Residuals

$\hookrightarrow \gamma_{im} = -\left[\frac{\partial L(y, f(x_i))}{\partial f(x_i)}\right] \rightarrow -\frac{\partial L}{\partial \hat{y}}$ [Der of L wrt to $\hat{y}$]

eg $\rightarrow \frac{1}{2}(y-\hat{y})^2 = L$

$\frac{\partial L}{\partial \hat{y}} = -(y-\hat{y})$

$\left[-\frac{\partial L}{\partial \hat{y}} = y - \hat{y}\right]$

R_B

$\gamma_{11} = a$
$\gamma_{21} = b$
$\gamma_{31} = c$

$\boxed{2.2} \rightarrow$ fit the DT, with i/p $\{x_i, \gamma_{im}\}$.

$\boxed{3}$ $\gamma_m = \underset{\gamma}{\text{argmin}} \sum_{i=1}^{n} L(y_i, F_{m-1}(x_i) + \gamma)$

γ — output of DT,

$F_{m-1}(x_i)$ — output of BM

output of prev ones + output of recent

The loss $\rightarrow \frac{1}{2}(y_i - (60 + \hat{y}))$
$\hat{y}$ — R_1

minimize it (follow rule 1)

0 - 1

4 Update Model learning Rate.

$f_m(x) + f_{m-1}(x) + \alpha\, \gamma_m\, h_m(x)$.

• Update Probability ← **XG BOOST** / update Residual

1) Classifier → To calc Res → 0.5 $\left(\frac{0+1}{2}\right)$

Probab = 0.5

• Ref Data

Sal (D)	Credit (D)	APPROVAL (ID)	Residue (Out - Res)
<=50K	B	0	-0.5
<=50K	G	1	0.5
<=50K	G	1	0.5
>50K	B	0	-0.5
>50K	G	1	0.5
>50K	N	1	0.5
<=50K	N	0	-0.5

[Base model Output Based on $\log\left(\frac{P}{1-P}\right)$ and then output of DTs]

STEP 1 - Construct Base Model [Construct Tree with Root] DT₁

→ Binary split

③ -0.14 Salary (one of the D feat)

<=50K: ① [-.5, .5, .5, -.5] 0

>50K: [-.5, .5, .5] ② -0.33

---

② STEP 2 → Calculate SIMILARITY WEIGHT

$$= \frac{(\Sigma(\text{Residual}))^2}{\Sigma(P_r(1-P_r))}$$

No. of Residuals

$$\to ① \to \frac{[-.5+.5+.5-.5]^2}{[[0.5(1-0.5)) + 0.5(1-0.5) + 0.5(1-0.5) + 0.5(1-0.5)]}$$

① → 0

$$② \to \frac{0.25}{0.75} \to \frac{25}{75} = \frac{1}{3} = 0.33$$

② → 0.33

$$③ \to \frac{0.25}{1.75} = \frac{25}{175} \Rightarrow \frac{5}{35} = \frac{1}{7} = 0.14$$

③ → 0.14

STEP-3 Calculate Gain

SW of L + SW of R - SW of Root = 0.19

Why → ? → We can also start spliting from another Dept feature.
↳ happens in DT → Based on IG → happens in all Boosting.

→ Split from credit feat in a Binary way (B, NB) (B, NG)
and compare the gain.

→ Sal was Best → High Gain → 0.19

Covering All features → DT_1

★ So 2nd split from {Sal, Credit} − {Sal} = Credit

Left = Right step

Sal

<=50K ← <=0.5 | >50K → 0.33

0

CREDIT

B | GN

(1) ← -0.5 | 0.5, 0.5, -0.5 → 0.33

credit split here and steps repeat

credit

BG | N

Compare split and choose the **Best** one.

$IG \Rightarrow 1 + 0.33 - 0 = \underline{1.33}$

—— x —— x —— x ——

To split more? → decided By Post Pruning

calc. By → Cover value

↳ $Pr(1-Pr)$ = $\boxed{0.25}$

if Gain < 0.25 → cut the Branch

—— x —— x ——

$\boxed{\text{New - Data}}$ arrives

↓

Find Base model output Based on $Pr = 0.5$

$\log(\text{odds}) = \log\left(\frac{P}{1-P}\right) = \boxed{BM = 0}$

↓

Output from $DT_1 \rightarrow D_1$

$var = \sigma(0 + \alpha(D_1)) = \frac{1}{1+e^{-var}}$ (sigmoid).

eq → $\sigma(Bm + \alpha(DT_1) + \alpha(DT_2) + \dots)$
sigmoid → New Probab

→ We got our New updated Probability

x —— x —— x

- upd Prob → (out − upd Probab) → New Residuals
  New DT
  Imp

x —— x —— x

After DT_2 → Upt Prob → $\sigma(Bm + \alpha(DT_1) + \alpha(DT_2))$
New DT_3 ⟵ New Residuals

x —— x —— x

sim weight

$\sigma(0 + \alpha(DT_1 \text{ output } 1))$
$\sigma(0 + \alpha(DT_1 \text{ output } 2))$

create like a list

x —— x —— x

— STOP → value of Probab where Res is very less.

x —— x —— x

2] Regressor

Avg = 51K

Exp	Gap	Sal	Res
2	yes	40K	-11
2.5	yes	42K	-9
3	NO	52K	1
4	NO	60K	9
4.5	yes	62K	11

① Base Model
└ train with output
└ Avg Sal
└ find the Residual value

Train the DT with Res {x_i, Res}.

☆ $\boxed{\lambda \text{ can be } 0, 1.}$

SW → $\boxed{\dfrac{\Sigma(\text{Residual})^2}{\text{No. of Res} + \lambda}}$ | IG → SWL + SWR − SW Root

---

☆ Points To be Remember – IMP ←

→ Split from which feat and on what basis happens for all → on the Basis of SW and IG The final decision happens. and then further Spliting happens Based on Other features which are remaining.

☆ class → Probab → Res → train → New Prob (log func 'sigmoid)'

New Res

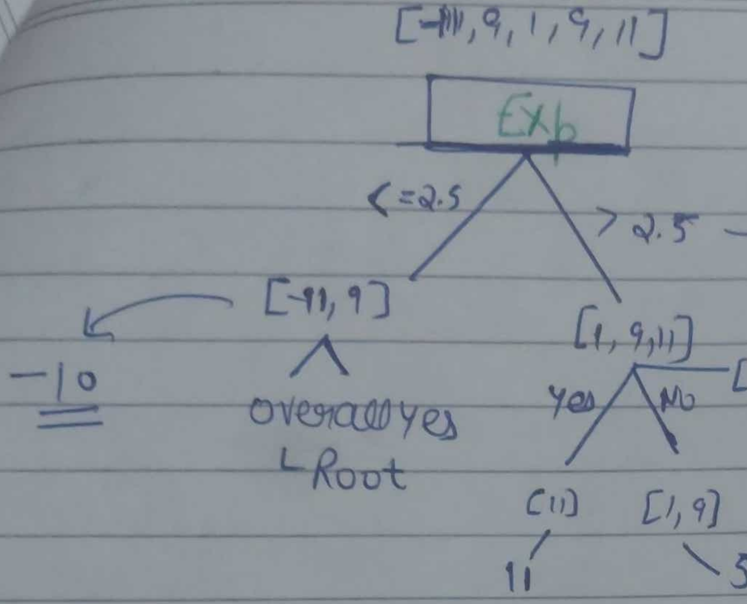
$\sigma(B + \alpha D_1 + \alpha D_2)$

☆ Reg → Avg value (or val at which lossfunc is low) — Residual → train → New output = Residue

• GB/XGB are similar → Diff is in SW and Gain used in XGB whereas Entropy + IG/Gini Index used in GB

---

let's say split from Exp gives Better Gain (SW → G)



Let's say split on this Basis is Better than split on <2, >2 and etc.

Compute value of DT → avg value.

Final output → $\boxed{Bm + \alpha DT_1 + \alpha DT_2 + \alpha DT_3}$

$DT_2 \rightarrow Dm + \alpha DT_1 + \alpha DT_2$ (current + previous).

$\alpha DT_1 \rightarrow \alpha(DT_1 - \text{output}1)$

$\alpha(DT_1 - \text{output}2)$ → New value → New Residual

$\alpha(DT_1 - \text{output}3)$

(DT_2 on R_2, DT_3 on R_3).

γ (Gamma) = 150.5 (let's say)

→ output − γ = −ve (Prune → cut)

output − γ = +ve (Continue).