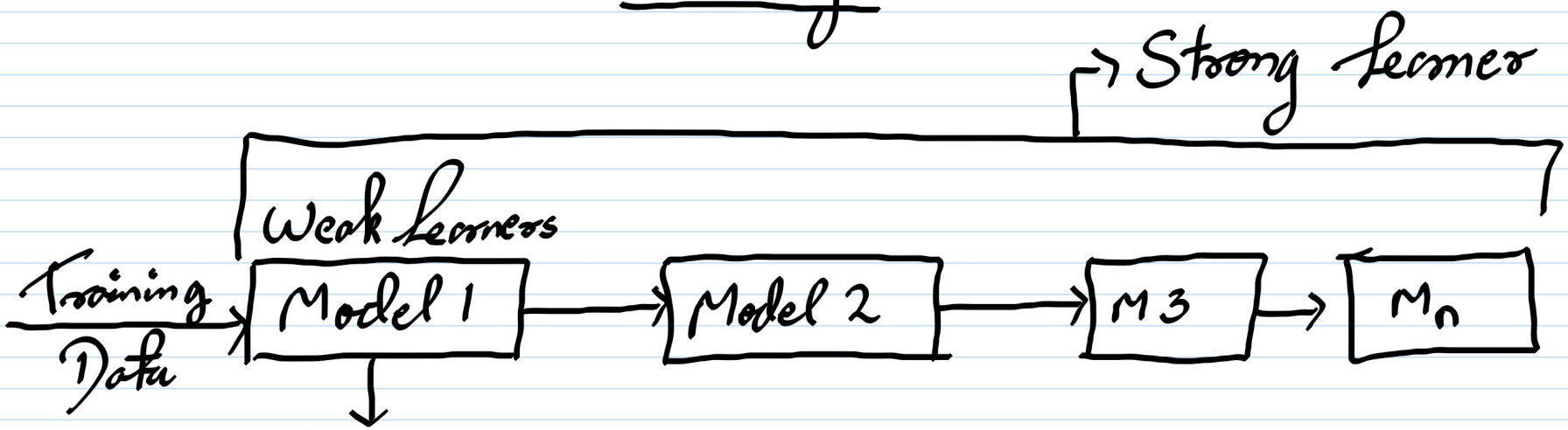
Boosting



Knowingly Wrongly Predict Results

Boosting, learns from the weak learners.

→ Combining all the weak learners, we get Strong Learners.

→ As we know, every model knowingly predicts the wrong result, that is the reason we call them as Weak Learners.

Gradient Boosting

e-g:

Independant Colums; Label

$F_0(x)$

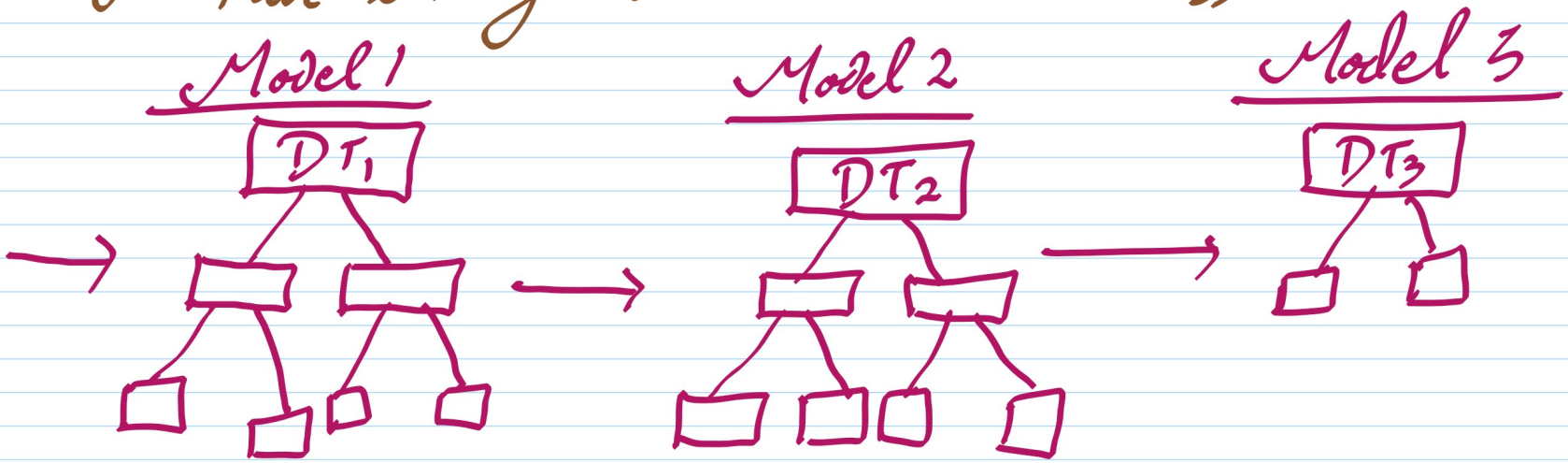
Exp	Degree	(K) Salary	$\hat{y}$	R_1	R_2	.....	R_n
2	BE	50	75	-25	-23		
3	Master	70	75	-5	-3		
5	Master	80	75	5	3		
6	Ph.d	100	75	25	20		
			↑ Label	↑ Label			

Step-1 Computing the base output i.e., average of model

Step-2 Compute Residual Error i.e., (Actual - Predict^n) = R_1

Step-3 Construct the Decision Trees, with new output feature (R_1)

Then, we will get the new result as (R_2)



Ex	Degree	(K) Salary	$h_0(x)$	1st model R_1	2nd model R_2	..... R_n
			G			
2	BE	50	75	-25	-23	
3	Master	70	75	-5	-3	
5	Master	80	75	5	3	
6	Ph.d	100	75	25	20	

Real label ↓ (K) Salary

$h_0(x)$ ↓ G

1st model ↓ R_1

2nd model ↓ R_2

$h_n(x)$ ..... R_n

Label ↑ G

Label ↑ R_1

For the 1st Record, ← Predict^{n}'s

$75 + (-23)$
$= 52$

2nd Record,

$75 + (-3)$
$= 72$

3rd

$75 + 3$
$= 78$

4th

$75 + 20$
$= 95$

This is a case of overfitting model, because our real target was 50 and my model predicted 52.

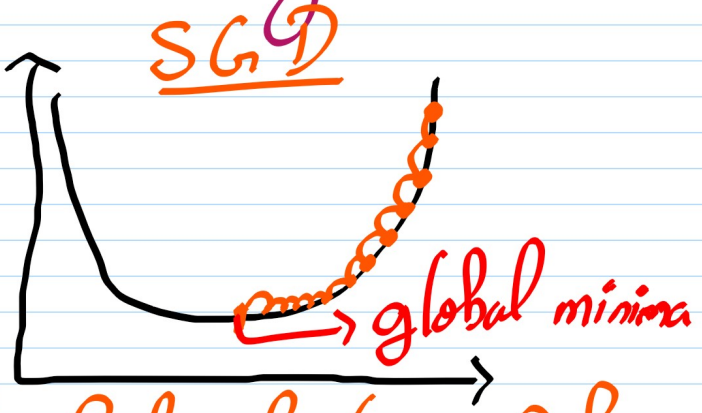
To make this generalized, we add a learning rate (α)

$\alpha = [0, 1]$

$= 75 + \alpha(R_2)$; $\alpha = 0.1$

$= 75 + (0.1)(-23)$

$= 75 - 2.3 = 73.7$



→ Used for finding best w and b.

$$y = w^T x + b$$

$$= w_1 x_1 + w_1 x_2 + \dots w_n x_n + b$$

$$F(x) = h_{w_0}(x) + \alpha_1 h_1(x) \rightarrow \text{Overfitt model}$$

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