Gradient Boosting : (For Both classification and Regression)

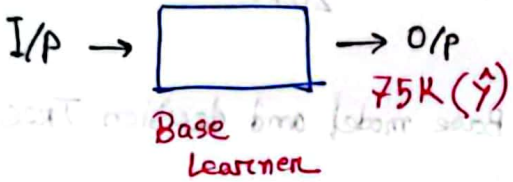
Dataset

Exp	Degree	Salary (target)
2	BE	50K
3	MS	70K
5	MS	80K
6	PHD	100K
		75K (Mean)

$$\text{Avg sal} \rightarrow \frac{[50K + 70K + 80K + 100K]}{4}$$

$$= 75K$$

Step 1: Create a Base Model



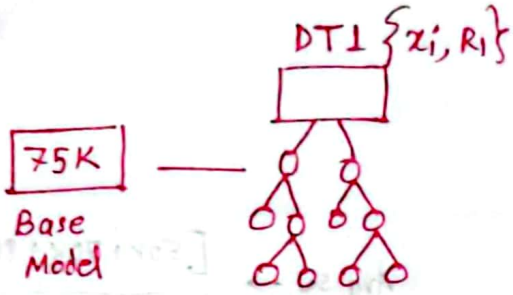
Step 2: Compute Residuals, Error

Dataset

Exp	Degree	Salary	R_1 $(y_i - \hat{y})$
2	BE	50K	-25K
3	MS	70K	-5K
5	MS	80K	25K
6	PHD	100K	25K

Step 3: Construct a decision tree. Consider inputs x_i and output R_1

(x_1, x_2) (y)



Dataset

Exp	Degree	Salary	R_1	Predicted(R_2)	
2	BE	50K	-25K	-23K	→ We assumed the output here
3	MS	70K	-5K	-3K	
5	MS	80K	5K	3K	
6	PHD	100K	25K	20K	

Final predicted output combining Base model and decision Tree:

Base Model output → 75K + (-23K) ← DT1 output

For, Exp = 2, Degree = BE

$= (75-23)$ K

$= 52$K → output for the first record

It's giving almost a correct prediction. But for now it's overfitting

For test data it won't perform well.

So, our predicted output should be:

$$75 + \alpha(-23)$$

α = learning rate

(0.01 val in general)

So, predicted output: $75 + 0.01(-23)$

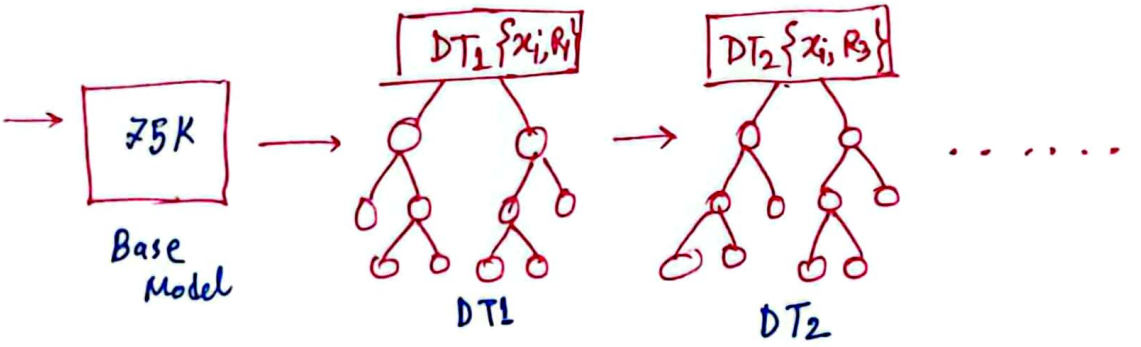
$= 74.77$ $(\hat{y})$ [For record 1]

predicted output: $75 + (0.01)(-3K)$

$= 74.97$ [For record 2]

Dataset:

Exp	Degree	Salary	$R_1(y_i - \hat{y})$	Predicted R_2	$\hat{y}_1$	Predicted R_3 $(y_i - \hat{y}_2)$
2	BE	50K	-25K	-23K	74.77K	-24.77
3	MS	70K	-5K	-3 K	74.97K	- 4.97
5	MS	80K	5 K	3 K	—	—
6	PHD	100 K	25K	20 K	—	—



Mathematical function:

$$F(x) = \overset{\alpha_0}{h_0}(x) + \overset{\alpha_1}{h_1}(x) + \overset{\alpha_2}{h_2}(x) + \dots + \overset{\alpha_n}{h_n}(x) \qquad [\alpha_0 = 1]$$

$\{\alpha_0, \alpha_1, \alpha_2, \dots \alpha_n\}$ = learning rate $\{0 \text{ to } 1\}$

$$F(x) = \sum_{i=0}^{n} \alpha_i h_i(x)$$