

Gradient Boosting Algorithm.

- Regression
- Classification

| Exp | Degree | Salary (y _{actual}) | $\hat{y}_{base\ pred}$ | R_1 | R_2 |
|-----|--------|-------------------------------|------------------------|-------|-------|
| 2 | Btech | 50k | 75 | -25 | -23 |
| 3 | M | 70k | 75 | -5 | -3 |
| 4 | M | 80k | 75 | 5 | 3 |
| 5 | Phd | 100k | 75 | 25 | 20 |

$Avg = \frac{50+70+80+100}{4} = 75k$

Step 1 → Create a base model

75

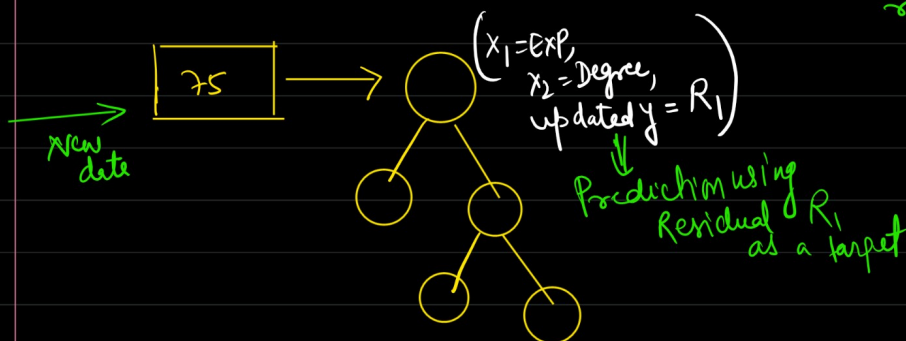
Step 2 → Compute the Residuals/error from y_{actual}

Step 3 → Make a decision

regression tree considering input as the independent variables (Exp, degree) and target variable as R_1

regression DT using highest Variance reduction.

Using this decision tree, make prediction for the input variables/rows



* Prediction $\Rightarrow 75 + (-23) \Rightarrow 52$

* The correct prediction would be $\Rightarrow 75 + \alpha$ (Residual Decision tree prediction)

$$\alpha = 0.1$$

$$\Rightarrow 75 + 0.1 \times (-23)$$

$$\Rightarrow 75 - 2.3$$

$$\Rightarrow 72.7$$

Residual Decision tree prediction

To avoid overfitting, we will add a learning rate (0 to 1)

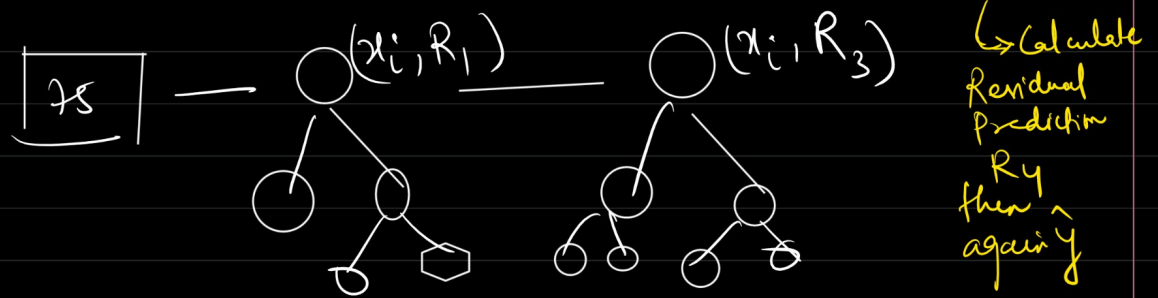
Gradient descent
So Gradient Boosting

2nd row $\Rightarrow \hat{y}$
 $\Rightarrow 75 - 0.1(-3)$
 $\Rightarrow 75 - 0.3 \Rightarrow 74.7$

| Exp | Degree | Salary (y _{actual}) | $\hat{y}_{base\ pred}$ | R_1 | R_2 | \hat{y} | R_3 | R_4 | \hat{y} |
|-----|--------|-------------------------------|------------------------|-------|-------|-----------|-----------|-------|-----------|
| 2 | Btech | 50k | 75 | -25 | -23 | 72.7 | 50 - 72.7 | - | - |
| 3 | M | 70k | 75 | -5 | -3 | 74.7 | 70 - 74.7 | - | - |
| 4 | M | 80k | 75 | 5 | 3 | 74.5 | - | - | - |
| 5 | Phd | 100k | 75 | 25 | 20 | - | - | - | - |

It is comp from base pred + α Residual DT pred
 $\Rightarrow 75 + 0.1 \times (-23)$
 $\Rightarrow 72.7$

again taking Exp & degree as IV and R_3 as target variable, Create a DT₂



$$f(x) = \alpha_1(RDT_1) + \alpha_2(RDT_2) + \alpha_3(RDT_3) \dots \alpha_n(RDT_n)$$

Learning rate α .

$$f(x) = \sum_{i=0}^n \alpha_i(RDT_i)$$

Final function for gradient descent.

* Gradient Boosting classifier

→ Base prediction 0.5.

→ Use Gini/Entropy to calculate information gain → Create the decision tree.