

XGBoost Regressor

x_1	x_2	y	\hat{y}_1	$(y - \hat{y}_1)$ R_1
exp.	gap	salary		
2	Y	40	51	$40 - 51 = -11$
2.5	Y	42	51	$42 - 51 = -9$
3	N	52	51	$52 - 51 = 1$
4	N	60	51	$60 - 51 = 9$
4.5	Y	62	51	$62 - 51 = 11$

Step-1 Base model, Average of o/p column

$$= \frac{40 + 42 + 52 + 60 + 62}{5}$$

$$= 51$$

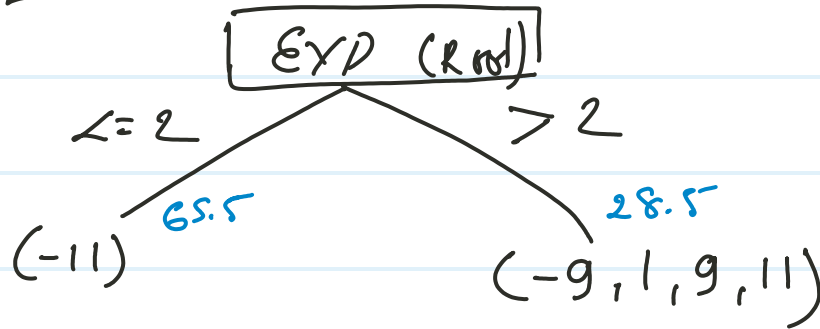
Based on 51k will find Residual

$$\text{Similarity weigh} = \frac{\sum (\text{Residual})^2}{\text{No of Residual} + 1}$$

$$[-11, -9, 1, 9, 11] \quad 0.16$$

Total
gain
93.84

$$\lambda = 1$$



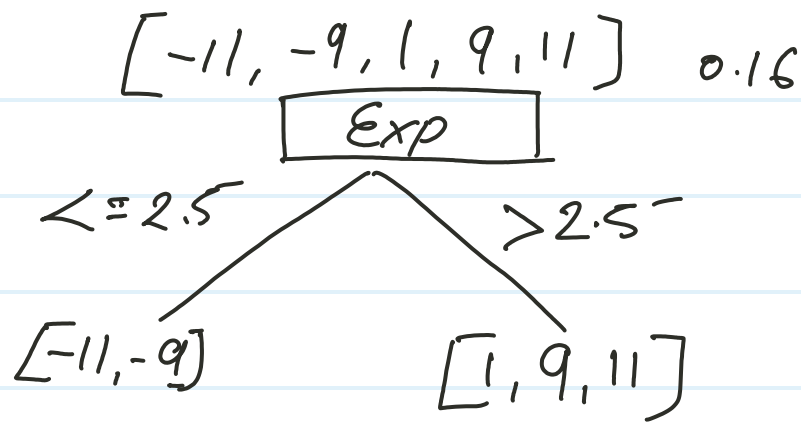
① similarity weight

$$Sm\ wt_1 = \frac{(-11)^2}{1+1} = \frac{121}{2} = 65.5$$

$$\begin{aligned} \text{② } Sm\ wt_2 &= \frac{(-\cancel{9} + 1 + \cancel{9} + 11)^2}{4+1} \\ &= \frac{144}{5} = 28.5 \end{aligned}$$

$$\begin{aligned} \text{③ } Sm\ wt_{root} &= \frac{(-\cancel{11} - \cancel{9} + 1 + \cancel{9} + \cancel{11})^2}{5+1} \\ &= 0.16 \end{aligned}$$

$$\begin{aligned} \text{Gain} &= (Sm\ wt_1 + Sm\ wt_2) - Sm\ wt_{root} \\ &= (65.5 + 28.5) - 0.16 \\ &= 94 - 0.16 \\ &= 93.84 \end{aligned}$$

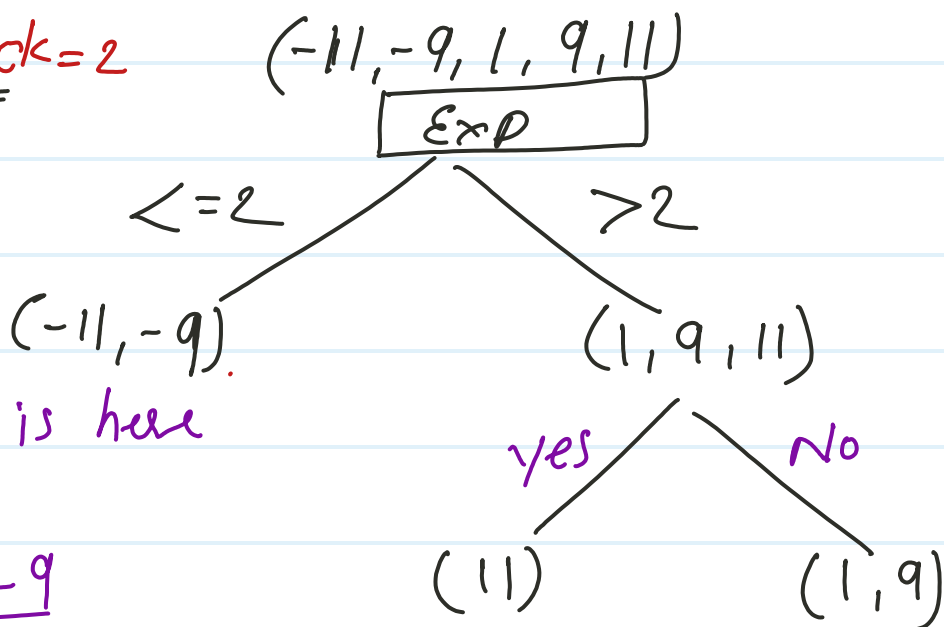


$$Sm\ wt_1 = \frac{(-11 - 9)^2}{2 + 1} = 133.33$$

$$Sm\ wt_2 = \frac{(1 + 9 + 11)^2}{3 + 1} = 110.25$$

$$\begin{aligned} \text{Gain} &= (133.33 + 110.25) - 0.16 \\ &= 143.42 \end{aligned}$$

check=2



if out put is here
avg.

$$\frac{-11 - 9}{2} \Rightarrow -\frac{20}{2} \Rightarrow -10$$

✓
11

$$\frac{1 + 9}{2} = 5 \checkmark$$

If we check with data point —

exp	gap	salary	R_1	\hat{y}_2 O/p	R_2
2	Y	40	-11	46	$40 - 46 = -6$
3	N	52	1	53.5	$52 - 53.5 = -1.5$

$$\text{for } 2 = B_m + \alpha [DT_1]$$

$$= 51 + 0.5(-10)$$

$$= 51 - 5$$

$$= 46$$

It will be new value.

$$\text{for } 3 = 51 + 0.5(5)$$

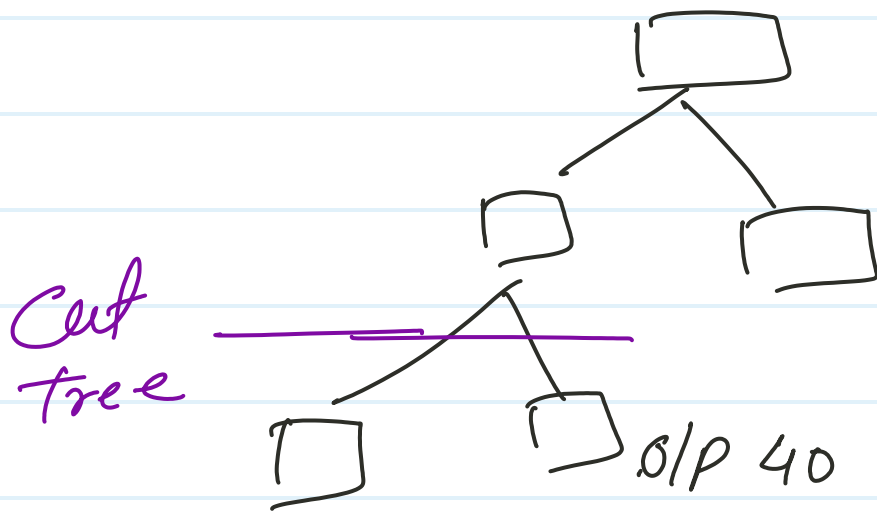
$$= 51 + 2.5$$

$$= 53.5$$

Final Formula —

$$= B_m + \alpha_1 DT_1 + \alpha_2 DT_2 + \dots + \alpha_n DT_n$$

For Post Pruning γ will be hyperparameter



Assume
 $\gamma = 50$

$$= 40 - 50$$

$$\Rightarrow -10$$

if this value is negative then we will cut the tree.

pip install xgboost

import xgboost as xgb

sklearn
↓