Xgboost Classification Algorithm:

Dataset:

Salary	Credit	Approval
<= 50K	B	0
<= 50K	G	1
<= 50K	G	1
> 50K	B	0
> 50K	G	1
> 50K	N	1
<= 50K	N	0

Steps:

Step 1: Construct a Base model

Step 2: Construct a decision tree with root ~~mod~~ node.

Step 3: Calculate Similarity weight

$$\cancel{SW =} \quad S.W = \frac{\sum (\text{Residuals})^2}{\sum Pr(1 - Pr)}$$

pr = probability

Step 4: Calculate gain.

Step 1: Construct a base mode whose output probability will be 0.5 or the output of the base model will be 0.

I/P → Base Model → O/P = 0

↳ pr = 0.5

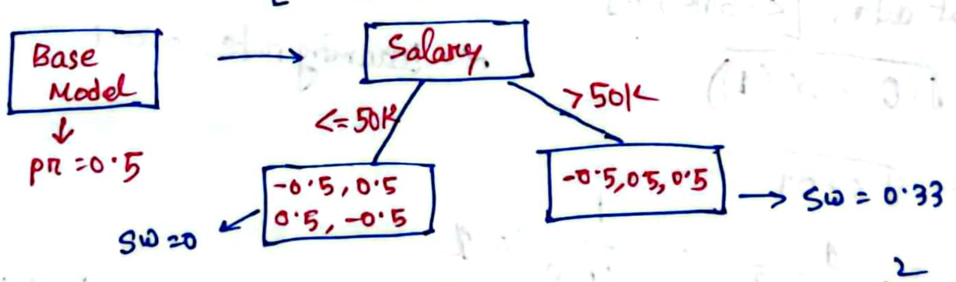
~~Step 2: Construct a dT with root node.~~

Dataset:

Salary	Credit	Approval (y_i)	$\boxed{R1\ (y_i - 0.5)}$
<= 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

Step 2: Construct a dT with root node

[−0.5, 0.5, 0.5, −0.5, 0.5, 0.5, −0.5] → SW = 0.14



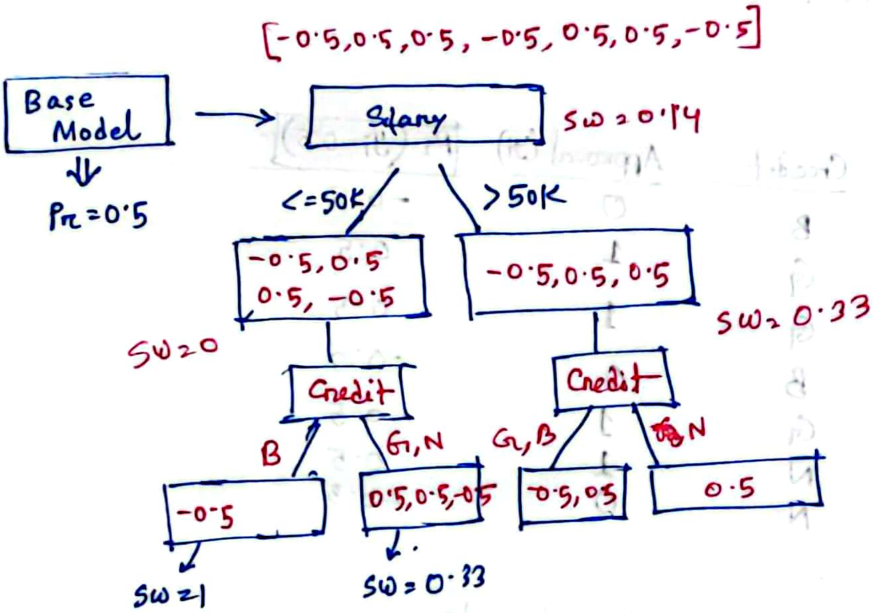
$$S.W\ (\text{Left child}) = \frac{\Sigma(\text{residual})^2}{\Sigma Pr(1-Pr)} = \frac{(-0.5+0.5+0.5-0.5)^2}{0.5(0.5)+0.5(0.5)+0.5(0.5)+0.5(0.5)} = 0$$

$$S.W\ (\text{right child}) = \frac{\Sigma(\text{residual})^2}{\Sigma Pr(1-Pr)} = \frac{(-0.5+0.5+0.5)^2}{0.5(0.5)+0.5(0.5)+0.5(0.5)}$$

$$= 0.33$$

$$Gain = \overset{\text{left}}{0} + \overset{\text{right}}{0.33} - \overset{\text{root}}{0.14} = 0.19$$

Suppose we would use decision tree on Credit column and get a better gain. So, we will use that decision Tree instead to go further.
Now, we are assuming that we have got better gain on Salary column



For test data, [<= 50K, B] →

$$Output = \sqrt{(0 + \alpha(1))} \qquad \alpha = \text{Learning rate} = 0.1$$

$$= \sqrt{0 + 0.1}$$

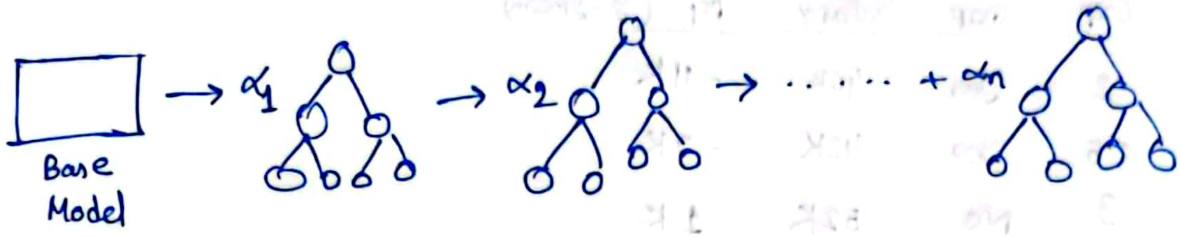
$$= \frac{1}{1+e^{-2}} = \frac{1}{1+e^{-0.1}}$$

$$= 0.52 \Rightarrow \text{if Threshold} = 0.6 \text{ [Set by Domain expert]}$$

$$= 0.52 < 0.6$$

$$\approx 0 \rightarrow \text{output (not approved)}$$

Xgboost summary:



$$\text{output} = \sigma\left(\text{Base learner} + \alpha_1(DT_1) + \alpha_2(DT_2) + \dots \alpha_n(DT_n)\right)$$

$$\boxed{\sigma = \text{sigmoid}}$$

Xgboost Regression:

Dataset

Exp	Gap	Salary
2	Yes	40K
2.5	Yes	42K
3	No	52K
4	No	60K
4.5	Yes	62K
		52K (Mean)

The formula of calculating similarity weight is a bit different than classification

$$\text{Similarity weight} = \frac{(\Sigma \text{ Residuals})^2}{\text{No. of Residuals}}$$

Steps:

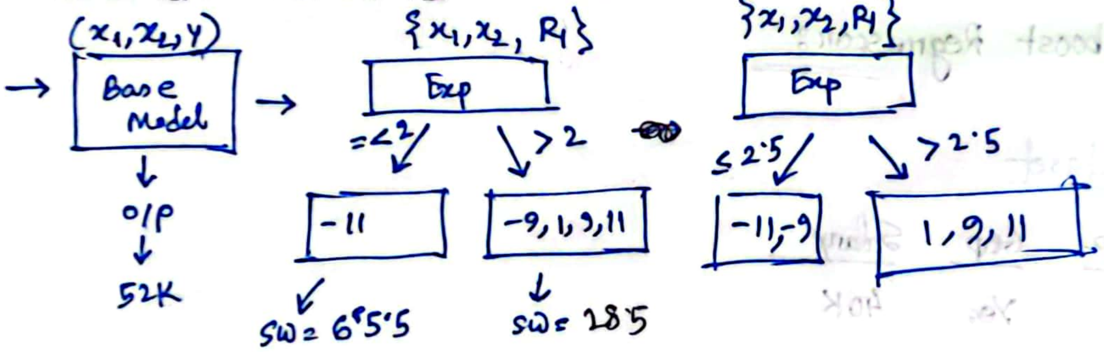
① Create a base model. Base model output will be the mean of the target feature.

$$I/P \rightarrow \boxed{\text{Base Model}} \rightarrow O/P = \text{Mean of } y = 52K$$

Dataset

x_1 Exp	x_2 Gap	y Salary	R_1 $(y - y_{mean})$
2	Yes	40K	-11K
2.5	Yes	42K	-9K
3	No	52K	1K
4	No	60K	9K
4.5	Yes	62K	11K
		≈ 52K	

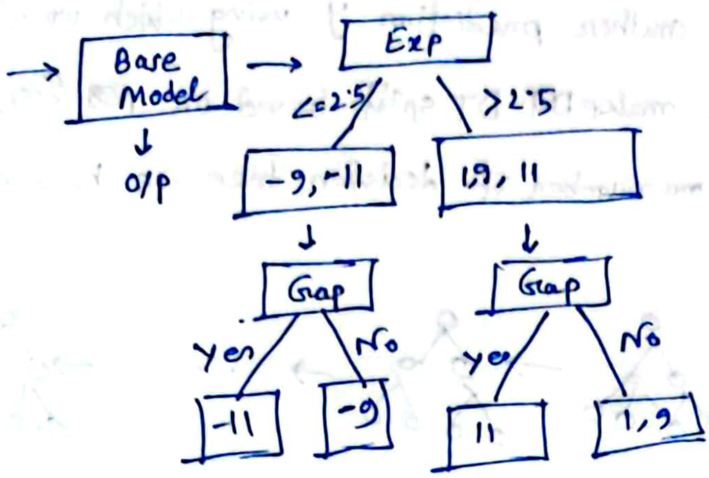
→ Construct dT based on R_1 SW = 0.16



gain from 1st dT split → 65.5 + 28.5 − 0.16
= 93.34

gain from the 2nd dT split = Suppose 99

So, we will proceed further with the greatest gain split which is ~~1st~~ 2nd dt split.



For test data (Exp, Gap) output →
(2.1, Yes)

Base learner output → ~~(9+11)/2 = 0.00~~ 51

dT output → ~~(-9-11)/2 = -10 5.5~~ 11/1 = 11

→ $51 + \alpha(\cancel{-10}\ 11)$ | $\alpha = 0.1$

→ $51 - \cancel{0.55}\ 1.1$

→ ~~50.945~~ 49.9

~~Another new data (2.7, Yes) → output~~

~~dt output → (1+9+11)/3 =~~

→ ~~51 + α(~~

Another new data → (2.7, No)

output → $51 + \alpha\left(\frac{1+9}{2}\right)$

→ $51 + 0.1(5)$

→ 51.5

for, $\{x_1, x_2, R_1\}$ we will get another prediction $\hat{y}$ using which we will make R_3, then we will make ~~DT~~ DT split based on ~~R3~~ $\{x_1, x_2, R_3\}$

It will continue till the n ~~nu~~ number of decision tree we have chosen