20/11/22

G

Gradient Boosting Algorithm [Boosting algorithm]

Regression :-

Exp (x_1)	Degree (x_2)	Salary (O/P y)	$\hat{y}$	R_1 $(y-\hat{y})$ residual	R_2 (O/P of decision Tree, assume)	$\hat{y}$ updated	R_3	R_4 .....
2	BE	50K	75	-25	-23	72.7	-22.7	
3	MTech	70K	75	-5	-3	74.7	-4.7	
5	MTech	80K	75	5	3	75.3	4.7	
6	PHD	100K	75	25	20	77	23	

Avg 75K

($\hat{y}$ updated → y - updated $\hat{y}$)

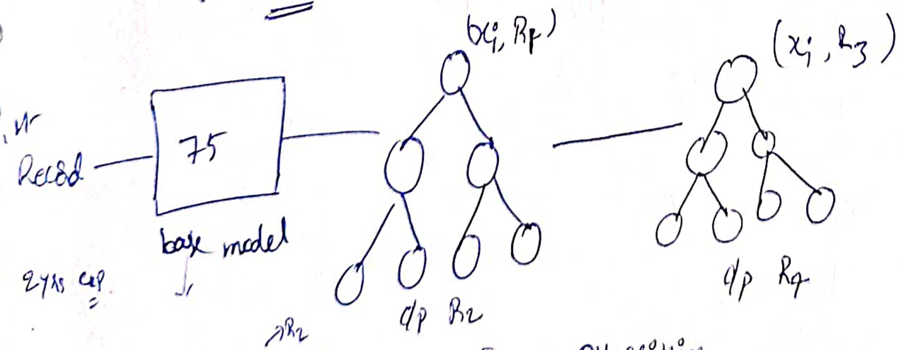
Step 1 :- Create a base model ⇒ avg. of O/P variable [Salary]

$$\text{Average} : \frac{50+70K+80K+100K}{4} = 75K \Rightarrow \hat{y}$$

Avg 75K

Step 2: Compute Residuals (or) errors ⇒ $R_1 = (y-\hat{y})$.

Step 3: we construct next sequential Decision Tree with i/p x_i and O/P residuals [R_1]



$75 + (-23) = 75-23 = 52 \Rightarrow$ Overfitting

α = learning rate, 0.1

$\alpha = 0.10$

$$\text{Predicted} = 75 + \alpha(-23)$$

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

$$= 72.7$$

2nd record = $75 + 0.1(-3)$

$= 74.7$

3rd record : $75 + 0.1(3) = 75.3$

4th record : $75 + 0.1(20) = 77$

Final function :- base learner

$M_1 \rightarrow$ 1st model, $M_2 \rightarrow$ 2nd model.

$$F(x) = h_0(x) + \alpha_1 h_1(x) + \alpha_2(h_{M_2}(x)) + \alpha_3(h_3(x)) + \dots + \alpha_n(h_n(x))$$

$$F(x) = \sum_{i=1}^{n} \alpha_i h_i(x)$$ → Gradient Boost

→ Black box Model

XGradient Boosting classifier :- Extream Gradient Boost

XG boost

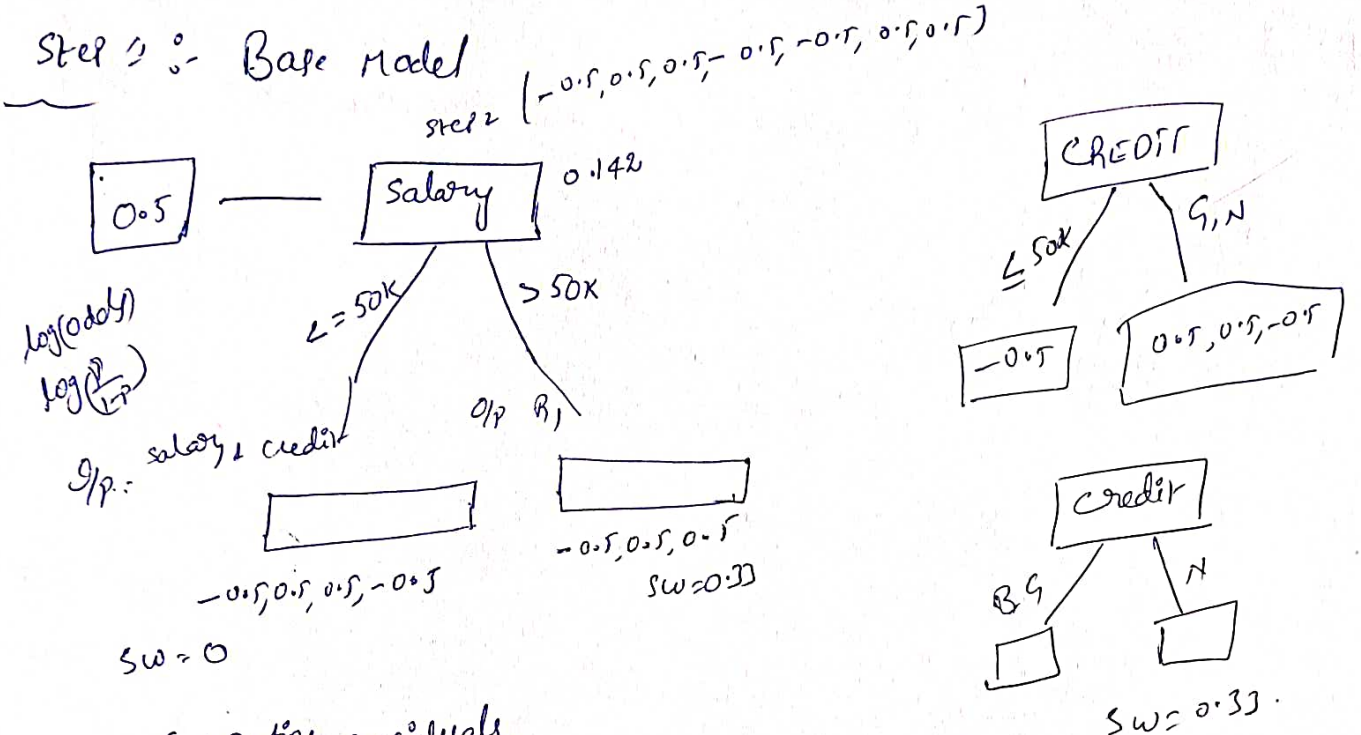
based on Similarity Model

Dataset.

Salary (x1)	Credit (x2)	y Approval	R_1	$\hat{y}$	$y - \hat{y}$ R_2
<=50K	B	0	-0.5	0.52	-0.52
<=50K	G	1	0.5	0.58	+0.42
<=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	-	

Avg 0.5 → by default threshold we'll take

Step 1 :- Base Model



Step 2 :: Computing residuals

Step 3 :: Similiarity weight

$$\frac{\Sigma(\text{residuals})^2}{\Sigma P_3(1-P_3)}$$

Probability w.r.t to base learner.

$$sw\ (\leq 50K) = \frac{(-0.5+0.5+0.5-0.5)^2}{0.5(1-0.5)+0.5(1-0.5)+(0.5)(1-0.5)+(0.5)(1-0.5)} = \frac{0}{1}$$

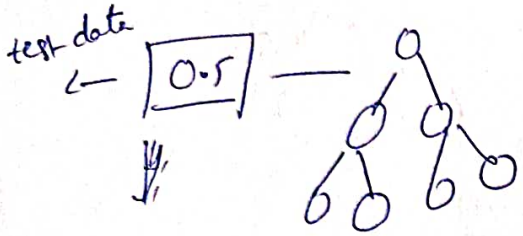
$$sw\ (>50K) = \frac{(+0.5+0.5+0.5)^2}{} = \frac{0.25}{0.75} = 0.33$$

$$\text{Gain} = 0 + 0.33 - 0.142$$

$$= 0.19.$$

Final O/P

Binary classification → Logistic Regression → log loss



$$\log(\text{odds}) = \log\left(\frac{p}{1-p}\right)$$

$$= \log\left(\frac{0.5}{0.5}\right)$$

$$= \log 1$$

$$= 0.$$

Similiary of 1st record ≤ 50K, B will come in left hand side

Model O/P: $\sigma[0 + \alpha(S\omega)]$ → log(odds)

$0 + \alpha(1)$

$= \frac{1}{1+e^{-0.1}} = 0.52$

$\sigma = \frac{1}{1+e^{-z}}$

$z = 0.1$

1st record

Since it's a binary classification Problem, we use sigmoid activation fn:

2nd record $\sigma[0 + 0.1(0.33)] = 0.58$

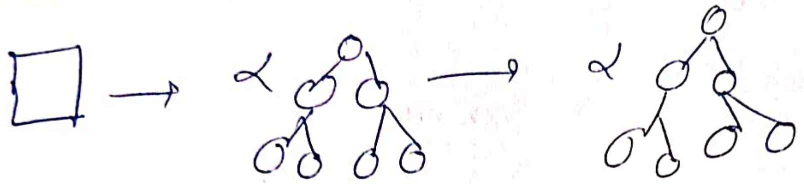
3rd record $\sigma[0 + 0.1$

XG Boost is a Black Box model :- as it's difficult to calculate.

we tend to do Pre Pruning.

Assignments :-
APS failure at Scania
Trucks Dataset.
UCI repository

Xg boost Classifier



→ the decision tree constructed based on independent features

$O/p : \sigma\ [\text{Base learner} + \alpha_1(DT_1) + \alpha_2(DT_2) + \dots + \alpha_n(DT_n)]$ (Base learner → log loss)

Practicaly :- If it's not installed in anaconda prompt pip install xgboost.

To see existed softwares/packages :-

Optuna → Hyper parameter tuning.

Xg boost Regression :-

Exp	Gap	Salary (O/p)		R_1
2	yes	40K	51K	−11K
2.5	yes	42K	51K	9K
3	No	52K	51K	1K
4	No	60K	51K	9K
4.5	yes	62K	51K	11K

1. Base models → 51K avg.
2. Step 2: Residuals
   [−11, 9, 1, 9, 11] → S.W = 1/6
3. EXP
   ≤2 → [−11], S.W = 65.5
   >2 → [−9, 1, 9, 11], SW = 28.5
4. Calculate the Similiarity weight

$$S.W = \frac{\Sigma(\text{residuals})^2}{\text{No. of residuals} + (\lambda)}$$, here $\lambda = 0$.

Residual :-

$S.W\ \frac{(-11)^2}{1+\lambda} = 121$, → left side

if $\lambda = 1$ = $\frac{121}{2} = 65.5$

right S.W = $\frac{(-9+1+9+11)^2}{4+1} = \frac{144}{5} = 28.5$

similiarity weight

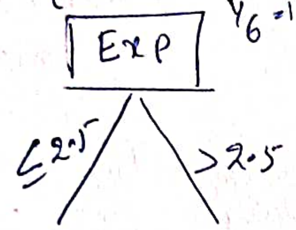
for root node = $\frac{(-11+9+1+9+11)^2}{5+1}$

$= \frac{1}{6}$

Information Gain :- $68.5 + 28.8 - 1/6 = 89.13$.

(DT1)

[−11, −9, 1, 9, 11]



$1/6 = 15.4$.

Sq root node = $1/$

left [−11, −9]   [1, 9, 11] right.

Information gain = $[133.33 + 110.25 - 1/6]$

(DT1).

$$SW = \frac{(-11-9)^2}{2+1} = \frac{(-20)^2}{3} = \frac{400}{3} = 133.33$$

$$SW_{right} = \frac{(1+9+11)^2}{3+1} = \frac{(21)^2}{4} = \frac{441}{4} = 110.25.$$

DT1 :- $51 + \alpha_1 \left[\frac{-11, -9}{2}\right]$ (left avg) $= 51 + \alpha_1(-10)$

$51 + \alpha_2 \left[\frac{1+9+11}{3}\right]$ (right avg) $= 51 + \alpha_2(-7)$

$$O/P = 51 + \alpha_1(-10) + \alpha_2(DT2) + \alpha_3(DT3) + \dots + \alpha_n(DTn)$$

it's a Blackbox model, there calculations are not visible.

XG BOOST.