Conclusion

① XG Boost example

dataset given:

salary	credit	approval	Residual	new pr	(Residual)² new Residual
<=50k	B	0	-0.5	0.6	-0.6
<=50k	G	1	0.5	0.4	0.6
<=50k	G	1	0.5	0.3	0.7
>50k	B	0	-0.5	0.7	-0.7
>50k	G	1	0.5	0.2	0.8
>50k	N	1	0.5	0.1	0.9
<=50k	N	0	-0.5	0.5	-0.5

credit = B = Bad, G = good, N = normal
approval = approval for loan if 1 then yes. 0 then no.

Step 1:

1. Calculate base model.
   we use the mean prob. of approval
   1+0 = 0.5 = pr. as base model.
   As there is 50% chance to get approval and 50% chance to not get approval.

2. Step 2:
   Calculate residual value (error) → diff b/w actual & predicted value. wrt base model.
   Residual value = error = approval - pr.
   If no = 0 - pr = 0 - 0.5 = -0.5
   If yes = 1 - pr = 1 - 0.5 = +0.5

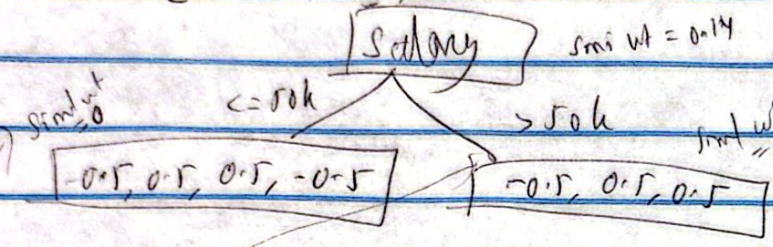
(2)

Step 3

create initial tree.

use data pts from residual values. salary

[-0.5, 0.5, 0.5, -0.5, 0.5, 0.5, -0.5]



In xg boost even if more than 2 categories have binary classifier.

Step 4

calculate similarity weight.

Similarity weight → quantifies how well residuals minimized within a leaf node & guides algo in chosing best splits to improve model. measure of leaf node. Larger similarity weight → node else well-explained. (prediction are accurate)

$$\frac{\Sigma(\text{Residuals})^2}{\Sigma pr(1-|prob|) + \lambda}$$ ← per leaf calc.

Split 1 = $\frac{[-0.5 + 0.5 + 0.5 - 0.5]^2}{0.5(1-0.5) + 0.5(1-0.5) + 0.5(1-0.5) + 0.5(1-0.5)}$

= 0

Split 2 = $\frac{[-0.5 + 0.5 + 0.5]^2}{0.5(1-0.5) + 0.5(1-0.5) + 0.5(1-0.5)}$ = $\frac{(0.5)^2}{0.75}$

(3)

Similarity wt of root :-

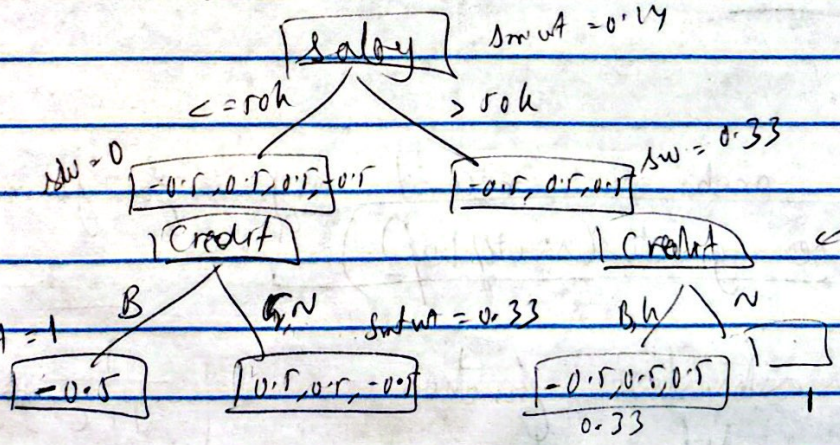
$$\frac{[-0.5+0.5+0.5-0.5+0.5+0.5-0.5]^2}{(7\times 0.25)} = \frac{0.25}{1.75} = \frac{1}{7} = 0.14$$

Step 5: Gain or info gain = $\sum$ leaf sim wt. − root sim wt.

Gain = 0 + 0.33 − 0.14 = 0.21

☆ Split can also be done using credit too so gain needs to be done per feature & best gain feature is taken into consideration for main node.

Step 6: Assume : salary had better gain wrt credit. so salary is root, next we add credit to the tree.



← can also split like this

sim wt for credit leaf B = $\frac{(-0.5)^2}{0.5(1-0.5)} = \frac{0.25}{0.25} = 1$

sim wt " " leaf N = $\frac{(0.5)^2}{0.25\times 3} = \frac{0.25}{0.75} = \frac{1}{3} = 0.33$

Gain = $\underset{\text{wt}}{\overset{\text{leaf 1}}{1}} + \underset{\text{wt}}{\overset{\text{leaf 2}}{0.33}} - \overset{\text{root}}{0} = 1.33$

☆ which ever split gives better gain we consider that split

Step 7:

whether to split it more? is post-pruning calc in xgboost using cover value = pr(1-prob) . = 0.5 × 0.5 = 0.25 mostly last step. here is <0.25 then split.

so decision tree 1 is created.

* Next step for boosting is to calculate new predictions using old residuals, creating new decision tree.

Step 1:

calculate new prob of base model.

using formula $\log(\text{odds}) = \log\left(\frac{Pr}{1-Pr}\right) = \log 1 = 0$

so new Pr = 0.

Step 2:

create new prob per record using formula.

new pr = σ (Pr + learning rate (sim wtg sof)) × per record

where σ = activation function = $\frac{1}{1+e^{-\text{log(odds)}}}$

Pr = base model prob.

learning rate = b/w 0 & 1.

new column added to dataset new Pr. See page 1

ex: for record 1: $\sigma[0 + 0.1(1)] = \sigma(0.1) = 0.6$

let learning rate = 0.1

so, $\sigma = \frac{1}{1+e^{-0.1}} = 0.6$. Assume we calc all record similarly

(F)

Step 3

calc new residuals using formula.

[approval - new pr.] per row

first record = 0 - 0.6 = -0.4

2nd " = 1 - 0.4 = 0.6

(see page 1 for new residual column)

Step 4

create new decision tree using new residuals as data points. same like first decision tree step

D#3

Next for new decision tree (3)

use formula to get new prob using below:

$$\sigma\left(Pr + \alpha_1(T_1^{sw}) + \alpha_2(T_2^{sw}) \ldots\right)$$

finally stop when residual is very less.