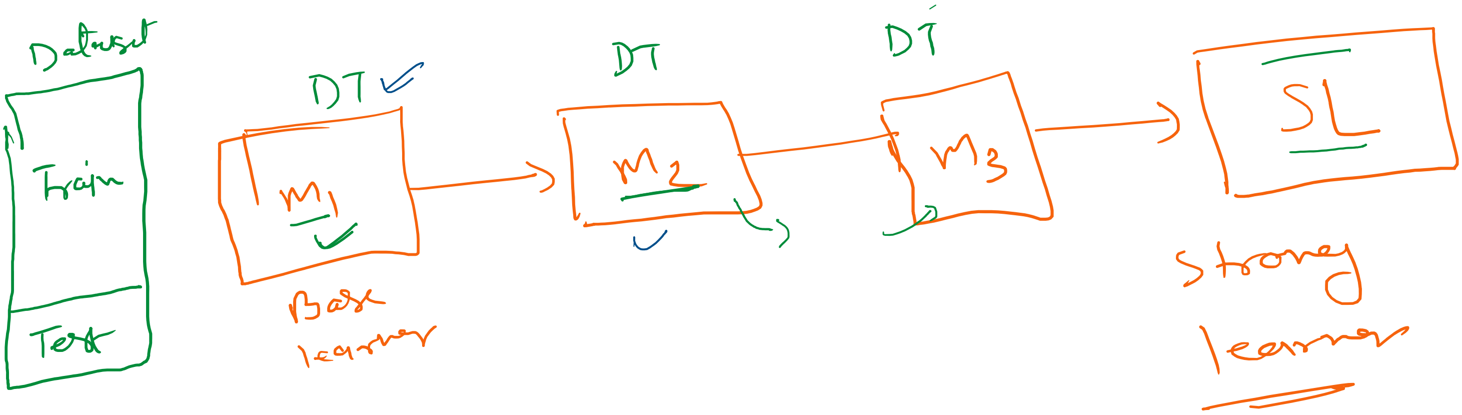
Session 36 ML

13 October 2024 07:05 PM

XGBoost

→ It works by sequentially adding simple models to correct the errors made by previous models.



② It is optimised to be highly computationally efficient and can handle large datasets.

③ XGBoost has an inbuilt routine to handle missing data.

Note → XGBoost is used for both Regression and classification.

Xgboost Algorithm

Steps

1. Construct base model
2. calculate Pseudo Residuals.
3. Building XGBoost tree with root node.
4. calculate similarity weight

for classification

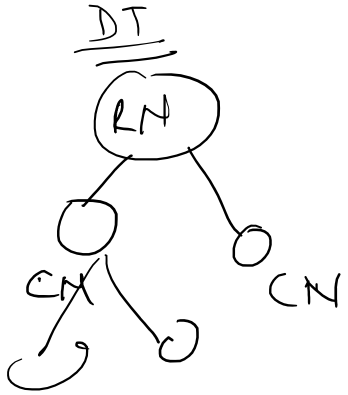
$$\boxed{S.W. = \frac{(\Sigma \text{ Residuals})^2}{\Sigma Pr(1-Pr)}}$$

for Regression ✓

$$\boxed{S.W. = \frac{(\Sigma \text{ Residual})^2}{\text{No. of Residuals}}}$$

5. Calculate Gain

DT



Dataset

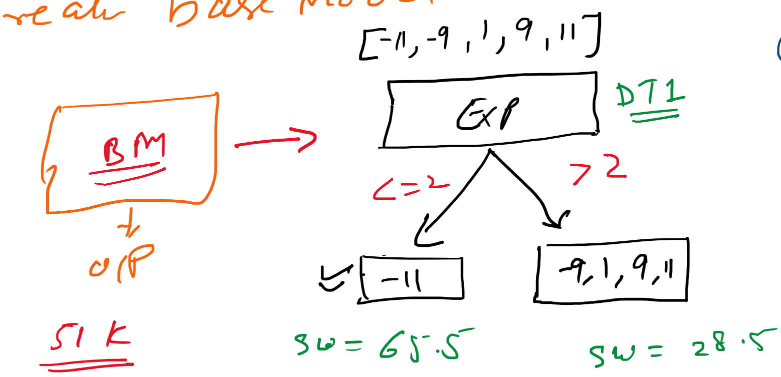
Exp	Gap	Salary	R_1
2	Yes	40k	-11
2.5	Yes	42K	-9

→	3	No	52K	1
→	4	No	60K	9
	4.5	Yes	62K	11

avg ≈ 51K

Step

① Create base model



$(SW)_{RM} = \cancel{-11} \cancel{-9} +1 \cancel{-9} \cancel{+11}$

$= \frac{1}{6+1} = \frac{1}{7} = 0.16$

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

↳ Hyperparameter tuning

$\boxed{\lambda = 1}$

$$SW_{(LS)} = \frac{121}{1+1} = \frac{121}{2} = 65.5$$

$$(SW)_{(RS)} = \frac{(\cancel{-9}+1+\cancel{9}+11)^2}{4+1} = \frac{144}{5} = \boxed{28.5}$$

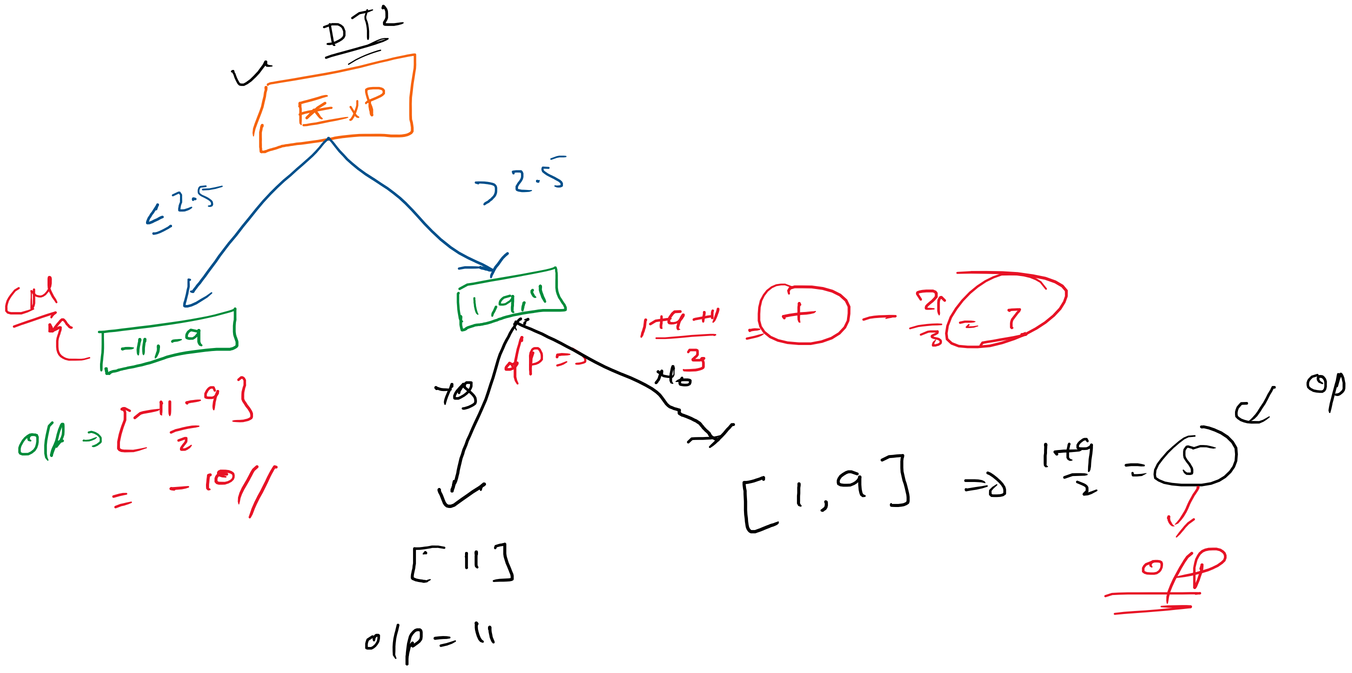
⑤ Calculate Gain

$$\text{Gain} = 65.5 + 28.5 - 0.16$$

$$= (SW)_{LS} + (SW)_{RS} - (SW)_{RM}$$

$$= (SW)_{LS} + (SW)_{RS} - (SW)_{RN}$$

$$\boxed{\text{Gain} = 98.34}$$



$$\text{XGBoost} = \underline{\text{Base learner}} + \alpha_1 (DT_1) + \alpha_2 (DT_2) \; \text{- - -} \; \alpha_n (DT_n)$$

$\alpha \to 0.1$

Ls learning rate (Hyperparameter)

$$\text{XGBoost} = 51K + 0.1 (5$$

$$= \underline{51.5}$$