

# Day 7 : Xgboost CLASSIFIER AND REGRESSOR

## Agenda

- ① Xgboost classifier
- ② Xgboost Regressor
- ③ SVM
- ④ SVR

$$\log\left(\frac{p}{1-p}\right) = \log\left(\frac{0.5}{0.5}\right) = 0$$

## ① Xgboost classifier

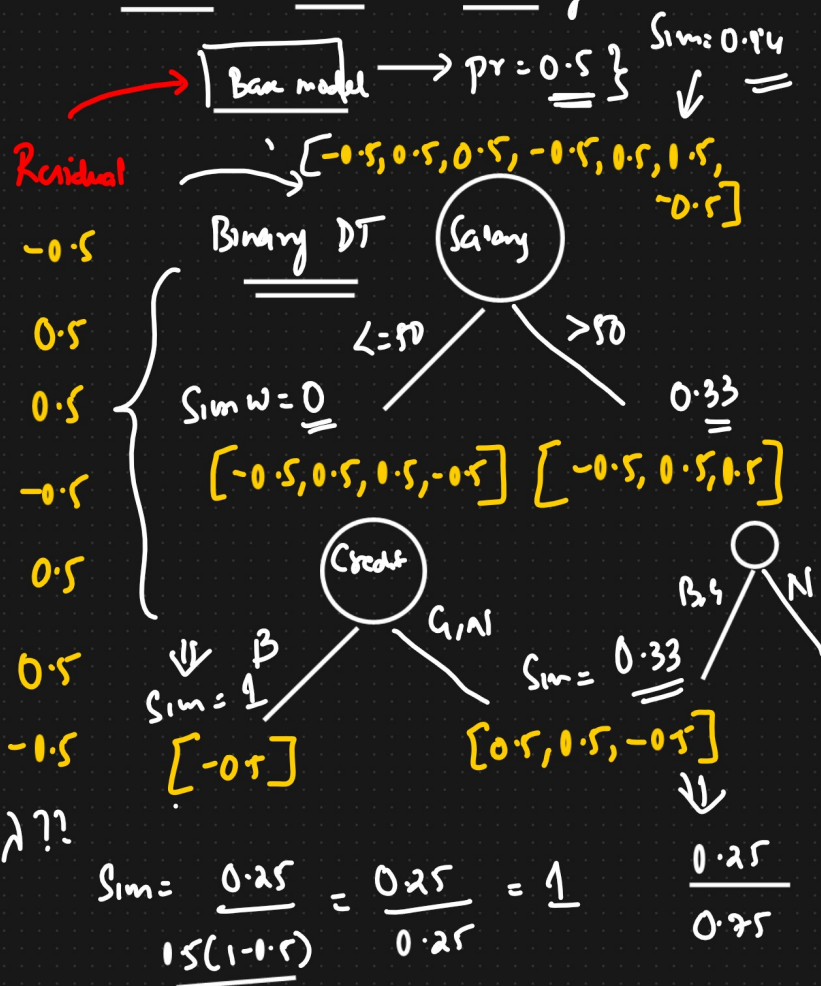
{ Dataset }

Salary	Credit	Approval
$\leq 50$	B	0
$\leq 50$	G	1
$\leq 50$	G	1
$> 50$	B	0
$> 50$	G	1
$> 50K$	N	1
$\leq 50K$	N	0

$$\lambda = 0$$

$\lambda ??$

## Extreme Gradient Boosting



① Create a Binary Decision Tree using the feature

② Calculate Similarity weight

$$= \frac{\sum (Residual)^2}{\sum (Pr(1-Pr) + \lambda)}$$

$$1 + 0.33 - 0 = 1.33$$

$$\sigma [0 + \alpha(1)]$$

$$\downarrow$$

$$[0 + 1] \checkmark$$

$$\alpha = 0.01$$

Sigmoid

③ Information gain

New  
Reward

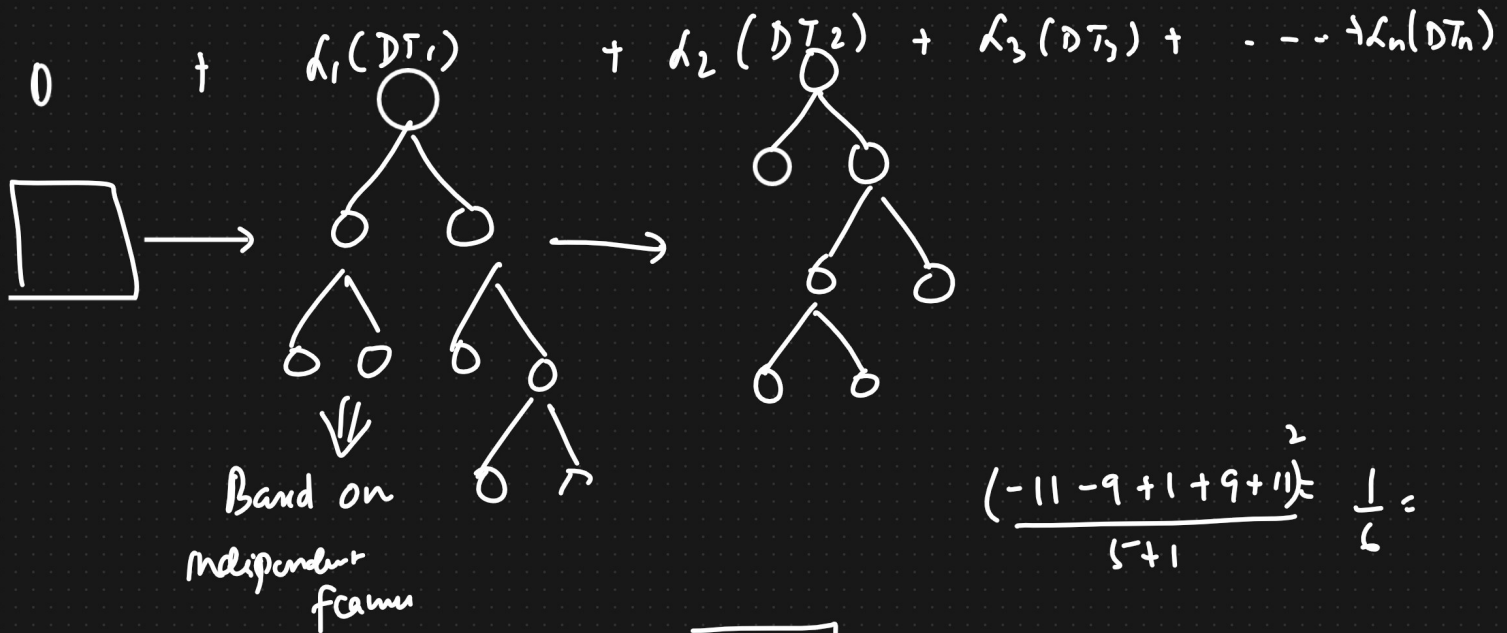
O/p

$$\sigma \left[ 0 + \alpha_1(\overline{DT_1}) + \alpha_2(\overline{DT_2}) + \alpha_3(\overline{DT_3}) + \alpha_4(\overline{DT_4}) + \dots + \alpha_n(\overline{DT_n}) \right]$$

Xgboost  $\rightarrow$  BLACK BOX Model

Pre pruning

$\lambda \rightarrow$  cross validation



$$\frac{(-11 - 9 + 1 + 9 + 11)^2}{5 + 1} \cdot \frac{1}{6} =$$

② Xgboost Regressor

Exp	Gap	Salary	Res
2	Yes	40K	-11K
2.5	Yes	42K	-9K
3	No	52K	1K
4	No	60K	9K
4.5	Yes	62K	11K

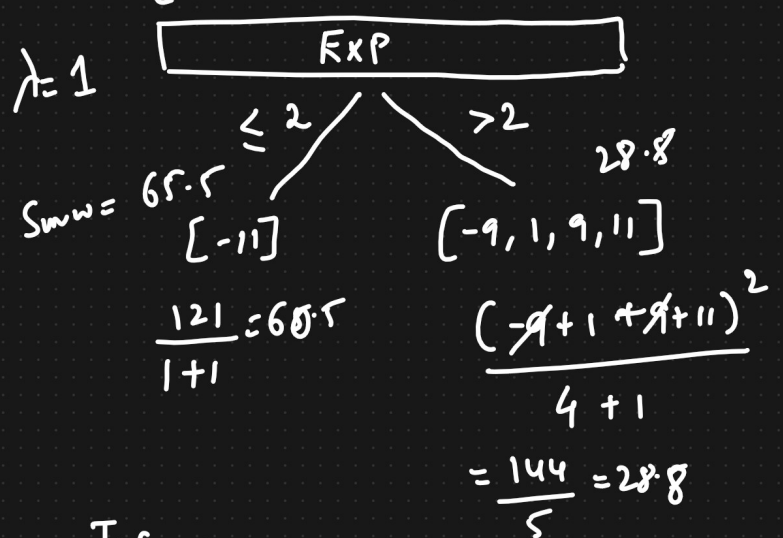
Similarity weight =  $\frac{\sum (\text{Residuals})^2}{\text{No. of Res} + \lambda}$

Bin Model

$\rightarrow S1$

Sim = 16

[-11K, -9, 1, 9, 11]



Information

$$= 60.5 + 28.8 - \frac{1}{6} = 89.13$$

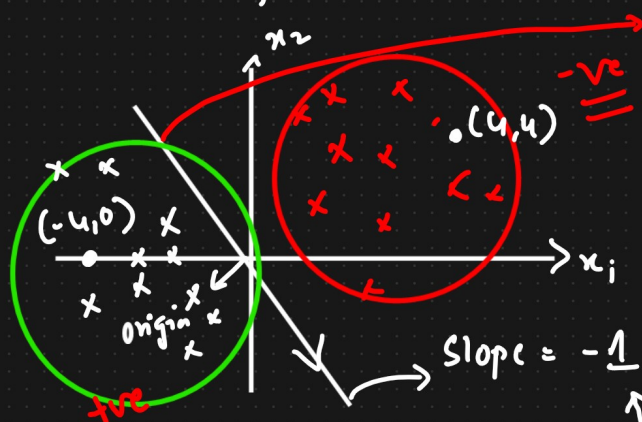
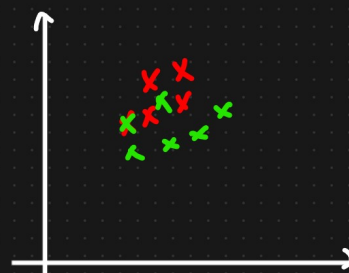
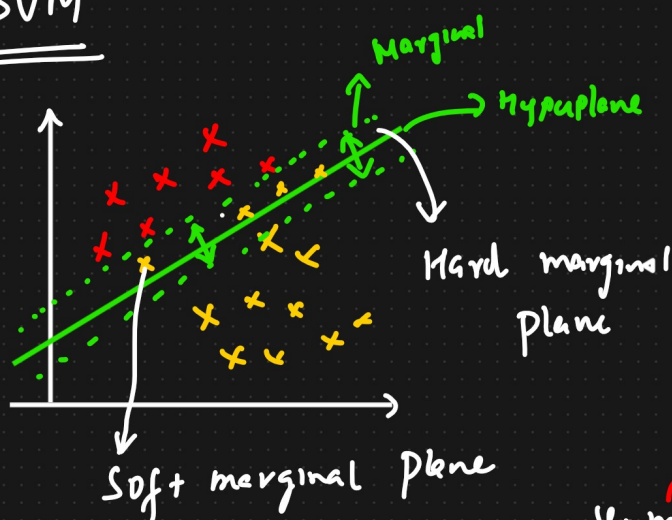
$$\frac{-11-9}{2} = -10 \quad \text{133.33} \quad \begin{matrix} \swarrow \leq 2.5 \\ \searrow > 2.5 \end{matrix} \quad \begin{matrix} \text{Exp} \\ \text{1/6} \end{matrix} \quad \begin{matrix} \swarrow \text{DT1} \\ \searrow \end{matrix}$$

$$[-11, -9] \quad [1, 9, 11] \quad \Rightarrow \quad \frac{1+9+11}{3} = \frac{21}{3} = 7$$

$$O/p = 51 + \alpha_1(-10) + \alpha_2(DT_2) + \alpha_3(DT_3) + \dots + \alpha_n(DT_n)$$

① {ZDA & Feature Engineering} ✓

③ SVM



$$y = mx + c$$

$$ax + by + c = 0 \quad \text{Equation of a straight line}$$

$$m = -a/b \quad c = -c/b$$

$$y = \frac{-c}{b} - \frac{ax}{b}$$

$$y = mx + c$$

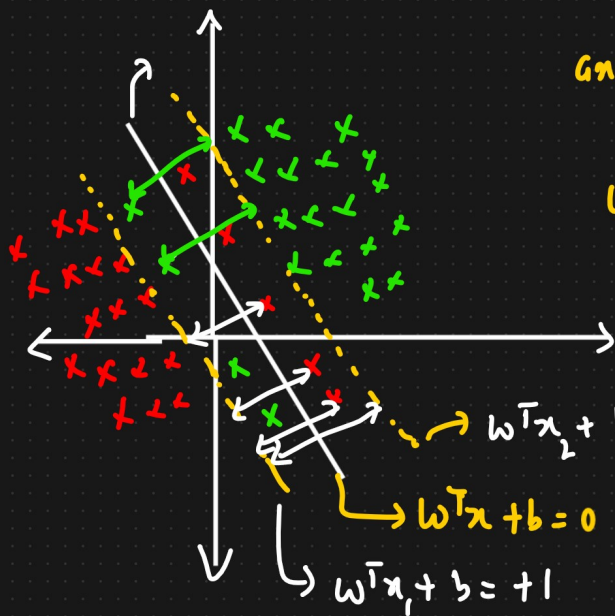
$$y = w_1x + w_2x_2 + \dots + b$$

$$y = \begin{bmatrix} -1 \\ 0 \end{bmatrix} \begin{bmatrix} -4 & 0 \end{bmatrix} \quad \omega \rightarrow -1 \quad \rightarrow \quad \boxed{y = \omega^T x + b}$$

$$= 4 \Rightarrow +ve \text{ value}$$

$$y = \begin{bmatrix} -1 \\ 0 \end{bmatrix} [4, 4]$$

$$= -4 + 0 = -4$$



$$\left. \begin{aligned} ax+by+c &= 0 \\ \Downarrow \\ w^T x + b &= 0 \end{aligned} \right\}$$

$\rightarrow \underline{w}$  vector + magnitude

$$w^T x_1 + b = 1$$

$$\frac{(-)w^T x_2 + b = -1}{(+)}$$

$$\frac{w^T (x_1 - x_2)}{|w|} = \frac{2}{|w|}$$

$$\left\{ \begin{array}{l} \text{Maximize } (w, b) \\ \frac{2}{|w|} \Rightarrow \text{margin} \\ \text{Such that } y_i \begin{cases} +1 \\ -1 \end{cases} \begin{cases} w^T x + b \geq 1 \\ w^T x + b \leq -1 \end{cases} \end{array} \right\}$$

$$\begin{aligned} \Downarrow \\ \text{Major Aim } y_i * (w^T x_i + b) &\geq 1 \\ \Downarrow \\ \text{for correct point} \end{aligned}$$

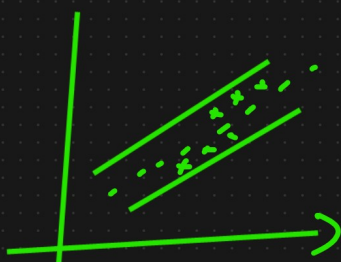
$$\boxed{\begin{array}{ccc} \text{Maximize} & \frac{2}{|w|} & \Leftrightarrow \text{Min} \\ (w, b) & & (w, b) \end{array} \quad \frac{|w|}{2}}$$

$$\text{Min } (w, b) \quad \frac{|w|}{2} + C_i \sum_{i=1}^n \xi_i \rightarrow \text{Summation of the distance of the wrong datapoints}$$

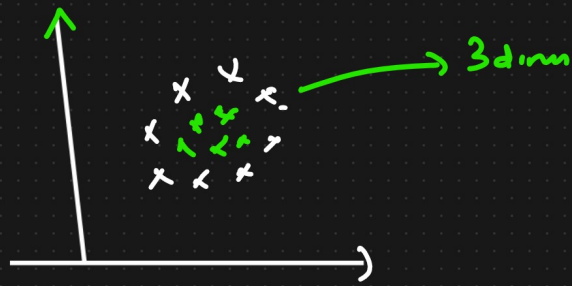
96%

$\left\{ \begin{array}{l} \text{How many} \\ \text{Errors we can} \\ \text{have} \end{array} \right\}$

$$\text{SVR} = \underline{\underline{\text{Expense}}}$$



SVM Kernel



Practically

