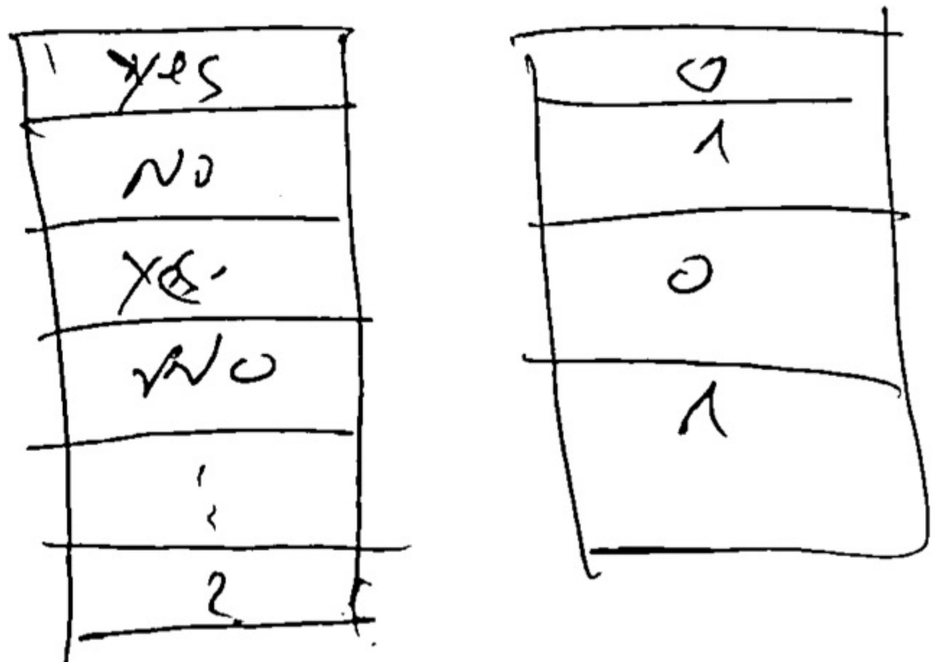
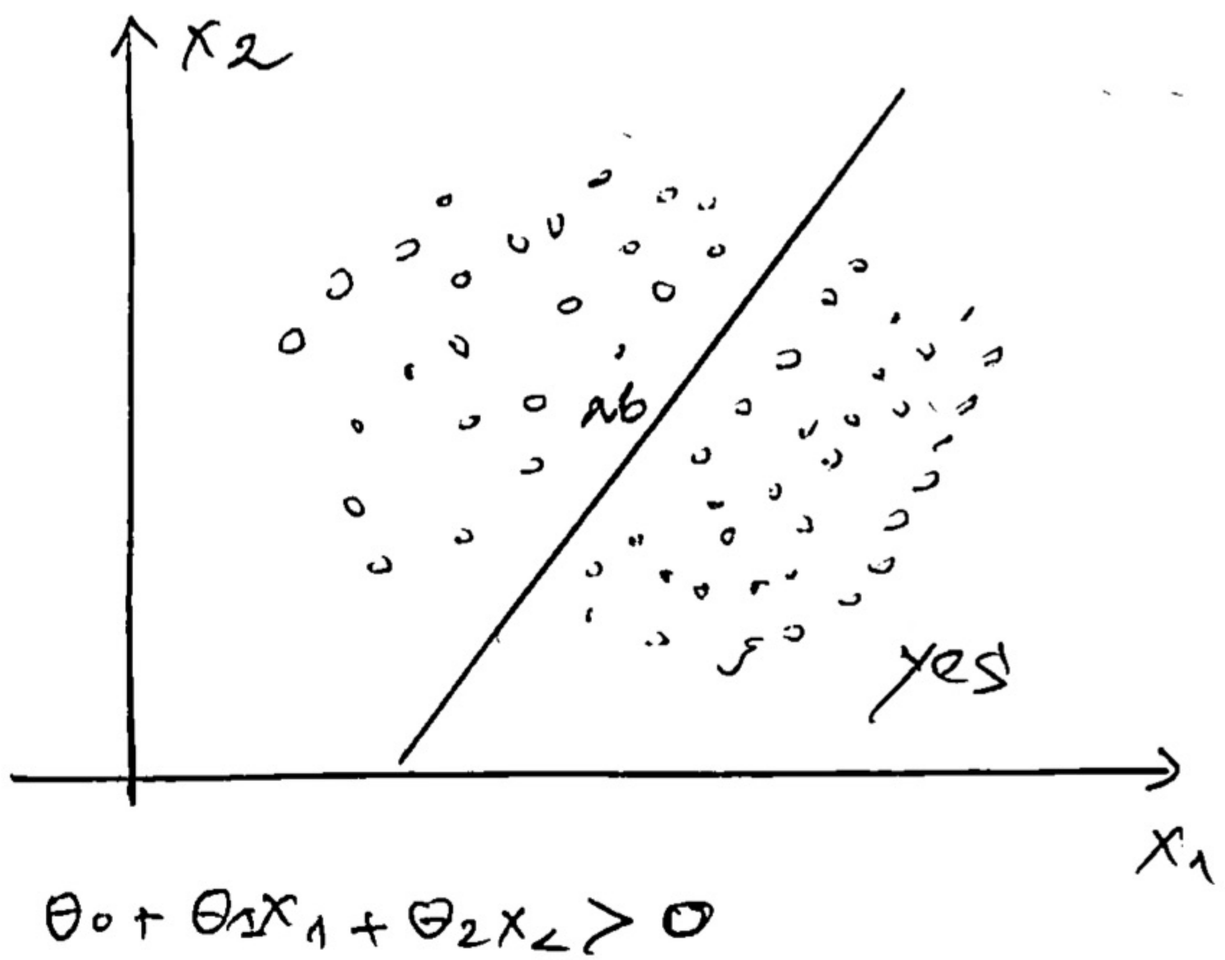
Logistique regression is a classification algorithm for categorical variables





$\theta_0 + \theta_1 x_1 + \theta_2 x_2 > 0$

x: dependent variables
y: Independent

$X \in \mathbb{R}^{m \times n}$
$y \in \{0, 1\}$

prediction

$$\hat{y} = P(y=1|x)$$

$$P(y=0|x) = 1 - P(y=1|x)$$

→ Logestique Regression Vs Linear Regression.

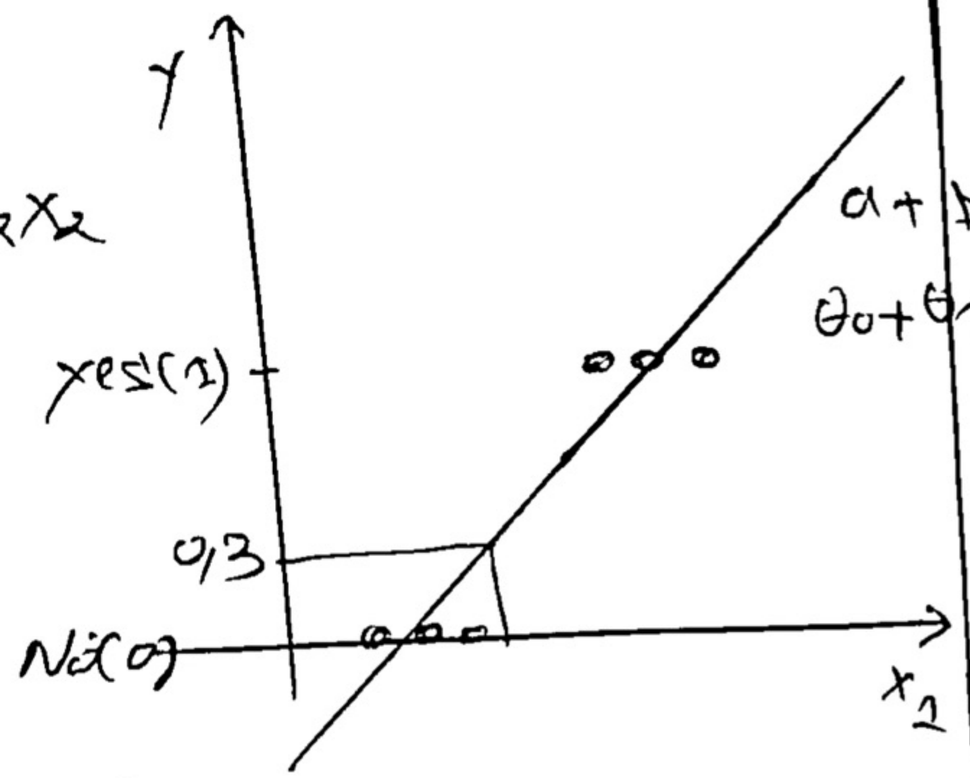
→ Linear Regression

$\theta^T X = \theta_0 + \theta_1 x_1$

$\theta^T X = \theta_0 + \theta_1 x_1 \cdots \theta_2 x_2$

$\theta^T = [\theta_0, \theta_1, -\theta_n]$

$X = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}$



example

soit $\theta^T x = -1 + 0,1x$

$x_1 = [13] \rightarrow \theta^T x = -1 + 0,1 x_1$

$= -1 + 0,1 \times 13$

$= 0,3$

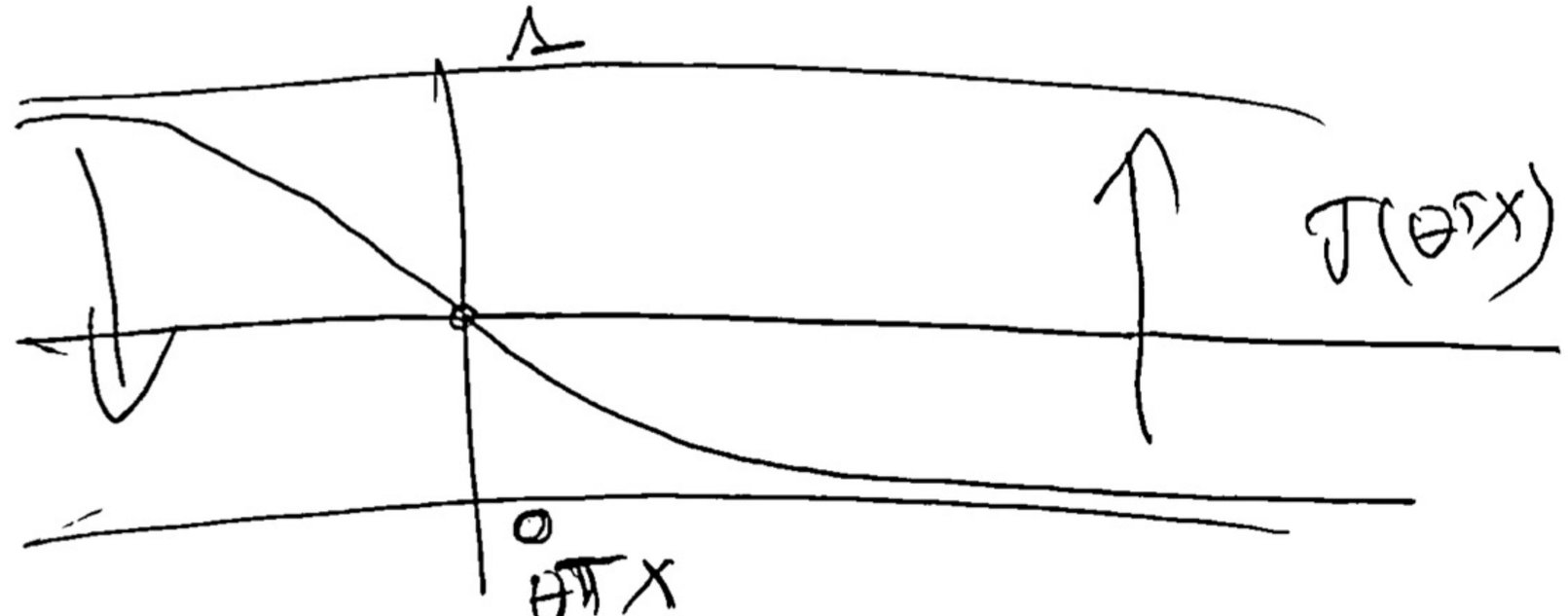
dlou

$$\hat{y} \begin{cases} 0 \text{ if } \theta^T X < 0,5 \rightarrow \text{class 0} \\ 1 \text{ if } \theta^T X \geq 0,5 \rightarrow \text{class 1} \end{cases}$$

→ always Number ~~Regression~~ classification not for

Solution

$\sigma(\theta^T X) = \sigma(\theta_0 + \cdots \theta_n x_n)$



$\hat{y} = \sigma(\theta^T X)$ $P(y=1|x)$

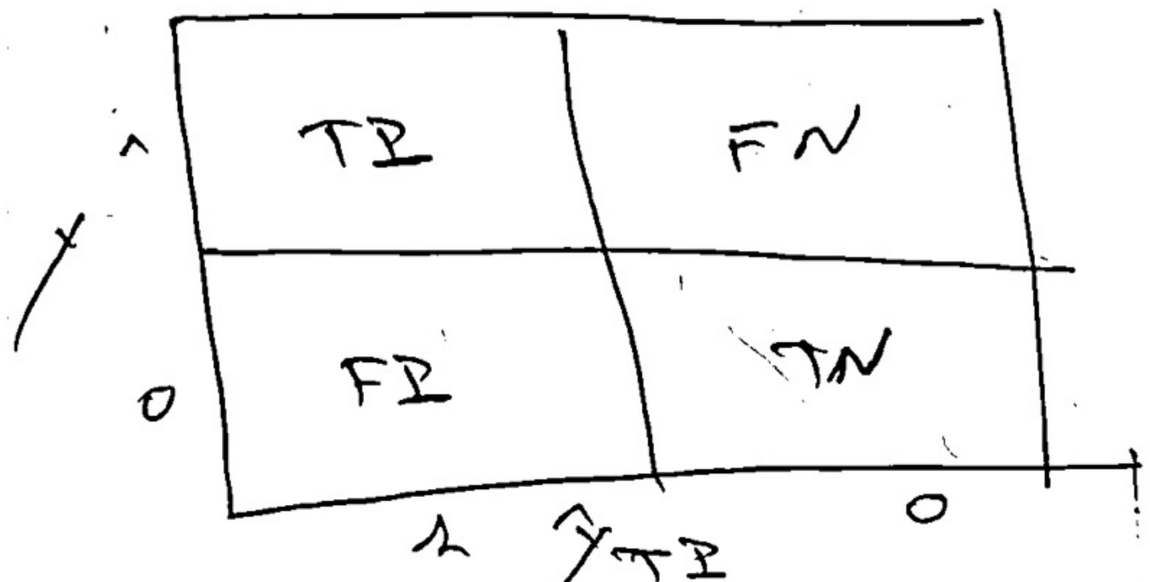
→ logestique function

$$\sigma(\theta^T X) = \frac{1}{1 + e^{-\theta^T X}}$$

$\sigma(\theta^T X) = 1$; $\sigma(\theta^T X) =$

F1 - Score :

- Confission Matrix



- precision = $\frac{TP}{TP + FP}$

- Recall = $\frac{TP}{TP + FN}$

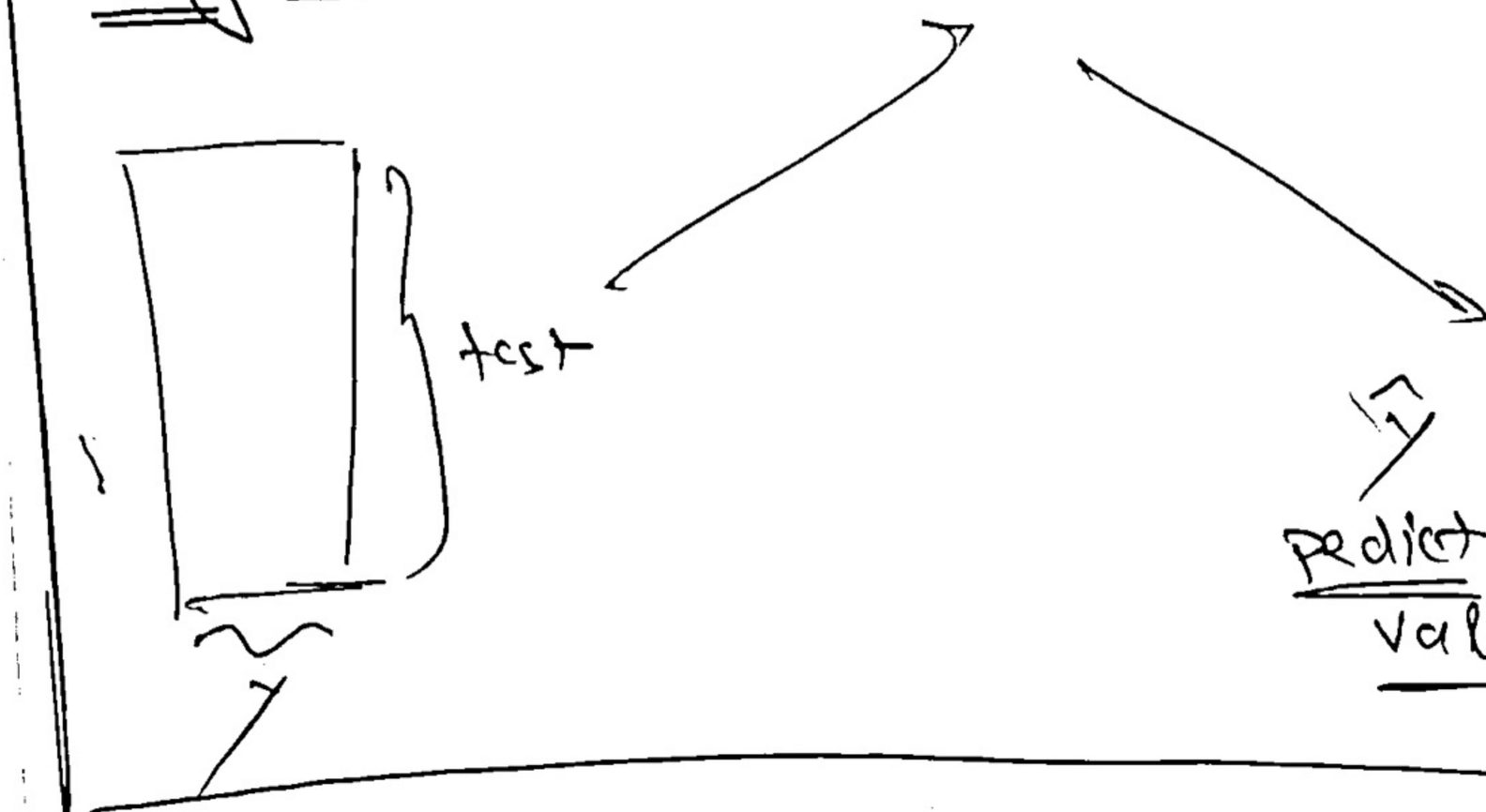
$$F1\text{-Score} = \frac{2 \times (Pre * Rec)}{Pre + Rec}$$

F1 - Score = 0,00 ⟶ 1

⟶ High accuracy

Ex:	Precision	recall	F1-Score
Churn = 0	0,93	0,96	0,83
Churn = 1	0,86	0,40	0,55
		Avg = 0,89	69%

Log Loss



Log Loss: 0,00 ⟵ 1,00

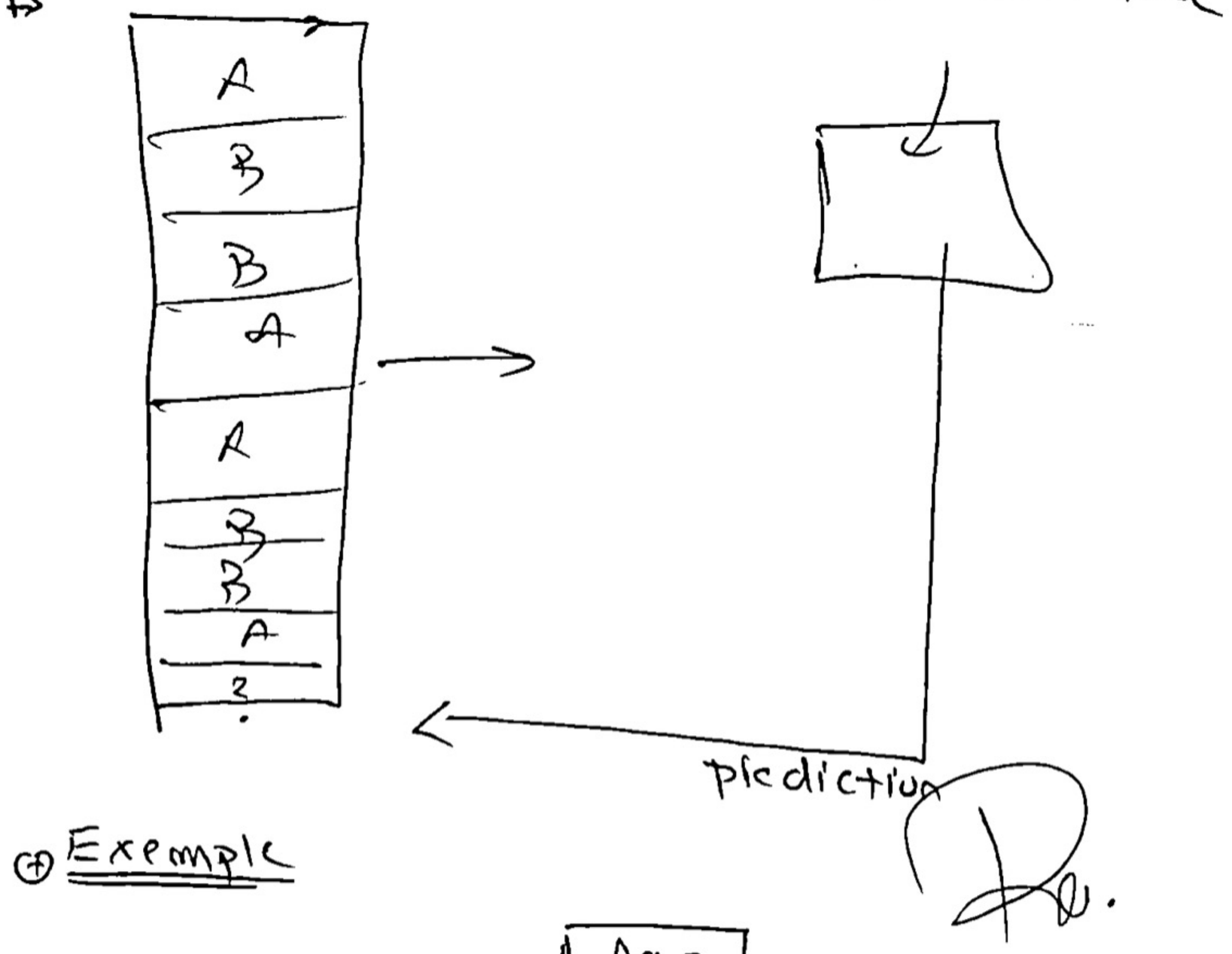
High accuracy

$$LogLoss = -\frac{1}{n}\sum y \log(\hat{y}) + (1-y)\log(1-\hat{y})$$

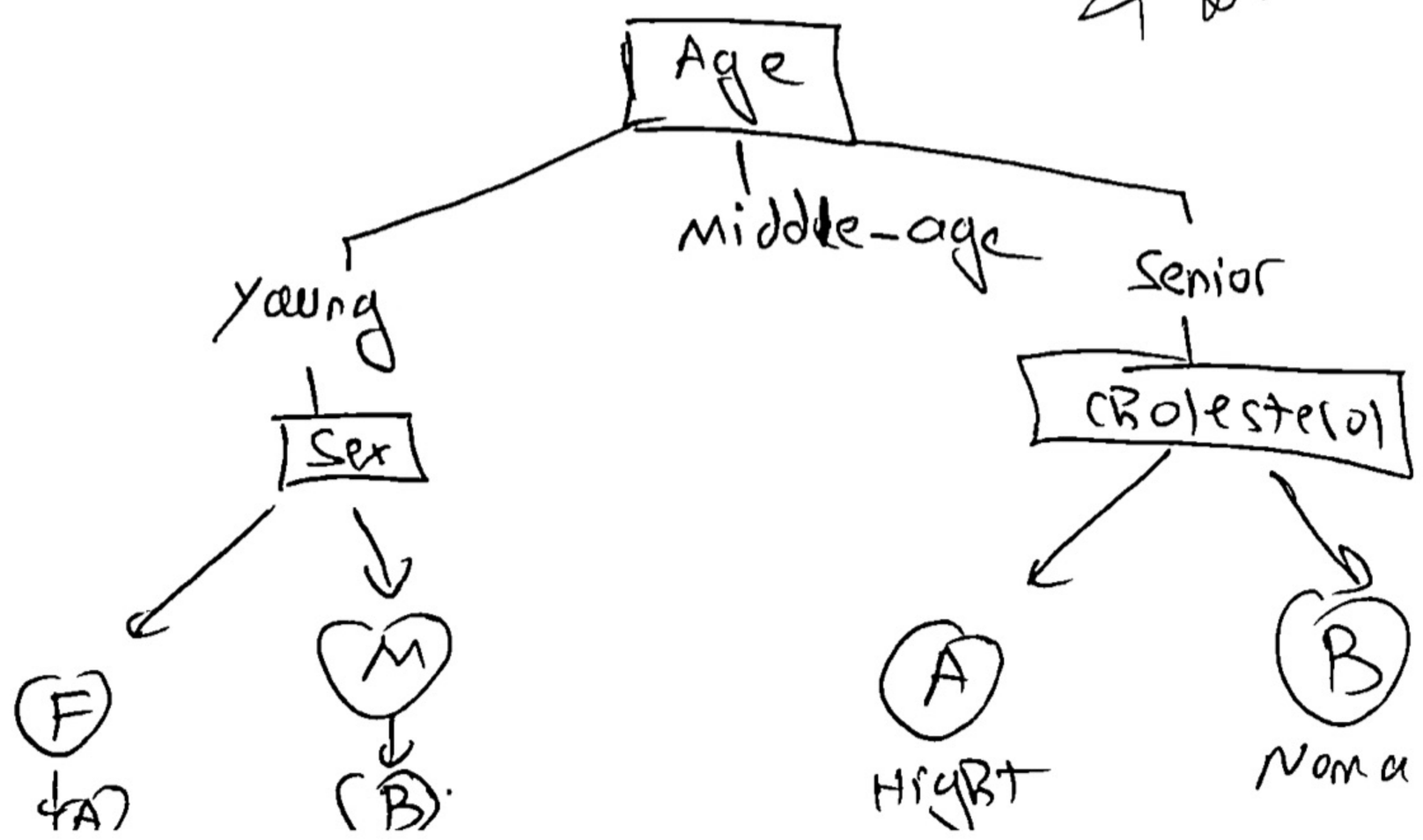
→ Introduction to :

« Decision Trees »

⇒



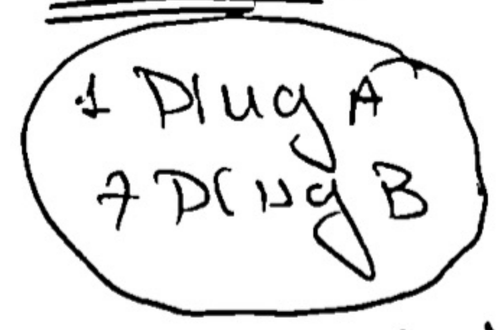
⊕ Exemple



• Entropy :

◦ Mesure of randomness or uncertainty : ≤

.exemple :



Entlopy ↓ ⩒

≤ ⇈ X

≤ = 0 V

$$S = -P(A)\log_2(P(A) - P(B)\log_2(P(B))$$

⊕ information gain ⊕

is the information that can increase the level of certainty after spliting.

Information Gain = (Entropy befor split) − (Entropy after split)

→ >>> more accuracy

$R^2 = 1 - RSE$

- 1 $\Rightarrow$ 100% acuracy
- 0 $\Rightarrow$ low

if $R^2 \in [0, 1] \simeq$ GOOD

→ Multiple Linear Regression

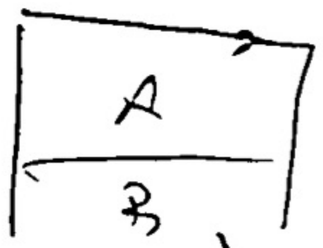
$\hat{y} = \theta_0 + \theta_1 x_1 + \theta_2 x_2 \dots \theta_n x_n$

$\hat{y} = \theta^T X$

$\theta^T = [\theta_0, \theta_2 \text{ — } \theta_n]$; $X = \begin{bmatrix} 1 \\ | \\ x_n \end{bmatrix}$

↳ the same Error

- How to estimate θ ?
  - Linier algebla opperation
  - Gradient Descent ⚠
  - Proper approacB big Data
    - MatB prouve → sun

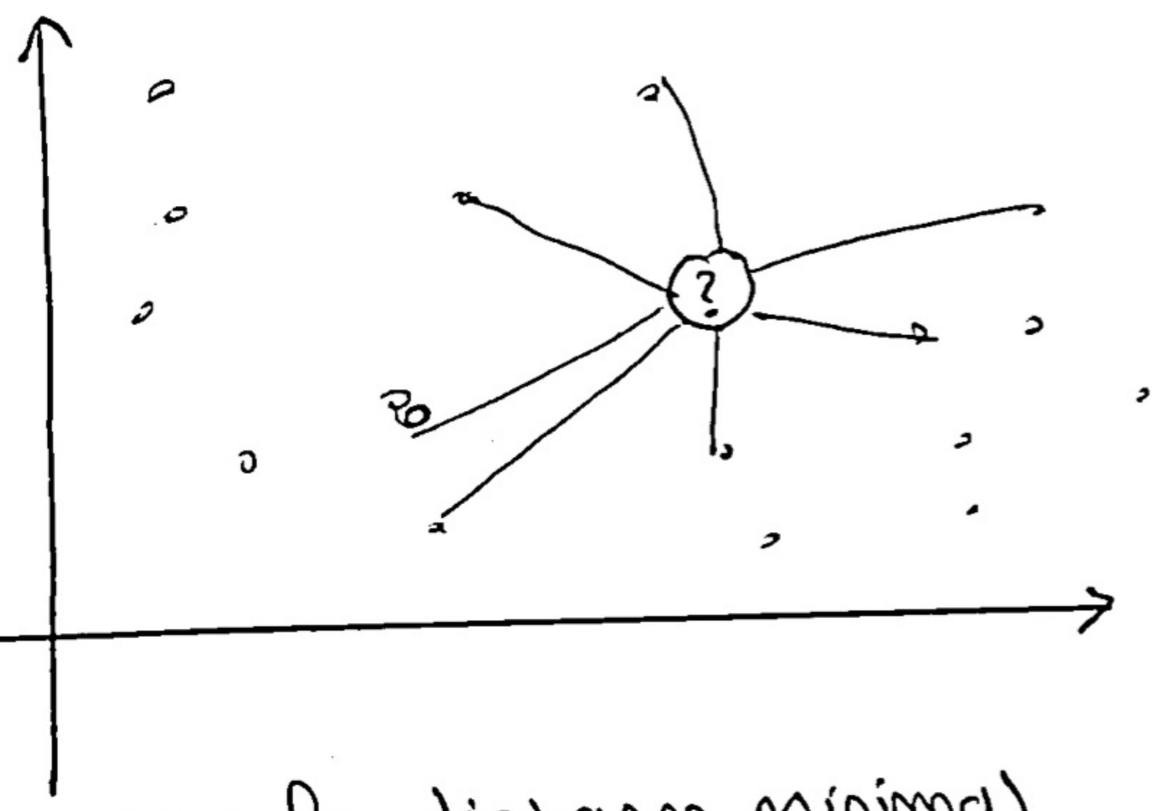


Dessision tree

→ Introduction to classification

" _____ ", ND

① → K - Nearest Neighbours K^{nn}



→ on choisir la distance minimal

$$Dis(x_1, x_2) = \sqrt{\sum_{i=0}^{n} (x_{1i} - x_{2i})^2}$$

$K = n$ $n \in \mathbb{N}$ $n \in class = [1, n]$

• Mesure of randomness of uncertainty : ①

• exemple :

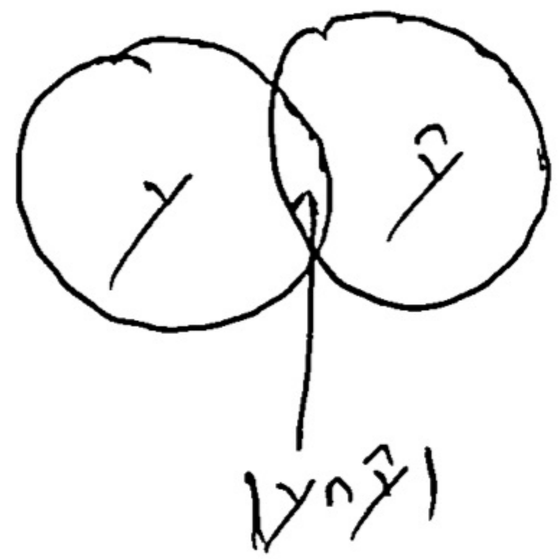
y : Actual labels

$\hat{y}$: Predict labels

→ Error

→ Jaccard index

$$J(y, \hat{y}) = \frac{|y \cap \hat{y}|}{|y \cup \hat{y}|}$$

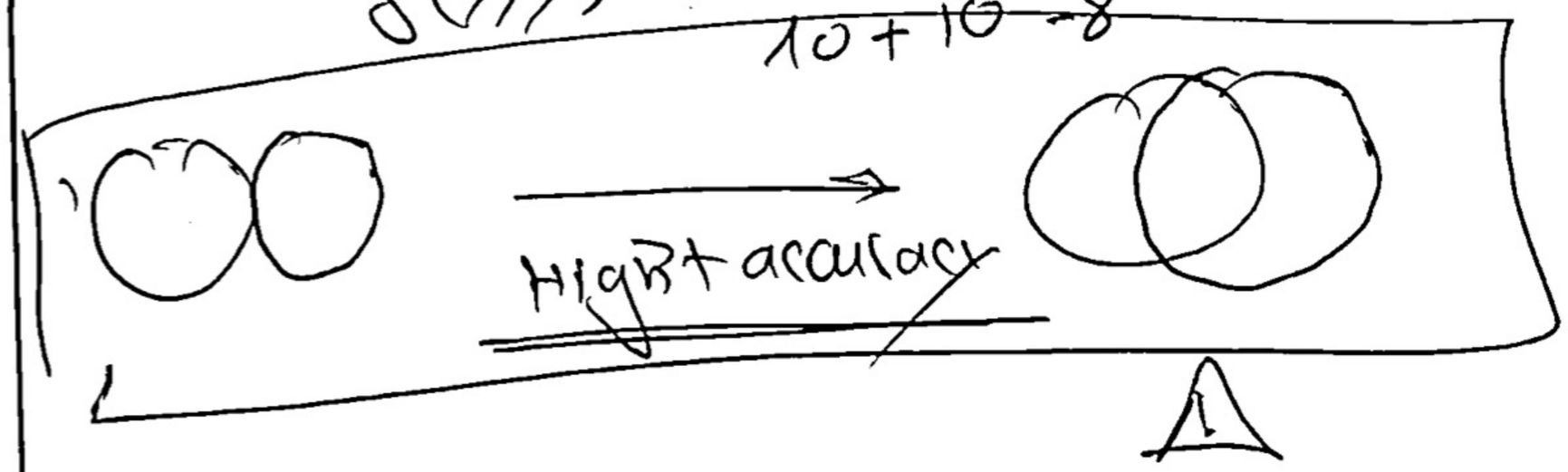


$$= \frac{|y \cap \hat{y}|}{|y| + |\hat{y}| - |y \cap \hat{y}|}$$

Exemple $y = [0, 0, 0, 1, 1, 1, 1]$

$\hat{y} = [1, 0, 0, 1, 1, 1, 1]$

$$J(y, \hat{y}) = \frac{8}{10 + 10 - 8} = 0,66 \Rightarrow 66\%$$



↦ Reda ouzidane "MacRine 2earning"

⊕2earning⊕

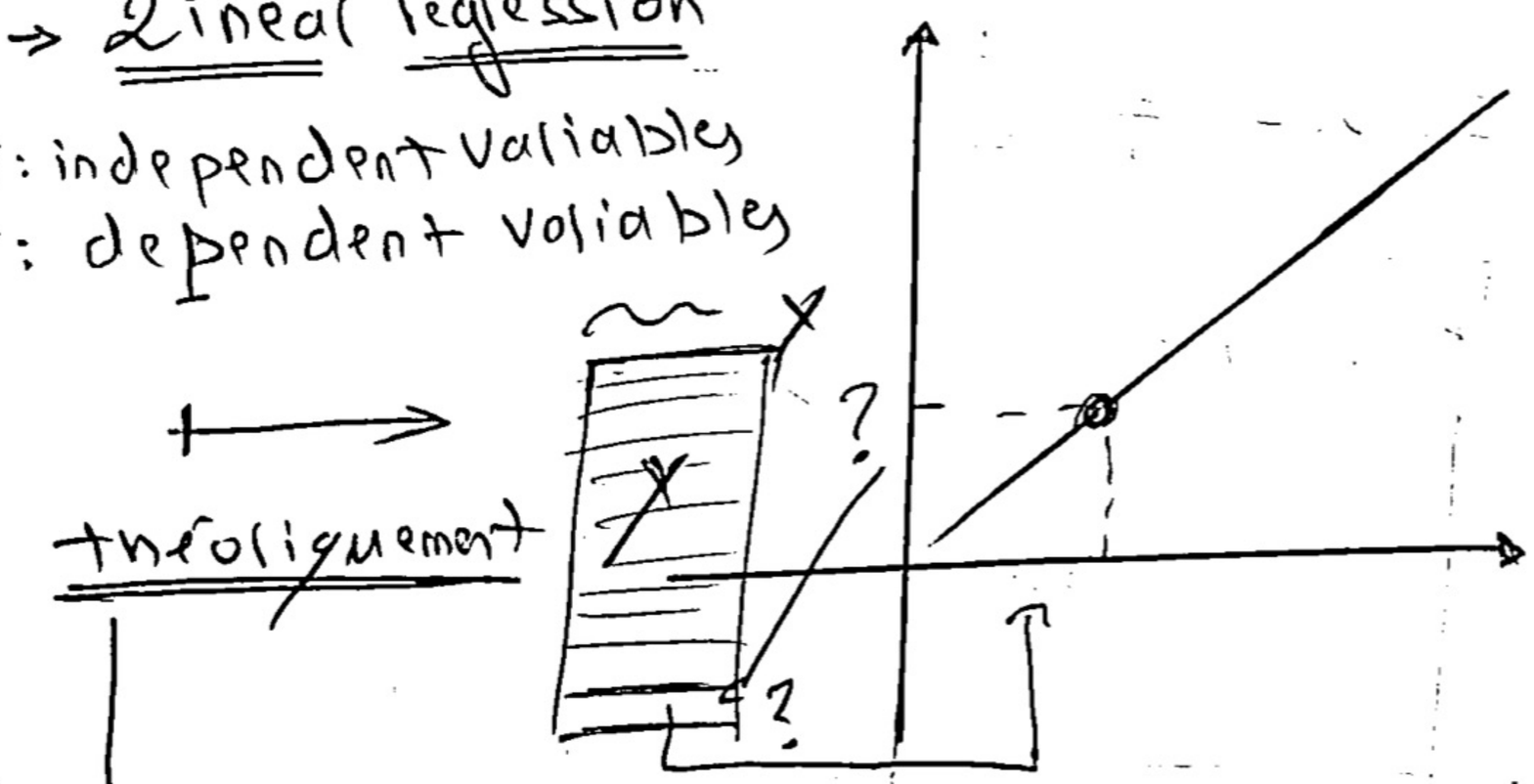
→ supevised learning

→ Regressiom : predicting continious values ; Nb.

$n \in \mathbb{R}, \mathbb{N}, \mathbb{Z}, \mathbb{C}$

① → 2inear regression

x : independent valiables
y : dependent valiables

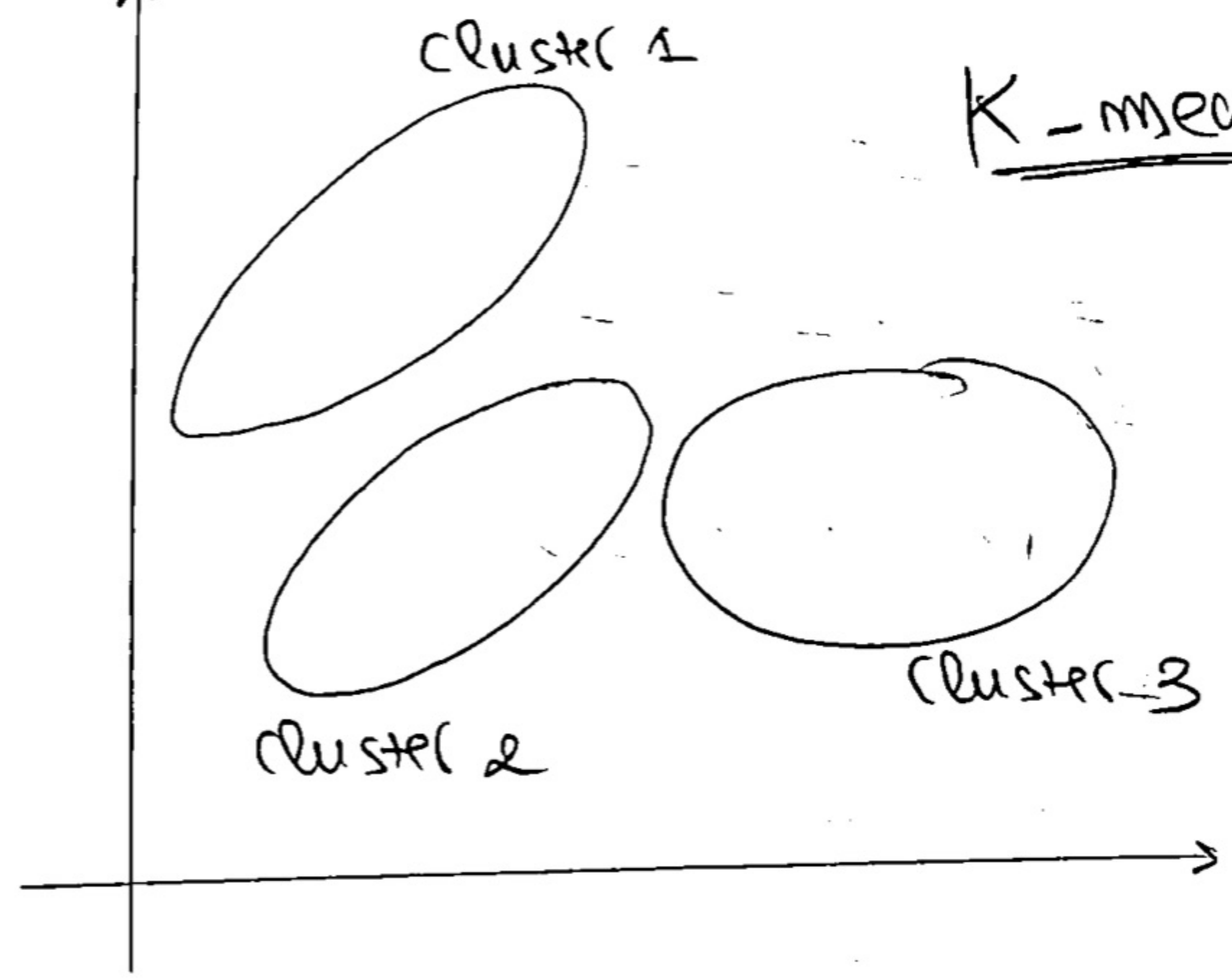


→ unsupervised learning

→ model discover info

→ Dimention reduction, Density estimation

→ clustering



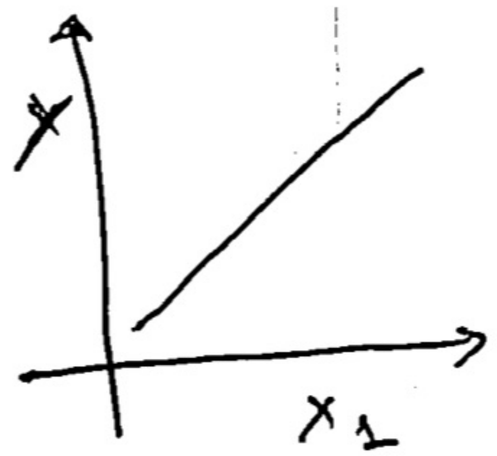
K-means

→ math (solo preuve) ⚠

$\hat{y} = \theta_0 + \theta_1 x_1$

we know

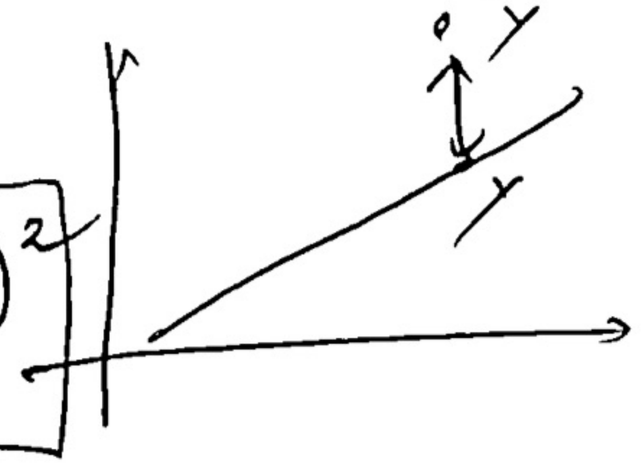
?



↳ calculate the Main squar error

$$MSE=(y-\hat{y})^2$$

$$MSE=\frac{1}{n}\sum_{i=1}^{n}(y_i-\hat{y}_i)^2$$



↳ finding the best palametels :

$$\hat{y}=\theta_0+\theta_1 x$$

$$\theta_1=\frac{\sum_{i=1}^{n}(x_i-\bar{x})(y_i-\bar{y})}{\sum_{i=1}^{n}(x_i-\bar{x})^2}$$

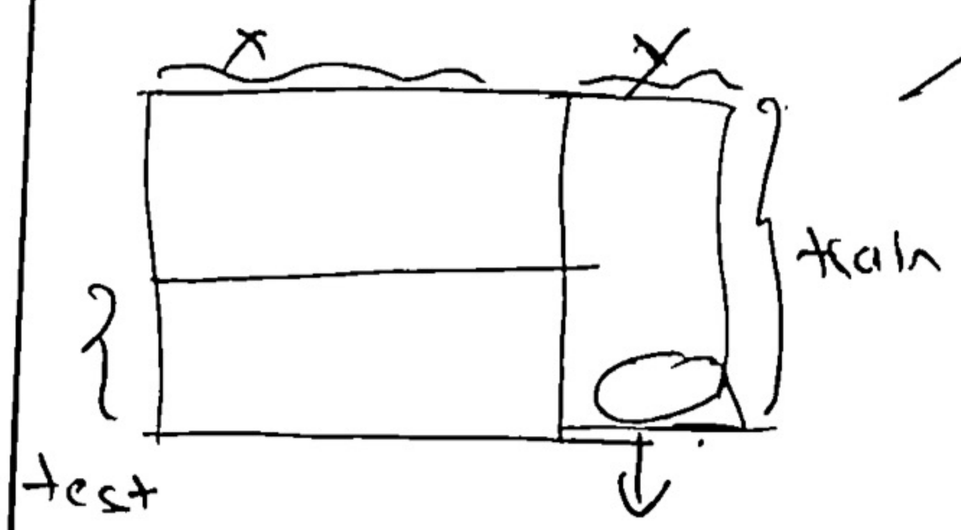
→ Pros of linear regression

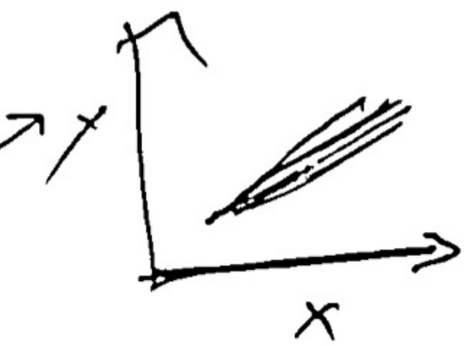
- very fast
- No palameter tuning
- Easy to understand
- High accuracy

Re..

→ Model Evaluation

- Train / Test Data



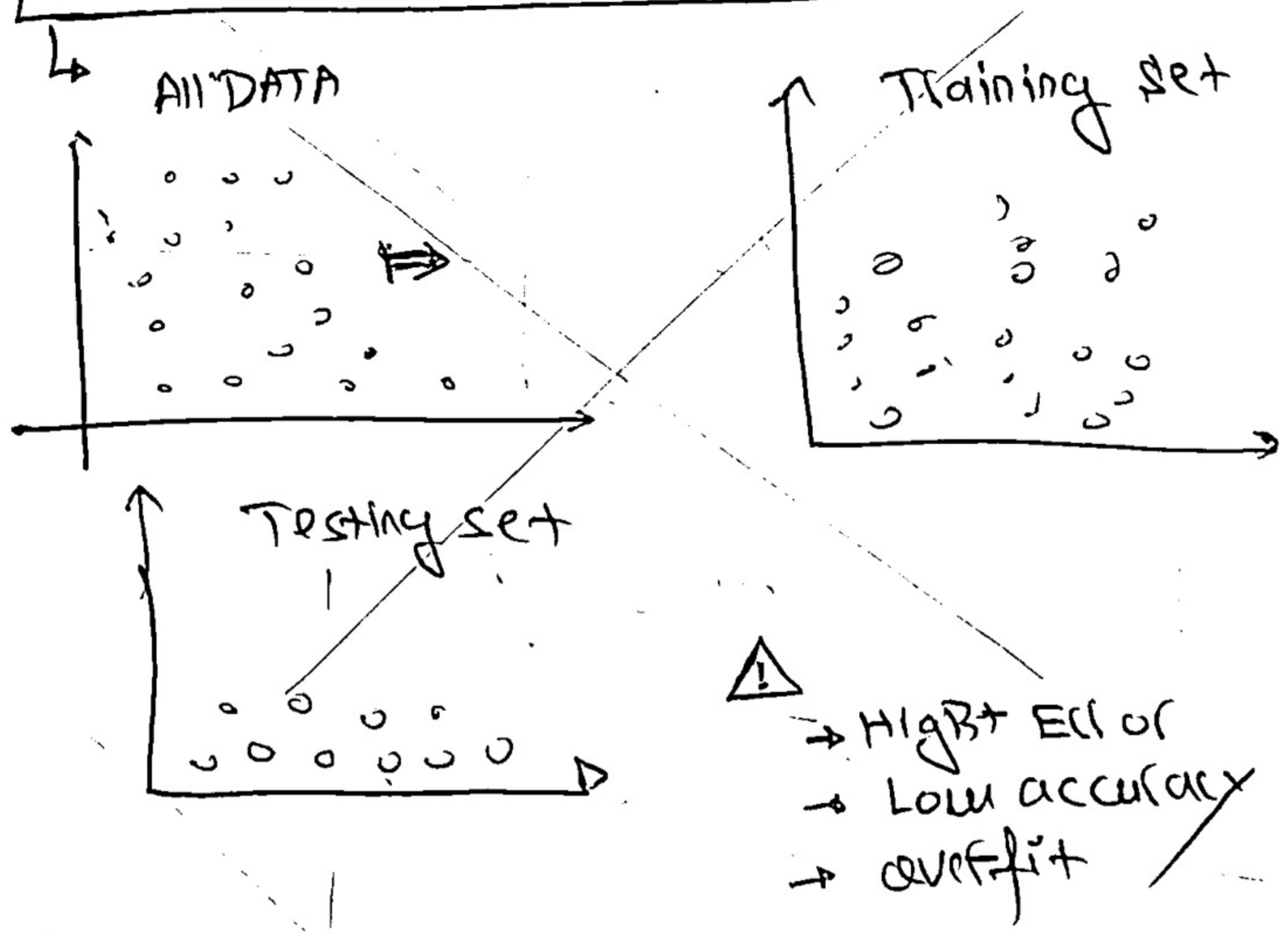


Prediction

→ calculate the Error

$$Error=\frac{1}{n}\sum_{j=1}^{n}|y_i-\hat{y}_i|$$

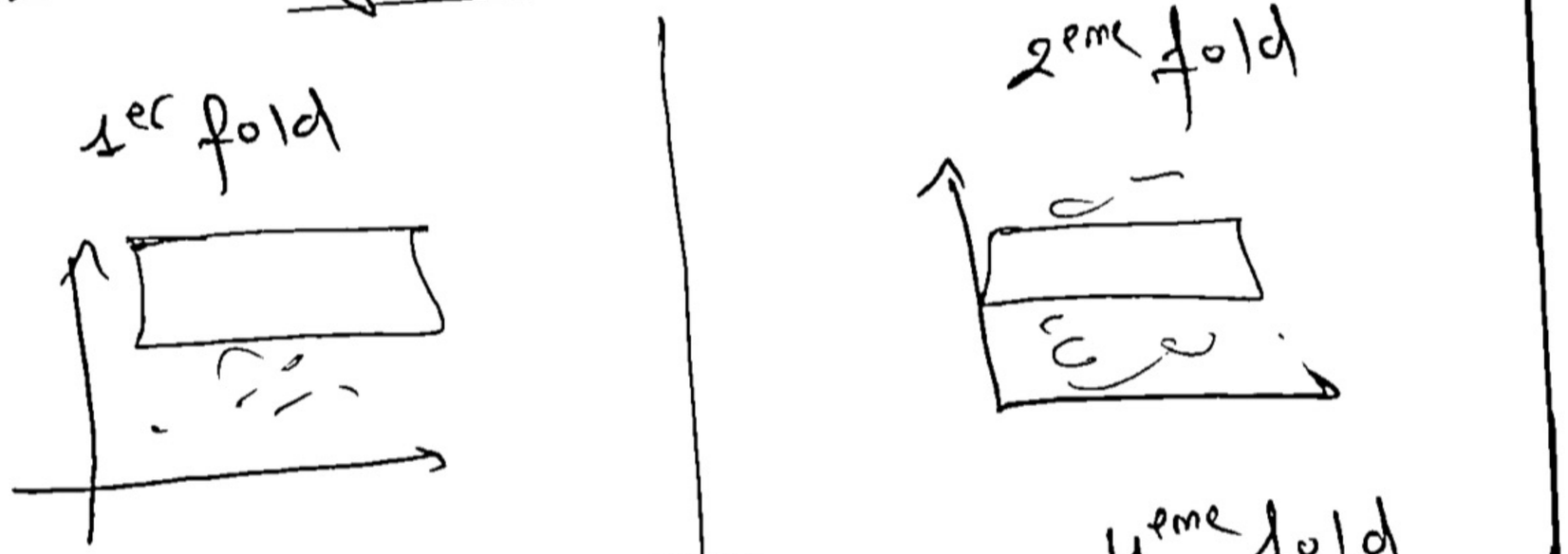
↳

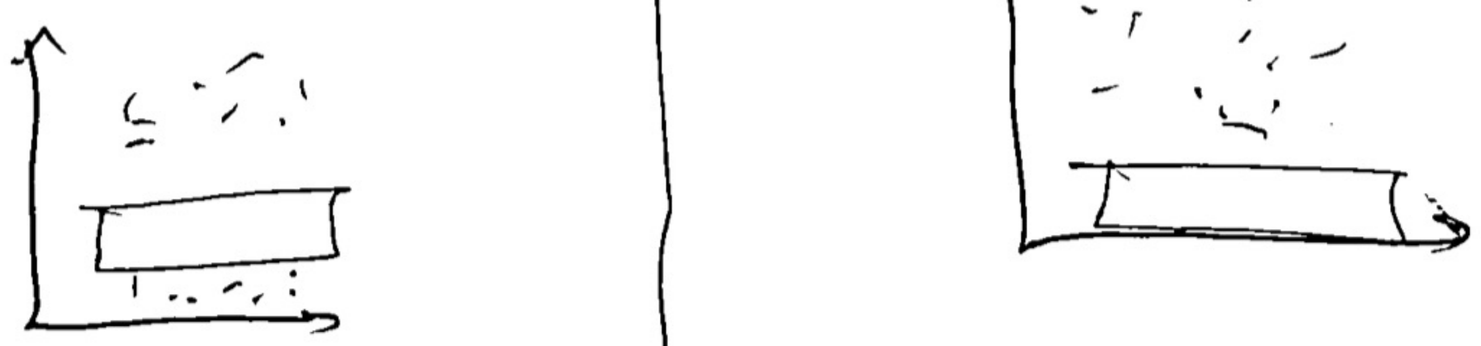


⚠
→ High Error
→ Low accuracy
→ overfit

- High accuracy

Algorithme k-fold cross-validation

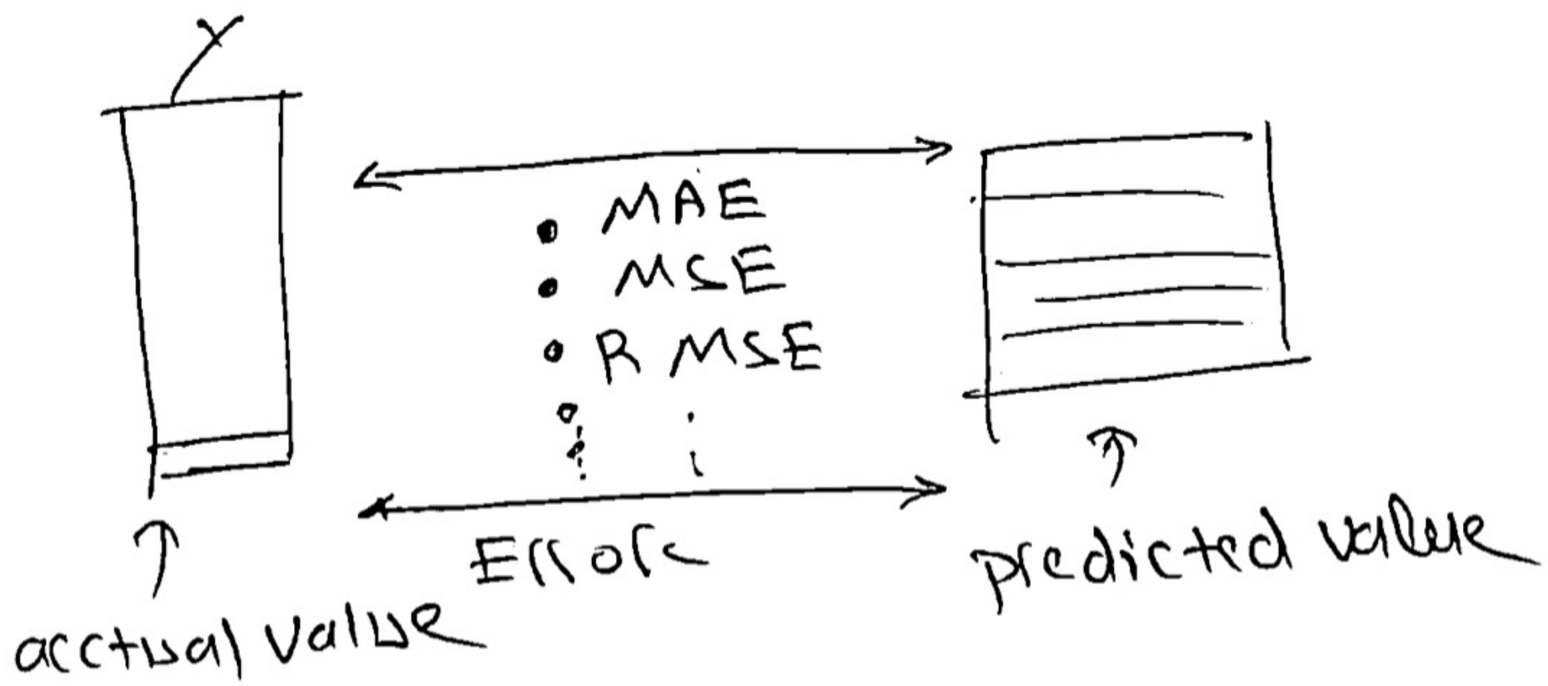




testing
training

→ High accuracy (80% → 90%)

- metrics in Regression Model



$$MAE = \frac{1}{n}\sum_{j=1}^{n} |y_i - \hat{y}_j|$$

$$MSE = \frac{1}{n}\sum (y_i - \hat{y}_i)^2$$

$$RMSE = \sqrt{\frac{1}{n}\sum_{j=1}^{n} (y_i - \hat{y}_i)^2}$$

$$RAE = \frac{\sum |y_i - \hat{y}_i|}{\sum |y_i - \bar{y}|}$$

$$RSE = \frac{\sum (y_i - \hat{y}_i)^2}{\sum (y_i - \bar{y}_i)^2}$$

more precision