Niveaux d'AI

Intelligence	vs	Artificial Intelligence
Sensoring	→	Data { signal, image, video } Processing
Optimization	→	Bio-inspired optimization
Learning	→	Machine Learning
Reasoning	→	Logique "floue"

Methodes de M. Learning

Training Process → CRISP-DM (Cross industry Standard Process for Data Mining): → Model

Experience
Data Set

1. Business and Data understanding
2. Data Preperation
3. Data Analysis
4. Machine Learning
5. Performance Eval
6. Deploiment

Testing Process → Projection → Model

ex: → sol
→ N. sol

Nouvelle experience

Types of M. Learning

- **Non annotated Data** → unsupervised Learning « clustering »
  - { $X_1 \cdots X_N$: ✓ ; Decision : ? }
- **Annotated Data** → supervised Learning
  - { $X_1 \cdots X_N$: ✓ ; Decision : ✓ }
  - → Numerical → Regression : Forecast (prevision)
  - → Categorical → Classification : prediction

• Example :

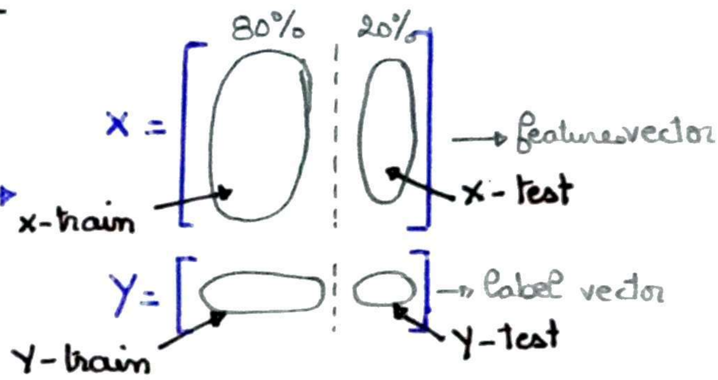
$\text{error} = (y\_test) - (y\_pred)$

$$\left.\begin{aligned} &\bullet\ MSE = \frac{1}{N} \sum e_i^2 \\ &\bullet\ MAE = \frac{1}{N} \sum |e_i| \\ &\bullet\ MAPE = \frac{1}{N} \sum \frac{|e_i|}{y\_test} \times 100 \end{aligned}\right\} \longrightarrow \text{Performance evaluation}$$

Forecasting

→ **Multivariate :**

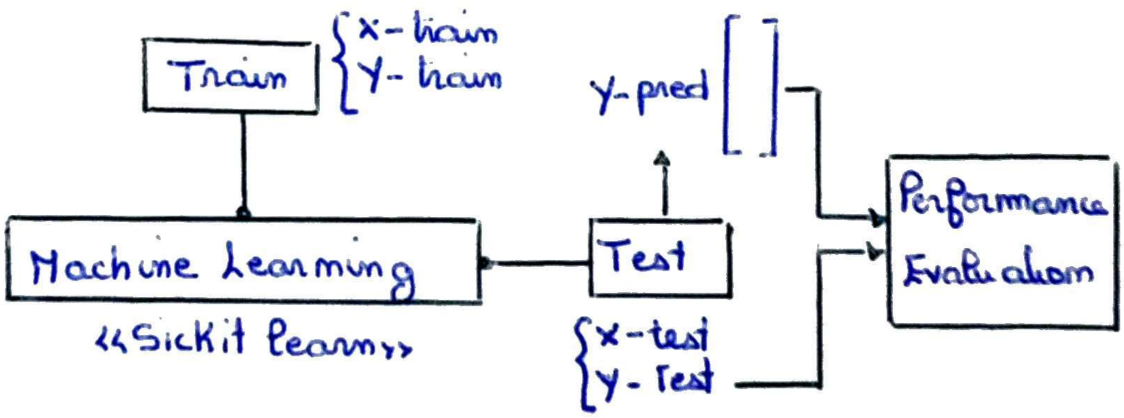
$$\begin{array}{c|c} x_1 \cdots x_N & \text{value} \\ \begin{bmatrix} & X & \end{bmatrix} & \begin{bmatrix} v_1 \\ \vdots \\ v_N \end{bmatrix} \end{array} \longrightarrow$$



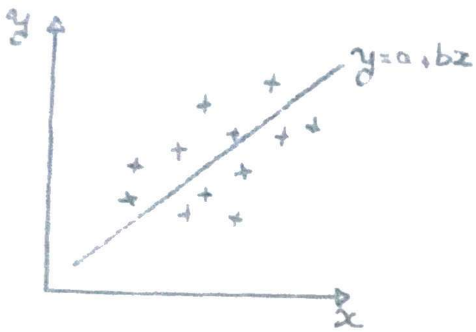
→ **Univariate :**

$$\begin{array}{c|c} \text{time} & \text{value} \\ D_1 & v_1 \\ \vdots & \vdots \\ D_N & v_N \end{array} \quad \left.\right\} \text{history : Seq-length} \longrightarrow X = \begin{bmatrix} v_1 & v_2 & v_3 \\ v_2 & v_3 & v_4 \\ v_3 & v_4 & v_5 \end{bmatrix} \qquad y = \begin{bmatrix} v_4 & v_5 & v_6 \end{bmatrix}$$

Load Data



Linear Regression Algorithm

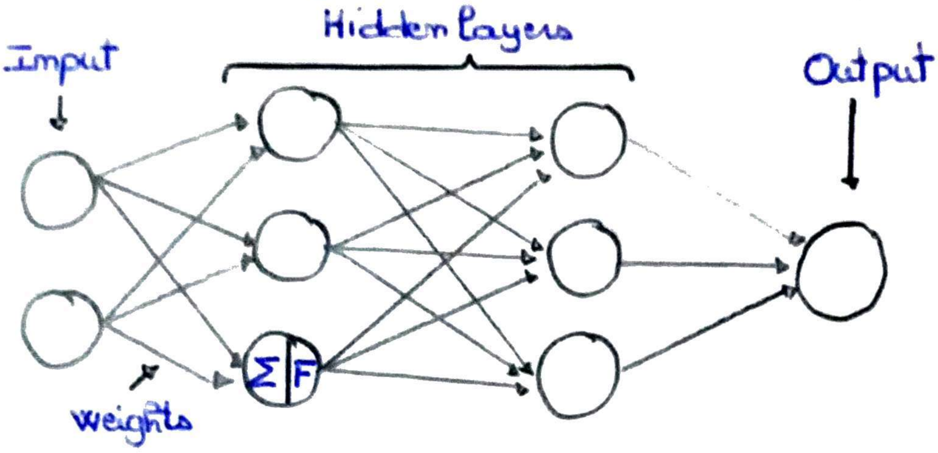


$$y = a + bx$$

$$\boxed{a = \bar{Y} - b\bar{X}}$$

$$\boxed{b = \frac{\sum x_i y_i - m\bar{X}\bar{Y}}{\sum x_i^2 - m\bar{X}^2}}$$

Artificial Neural Network

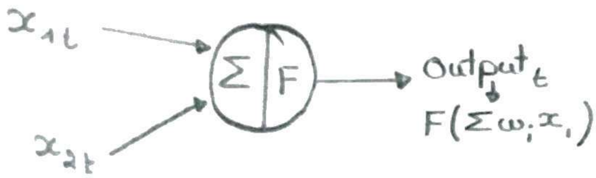


Input → weight → Sum → function of activation → output

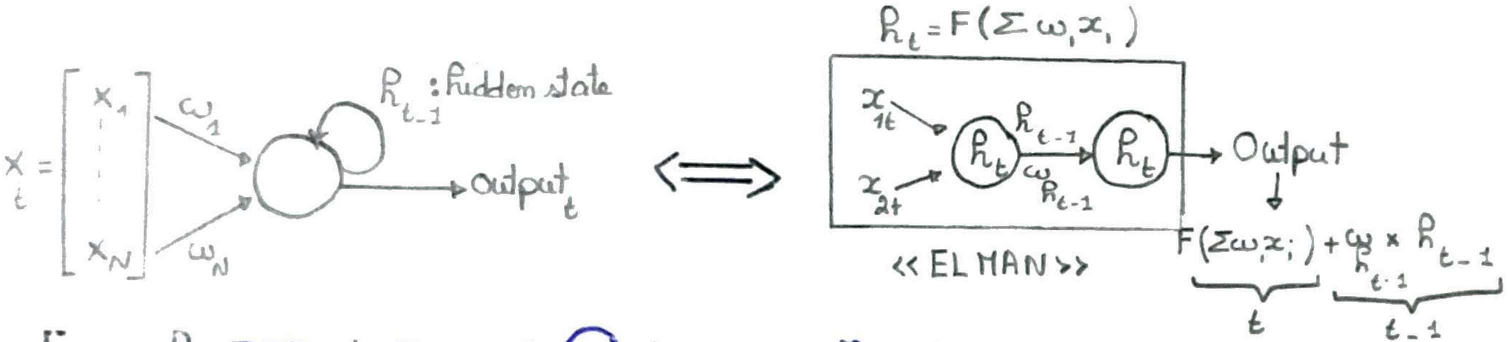
Machine Learning for Regression

- Statistical: Linear Regression
- Learning: MLP Regression

Feed - Forward



Recurrent Neural Network (RNN)

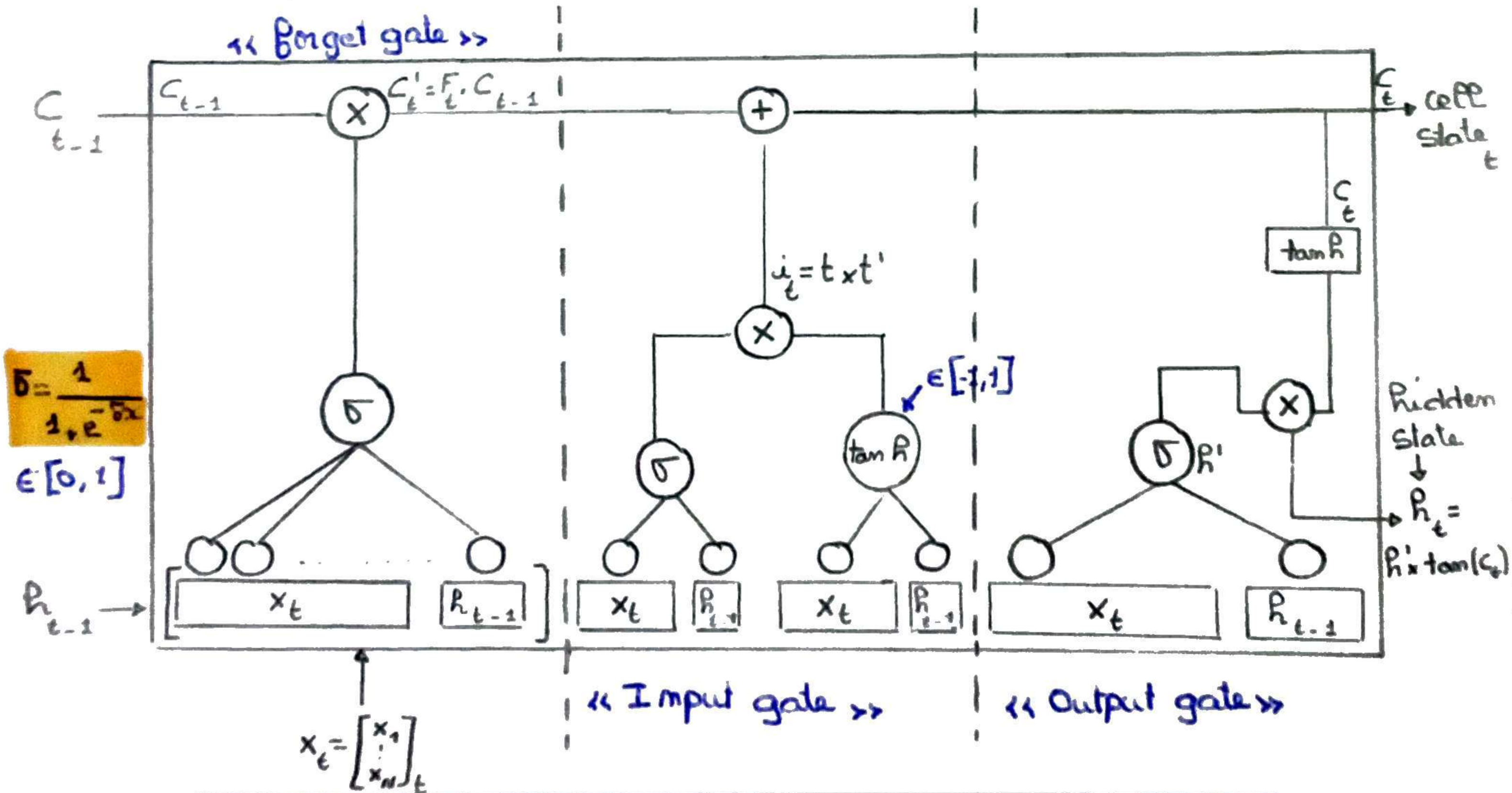


• Example : IHEC is the best . It is an excellent institute .

× context

« Vanishing Problem »

Long-short Term Memory (LSTM)



Clustering

K-means (K-moyenne)

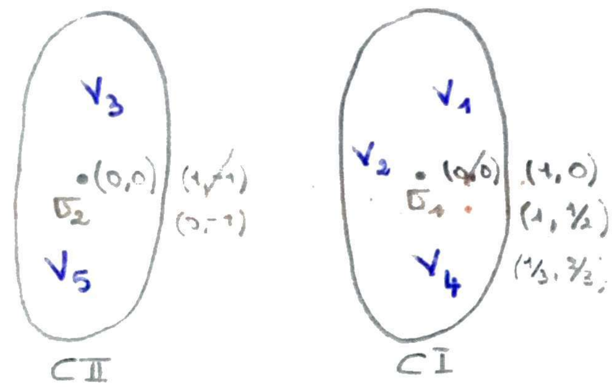
dataset =

	x_1	x_2
v_1	1	0
v_2	1	1
v_3	1	-1
v_4	-1	1
v_5	-1	-1

K : number of clusters

σ_i : the centroid of cluster C_i

1/ $K = 2$



Rule 1: Update the (C_i) after each assignment

2/ Assign V_1 randomly

3/ Assign next vectors based on Euclidean Distance

$$d = \sqrt{(x_1-y_1)^2+\dots\dots+(x_m-y_m)^2}$$

Rule 2: Assign randomly when it's equal

V_2: $d(V_2,\sigma_1) = \sqrt{(1-1)^2+(1-0)^2} = 1$

$d(V_2,\sigma_2) = \sqrt{(1-0)^2+(1-0)^2} = \sqrt{2}$

V_3: $d(V_3,\sigma_1) = \sqrt{(1-1)^2+(-1-1/2)^2} = \frac{3}{2}$

$d(V_3,\sigma_2) = \sqrt{(1-0)^2+(-1-0)^2} = \sqrt{2}$

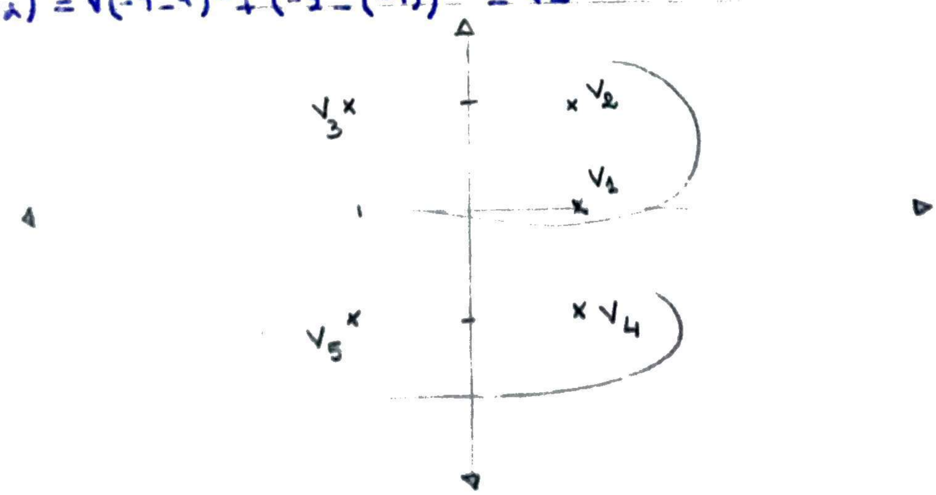
V_4: $d(V_4,\sigma_1) = \sqrt{(-1-1)^2+(1-1/2)^2} = \sqrt{\frac{17}{4}}$

$d(V_4,\sigma_2) = \sqrt{(-1-1)^2+(1-(-1))^2} = \sqrt{8}$

$-\frac{16}{9}+\frac{25}{9}$

V_5: $d(V_5,\sigma_1) = \sqrt{(-1-1/3)^2+(-1-2/3)^2} = \sqrt{\frac{41}{9}}$

$d(V_5,\sigma_2) = \sqrt{(-1-1)^2+(-1-(-1))^2} = \sqrt{2}$



Classification

Naive bayes :

X_1	X_2	X_3	Class
0	1	1	False
1	-1	1	False
0	-1	0	True
1	1	0	True
1	1	0	False

$P(A/B) = ?$

(X_i) are independant

Train

X_1: $P(0/True) = 1/2$ $\quad$ $P(0/False) = 1/3$

$P(1/True) = 1/2$ $\quad$ $P(1/False) = 2/3$

X_2: $P(1/True) = 1/2$ $\quad$ $P(1/False) = 2/3$

$P(-1/True) = 1/2$ $\quad$ $P(-1/False) = 1/3$

X_3: $P(1/True) = 0$ $\quad$ $P(1/False) = 2/3$

$P(0/True) = 2/2$ $\quad$ $P(0/False) = 1/3$

Test

$$X = \langle \overset{x_1}{1}, \overset{x_2}{-1}, \overset{x_3}{0} \rangle$$

$$P(X/True) = P(X_1 = 1/True) \times P(X_2 = -1/True) \times P(X_3 = 0/True)$$

$$= \frac{1}{2} \times \frac{1}{2} \times 1 \times \underbrace{\frac{2}{5}}_{P(True)} = P_T$$

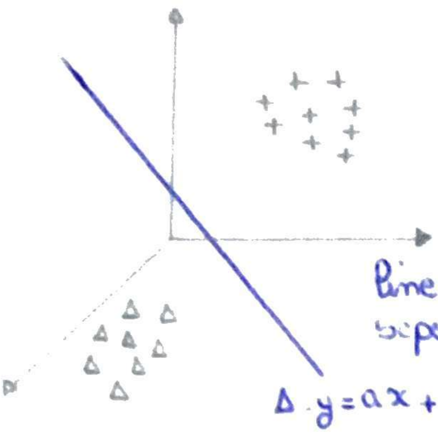
$$P(X/False) = P(1/False) \times P(-1/False) \times P(0/False)$$

$$= \frac{1}{3} \times \frac{1}{3} \times \frac{1}{3} \times \underbrace{\frac{3}{5}}_{P(False)} = P_F$$

$P_T < P_F \longrightarrow F$

$P_T > P_F \longrightarrow T$

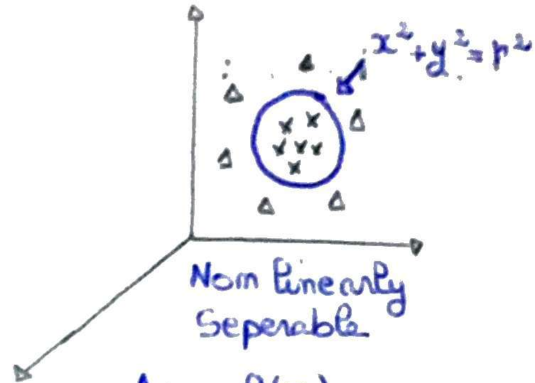
Support Vector Machine (SVM):



Linearly seperable → Linear

$\Delta\ y = ax + b$

VS

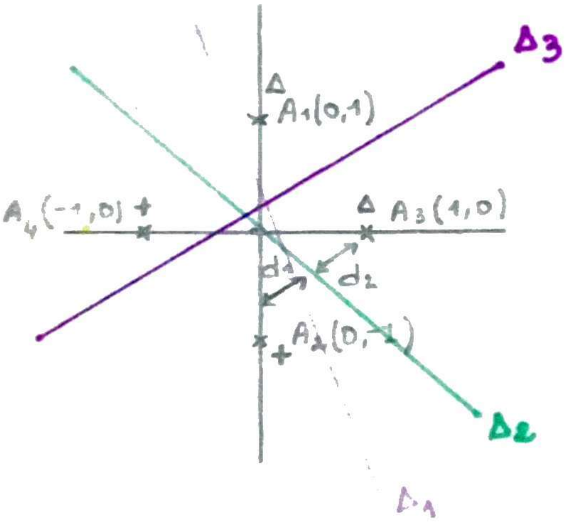


Non linearly Seperable

$\Delta : y = f(x)$

Eg non linear
- RBF
- Sigmoid
- Polynomial

• 1st case : Linear Separation



1/ Initial $(N-1)$ separator randomly (number of samples)

$\Delta_1 : a_1 x + b_1$

$\Delta_2 . a_2 x + b_2$

$\Delta_3\ a_3 x + b_3$

2/ Compute precision of each (Δ_i)

$\Delta_1 : 100\%$

$\Delta_2 : 100\%$

$\Delta_3 : 50\%$

3/ $P > 85\%$

Exclude Δ_3

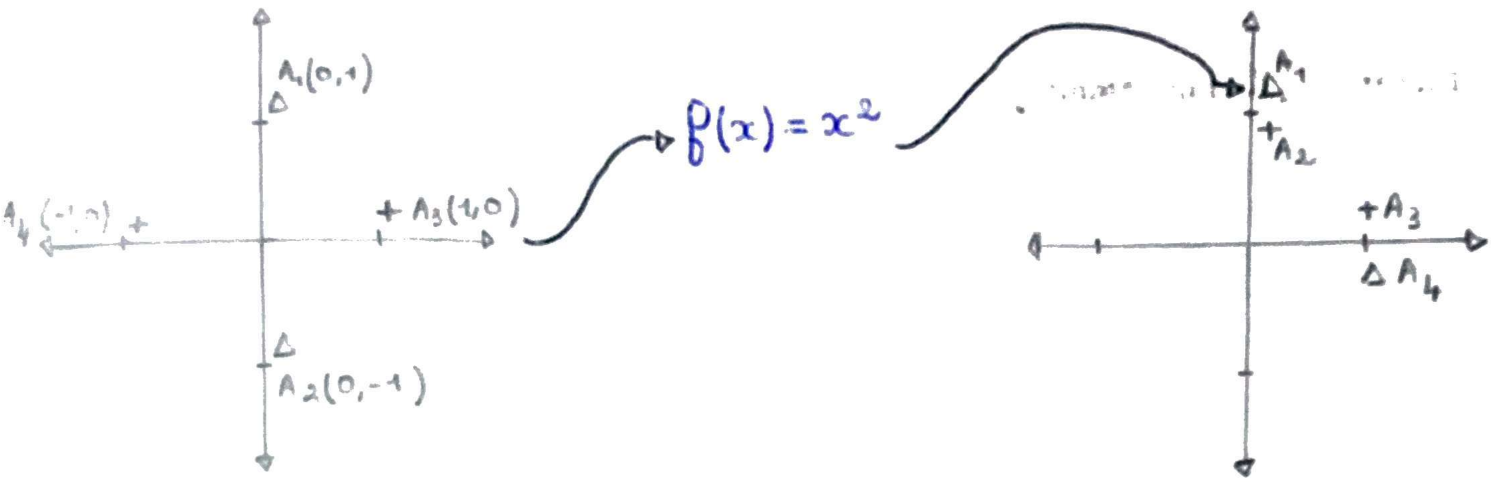
4/ Δ_1 vs Δ_2

M = margin distance $= d_1 + d_2$

Si $M_1 > M_2 \longrightarrow \Delta_1$

$M_1 < M_2 \longrightarrow \Delta_2$

- 2^{nd} case: Non linear separation.



Method of Machine Learning

1/ Business and Data Understanding

- Nbr of classes
- Nbr of variables and their quality
- Nbr of samples

2/ Data Preparation

- 2.1 Data → Missed values
- 2.2 Visualization (Data Analysis)
- 2.3 → Mixed values
- 2.4 Normalization
- 2.5 Split Data

3/ Machine Learning

4/ Performance Evaluation

Confusion Matrix

	+	−
+	45	1
−	5	49

100 vectors of Test : 50(+), 50(−)

$$Accuracy = \frac{45+49}{100} \times 100\%$$

$$P(+) = \frac{45}{50} \qquad R(+) = \frac{45}{45+1}$$

$$P(-) = \frac{49}{50} \qquad R(-) = \frac{49}{49+5}$$

$$F1\text{-}score = 2 \times \frac{R \times P}{R+P}$$

$$g\text{-}mean = \sqrt{R \times P}$$

5/ Deployment