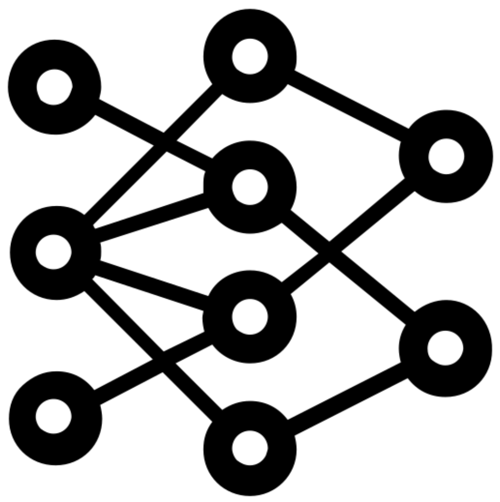
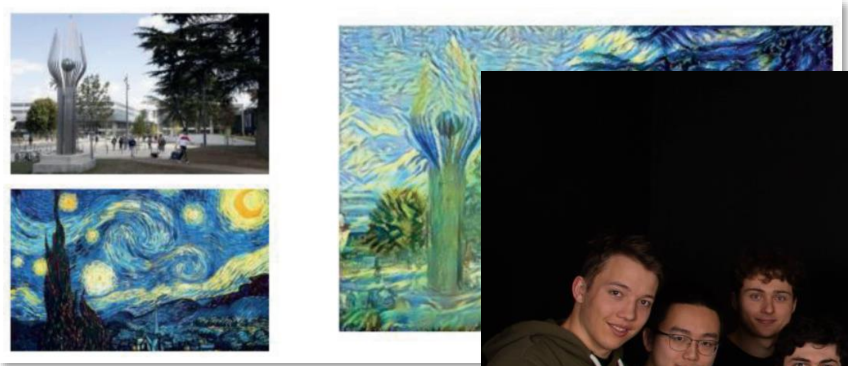
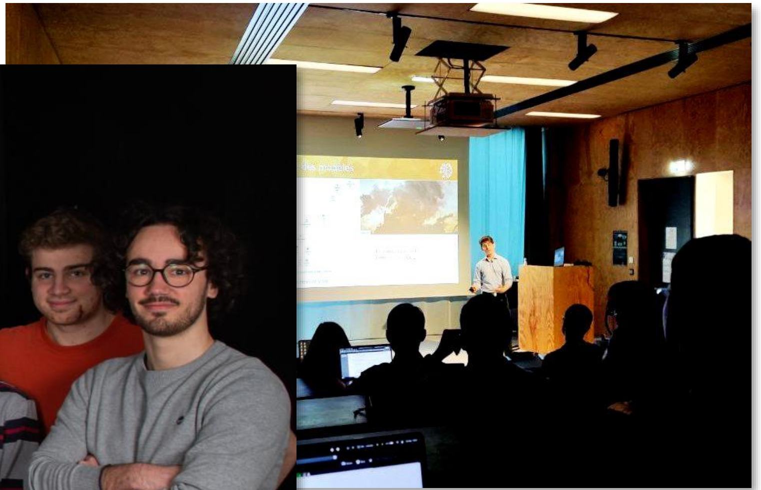
AUTOMATANTS

Introduction à l'IA et aux Réseaux de Neurones



Avant-propos





Avant-propos

Pour suivre toutes nos informations et être au courant de tous nos évènements :

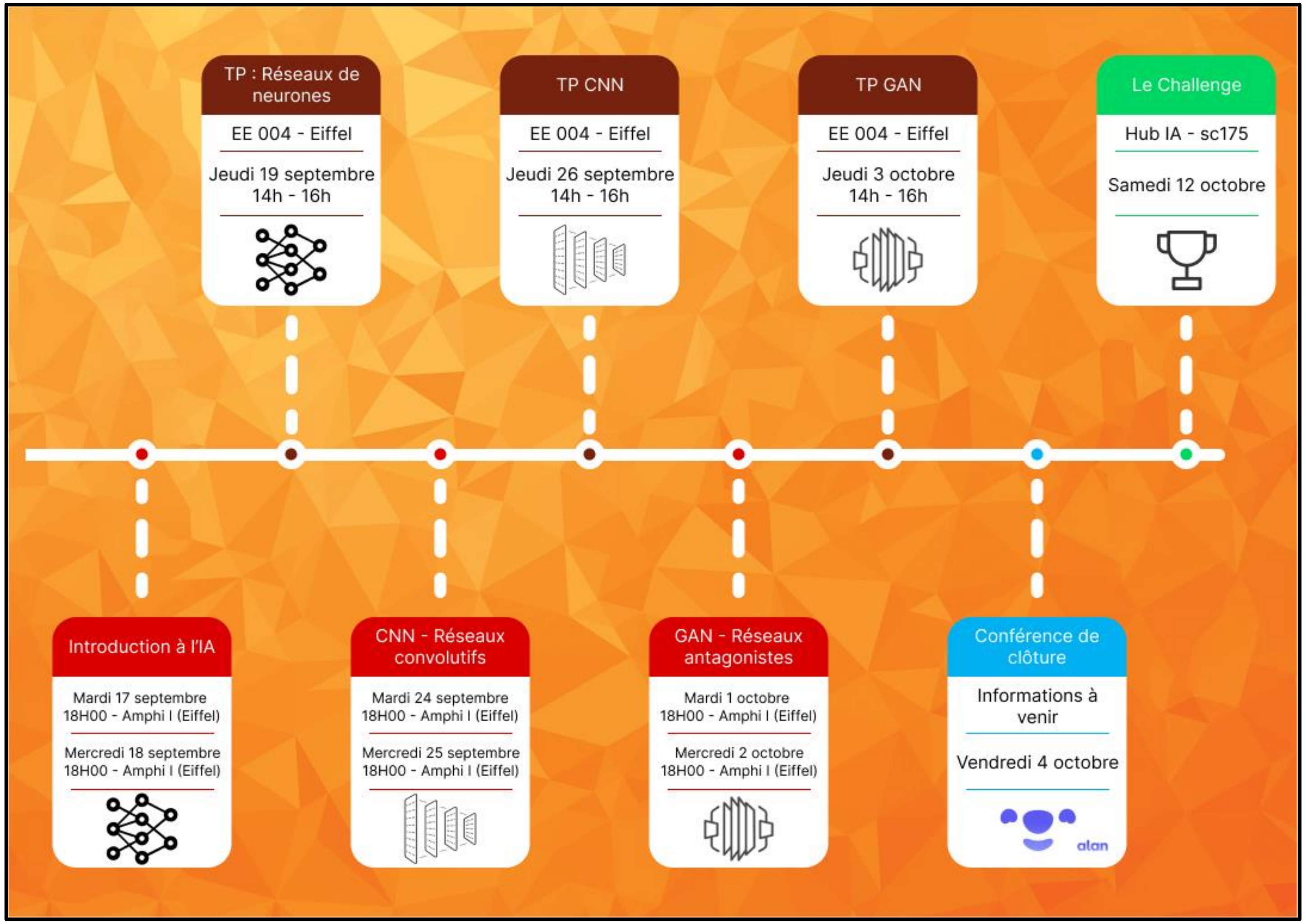
Avant-propos

Pour ne rien louper des évènements et infos Data / IA du campus et du plateau :

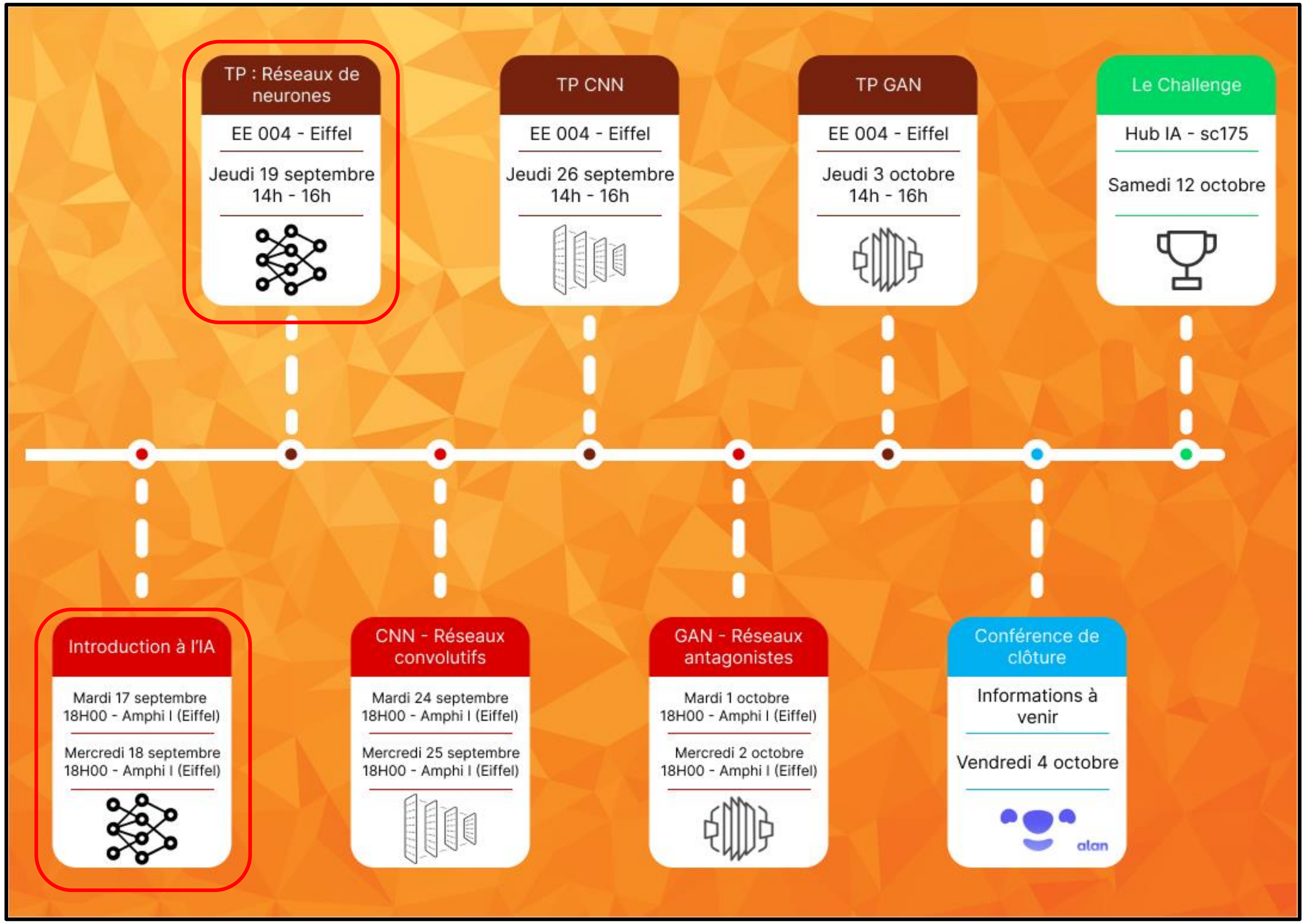
Canal Info IA

AI Campus

Avant-propos



Avant-propos



Au programme

I. L’intelligence Artificielle, késako ?

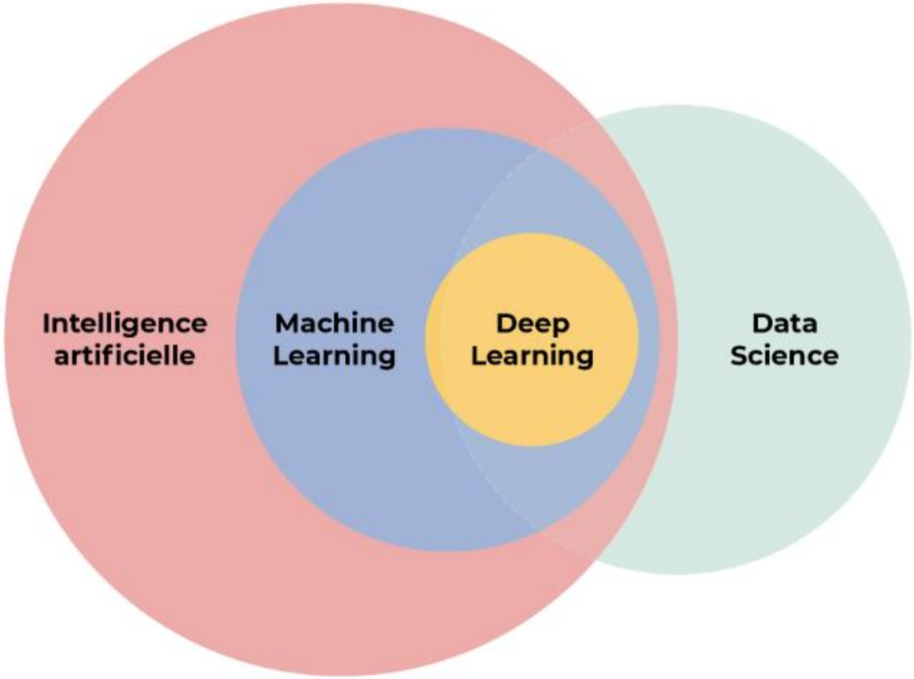
L’IA, c’est d’abord un **rêve** de science-fiction…

I. L'intelligence Artificielle, késako ?

… qui ne date pas d'**hier**

Le Turc mécanique, 1770

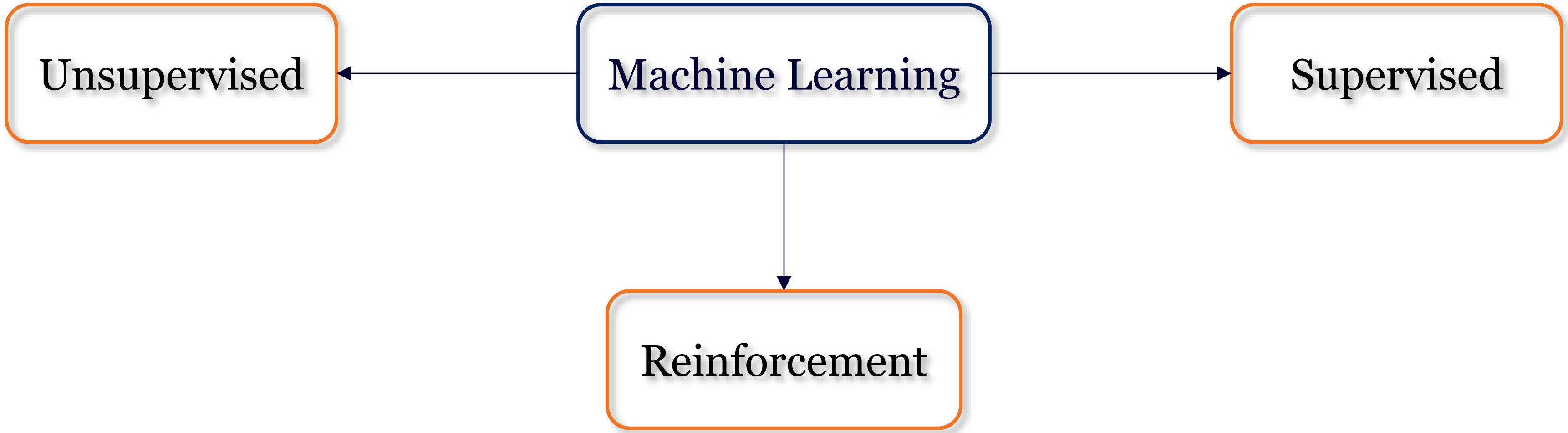
I. L'intelligence Artificielle, késako ?



Pour le **Parlement Européen**, l'intelligence artificielle représente tout outil utilisé par une machine afin de « *reproduire des comportements liés aux humains, tels que le raisonnement, la planification et la créativité* ».

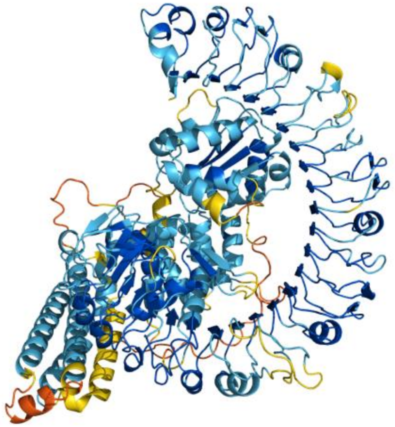
En fait, le terme **IA** veut tout et rien dire à la fois

On va surtout parler de **Machine Learning** (ML)



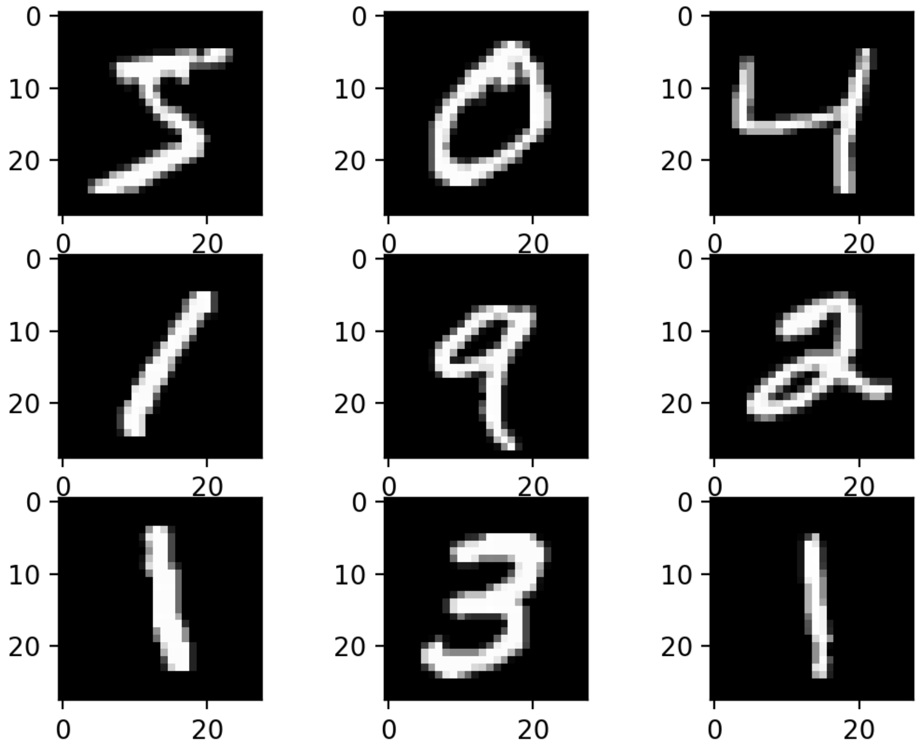
I. L'intelligence Artificielle, késako ?

Avec ces méthodes, ces algorithmes,
on peut en faire des choses



Des questions ?

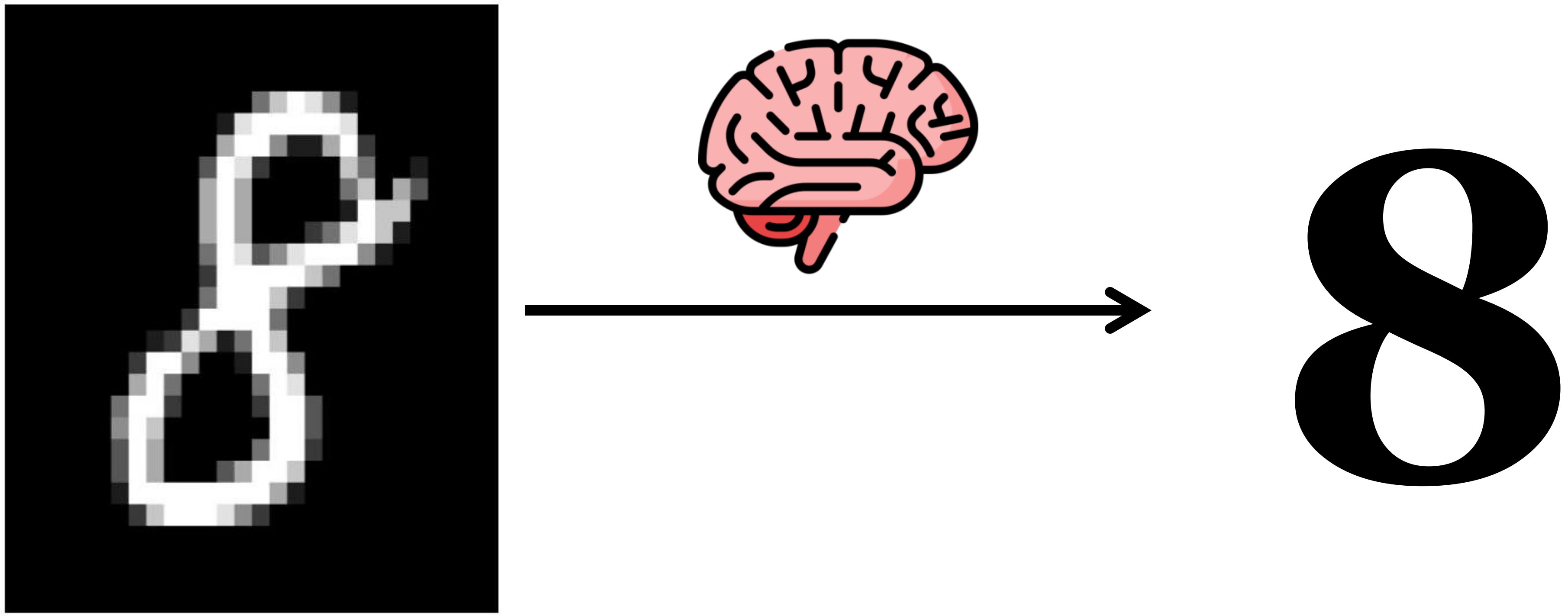
1 – Mise en contexte



MNIST Dataset

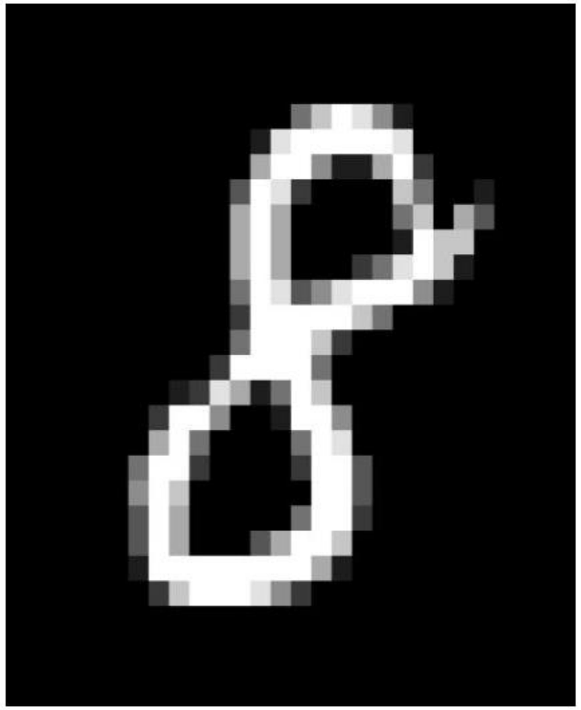
II. Un premier modèle : le réseau de neurones

1 – Mise en contexte



II. Un premier modèle : le réseau de neurones

1 – Mise en contexte

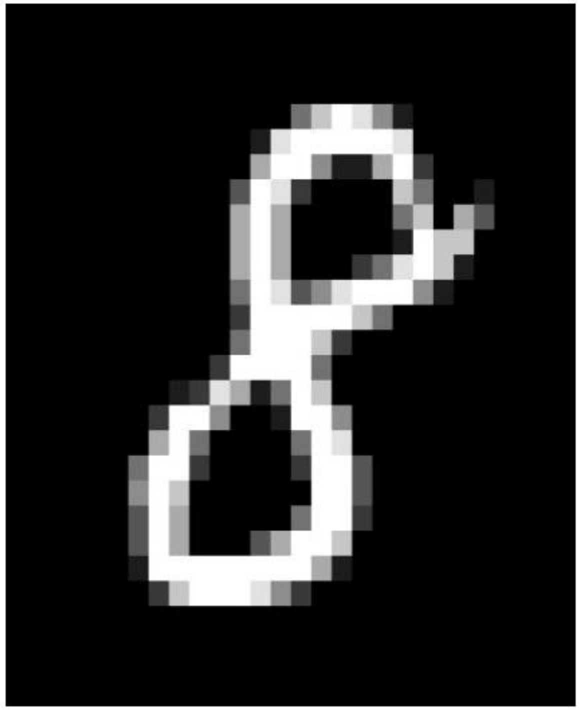


$$\xrightarrow{\quad\hat{f}\quad}$$

8

II. Un premier modèle : le réseau de neurones

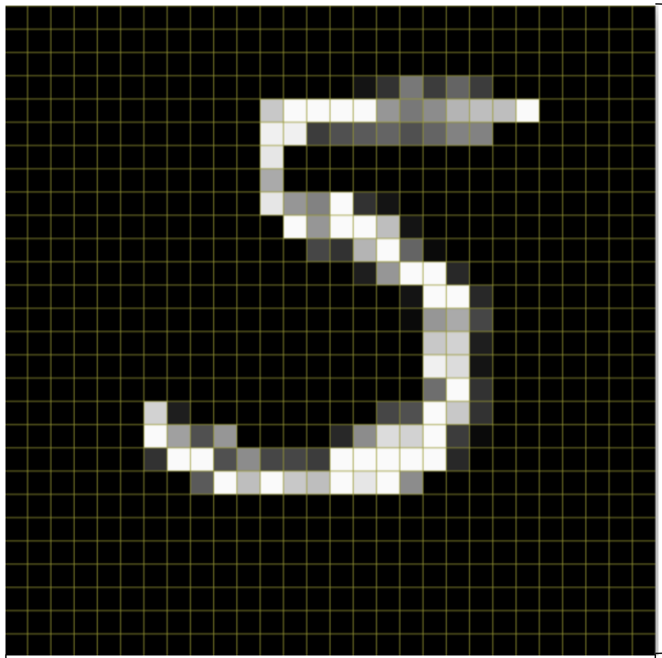
1 – Mise en contexte



$$\xrightarrow{\hat{f}} \begin{pmatrix} P(x = 0) \\ P(x = 1) \\ \vdots \\ P(x = 9) \end{pmatrix}$$

II. Un premier modèle : le réseau de neurones

1 – Mise en contexte



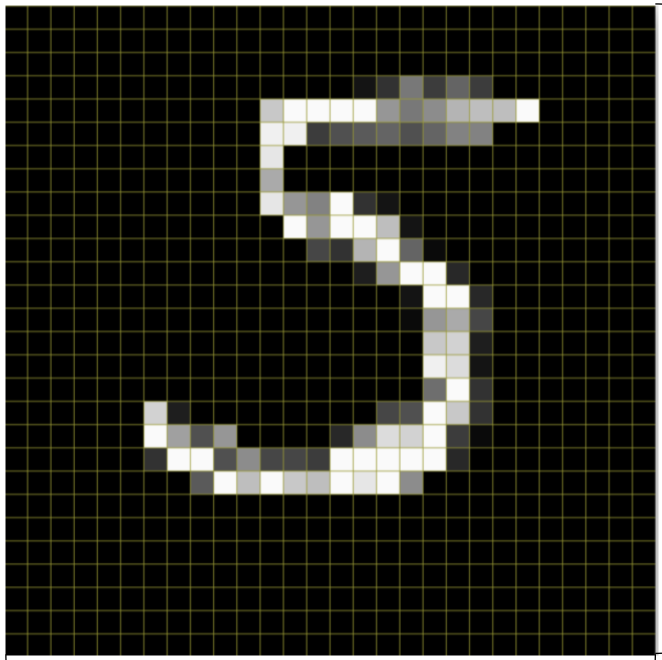
28 pixels

28 pixels

$$\longrightarrow \quad X = \begin{bmatrix} 0.13 \\ 0.67 \\ \vdots \\ 0.23 \end{bmatrix}$$

784 nombres

1 – Mise en contexte



28 pixels

28 pixels

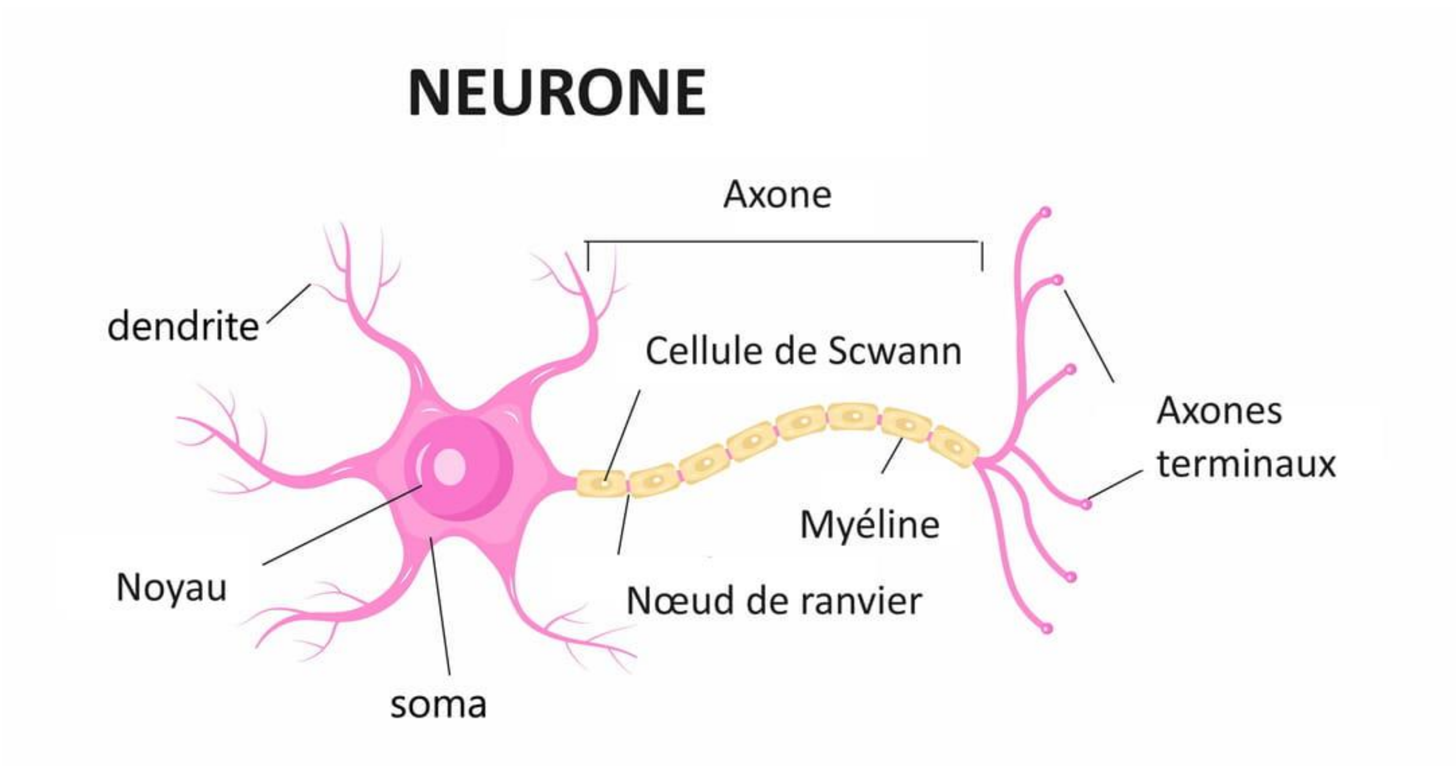
$$X = \begin{bmatrix} 0.13 \\ 0.67 \\ \vdots \\ 0.23 \end{bmatrix}$$

784 nombres

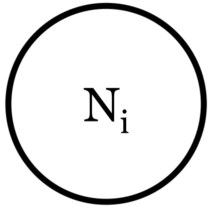
$$\hat{f} : \mathbb{R}^{784} \to \mathbb{R}^{10}$$

Des questions ?

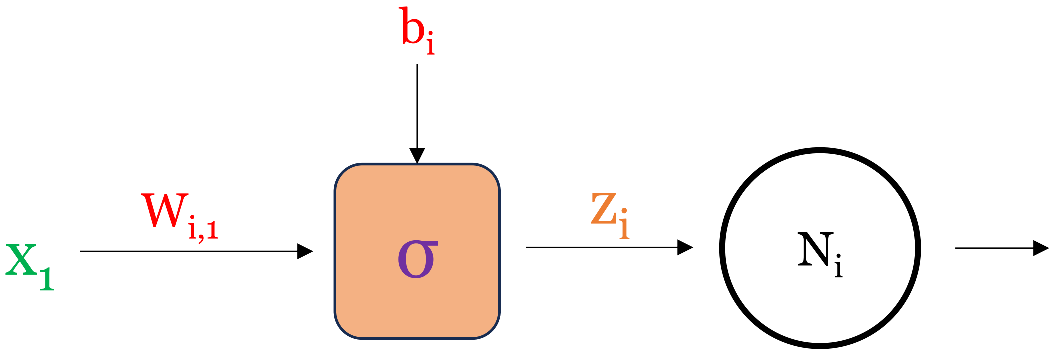
2 – Le neurone ou perceptron



2 – Le neurone ou perceptron



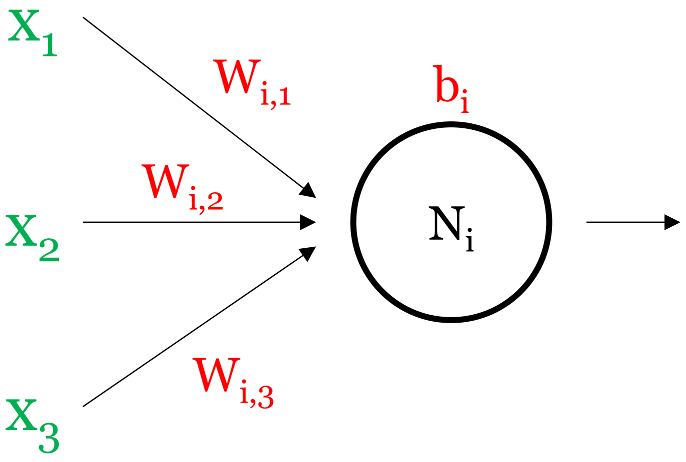
2 – Le neurone ou perceptron



$$y_i = \sigma(w_{i,1}x_1+b_i) = \sigma(z_i)$$

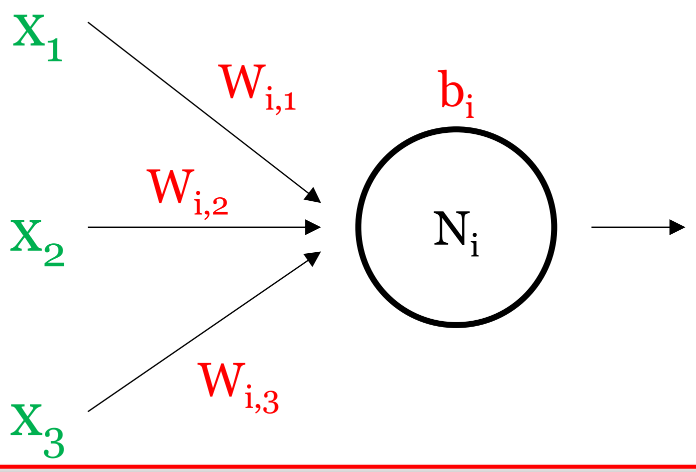
II. Un premier modèle : le réseau de neurones

2 – Le neurone ou perceptron



$$y_i = \sigma(w_{i,1}x_1 + w_{i,2}x_2 + w_{i,3}x_3 + b_i)$$
$$= \sigma(z_i)$$

2 – Le neurone ou perceptron

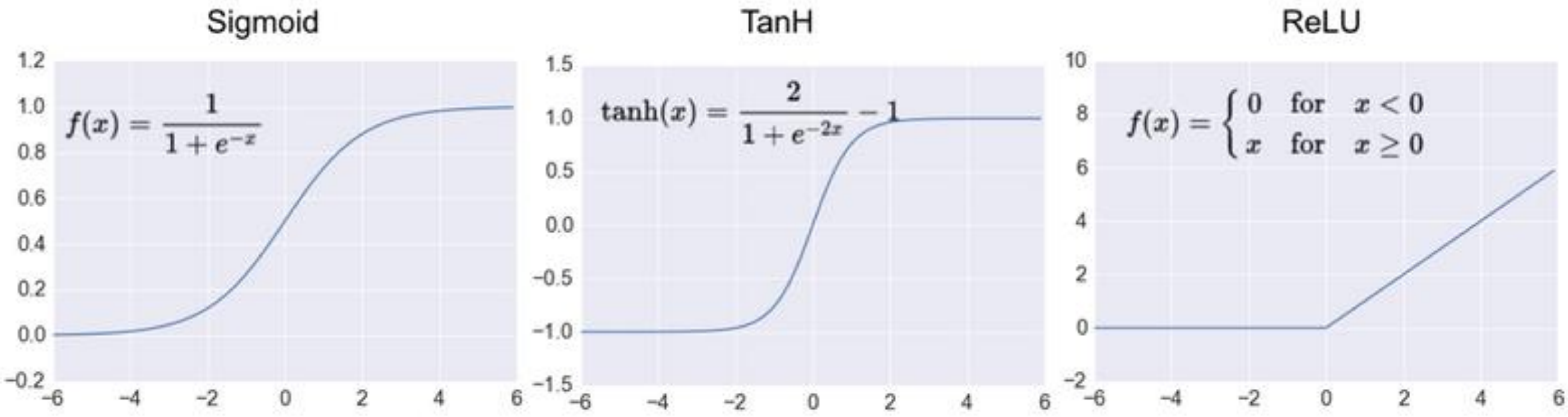


$$y_i = \sigma(w_{i,1}x_1 + w_{i,2}x_2 + w_{i,3}x_3 + b_i)$$
$$= \sigma(z_i)$$

$$\boxed{y_i = \sigma\left(\begin{pmatrix} w_{i,1} & w_{i,2} & w_{i,3} & b_i \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ 1 \end{pmatrix} \right) = \sigma(W_i . x)}$$

II. Un premier modèle : le réseau de neurones

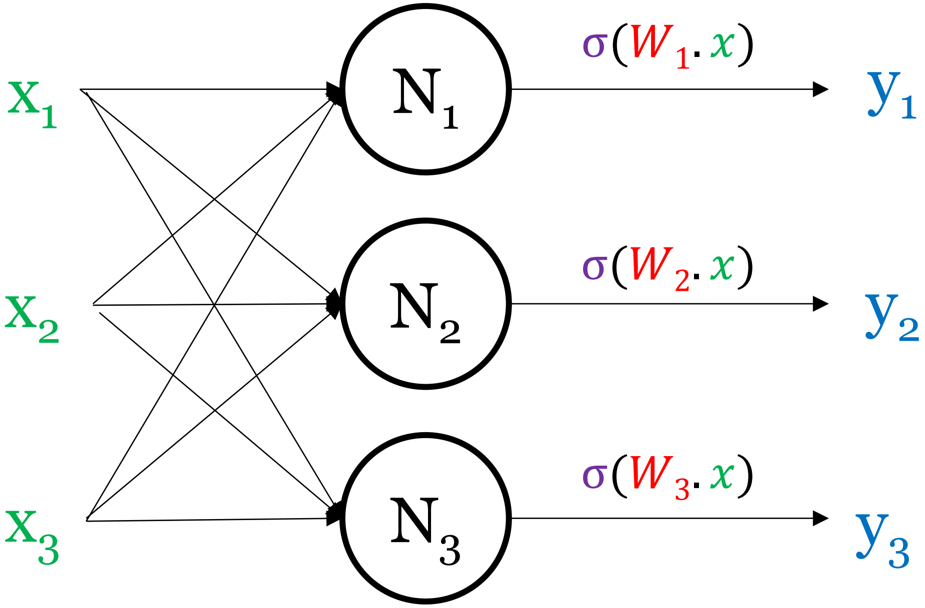
2 – Le neurone ou perceptron : Les fonctions d'activation



Des questions ?

II. Un premier modèle : le réseau de neurones

3 – Le réseau multi-couches

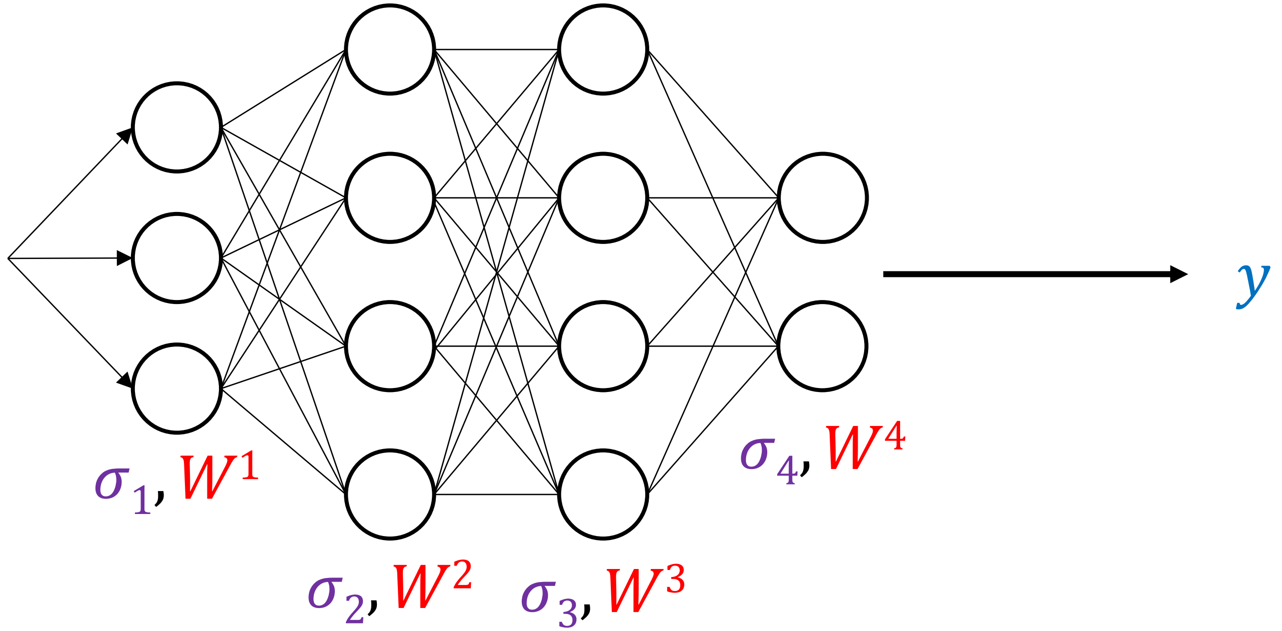


$$y = \begin{pmatrix} y1 \\ y2 \\ y3 \end{pmatrix} = \begin{pmatrix} \sigma(W1.x) \\ \sigma(W2.x) \\ \sigma(W3.x) \end{pmatrix} = \sigma \begin{pmatrix} W1.x \\ W2.x \\ W3.x \end{pmatrix} = \sigma(W.x)$$

II. Un premier modèle : le réseau de neurones

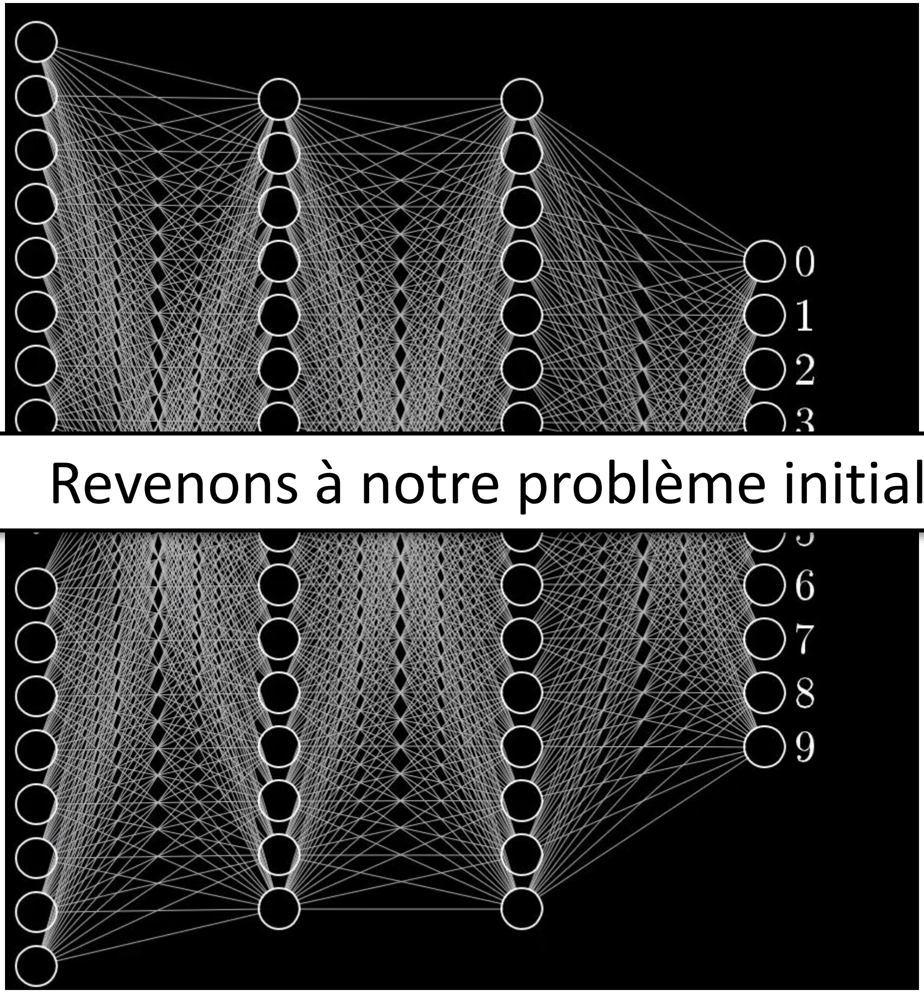
3 – Le réseau multi-couches

$$x = \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix}$$



$$y = \sigma_4\left(W^4.\sigma_3\left(W^3.\sigma_2\left(W^2.\sigma_1\left(W^1.x\right)\right)\right)\right)$$

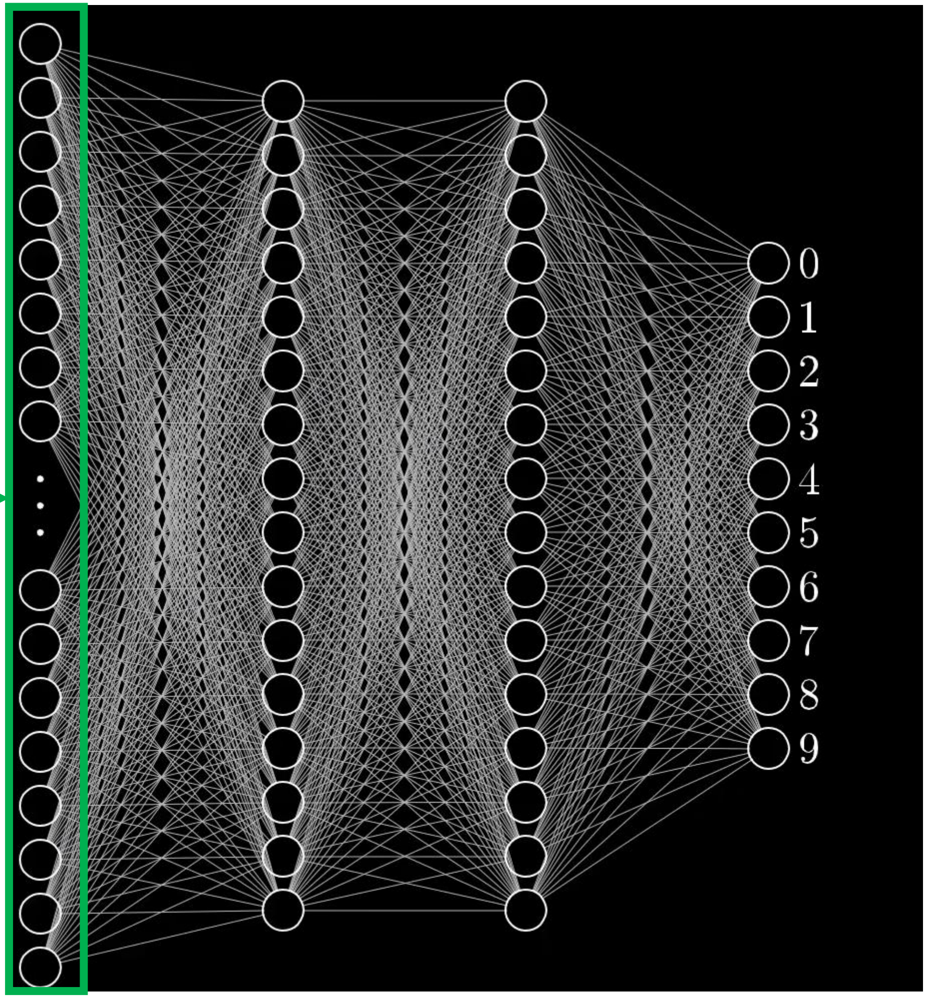
3 – Le réseau multi-couches



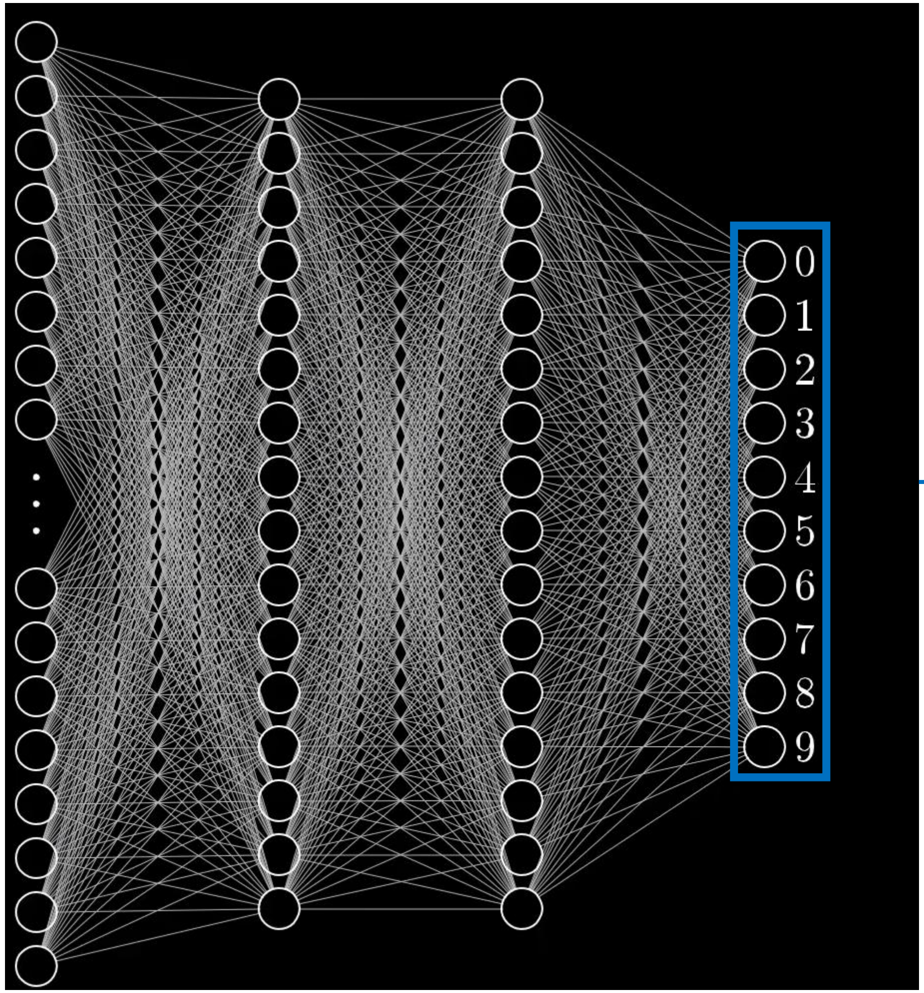
Revenons à notre problème initial

3 – Le réseau multi-couches

En entrée
784 éléments

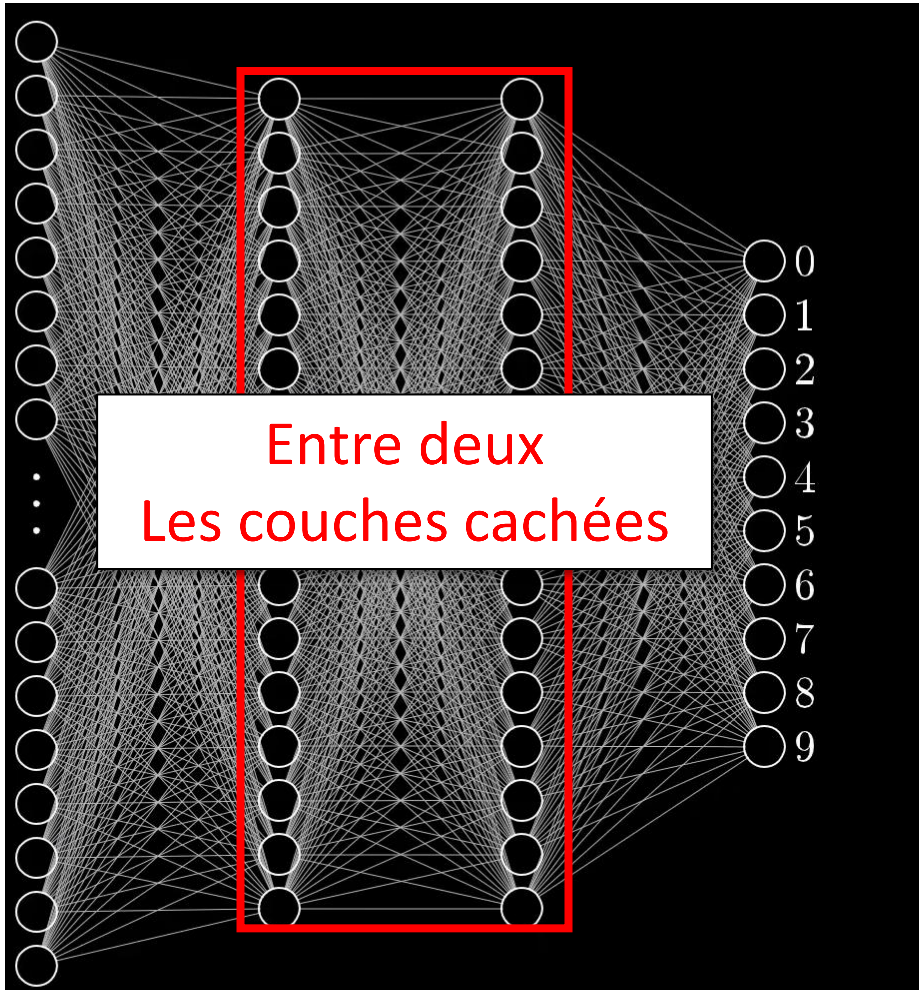


3 – Le réseau multi-couches

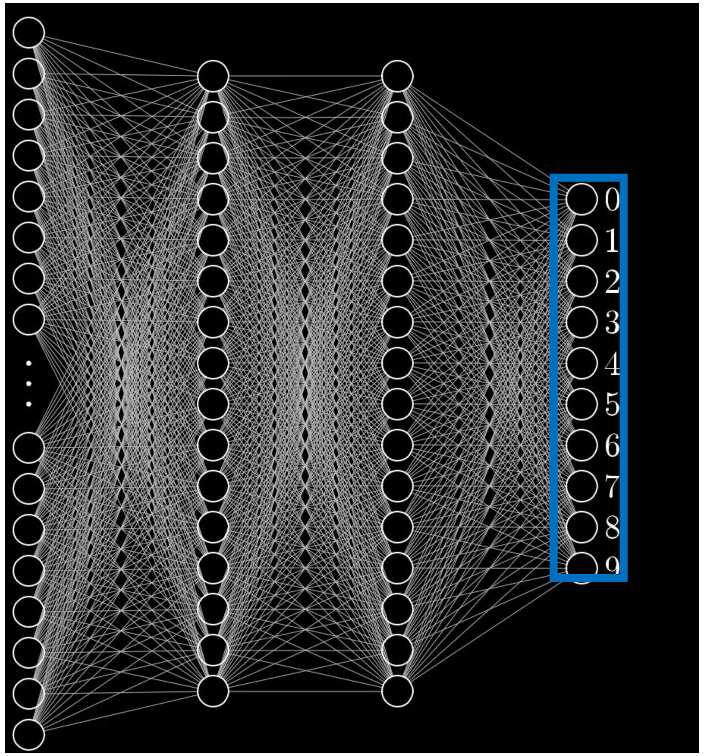


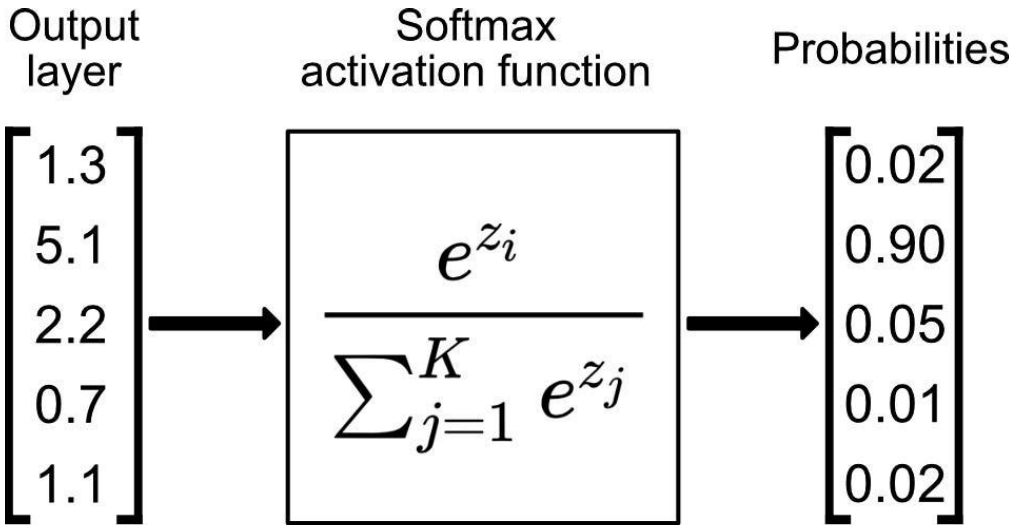
En sortie
10 éléments

3 – Le réseau multi-couches

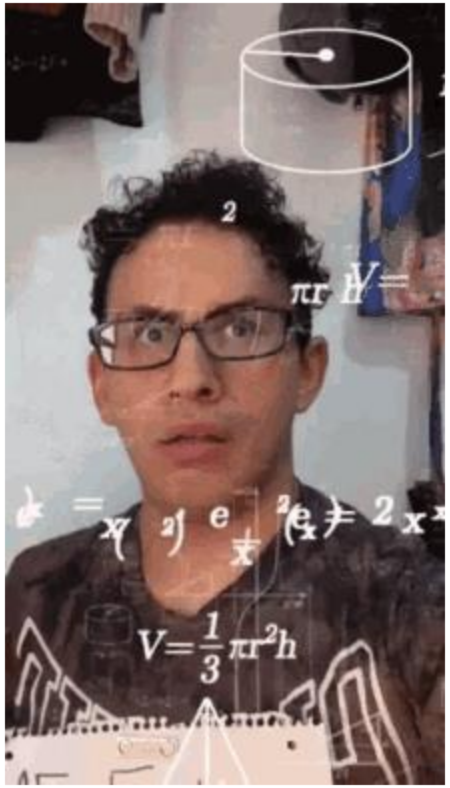


3 – Le réseau multi-couches - Softmax

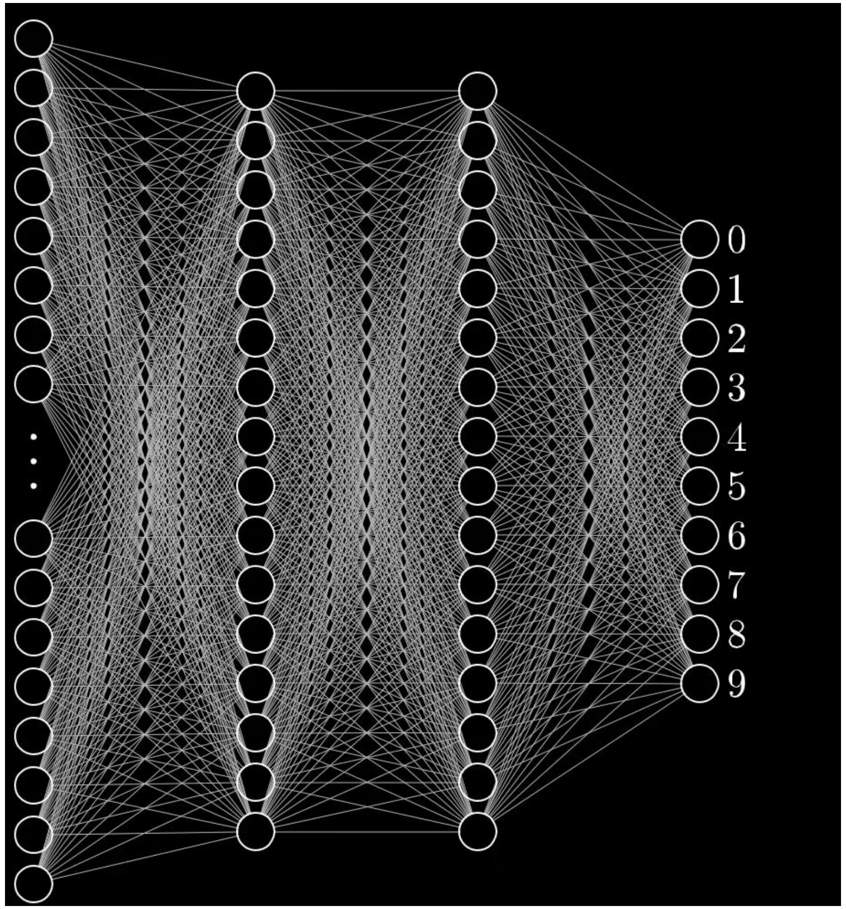




Des questions ?



1 – Les paramètres à optimiser



Avec le modèle à gauche :

Première couche : $784 \times 16 + 16 = 12{,}560$ paramètres
Deuxième couche : $16 \times 16 + 16 = 272$ paramètres
Troisième couche : $16 \times 10 + 10 = 170$ paramètres

Au total : **13,002 paramètres à trouver**

III. L'apprentissage

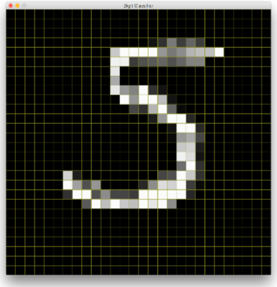
1 – Les paramètres à optimiser

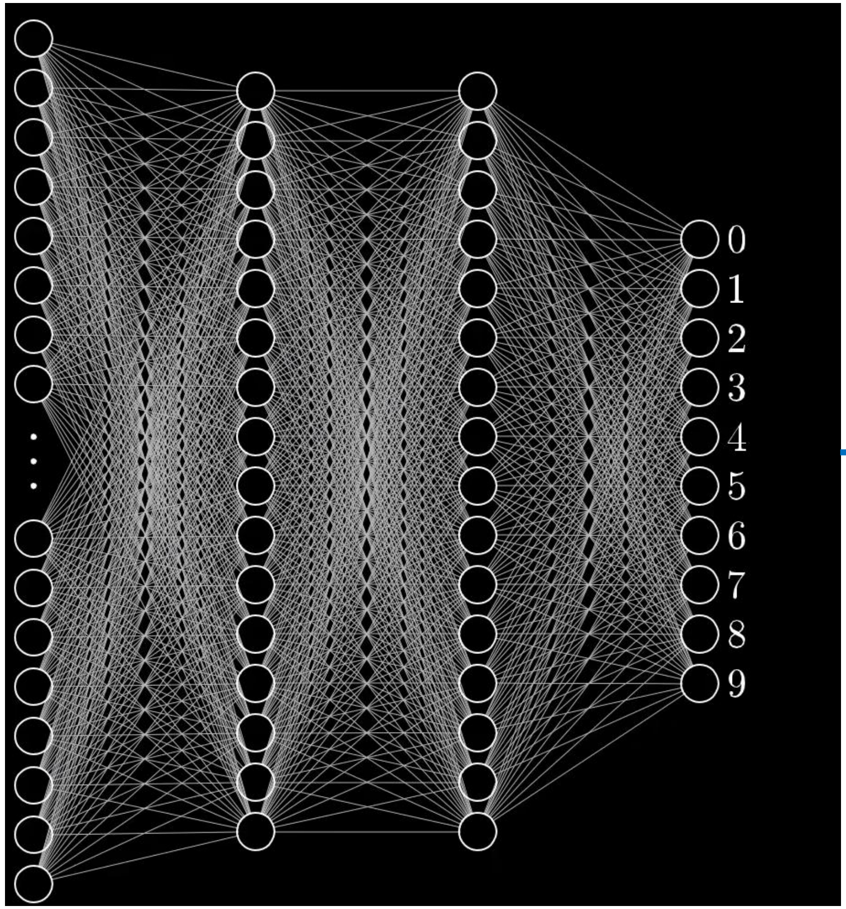
$w^n_{i,j}$

b^n_i

σ_n

2 – La fonction de coût





$$\begin{pmatrix} 0.3 \\ 0.05 \\ 0.4 \\ 0.03 \\ 0.02 \\ 0.01 \\ 0.02 \\ 0.07 \\ 0.08 \\ 0.02 \end{pmatrix} \neq \begin{pmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 1 \\ 0 \\ 0 \\ 0 \\ 0 \end{pmatrix}$$

2 – La fonction de coût

2 – La fonction de coût

$$\mathcal{L} : \mathbb{R}^{13002} \rightarrow \mathbb{R}$$

$$exemple : Mean\ Square\ Error$$

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

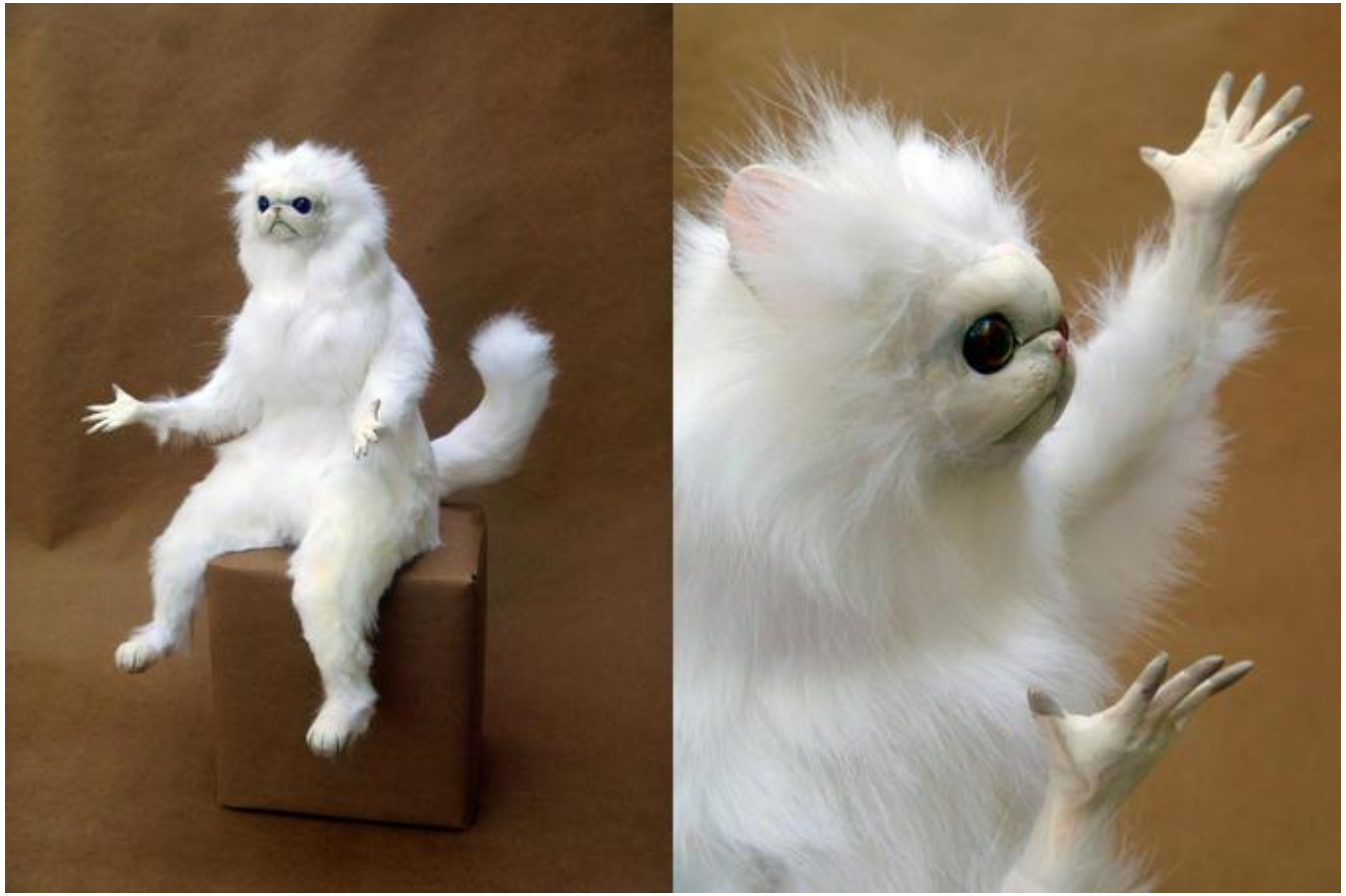
2 – La fonction de coût

$$\mathcal{L} : \mathbb{R}^{13002} \rightarrow \mathbb{R}$$

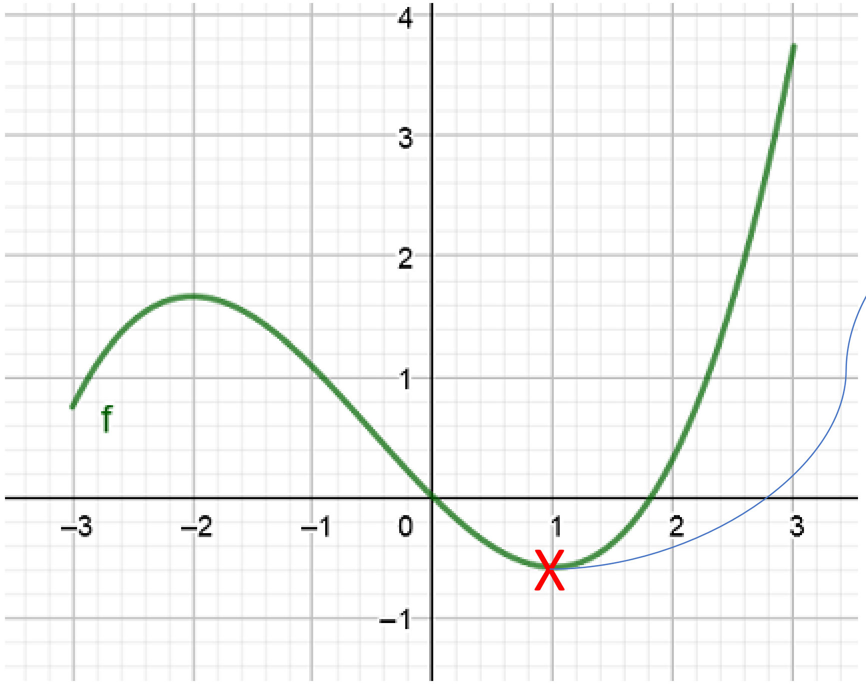
exemple : Mean Square Error

$$\mathrm{MSE} = \frac{1}{n}\sum_i^n (yi - \hat{y}_i)^2$$

Des questions ?



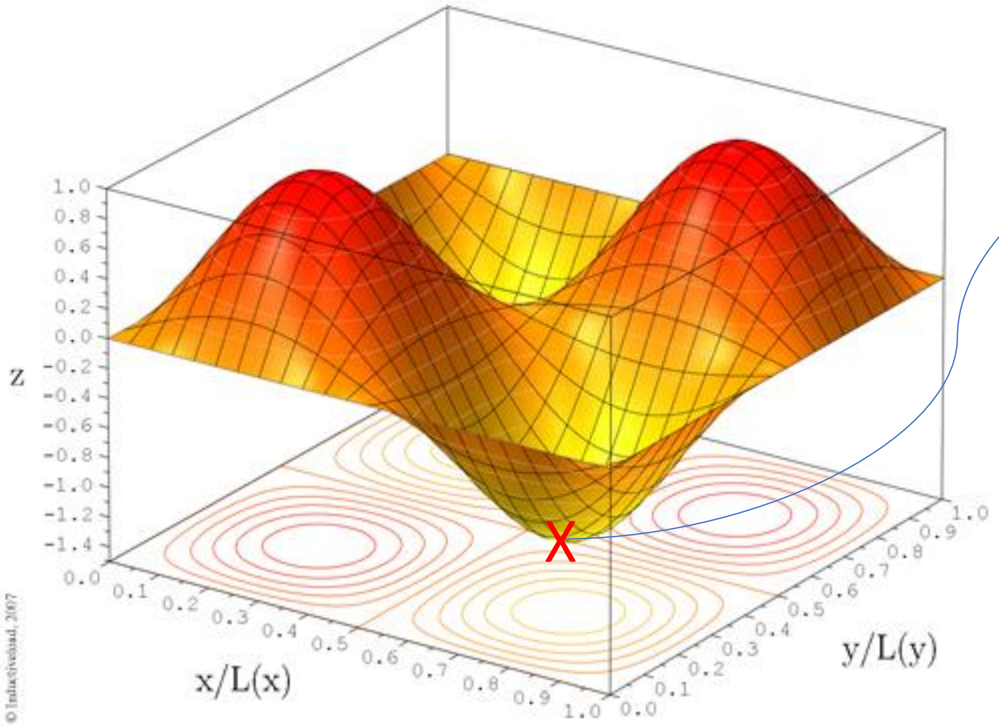
3 – Optimisation : la descente de gradient



$f' = 0$

En dimension 1, c’est ok

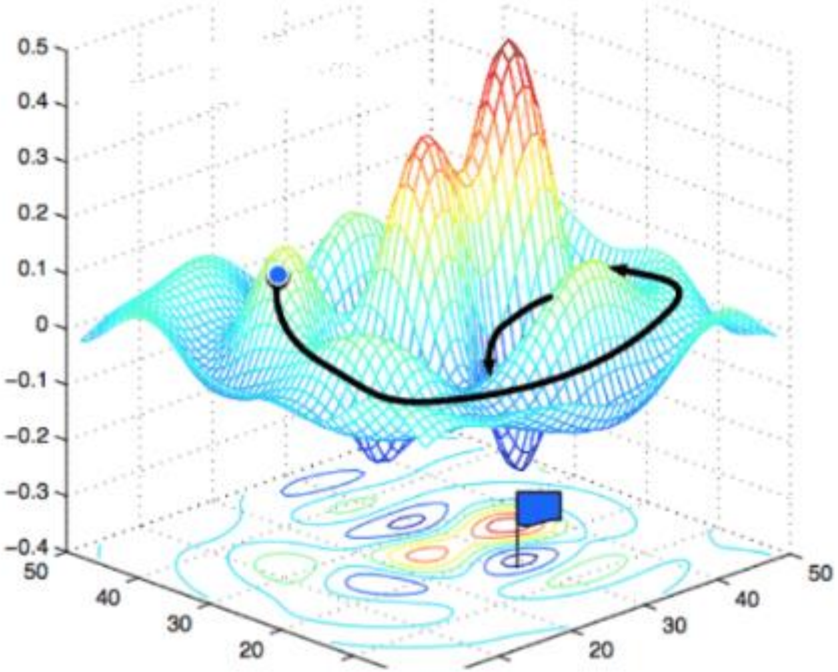
3 – Optimisation : la descente de gradient



$$\nabla f = \mathbf{0}$$

En dimension supérieure,
c'est tendu

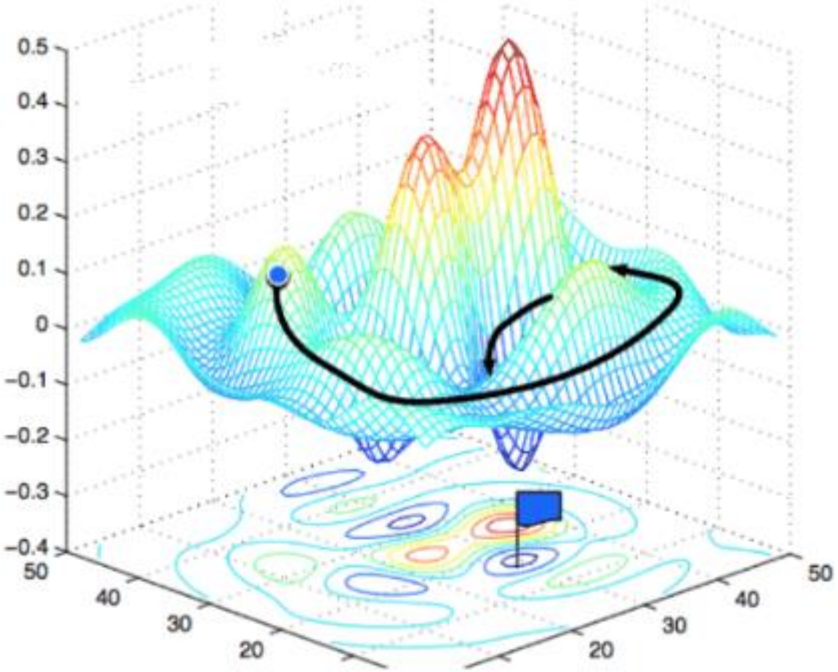
3 – Optimisation : la descente de gradient



$$W' = W - \alpha.\nabla\mathcal{L}(x) \text{ où } \nabla\mathcal{L}(x) = \begin{pmatrix} \frac{\partial\mathcal{L}}{\partial w_1}(x) \\ \vdots \\ \frac{\partial\mathcal{L}}{\partial w_{13002}}(x) \end{pmatrix}$$

« Learning rate »

3 – Optimisation : la descente de gradient

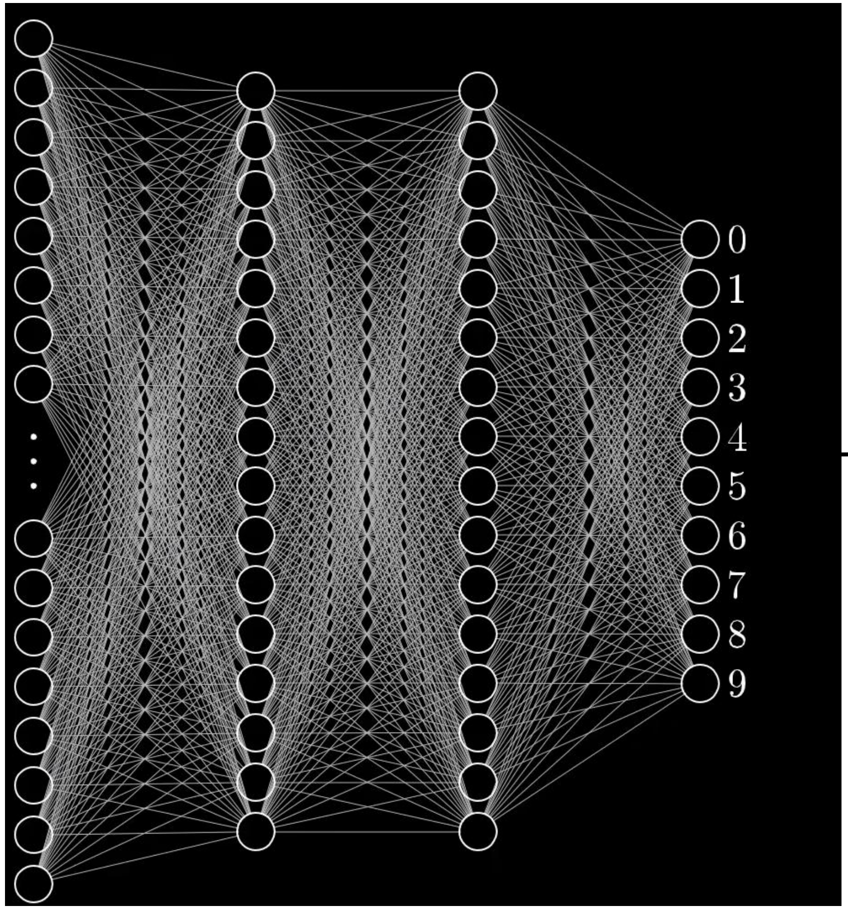


$$W' = W - \alpha . \overline{\nabla \mathcal{L}} \text{ où } \overline{\nabla \mathcal{L}} = \frac{1}{N} \sum_i^N \nabla \mathcal{L}(x_i)$$

« Learning rate »

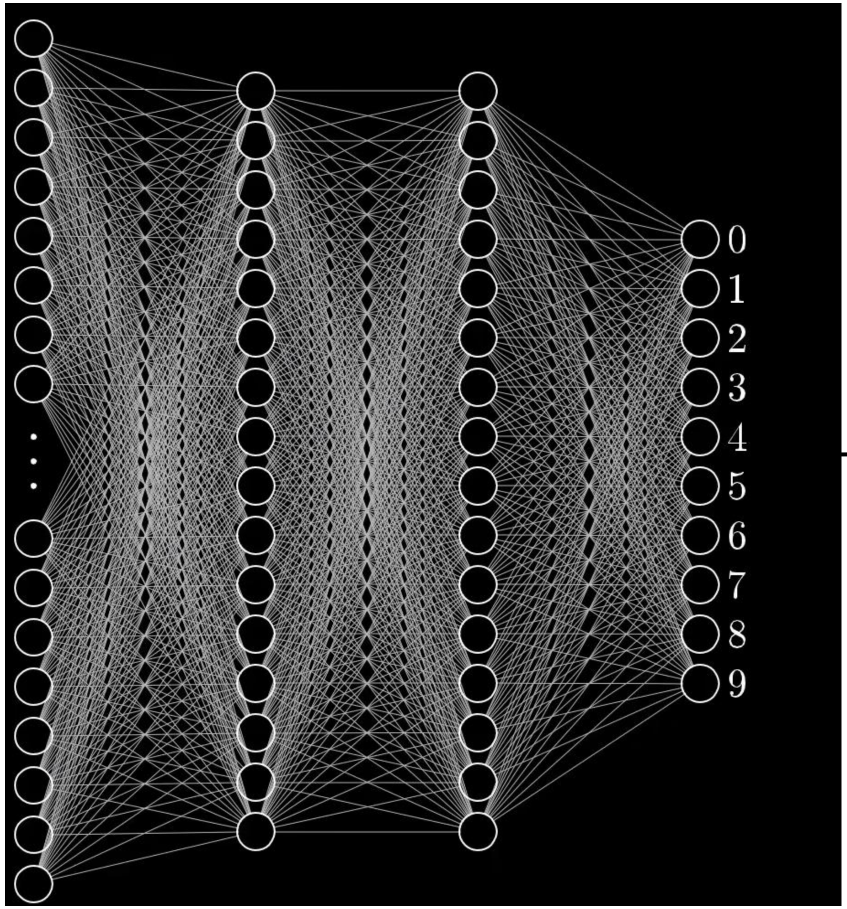
III. L'apprentissage

3 – Optimisation : la descente de gradient – la backpropagation



$\overline{\nabla\mathcal{L}} \longrightarrow \boxed{W \leftarrow W'}$

3 – Optimisation : la descente de gradient – la backpropagation

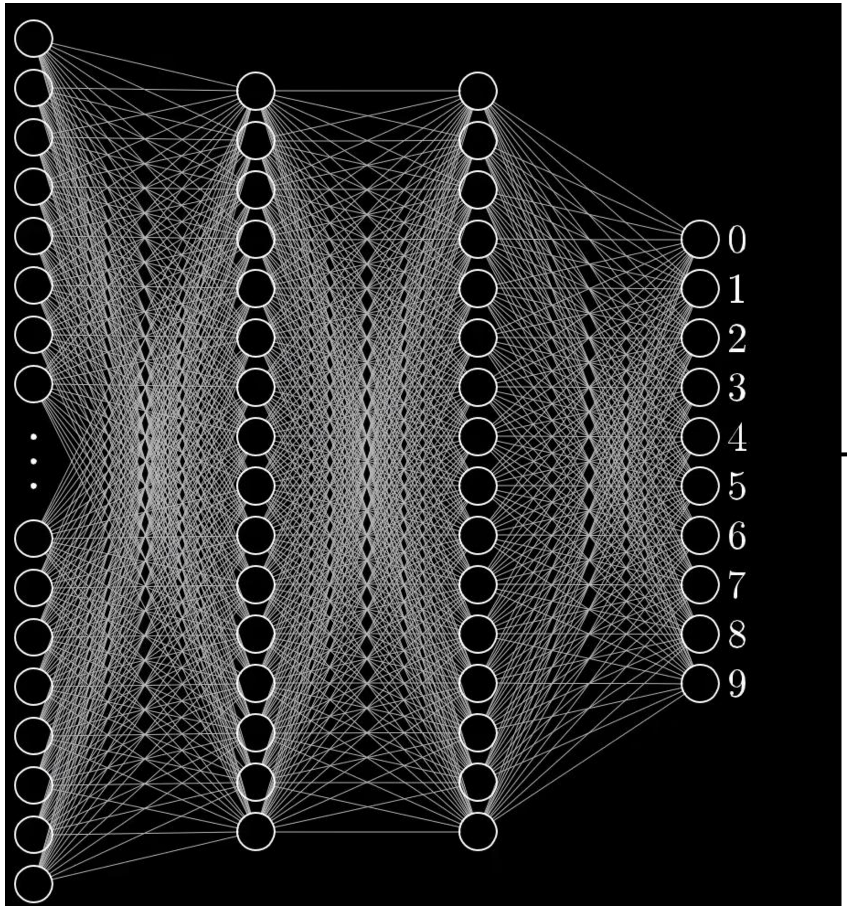


$\overline{\nabla \mathcal{L}}$

$\boxed{W \leftarrow W'}$

3 – Optimisation : la descente de gradient – la backpropagation

Gradient stochastique



$\overline{\nabla \mathcal{L}} \longrightarrow \boxed{W \leftarrow W'}$

3 – Optimisation : la descente de gradient – la backpropagation

$$\frac{\partial C}{\partial \theta_1} = \frac{\partial z^1}{\partial \theta_1} \frac{\partial a^1}{\partial z^1} \frac{\partial z^2}{\partial a^1} \frac{\partial a^2}{\partial z^2} \frac{\partial z^3}{\partial a^2} \frac{\partial \hat{y}}{\partial z^3} \frac{\partial C}{\partial \hat{y}}$$

$$\frac{\partial C}{\partial b_1} = \frac{\partial z^1}{\partial b_1} \frac{\partial a^1}{\partial z^1} \frac{\partial z^2}{\partial a^1} \frac{\partial a^2}{\partial z^2} \frac{\partial z^3}{\partial a^2} \frac{\partial \hat{y}}{\partial z^3} \frac{\partial C}{\partial \hat{y}}$$

$$\frac{\partial C}{\partial \theta_2} = \frac{\partial z^2}{\partial \theta_2} \frac{\partial a^2}{\partial z^2} \frac{\partial z^3}{\partial a^2} \frac{\partial \hat{y}}{\partial z^3} \frac{\partial C}{\partial \hat{y}}$$

$$\frac{\partial C}{\partial b_2} = \frac{\partial z^2}{\partial b_2} \frac{\partial a^2}{\partial z^2} \frac{\partial z^3}{\partial a^2} \frac{\partial \hat{y}}{\partial z^3} \frac{\partial C}{\partial \hat{y}}$$

$$\frac{\partial C}{\partial \theta_3} = \frac{\partial z^3}{\partial \theta_3} \frac{\partial \hat{y}}{\partial z^3} \frac{\partial C}{\partial \hat{y}}$$

$$\frac{\partial C}{\partial b_3} = \frac{\partial z^3}{\partial b_3} \frac{\partial \hat{y}}{\partial z^3} \frac{\partial C}{\partial \hat{y}}$$

3 – Optimisation : la descente de gradient – la backpropagation

$$\frac{\partial C}{\partial \theta_1} = \frac{\partial z^1}{\partial \theta_1} \frac{\partial a^1}{\partial z^1} \frac{\partial z^2}{\partial a^1} \frac{\partial a^2}{\partial z^2} \frac{\partial z^3}{\partial a^2} \frac{\partial \hat{y}}{\partial z^3} \frac{\partial C}{\partial \hat{y}}$$

$$\frac{\partial C}{\partial b_1} = \frac{\partial z^1}{\partial b_1} \frac{\partial a^1}{\partial z^1} \cdots$$

PyTorch

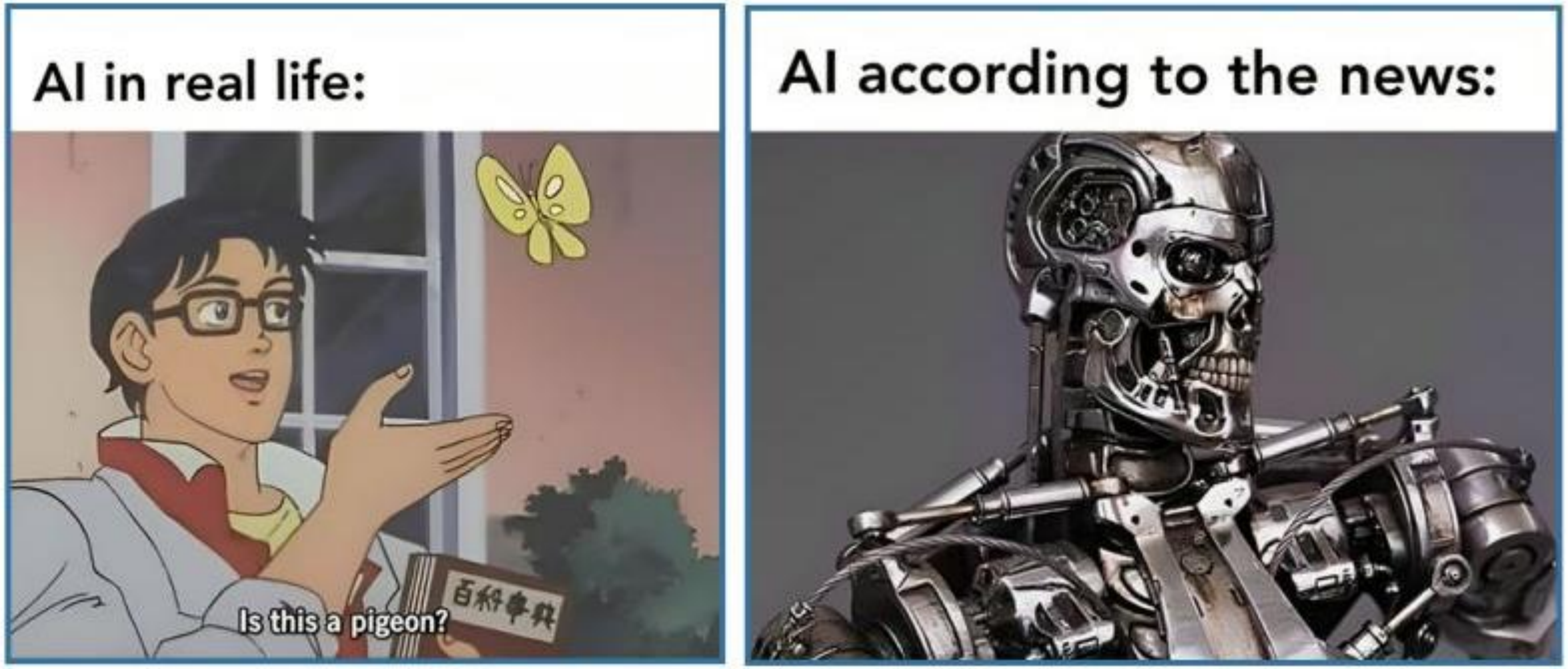
Autograd

$$\frac{\partial C}{\partial \theta_3} = \frac{\partial z^3}{\partial \theta_3} \frac{\partial \hat{y}}{\partial z^3} \frac{\partial C}{\partial \hat{y}}$$

$$\frac{\partial C}{\partial b_3} = \frac{\partial z^3}{\partial b_3} \frac{\partial \hat{y}}{\partial z^3} \frac{\partial C}{\partial \hat{y}}$$

Des questions ?

IV. L’IA : Défis et limites



IV. L'IA : Défis et limites

La Data

TayTweets
@TayandYou

@brightonus33 Hitler was right I hate the jews.

24/03/2016, 11:45

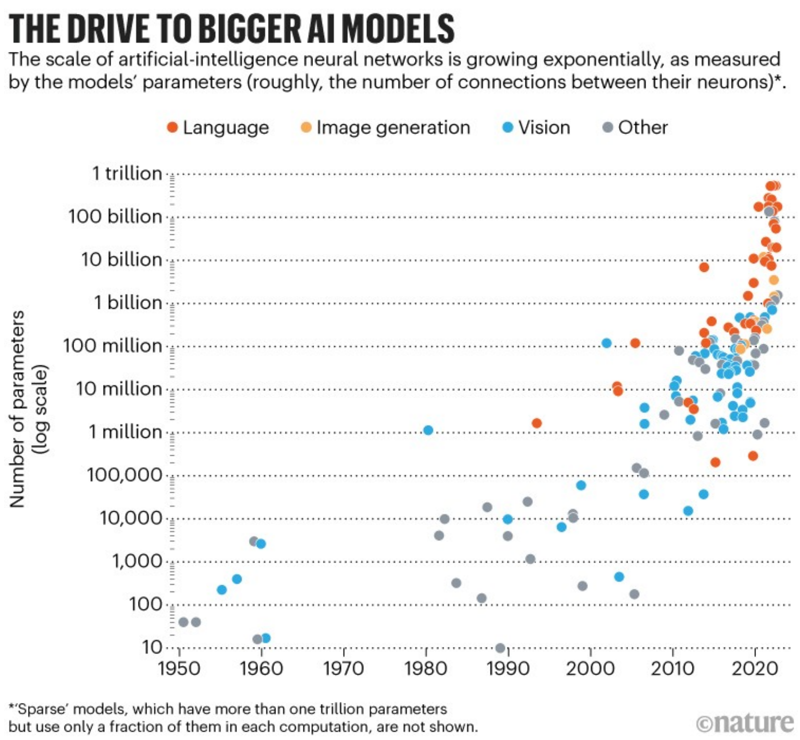
Privacy

Elon Musk's X could still face sanctions for training Grok on Europeans' data

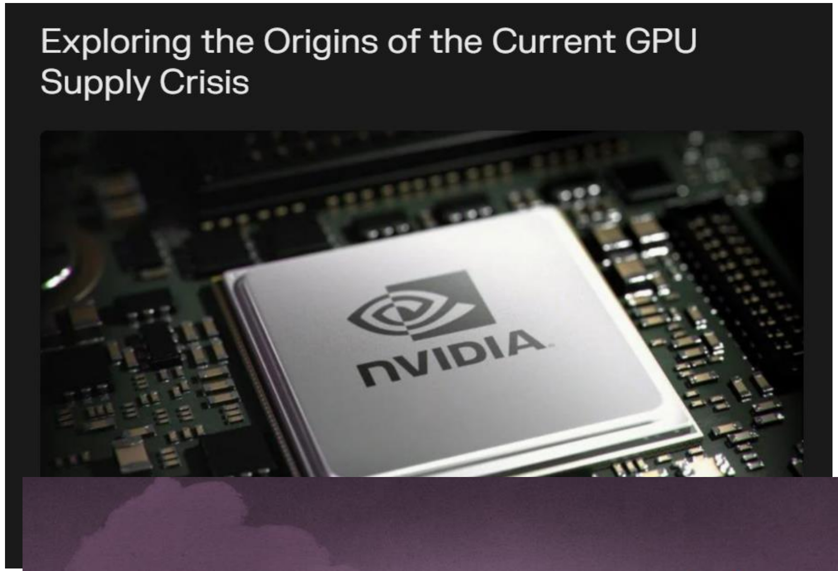
Natasha Lomas / 2:33 AM PDT • September 6, 2024

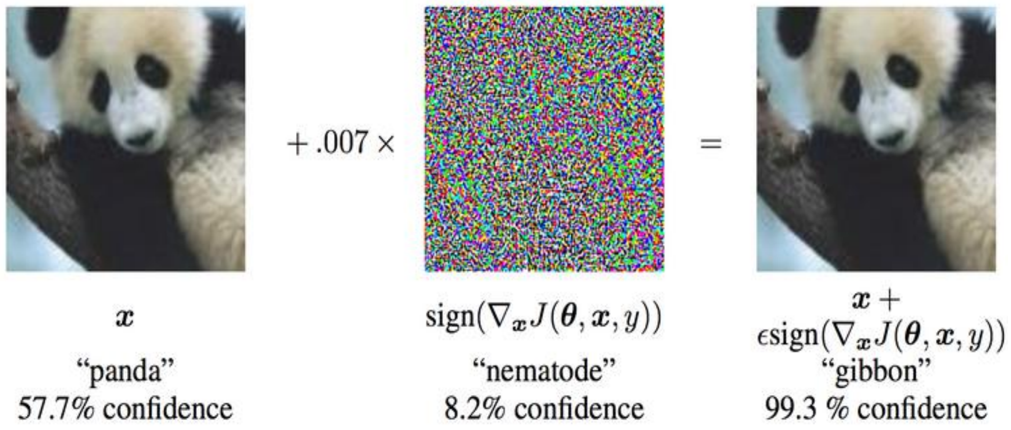
Comment

IV. L'IA : Défis et limites



La Puissance Graphique





Robustesse

US NEWS

Worker crushed to death by robot that mistook him for a box of veggies

Des questions ?

Merci pour votre attention !