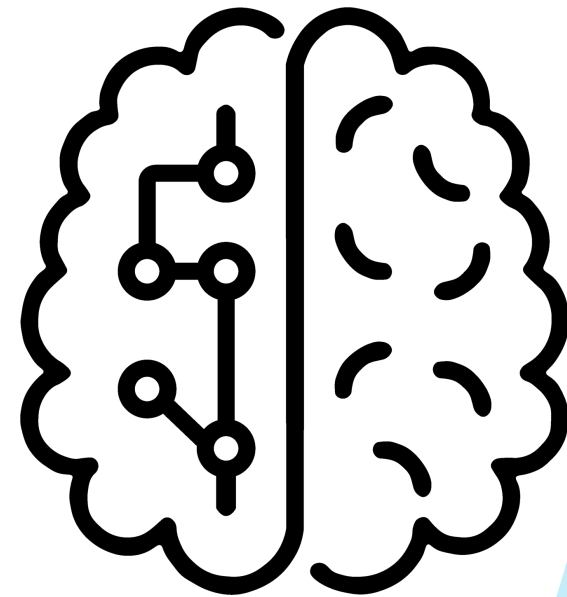


# Le deep learning

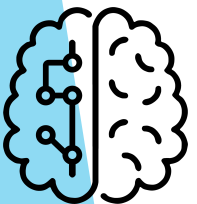
AIForYou - Morgan Gautherot



# Définitions

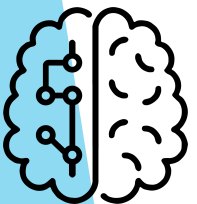
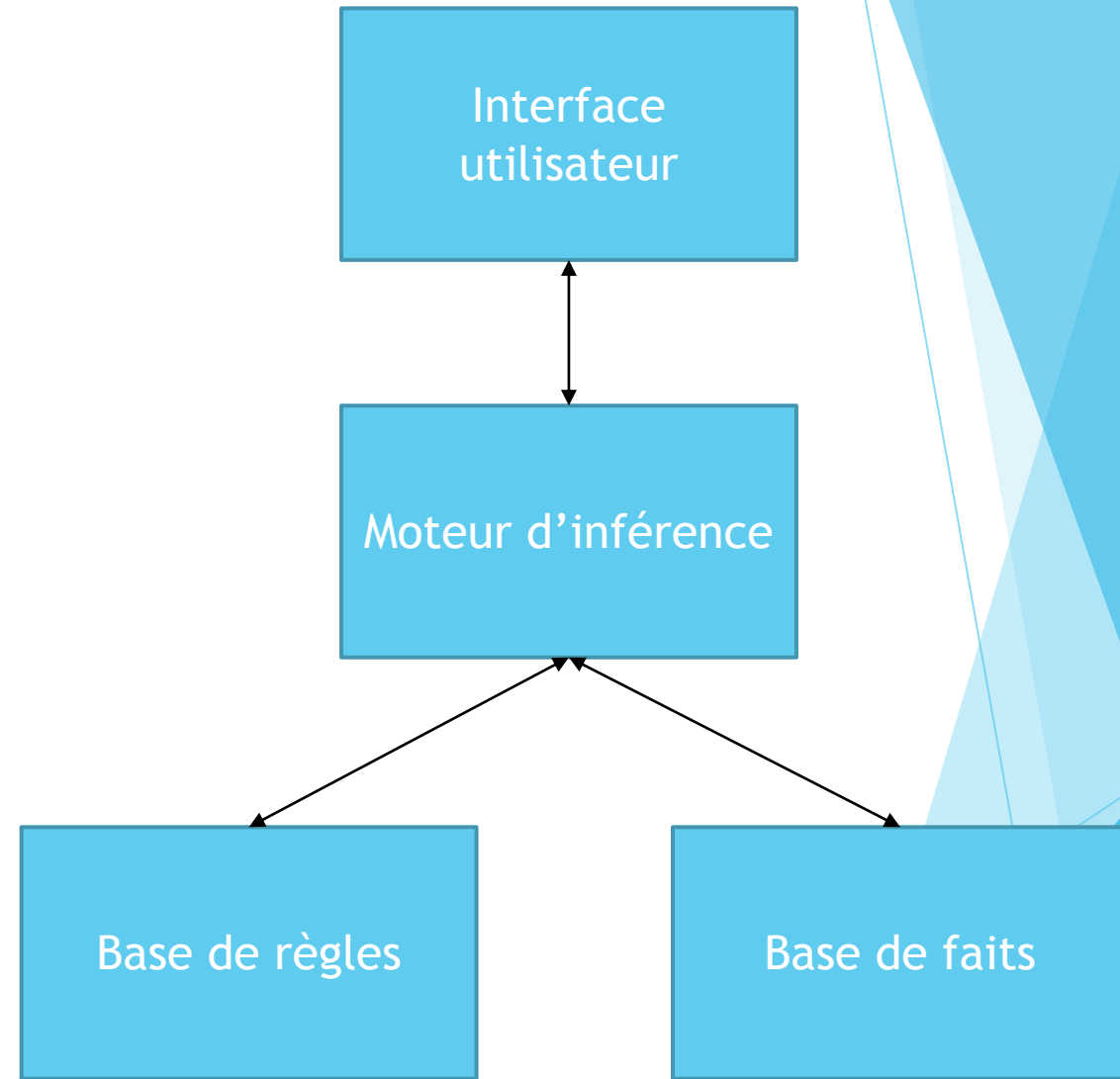
- ▶ **L'intelligence artificielle (IA)** est un domaine de l'informatique dont l'objectif est de créer des programmes qui **exécutent des tâches normalement assignées à l'intelligence humaine et de la simuler.**

Intelligence Artificielle

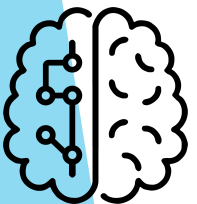
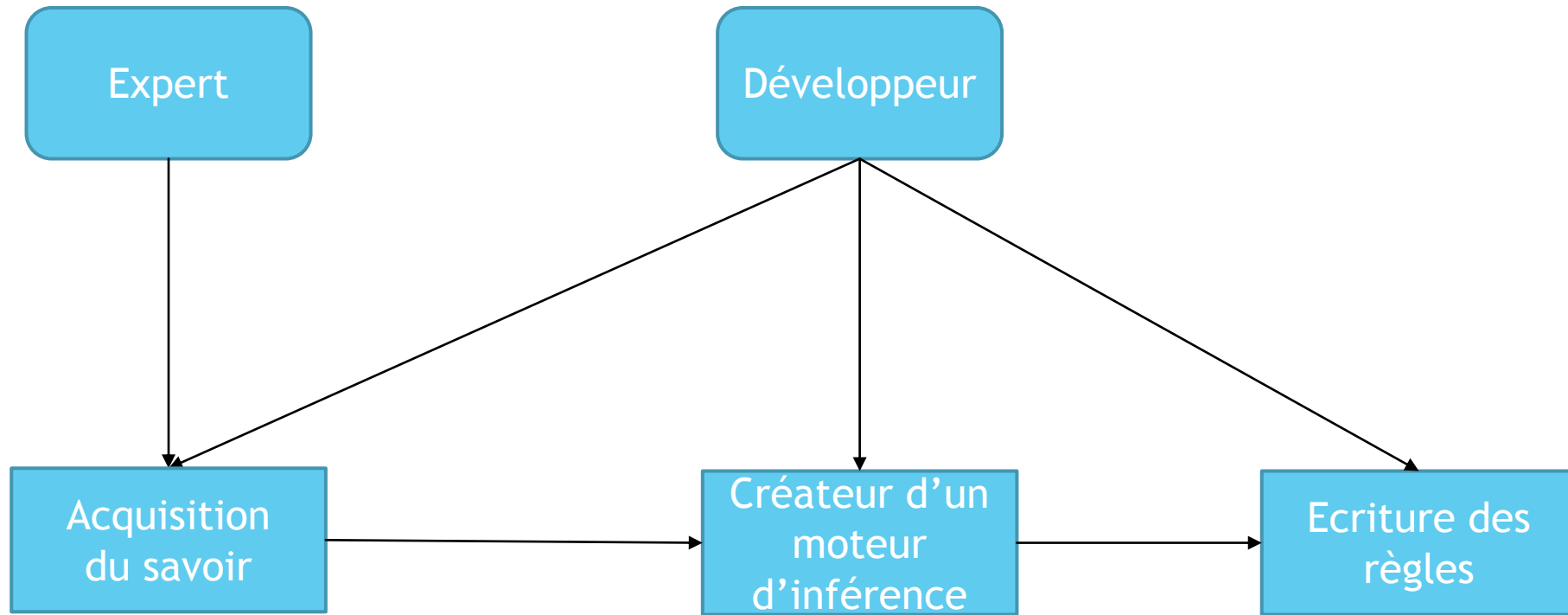


# Systeme expert

- Un système expert est un programme qui répond à des questions ou résout des problèmes dans un domaine de connaissance donné, à l'aide de règles logiques dérivées de la connaissance des experts humains de ce domaine.

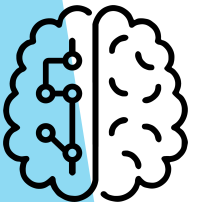


# Construction d'un système



# Applications

- ▶ Diagnostic d'analyse de sang
- ▶ Diagnostic de panne pour des voitures
- ▶ Vérification système d'avion avant le décollage
- ▶ Télémédecine
- ▶ Chaîne de montage automatisé



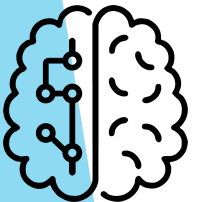
# Pour et contre



- ▶ Implémentation du sens commun
- ▶ Performant sur des environnements fermés
- ▶ Modèle interprétable

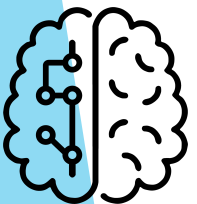
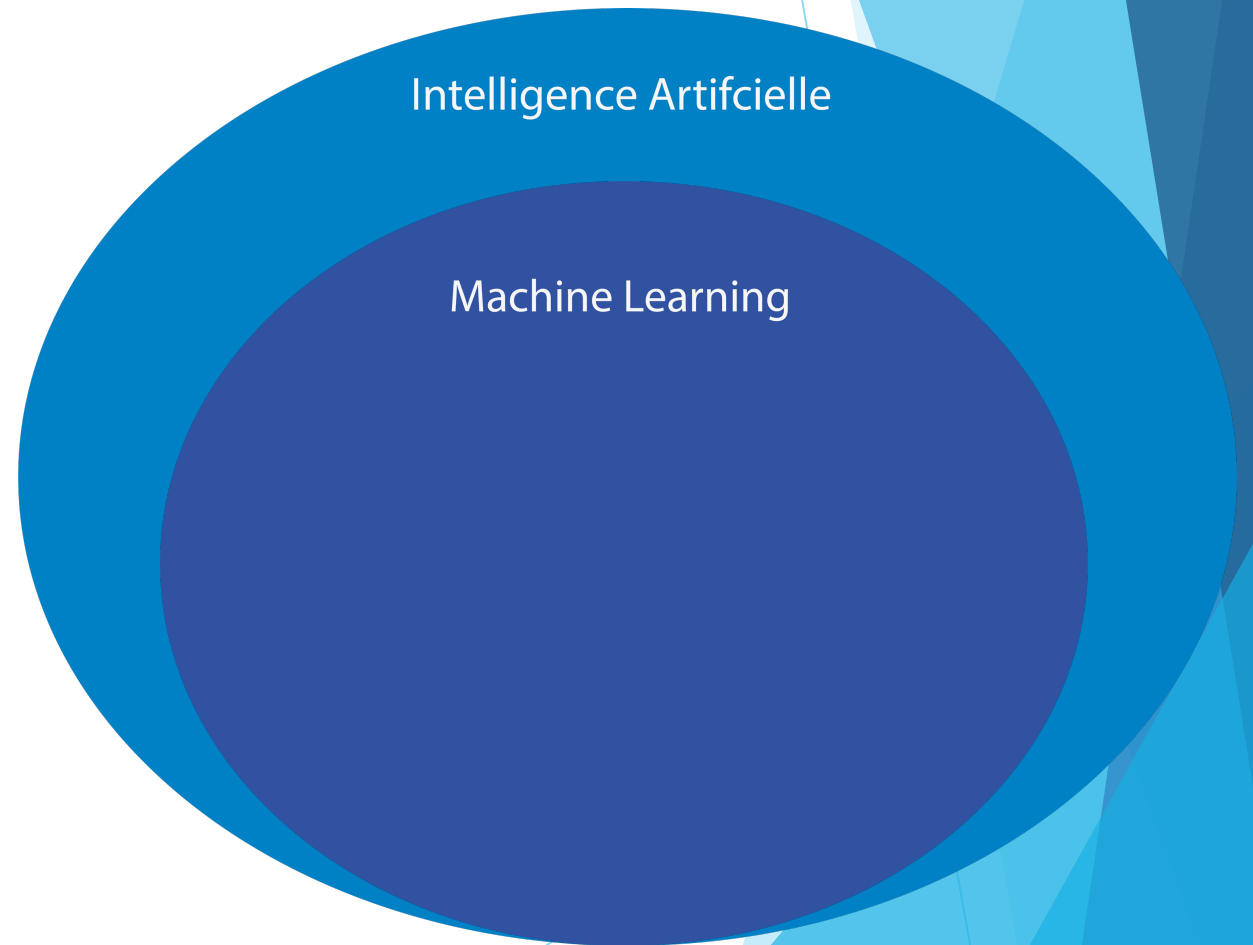


- ▶ Long à mettre en place
- ▶ Difficilement maintenable
- ▶ Mise-à-jour constante dans le cas d'environnements ouverts
- ▶ Approche biaisée



# Définitions

- ▶ **L'intelligence artificielle (IA)** est un domaine de l'informatique dont l'objectif est de créer des programmes qui **exécutent des tâches normalement assignées à l'intelligence humaine et de la simuler.**
- ▶ **Le machine learning (ML)** est une branche de l'IA. Il a la **capacité d'apprendre à partir de données** en utilisant un algorithme d'apprentissage dont l'objectif est d'effectuer des analyses explicatives, prédictives ou préventives.

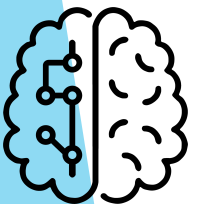


# Apprentissage supervisé

Variables explicatives

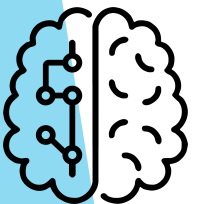
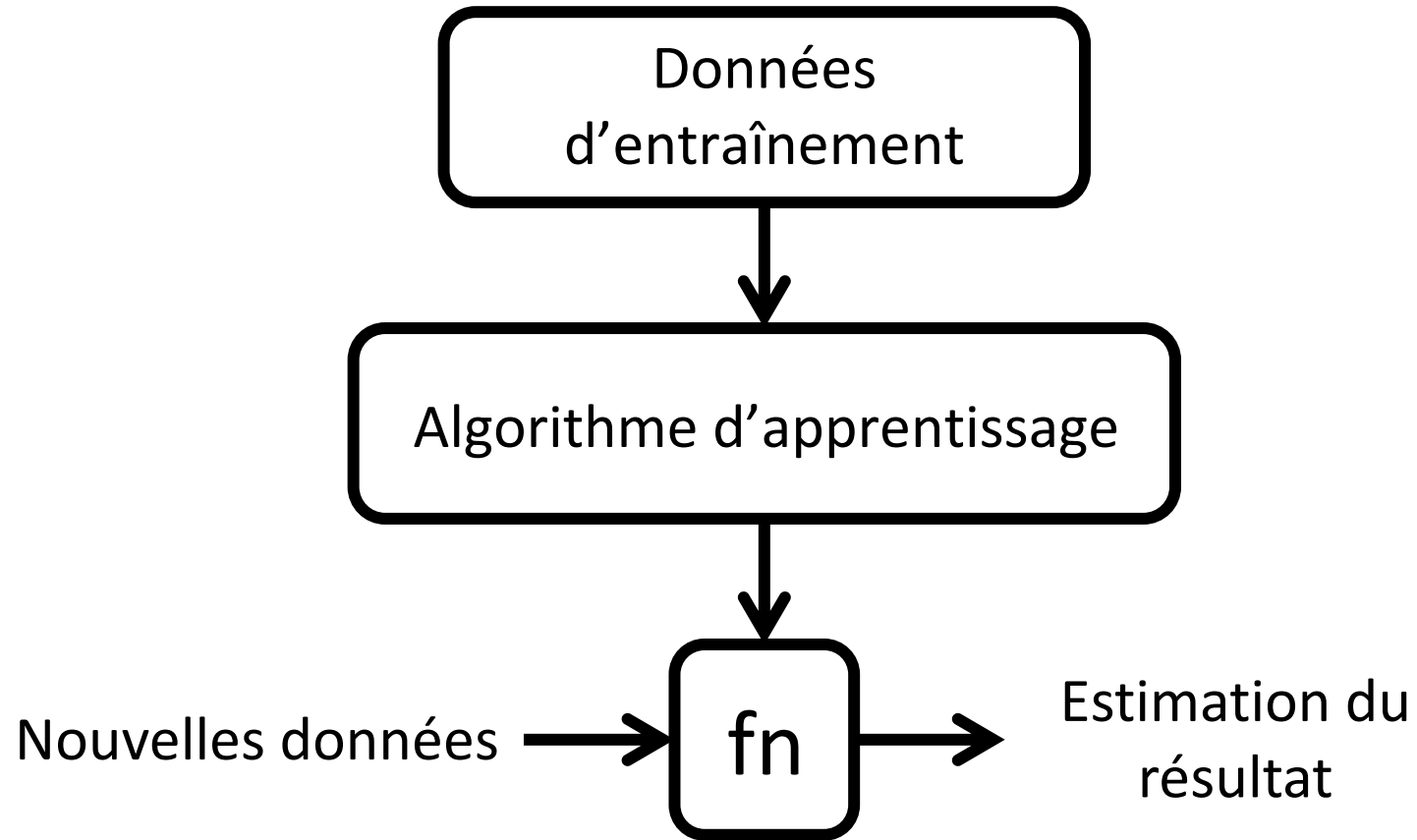


Variable cible

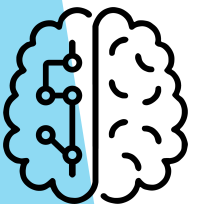
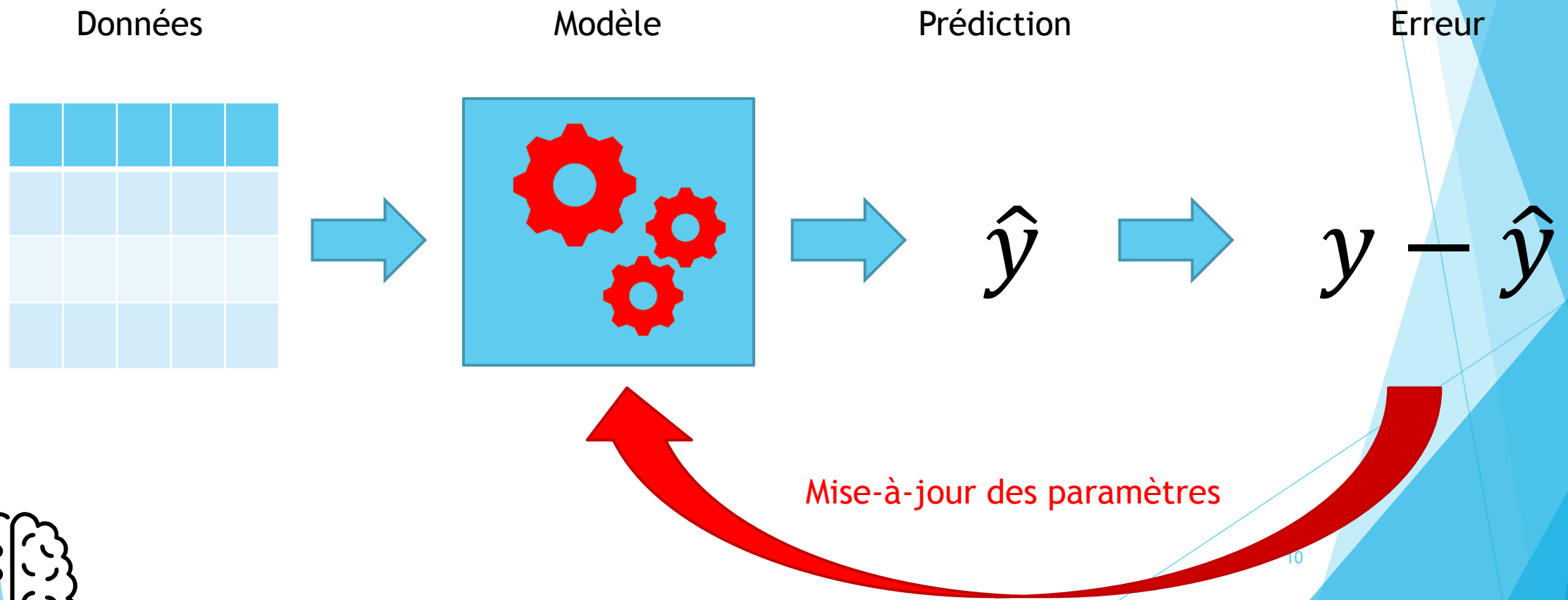




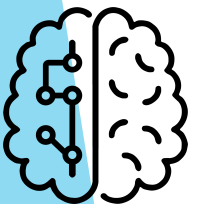
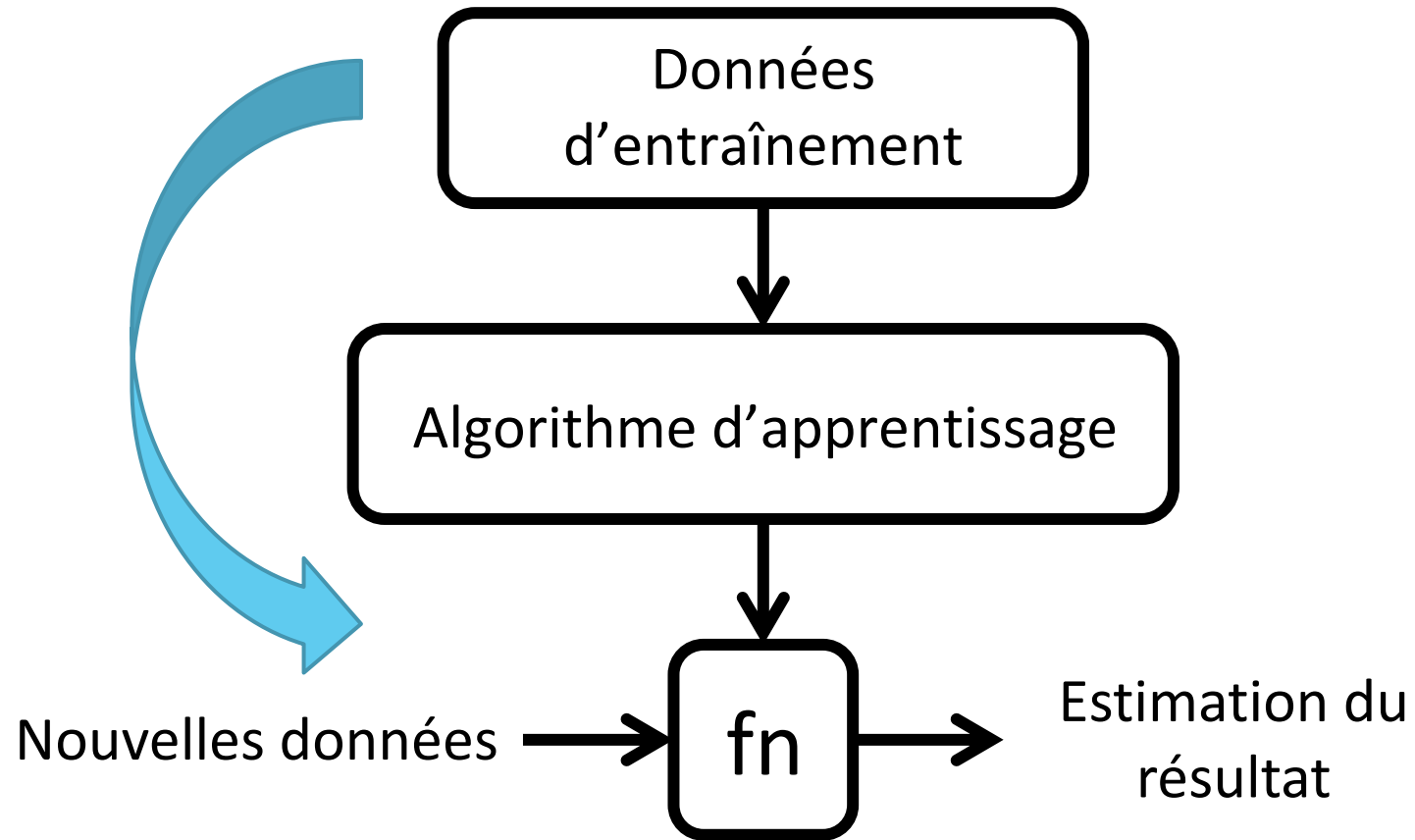
# Recherche de corrélations automatique



# Apprentissage d'un modèle

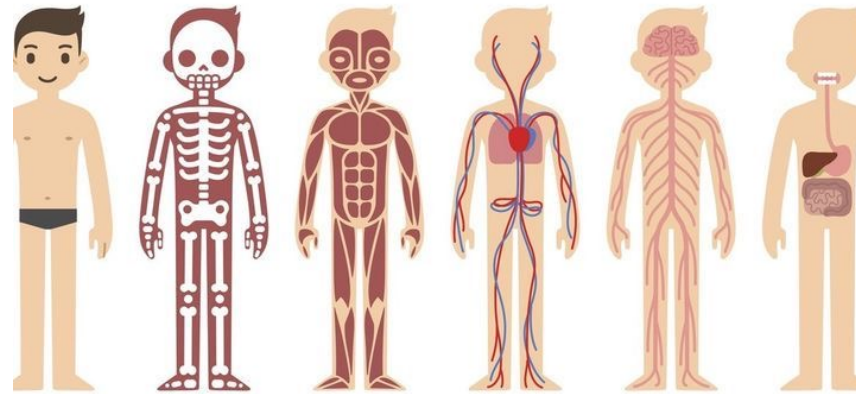


# Mise-à-jour facilité

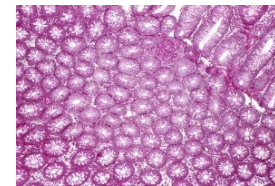


# Extraction des caractéristiques

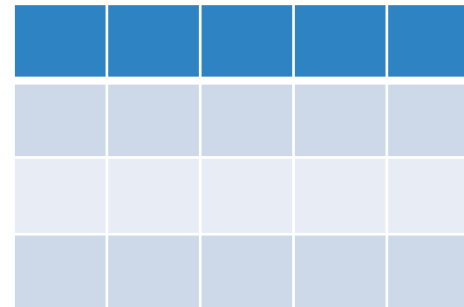
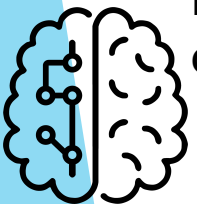
Objet de l'étude



Mesure et examens



Extraction des caractéristiques



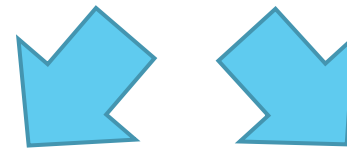
# Deux types de problèmes

## La régression



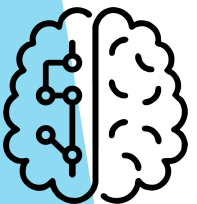
Une variable continue

## La classification



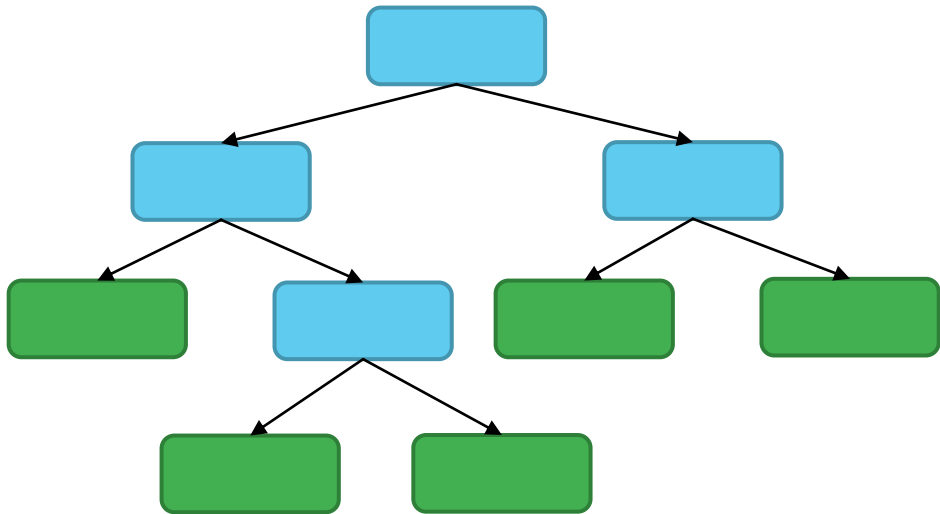
Diagnostic 1

Diagnostic 2

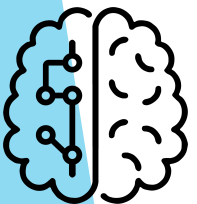
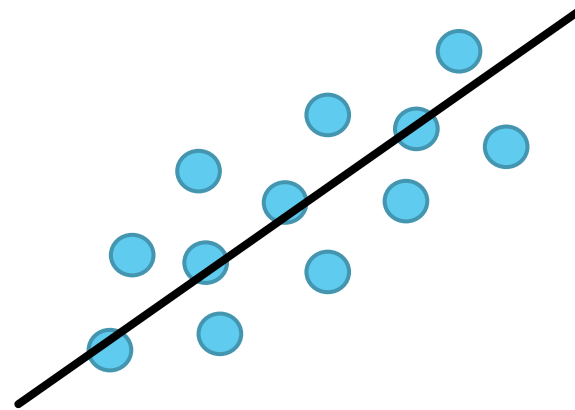


# Différent algorithmes d'apprentissage

Approche à base d'arbre



Approche linéaire



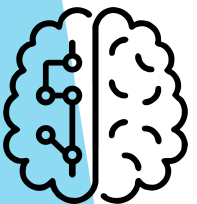
# Pour et contre



- ▶ Apprentissage de nouvelles connaissances
- ▶ Modèle interprétable
- ▶ Facilement maintenable

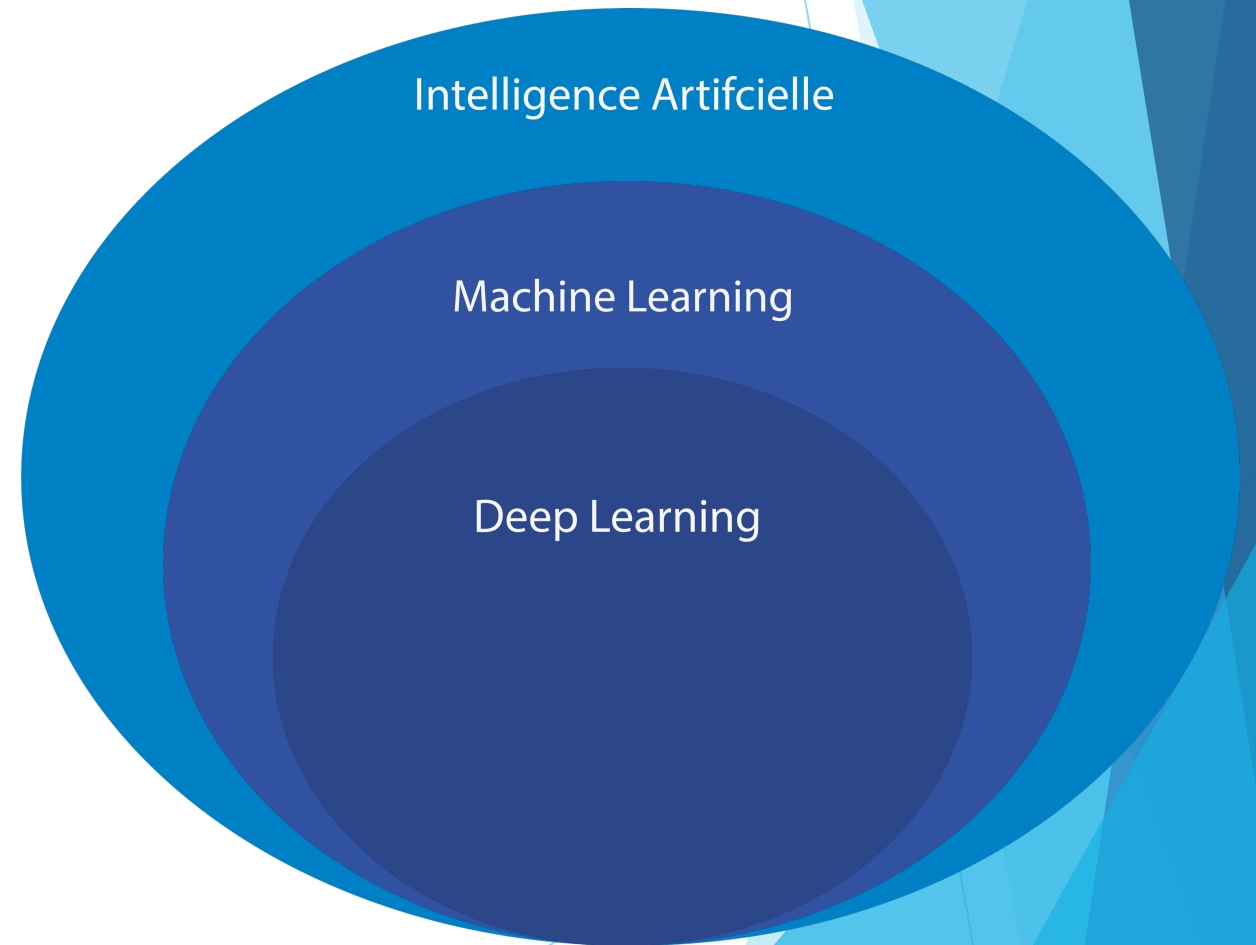
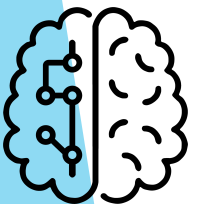


- ▶ Mauvaise performance sur les données non structurés
- ▶ Approche biaisée
- ▶ Perte du sens commun



# Définitions

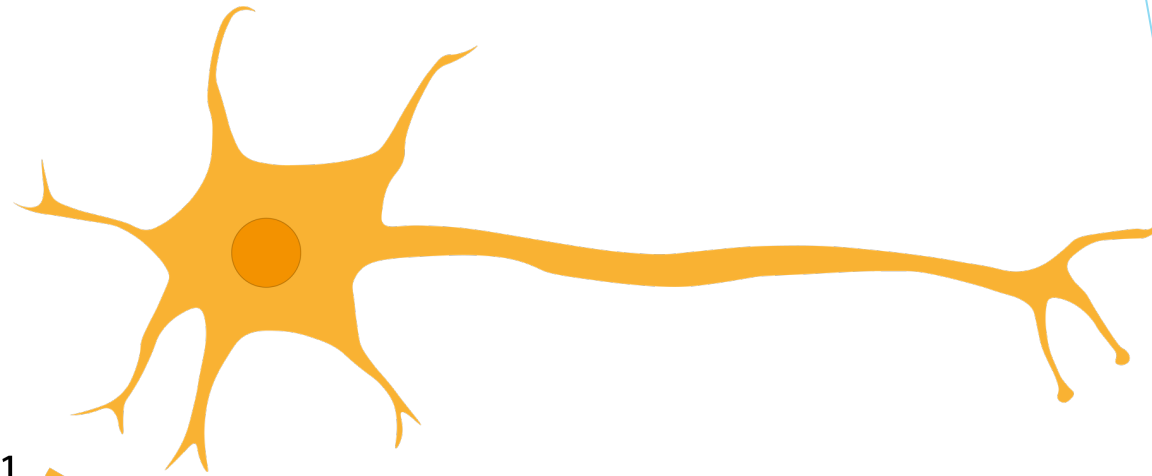
- ▶ **L'intelligence artificielle (IA)** est un domaine de l'informatique dont l'objectif est de créer des programmes qui **exécutent des tâches normalement assignées à l'intelligence humaine et de la simuler.**
- ▶ **Le machine learning (ML)** est une branche de l'IA. Il a la **capacité d'apprendre à partir de données** en utilisant un algorithme d'apprentissage dont l'objectif est d'effectuer des analyses explicatives, prédictives ou préventives.
- ▶ **Le deep learning (DL)** est une discipline d'apprentissage machine qui combine des méthodes basées sur **les réseaux neuronaux.**



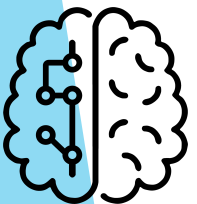
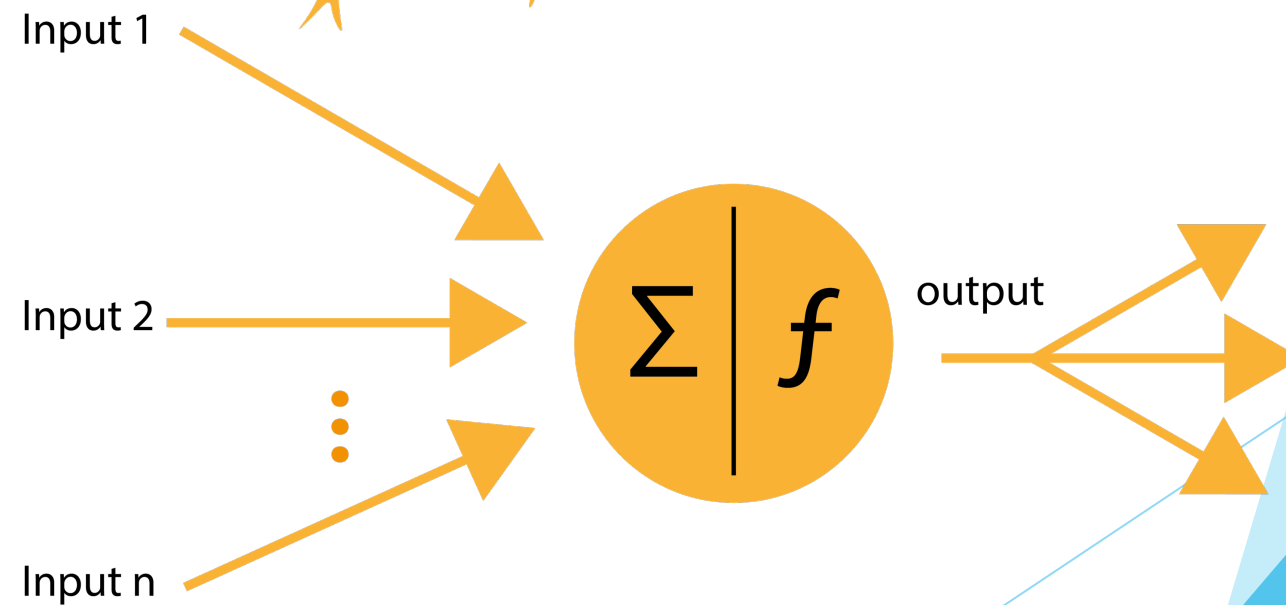


# Neurone artificiel

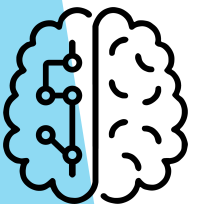
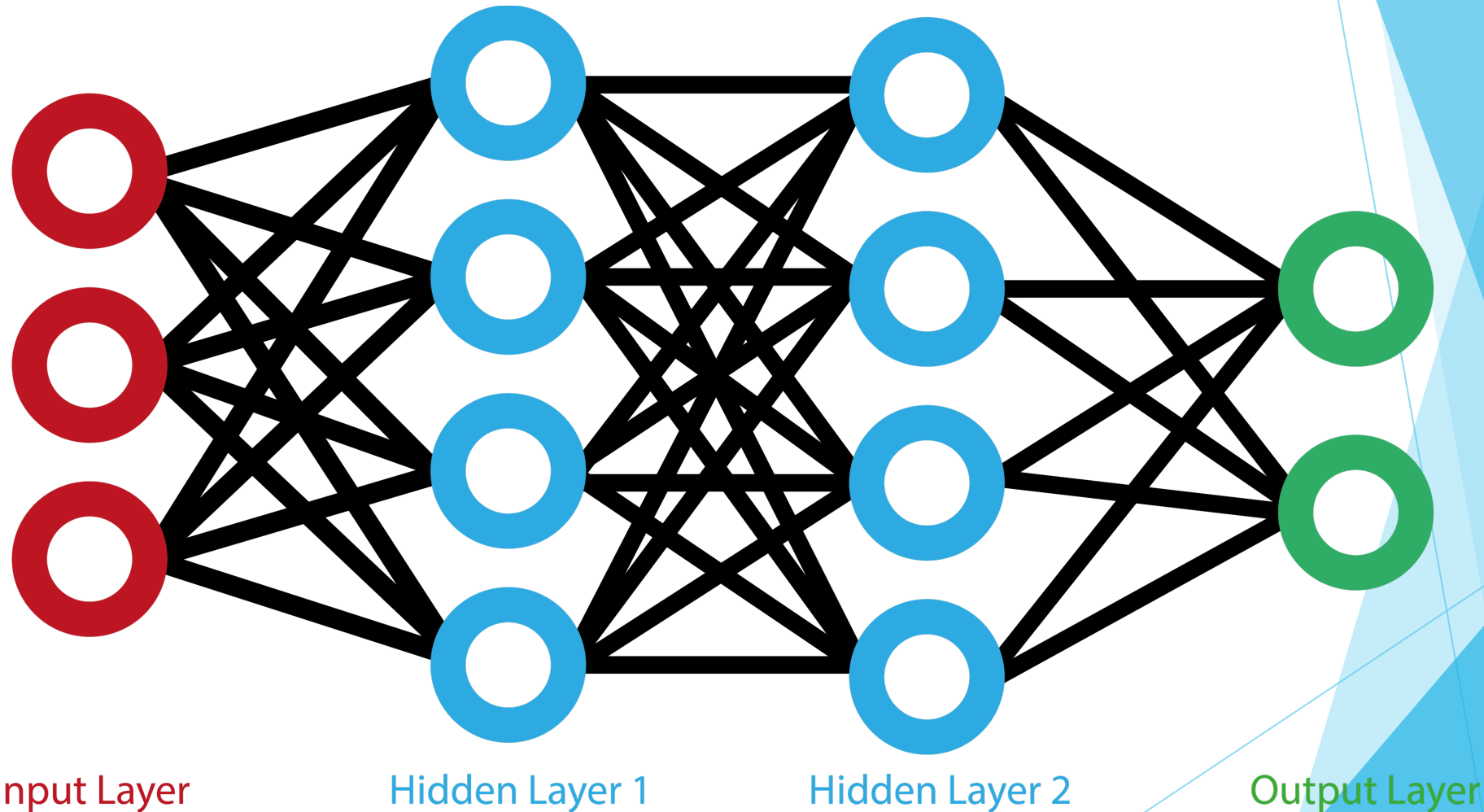
Neurone Biologique



Neurone Artificiel

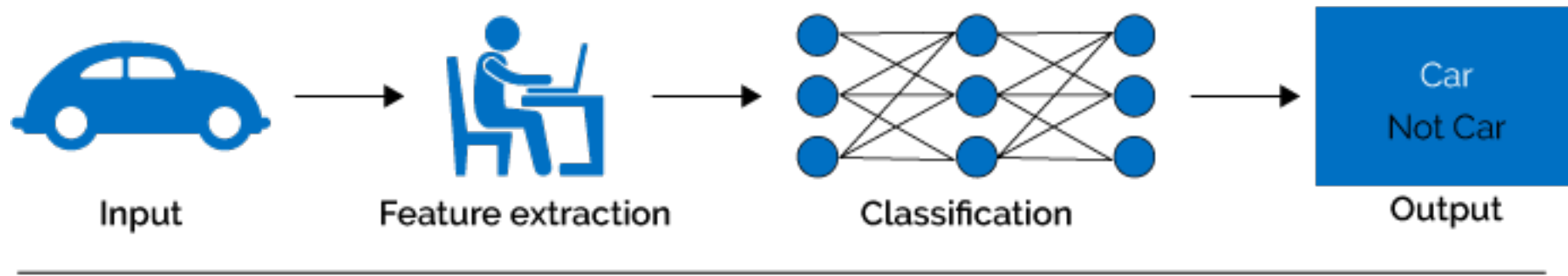


# Réseau de neurones

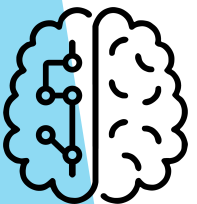
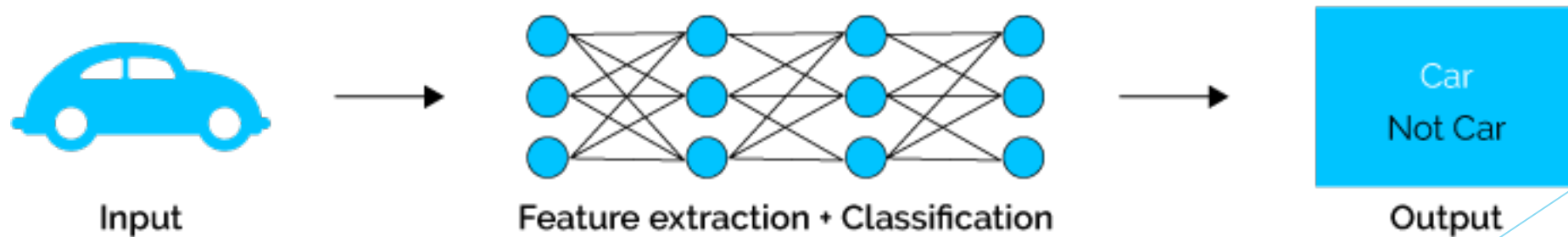


# Machine learning VS deep learning

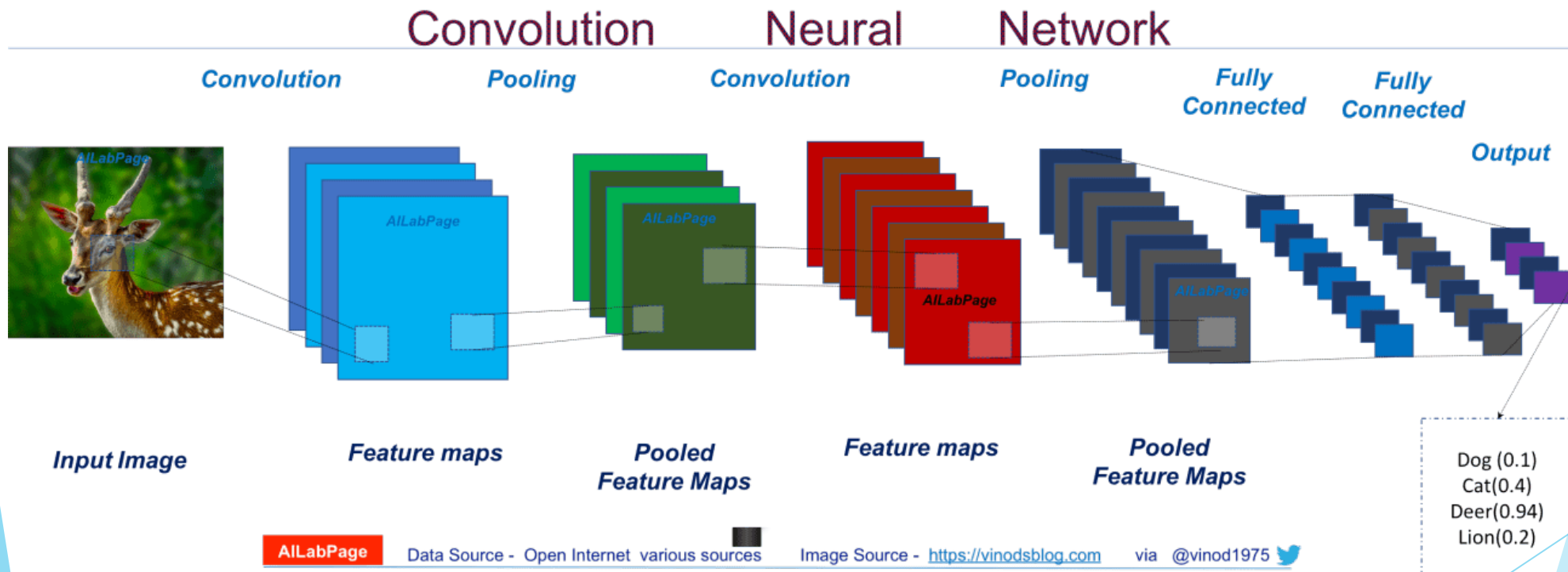
## Machine Learning



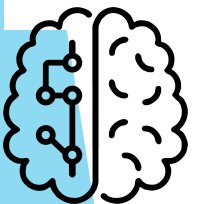
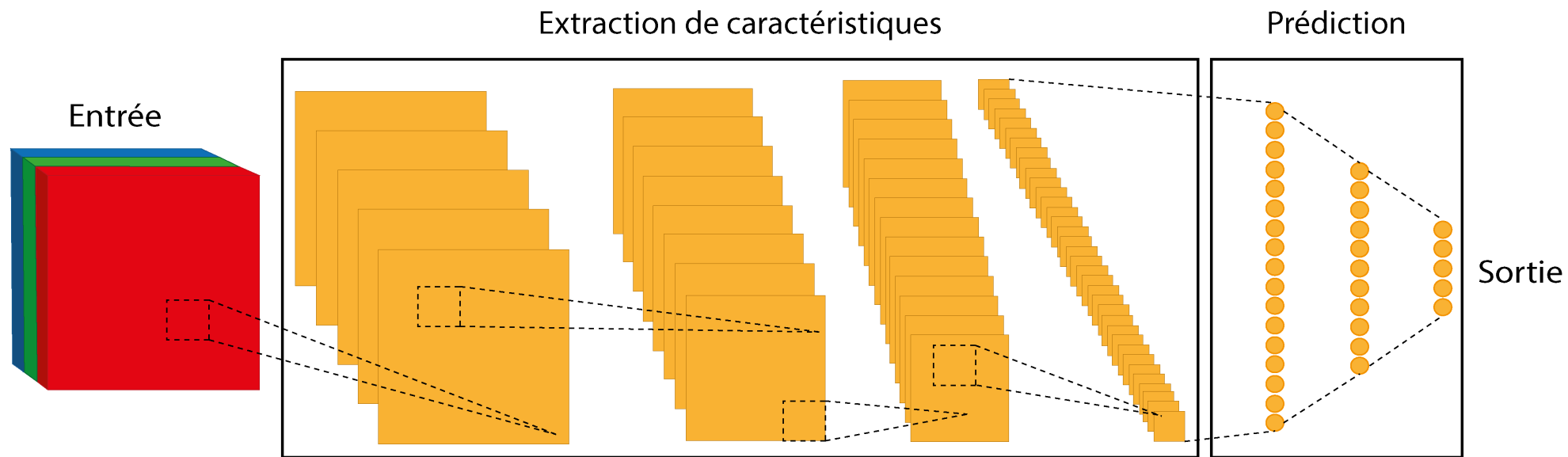
## Deep Learning



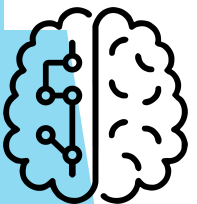
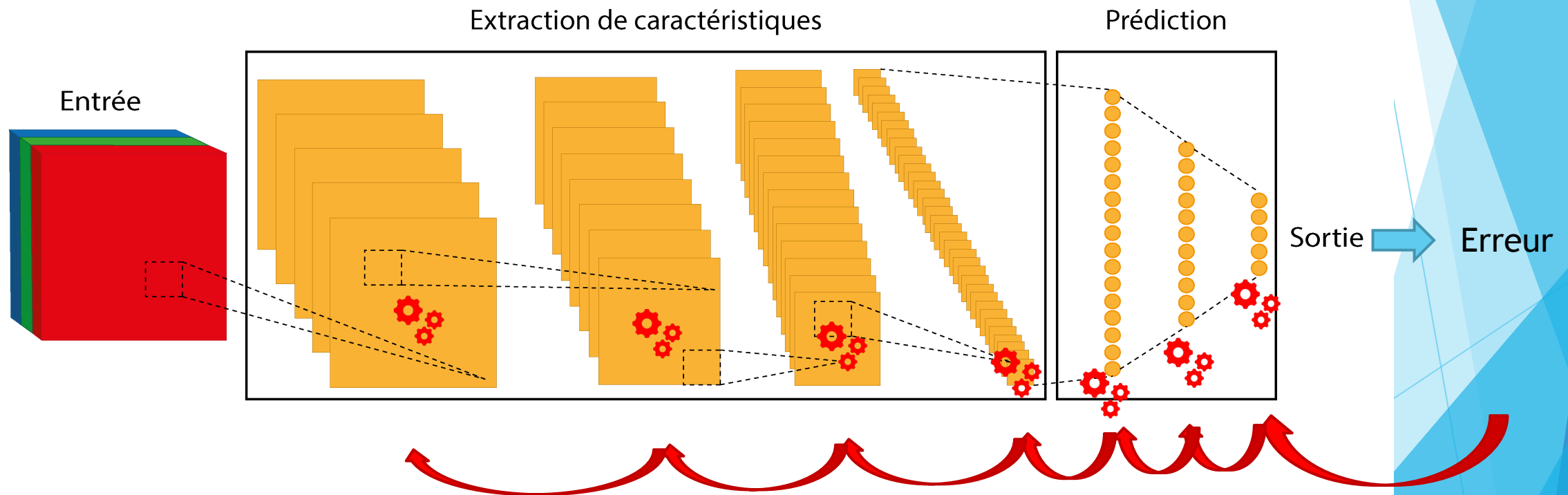
# Convolutional Neural Network (CNN)



# Extraction des caractéristiques automatiques

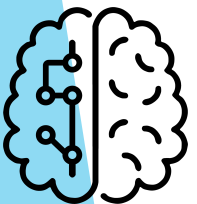
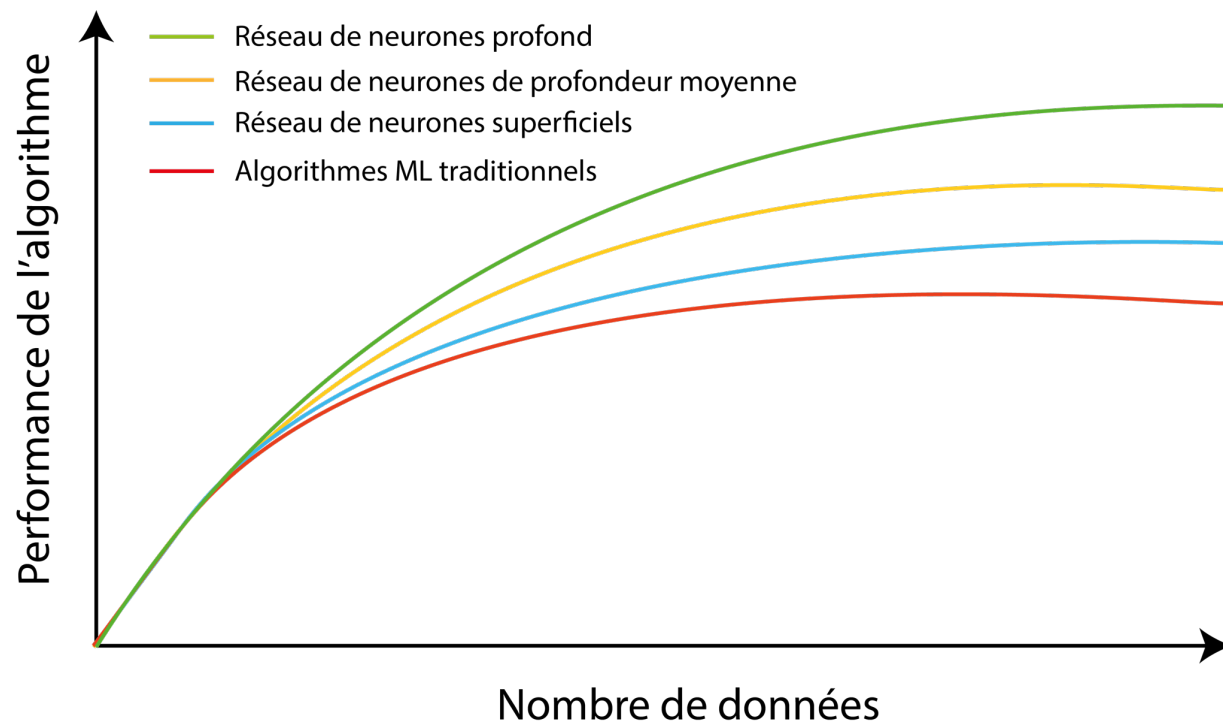


# Entraînement d'un modèle



# Performances du deep learning

Nous avons besoin de beaucoup de données labélisées pour entraîner un model de deep learning performant



# Les nouvelles possibilités

## Le son



## La vidéo



## Les images



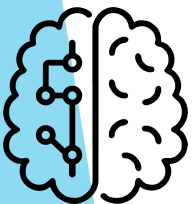
## Le texte

On t'a dit : « La vie est courte et tu vas vite en faire le tour. »  
On t'a dit : « Ne courbe jamais l'échine, cours, ma fille, cours... »  
Et dans cette course folle, t'as rencontré bien des malheurs.  
Car pour courir à mille à l'heure il fallait perdre des valeurs.  
Et puisqu'aimer, c'est prendre des risques, toi, tu braves les interdits,  
On t'a dit : « Faut bien te conduire », t'as répondu : « J'ai pas le permis ! »  
Insolente, lentement, je te vois dévaler la pente,  
Si tu te révais présidente, tu ne seras que figurante.  
Petite sœur, tu veux de l'or, des bolides et des sapes,  
Toujours dehors, toi, la vie, tu ne la vis plus, tu la snapes.  
Au-dessus de toi, tu vois les stars mais tu ne vois plus le ciel,  
Devenir celle dont tout le monde parle serait pour toi le casse du siècle.  
Envie de gloire obsessionnelle, de garde-robe exceptionnelle,  
Pour l'ascension professionnelle, l'honneur te sera optionnel.  
Dans ce monde que tu convoites, pour être un homme faut être prince,  
La place qu'il reste est si étroite que pour être star faut être mince.  
Tu vois des liasses, tu vois du rêve mais ça n'est que de la poussière,  
Certes ils brassent, mais ils n'aveuglent que ceux qui craignent la lumière.  
Tes modèles ne montrent d'eux que lorsqu'ils baignent dans l'opulence,  
Dans le silence ils taillent leurs veines et déambulent en ambulance.  
À trop courir après du vent, t'en oublieras le Divin,  
Au suivant ! Tu finiras par parler seule sur un divan.  
Crois-moi, quoi que tu gagnes, tu finis insatiable,  
Toi qui aimais que tout le monde t'aime, tu finiras associable.  
Contre un « je t'aime, je m'engage », tu donneras ton cœur en gage,  
Tu vieilliras jusqu'à ce qu'il parte avec une femme plus jeune en âge.  
Car c'est l'histoire qui se répète, toi qui voudrais changer de nom,  
Commence par te changer toi-même et tu pourras changer le monde.

- MÉLANIE DITE DIAM'S -

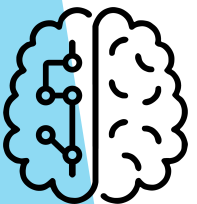
www.MELBYMEL.com f t i

Mel





# Clients 360°



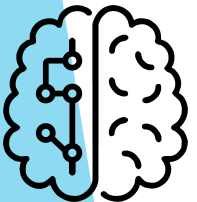
# Pour et contre



- ▶ Modèle le plus performant
- ▶ Extraction des caractéristiques automatiques

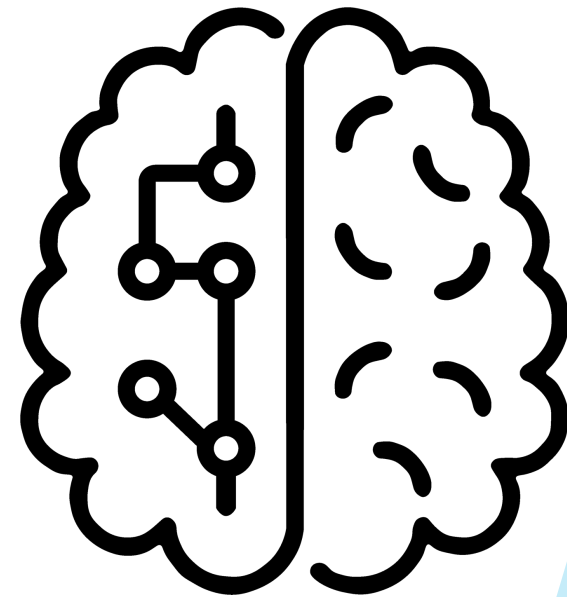


- ▶ Boîte noire
- ▶ Perte du sens commun
- ▶ Grande quantité de données nécessaire
- ▶ Apprentissage long et complexe



# Word embeddings

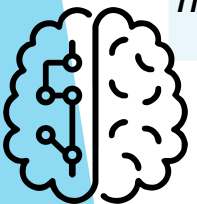
AIForYou - Morgan Gautherot



# Les matrices creuses

$V$	$D_1$	$D_2$	$D_3$	$D_4$	...	$D_n$
$mot_1$	0	<i>tfidf</i>	0	0	...	0
$mot_2$	0	0	0	<i>tfidf</i>	...	0
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
$mot_m$	0	<i>tfidf</i>	0	0	0	0

$V$	$D_1$	$D_2$	$D_3$	$D_4$	...	$D_n$
$mot_1$	0	1	0	0	...	0
$mot_2$	0	0	0	1	...	0
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
$mot_m$	0	1	0	0	0	0



# Collocation

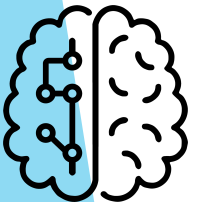
Lier les mots qui apparaissent souvent ensemble

« Je lis souvent le New-York times »



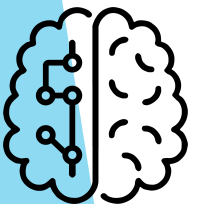
« New-York » + « times »

« New-York times »



# Word embeddings

- ▶ Représentent les mots et les documents par des vecteurs
- ▶ Est une représentation qui capture le sens du mot ou du document



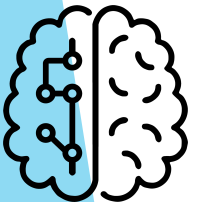
# Proximité des mots

Content = Papier = Heureux

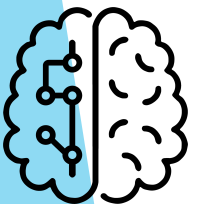
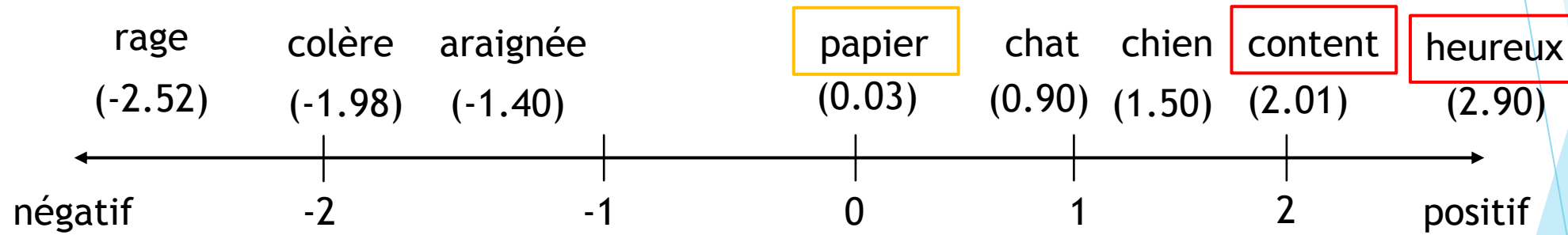
Content = [ 0, ... , 0 , 0, ... , 1, ... , 0, ... , 0, ... , 0]

Papier = [ 0, ... , 0 , 0, ... , 0, ... , 1, ... , 0, ... , 0]

Heureux = [ 0, ... , 1 , 0, ... , 0, ... , 0, ... , 0, ... , 0]

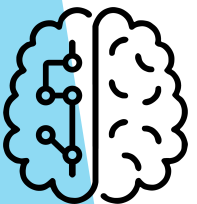
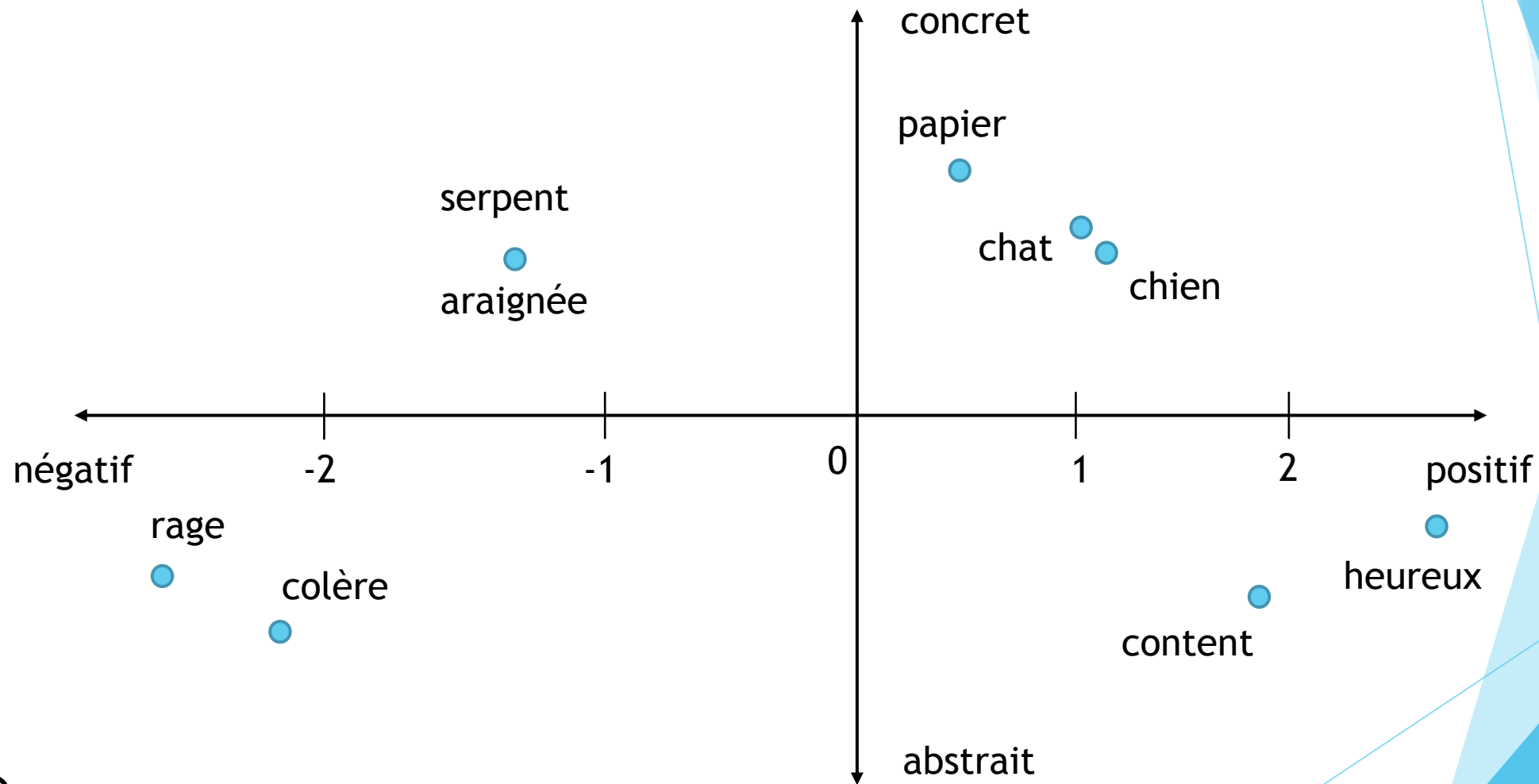


# Représentation sous forme vectorielle

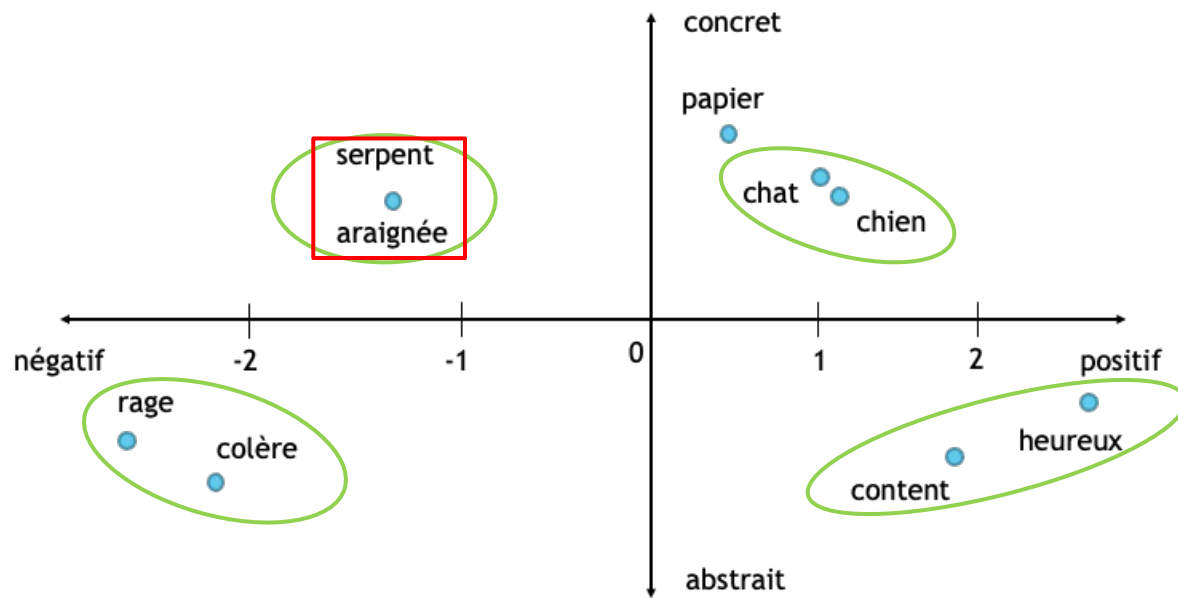




# Représentation sous forme vectorielle

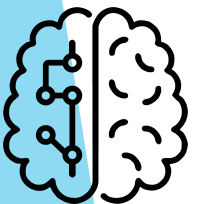


# Représentation sous forme vectorielle



↗ sens ↗ petite dimension  
↘ précision

Mots	$x_1$	$x_2$
araignée	-1.40	0.41
...	...	...
content	2.01	-0.32
...	...	...
serpent	-1.40	0.41
...	...	...



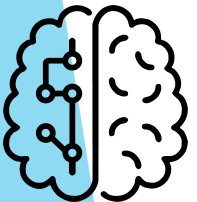
# Word2vec (Google 2013)

- ▶ Continuous bag-of-words (CBOW)

- ▶ « J'aime lire le ??????? en buvant mon café »

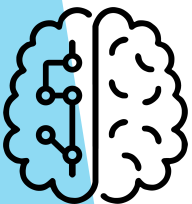
- ▶ Continuous skip-gram / skip-gram with negative sampling (SGNS)

- ▶ « ???? ?????? journal ????? ????? »



# GloVe (Stanford 2014)

- ▶ Factorisation du logarithme de la matrice de co-occurrence du corpus.

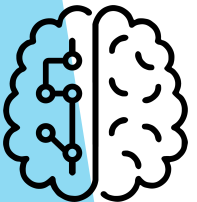


# FastText (Facebook, 2016)

- ▶ Prend en compte la structure des mots en les représentant par des n-gram
- ▶ Permet d'utiliser des mots non vus pendant l'entraînement (OOV, out-of-vocabulary)
- ▶ Se base sur les lettres qui composent les mots pour créer l'embedding

Exemple :

Chaton  $\approx$  Chat



# Technique plus avancées d'embeddings

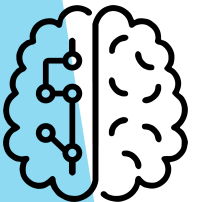
- ▶ Ces modèles ont des techniques de word embeddings qui change le vecteur d'un même mot suivant le contexte. (ex : orange, adjectif ou nom ?)

- ▶ BERT (Google, 2018)

- ▶ ELMo (Allen institute for AI, 2018)

- ▶ GPT-2 (OpenAI, 2018)

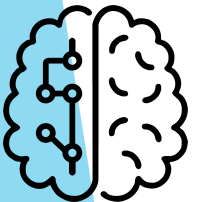
Des versions pré-entraîné de ces Modèles sont disponible sur internet



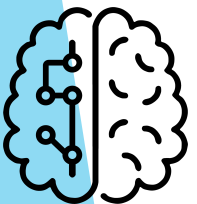
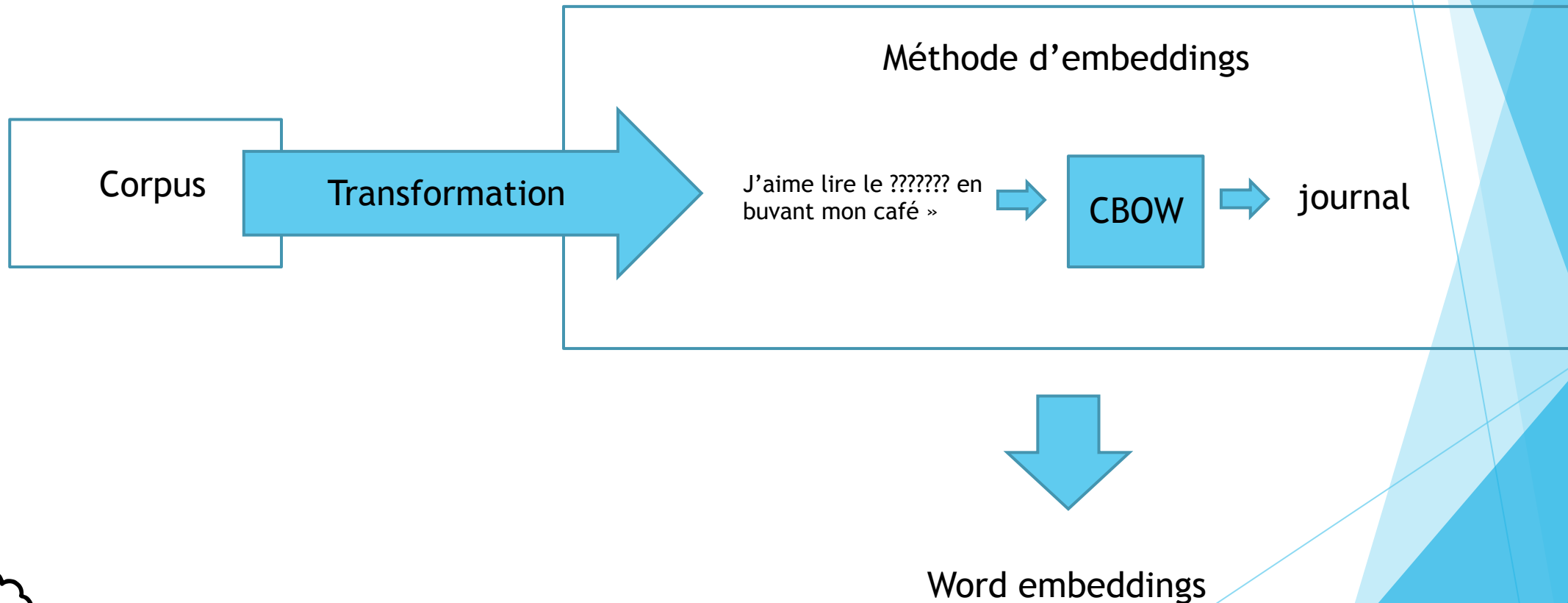
# Continuous bag-of-words

J'aime lire le ??????? en buvant mon café »

Self learning = unsupervised learning + supervised learning



# Continuous bag-of-words word embeddings





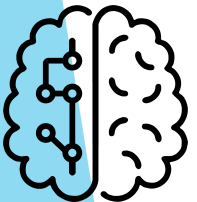
# Distance entre deux mots

La signification des mots sera déterminé d'après leur contexte

J'aime lire un journal en buvant mon café »

J'aime lire un livre en buvant mon café »

Livre  $\approx$  Journal



# Créer une observation du jeu d'entraînement

$C = 2$

Demi taille du contexte

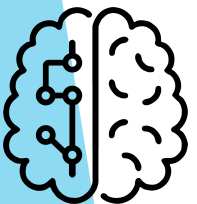
Taille de la fenêtre = 5

Fenêtre

Center word

J'aime lire un journal en buvant mon café »

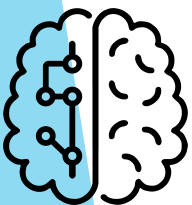
Context words



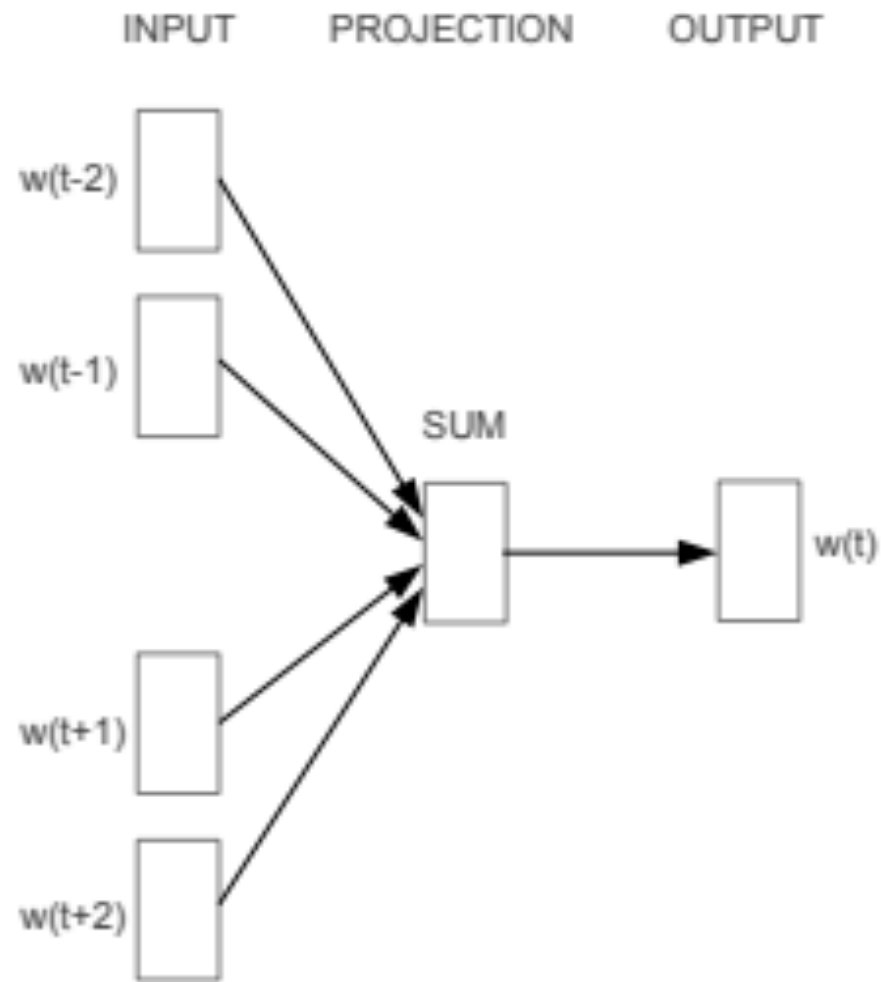
# Extraire les mots

$C = 2$

Context word 1 ( $x_1$ )	Context word 2 ( $x_2$ )	Context word 3 ( $x_3$ )	Context word 4 ( $x_4$ )	Center word ( $y$ )
lire	un	en	buvant	journal
journal	en	mon	café	buvant
...	...	...	...	...

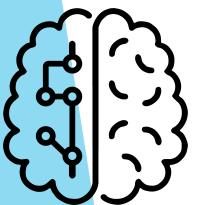


# CBOW



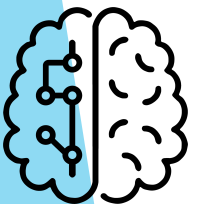
**CBOW**

Source : Efficient Estimation of Word Representations in Vector Space,  
Tomas Mikolov, Kai Chen, Greg Corrado,  
Jeffrey Dean



# Pré-traitement du corpus

- ▶ Tout mettre en minuscule
- ▶ Soit supprimer la ponctuation, soit tout remplacer par « . »
- ▶ Soit supprimer tout les nombres, soit tout remplacer par <NUMBER>
- ▶ Supprimer les caractères spéciaux \$ \* € ...

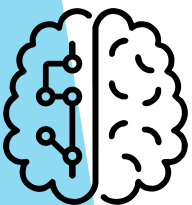


# Center word en vecteur

journal

$$V = \begin{bmatrix} arachide \\ \vdots \\ journal \\ \vdots \\ voiture \\ \vdots \\ zumba \end{bmatrix}$$

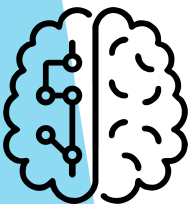
$$w_{journal} = \begin{bmatrix} 0 \\ \vdots \\ 1 \\ \vdots \\ 0 \\ \vdots \\ 0 \end{bmatrix}$$



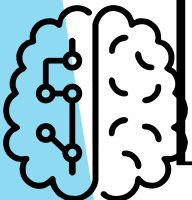
# Context words en vecteur

Lire, un, en, buvant

$$V = \begin{bmatrix} arachide \\ \vdots \\ buvant \\ \vdots \\ en \\ \vdots \\ lire \\ \vdots \\ un \\ \vdots \\ zumba \end{bmatrix}$$
$$\left[ \begin{array}{c} \text{lire} \\ \begin{bmatrix} 0 \\ \vdots \\ 0 \\ \vdots \\ 0 \\ \vdots \\ 1 \\ \vdots \\ 0 \\ \vdots \\ 0 \end{bmatrix} \end{array} + \begin{array}{c} \text{un} \\ \begin{bmatrix} 0 \\ \vdots \\ 0 \\ \vdots \\ 0 \\ \vdots \\ 0 \\ \vdots \\ 1 \\ \vdots \\ 0 \end{bmatrix} \end{array} + \begin{array}{c} \text{en} \\ \begin{bmatrix} 0 \\ \vdots \\ 0 \\ \vdots \\ 1 \\ \vdots \\ 0 \\ \vdots \\ 0 \\ \vdots \\ 0 \end{bmatrix} \end{array} + \begin{array}{c} \text{buvant} \\ \begin{bmatrix} 0 \\ \vdots \\ 1 \\ \vdots \\ 0 \\ \vdots \\ 0 \\ \vdots \\ 0 \\ \vdots \\ 0 \end{bmatrix} \end{array} \right] / 4 = \begin{array}{c} \text{Context vecteur} \\ \begin{bmatrix} 0 \\ \vdots \\ 0.25 \\ \vdots \\ 0.25 \\ \vdots \\ 0.25 \\ \vdots \\ 0.25 \\ \vdots \\ 0 \end{bmatrix} \end{array}$$



## Données d'entraînement


$$V = \begin{bmatrix} arachide \\ \vdots \\ buvant \\ \vdots \\ en \\ \vdots \\ journal \\ \vdots \\ lire \\ \vdots \\ un \\ \vdots \\ zumba \end{bmatrix}$$

$$x = \begin{bmatrix} 0 \\ \vdots \\ 0.25 \\ \vdots \\ 0.25 \\ \vdots \\ 0 \\ \vdots \\ 0.25 \\ \vdots \\ 0.25 \\ \vdots \\ 0 \end{bmatrix}$$

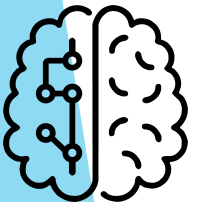
$$y = \begin{bmatrix} 0 \\ \vdots \\ 0 \\ \vdots \\ 0 \\ \vdots \\ 1 \\ \vdots \\ 0 \\ \vdots \\ 0 \\ \vdots \\ 0 \end{bmatrix}$$



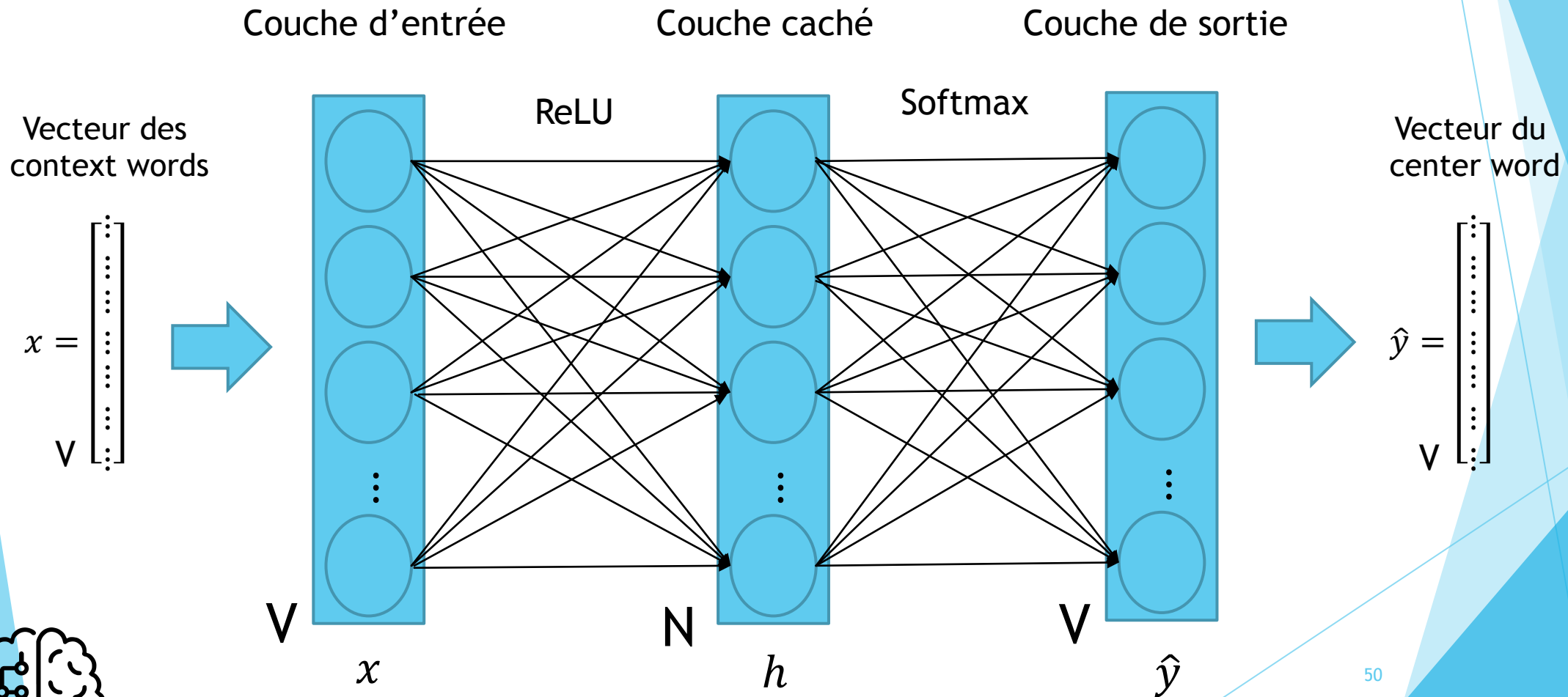
# Préparation des données TP 5

Objectifs :

- ▶ Pré-traiter les données textuelles
- ▶ Extraire les éléments de contexte des mots centraux
- ▶ Transformer les mots en vecteurs

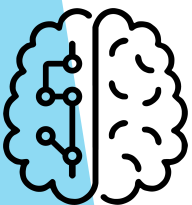
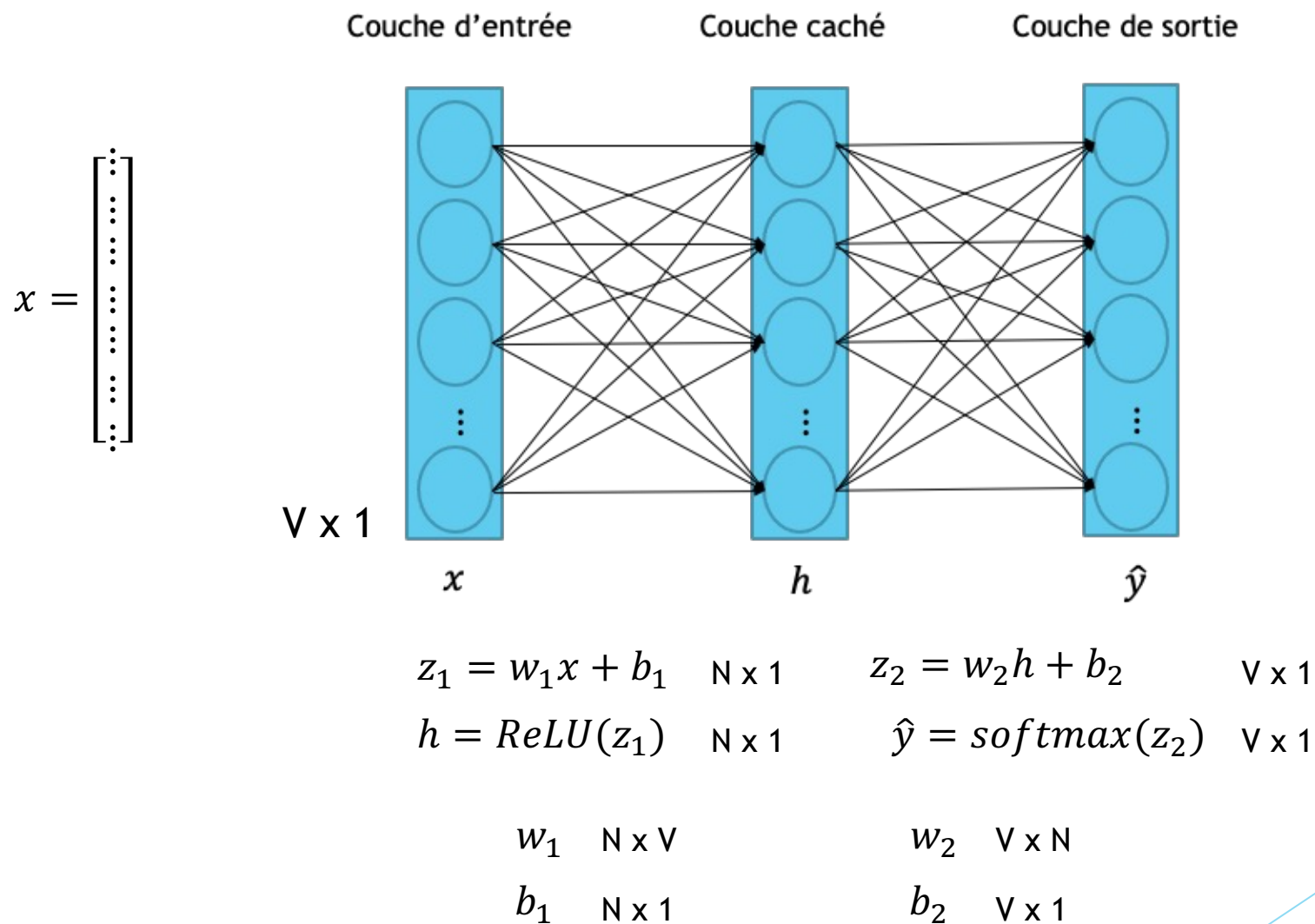


# L'architecture du modèle CBOW



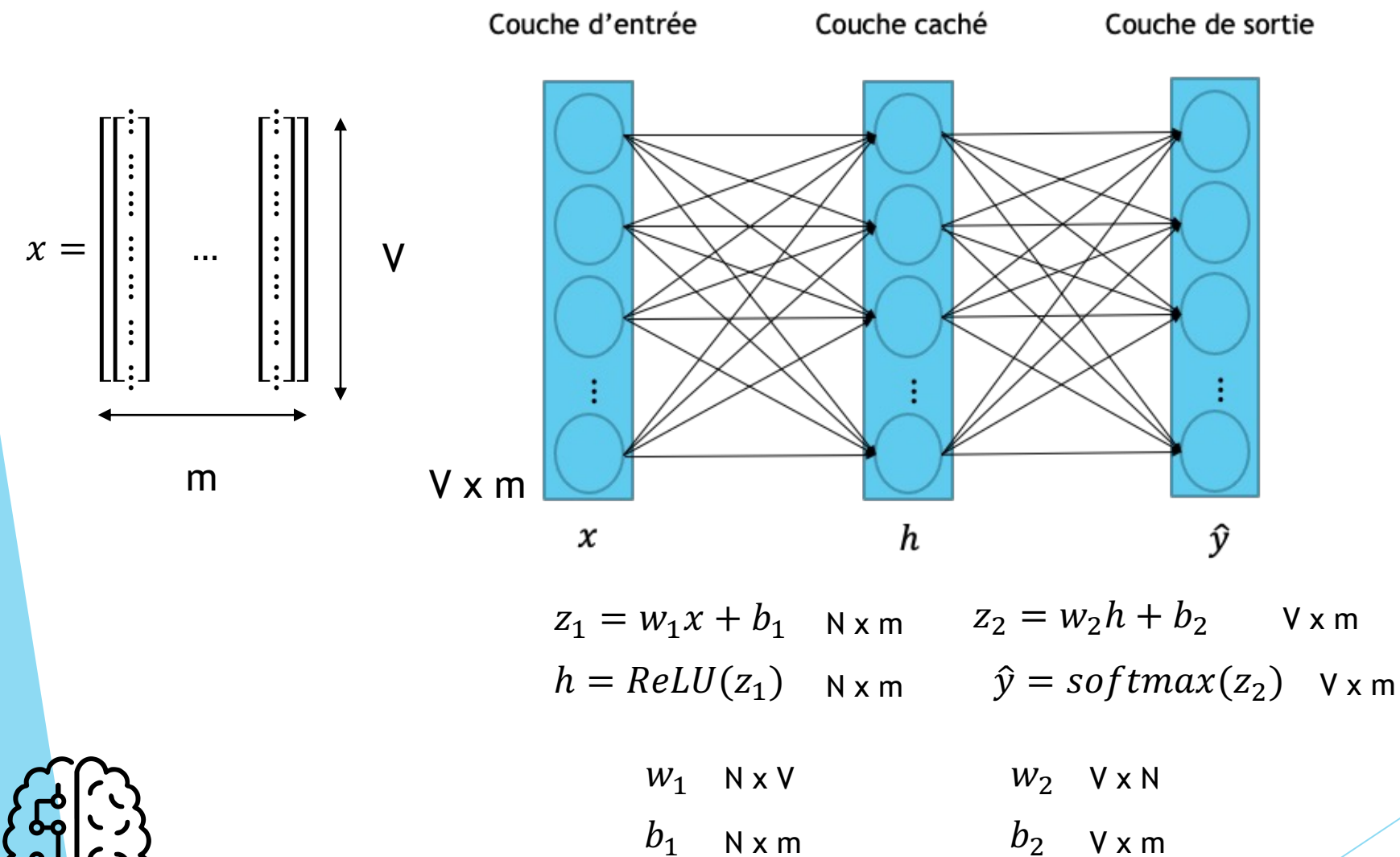
# Dimension de l'architecture

Avec un seul exemple

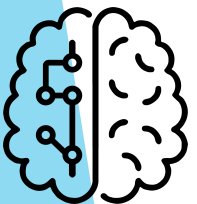
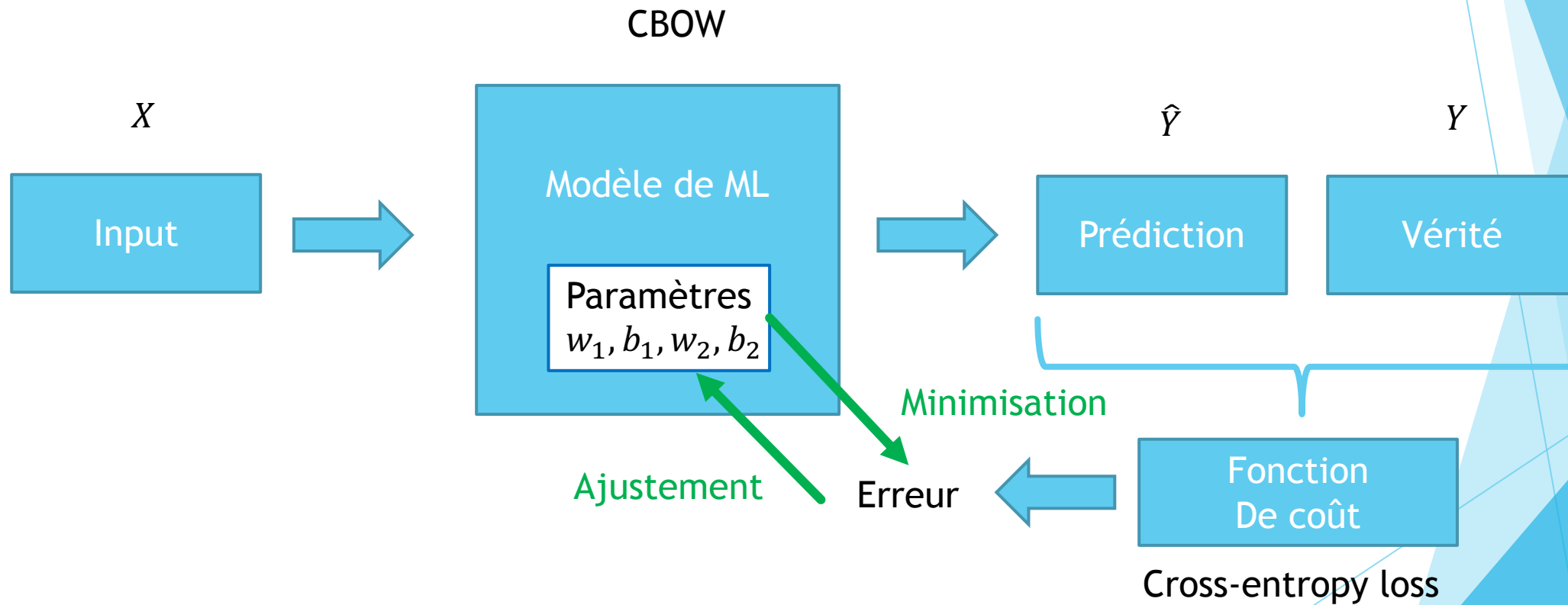


# Dimension de l'architecture

Avec  $m$  exemples

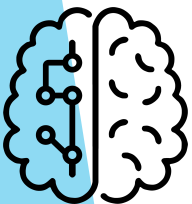


# Entraînement du CBOW



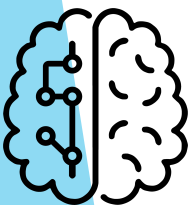
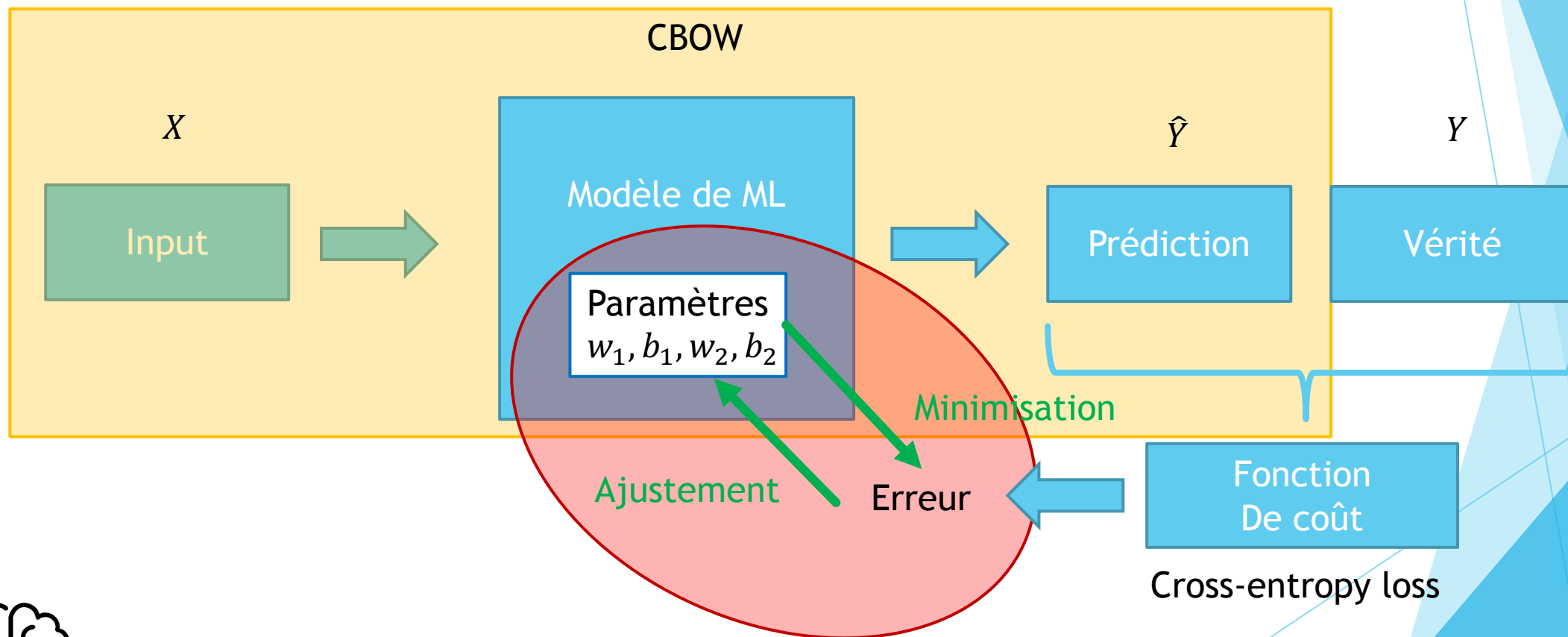
# Cross entropy loss

$$J = - \sum_{k=1}^V y_k \log \hat{y}_k$$

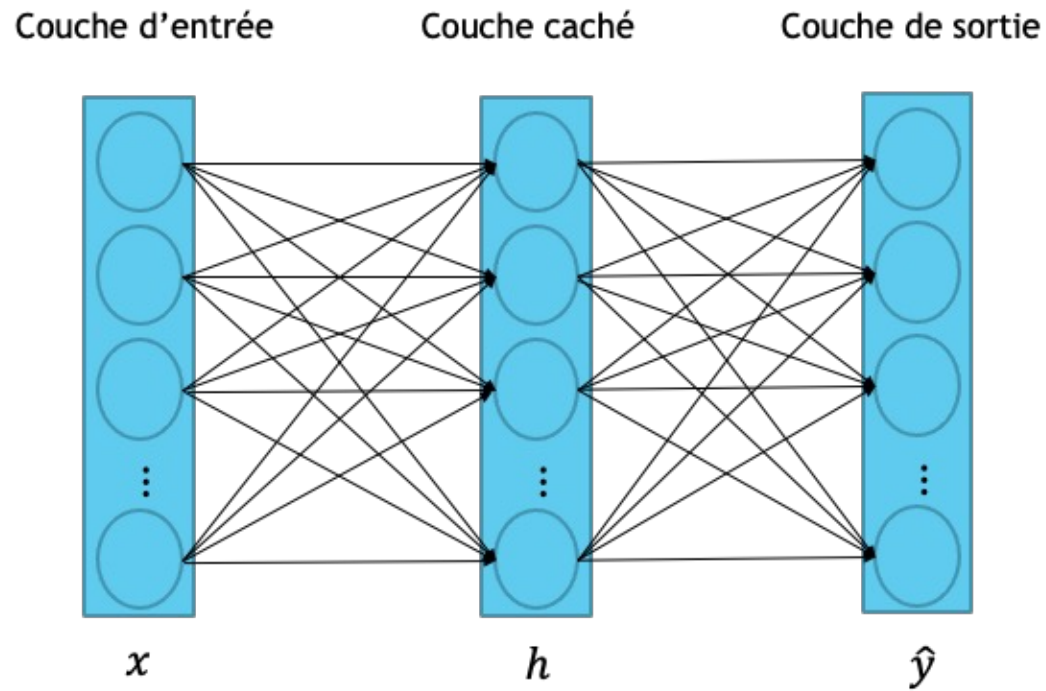


# Forward et backward propagation

Forward propagation

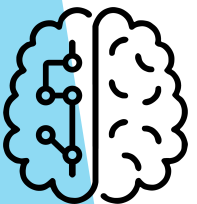


# Extraire la partie word embeddings option 1



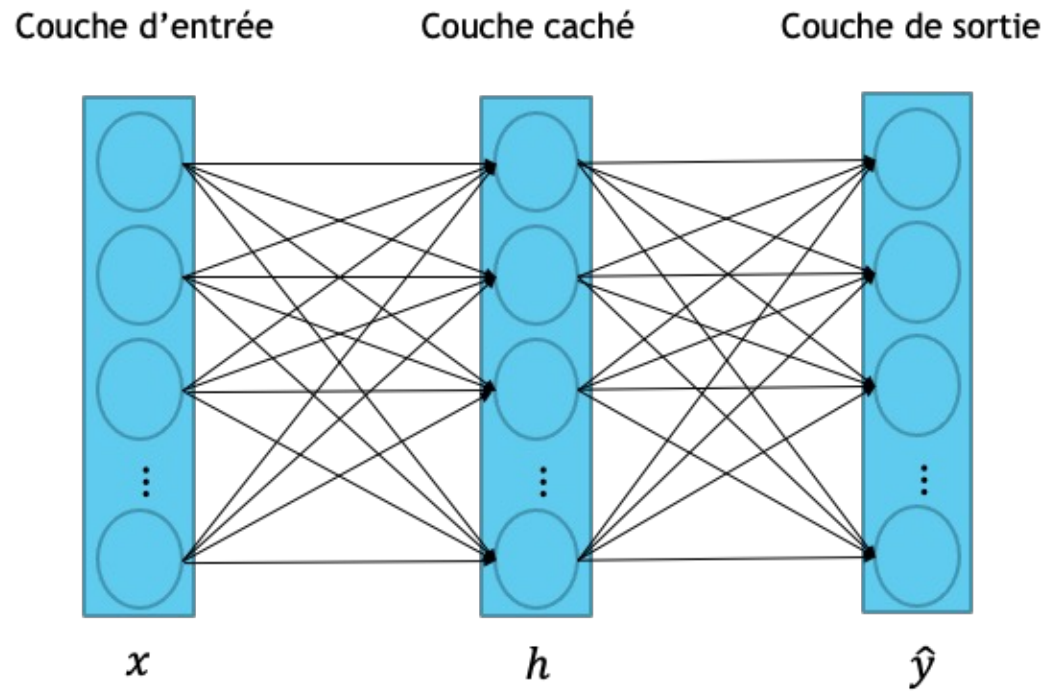
$$w_1 \quad N \times V$$

$$w_1 = \begin{bmatrix} [w^{(1)}] & \dots & [w^{(V)}] \end{bmatrix} \begin{matrix} \updownarrow N \\ \leftarrow V \rightarrow \end{matrix}$$



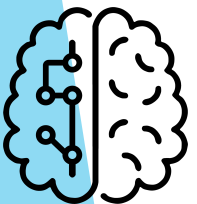


# Extraire la partie word embeddings option 2

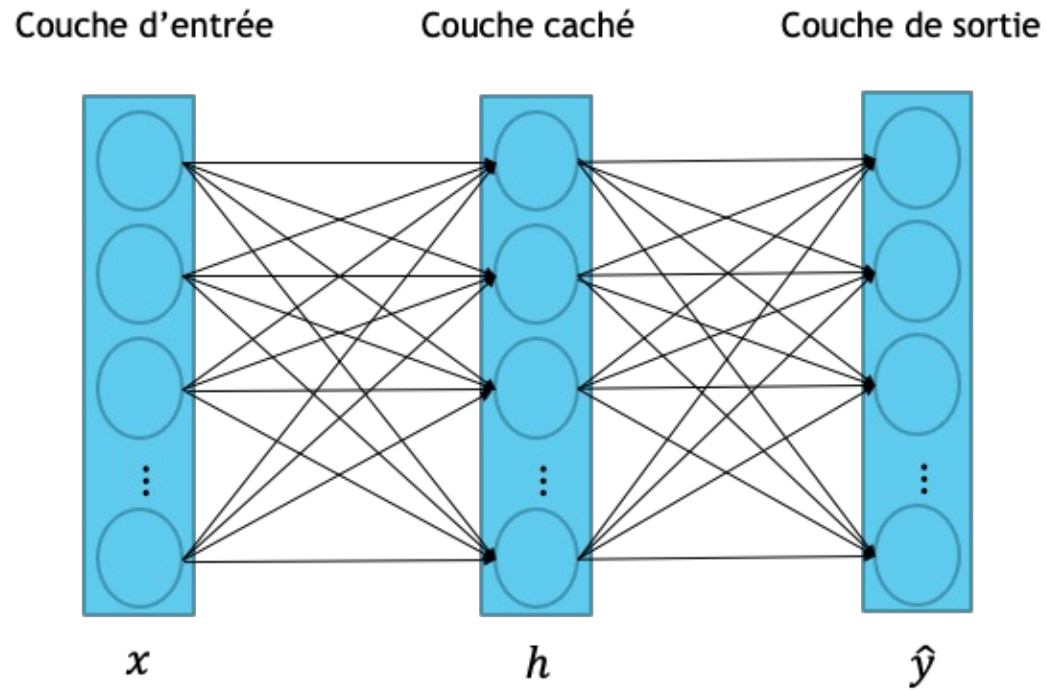


$$w_2 \quad V \times N$$

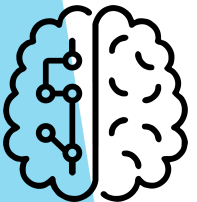
$$w_2 = \begin{bmatrix} [w^{(1)}] \\ \vdots \\ [w^{(V)}] \end{bmatrix} \begin{matrix} \updownarrow V \\ \leftarrow N \end{matrix}$$



# Extraire la partie word embeddings option 3



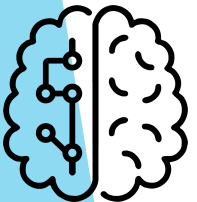
$$w_3 = 0.5(w_1 + w_2^T) = \underbrace{\begin{bmatrix} [w_3^{(1)}] & \dots & [w_3^{(V)}] \end{bmatrix}}_V \updownarrow N$$



# Implémenter le CBOW TP 6

Objectifs :

- ▶ Initialiser les poids du modèle
- ▶ Implémenter la fonction softmax
- ▶ Implémenter la forward propagation
- ▶ Implémenter la backward propagation
- ▶ Implémenter le gradient descent



# Entraîner un algorithme CBOW TP 7

Objectifs :

- ▶ Utiliser les fonctions du TP 5 pour préparer le jeu de données
- ▶ Initialiser une architecture de CBOW (avec Keras)
- ▶ Entraîner l'algorithme CBOW (avec Keras)
- ▶ Extraire le word embeddings pour les trois options

