

Deep Learning avec PyTorch

ALDEBARAN

Du Lundi 04 au Jeudi 07 novembre 2024



ORSYS
formation

Deep learning avec Pytorch

Partie 1 : Introduction à l'IA



Présenté par **Morgan Gautherot**



Plan des trois prochains jours

- Introduction à l'IA
- Les modèles d'IA
- Les outils à utiliser pour développer de modèles
- Comprendre le deep learning
- Manager un projet de deep learning
- La convolutional neural network
- Le transfer learning
- Le recurrent neural network
- Les transformers

Chaque cours sera accompagné de travaux pratiques pour appliquer les nouvelles notions apprises.



Qui je suis ?



Morgan Gautherot, PhD 

Youtuber | Data Scientist Freelance

Paris, Île-de-France, France · [Coordonnées](#)

[Newsletter Data & AI](#) 

13 513 abonnés · [Plus de 500 relations](#)

Lille Big Data and Machine Learning Meetup

 (363) 

 Lille, France

 2566 membres · Groupe public 



AI for you - Morgan Gautherot

@AlforyouMorganGautherot · 7,54 k abonnés · 174 vidéos

A travers cette chaîne YouTube, je mets à disposition du contenu de qualité. ...plus
fr.tipeee.com/aiforyou et 2 autres liens

[Personnaliser la chaîne](#)

[Gérer les vidéos](#)



Evaluation des connaissances

Introduction à l'IA



Partie 1 : Pourquoi utiliser le deep learning ?



Technologie de rupture

- L'ordinateur
- L'internet
- Le mobile
- L'intelligence artificielle

"Mieux vaut prendre le changement par la main avant qu'il ne vous prenne à la gorge"
Winston Churchill



Rendre votre entreprise plus efficace

Améliorer l'ensemble de vos procédures.

Cela touche toutes vos activités:

- Achats,
- Ressources humaines,
- Distributions,
- Productions,
- ...



La digitalisation au service de l'IA

Courrier papier

Information sur la réception

Email

L'ouverture

Les clics

Temps de lecture

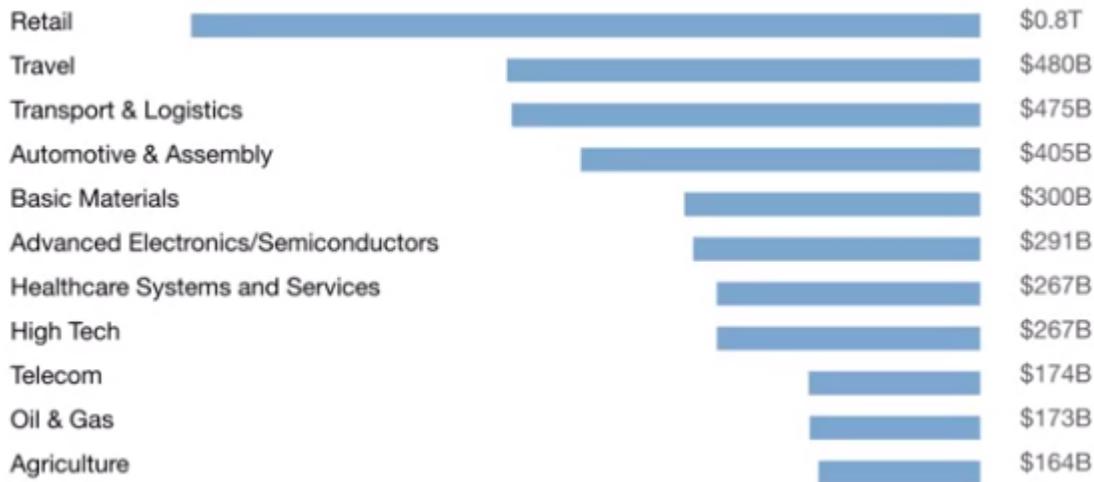
Pourcentage de l'email lu

...



Impact de l'IA

- Création de valeur par l'IA d'ici 2030

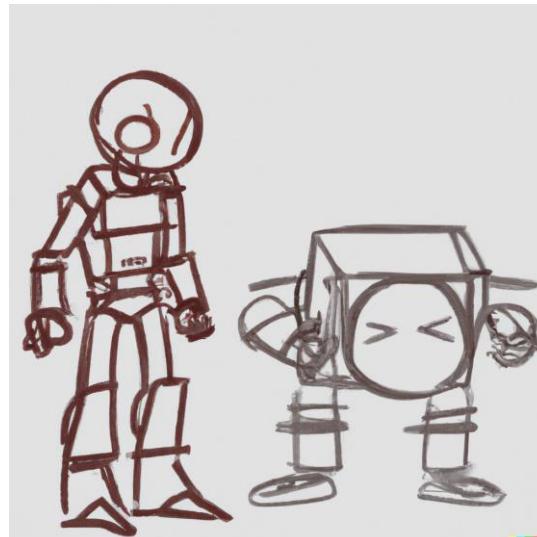


13 trillions de dollars



IA dans les grandes lignes

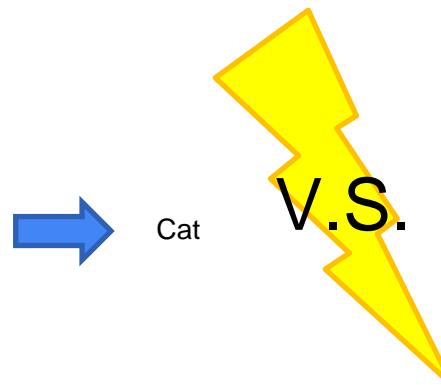
Domaine qui permet aux ordinateurs de copier le comportement humain





Intelligence Artificielle VS Intelligence Artificielle Générale

Artificial Narrow Intelligence



Intelligence Générale Artificielle



- Résoudre un problème très ciblé
- Ne peut pas être adapté à d'autres problèmes

Faire tout ce qu'un humain peut faire



Données structurées

Structured

1	0	0	1
1	0	1	0
1	1	1	0

1	0	0	1
1	1	1	0
0	1	1	0

0	1	0	1
1	1	1	0
1	0	0	0

1	0	0	1
1	0	1	0
1	0	0	0

1	0	0	1
0	0	1	0
1	1	1	0

1	0	0	1
1	0	1	0
1	1	1	0

1	0	0	1
1	0	1	0
1	1	1	0

1	0	0	1
1	0	1	0
1	0	0	0

0	1	0	1
1	1	1	0
1	0	0	0



Exemples d'applications

- Prédition du prix des maisons
- Prédition du contrôle d'un colis
- Détermination de la performance énergétique (DPE) d'un logement
- Diagnostique automatique des maladies
- Calcul du score d'appétance d'un produit
- Calcul du score de risque
- Segmentation d'une base clients



Traitement du langage naturel



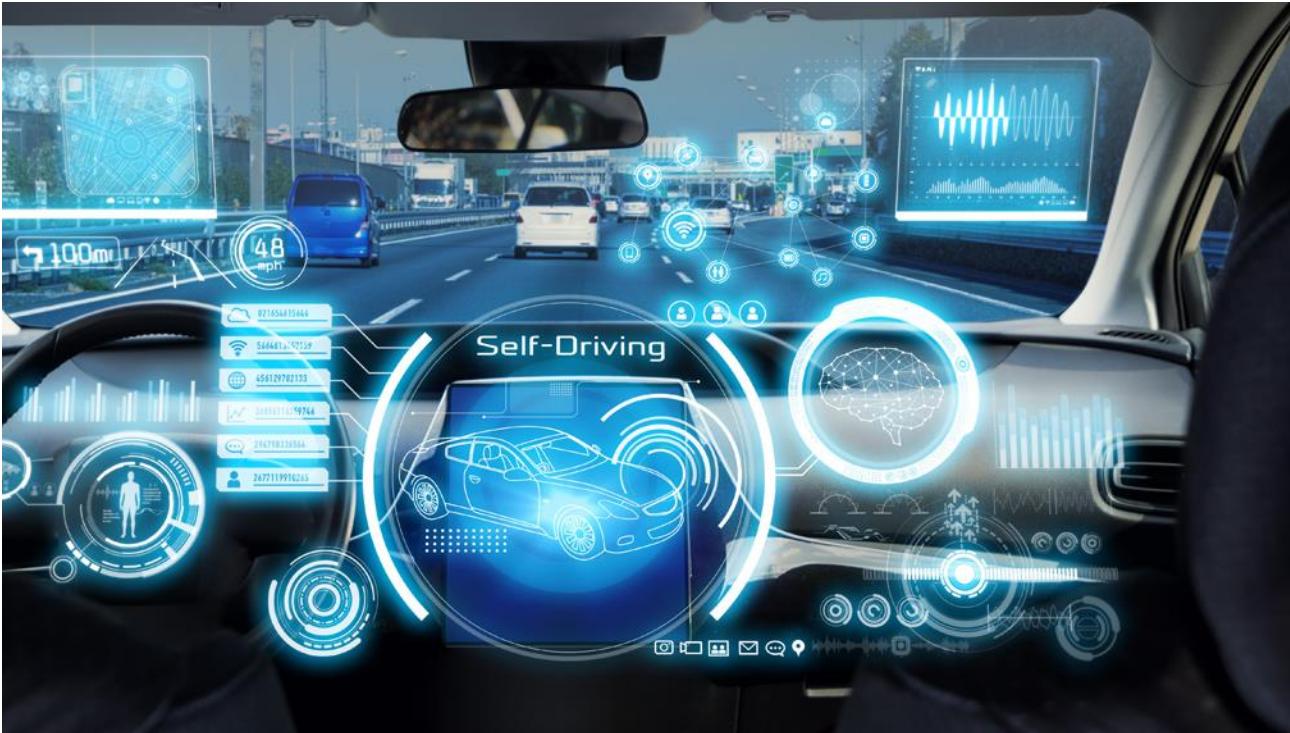


Exemples d'applications

- Auto-completion
- Traduction automatique
- Question/Answering
- Chatbot
- Résumé de text
- Analyse de sentiments
- Détection de commentaires toxiques



Vision par ordinateur





Exemples d'applications

- Validation automatique de documents
- Extraction d'information à partir d'un PDF
- Segmentation automatique
- Conduite autonome
- FaceID
- Détection de casses sur une chaîne de montage
- Détection de comportements anormaux



L'analyse du son





Exemples d'applications

- Transcription d'une conversation audio
- Traduction automatique
- Composition automatique
- Augmentation de la qualité audio
- Voix de synthèse
- Détection de tags pour une conversation audio



Système de recommandation



as



TV Shows





Exemples d'applications

- Catalogue personnalisé
- Accueil personnalisé
- Prochain visionnage
- Upsell
- Promotion personnalisée



Apprentissage par renforcement





Exemples d'applications

- Conduite autonome
- Jouer à Tetris
- Trouver de nouveaux algorithmes



IA génératif





Passage d'une modalité à l'autre

Textes



Images

- Midjourney,
- Stable diffusion,
- Dall-e

Textes



Textes

- ChatGPT,
- Bard

Textes



Code

- GitHub copilot



La réalité terrain

- Données structurées (80%)
- Traitement naturel du langage (10%)
- Vision par ordinateur (8%)
- Analyse du son (0.01%)
- Système de recommandation (1.99%)
- Apprentissage par renforcement (0%)
- IA génératif (?)



Principales applications

- Economie de temps
 - Planning automatique / Diagnostique automatique



Planning d'un service d'urgence

Créer un planning qui minimise le besoin d'agents en respectant les contraintes.

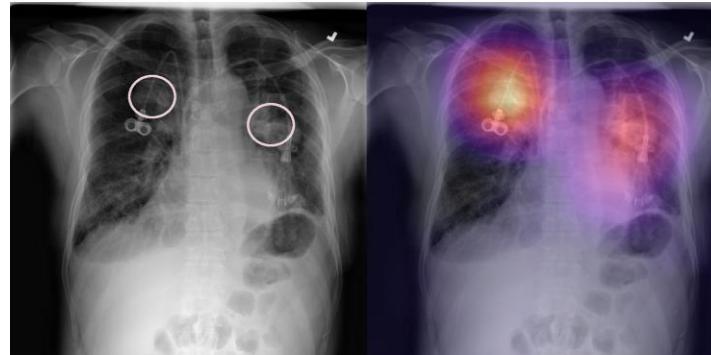
Contraintes:

- Minimum de 2 jours consécutif
- Maximum de 5 jours consécutif
- Avoir toujours au moins 3 agents en poste
- Poste de 8h ou 10h

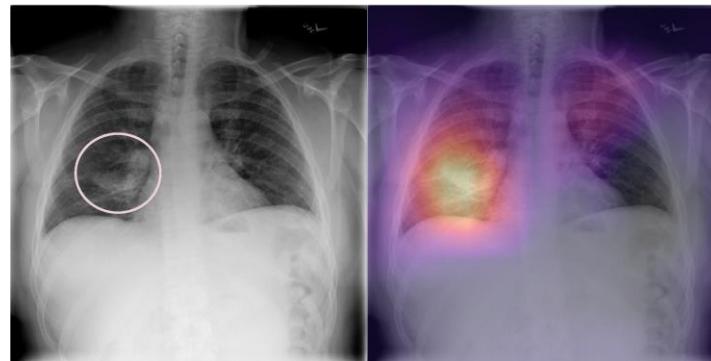


Diagnostique automatique

a



b





Diagnostique automatique

Pathology	Radiologists (95% CI)	Algorithm (95% CI)	Algorithm – Radiologists Difference (99.6% CI) ^a	Advantage
Atelectasis	0.808 (0.777 to 0.838)	0.862 (0.825 to 0.895)	0.053 (0.003 to 0.101)	Algorithm
Cardiomegaly	0.888 (0.863 to 0.910)	0.831 (0.790 to 0.870)	-0.057 (-0.113 to -0.007)	Radiologists
Consolidation	0.841 (0.815 to 0.870)	0.893 (0.859 to 0.924)	0.052 (-0.001 to 0.101)	No difference
Edema	0.910 (0.886 to 0.930)	0.924 (0.886 to 0.955)	0.015 (-0.038 to 0.060)	No difference
Effusion	0.900 (0.876 to 0.921)	0.901 (0.868 to 0.930)	0.000 (-0.042 to 0.040)	No difference
Emphysema	0.911 (0.866 to 0.947)	0.704 (0.567 to 0.833)	-0.208 (-0.508 to -0.003)	Radiologists
Fibrosis	0.897 (0.840 to 0.936)	0.806 (0.719 to 0.884)	-0.091 (-0.198 to 0.016)	No difference
Hernia	0.985 (0.974 to 0.991)	0.851 (0.785 to 0.909)	-0.133 (-0.236 to -0.055)	Radiologists
Infiltration	0.734 (0.688 to 0.779)	0.721 (0.651 to 0.786)	-0.013 (-0.107 to 0.067)	No difference
Mass	0.886 (0.856 to 0.913)	0.909 (0.864 to 0.948)	0.024 (-0.041 to 0.080)	No difference
Nodule	0.899 (0.869 to 0.924)	0.894 (0.853 to 0.930)	-0.005 (-0.058 to 0.044)	No difference
Pleural thickening	0.779 (0.740 to 0.809)	0.798 (0.744 to 0.849)	0.019 (-0.056 to 0.094)	No difference
Pneumonia	0.823 (0.779 to 0.856)	0.851 (0.781 to 0.911)	0.028 (-0.087 to 0.125)	No difference
Pneumothorax	0.940 (0.912 to 0.962)	0.944 (0.915 to 0.969)	0.004 (-0.040 to 0.051)	No difference

^aThe AUC difference was calculated as the AUC of the algorithm minus the AUC of the radiologists. To account for multiple hypothesis testing, the Bonferroni-

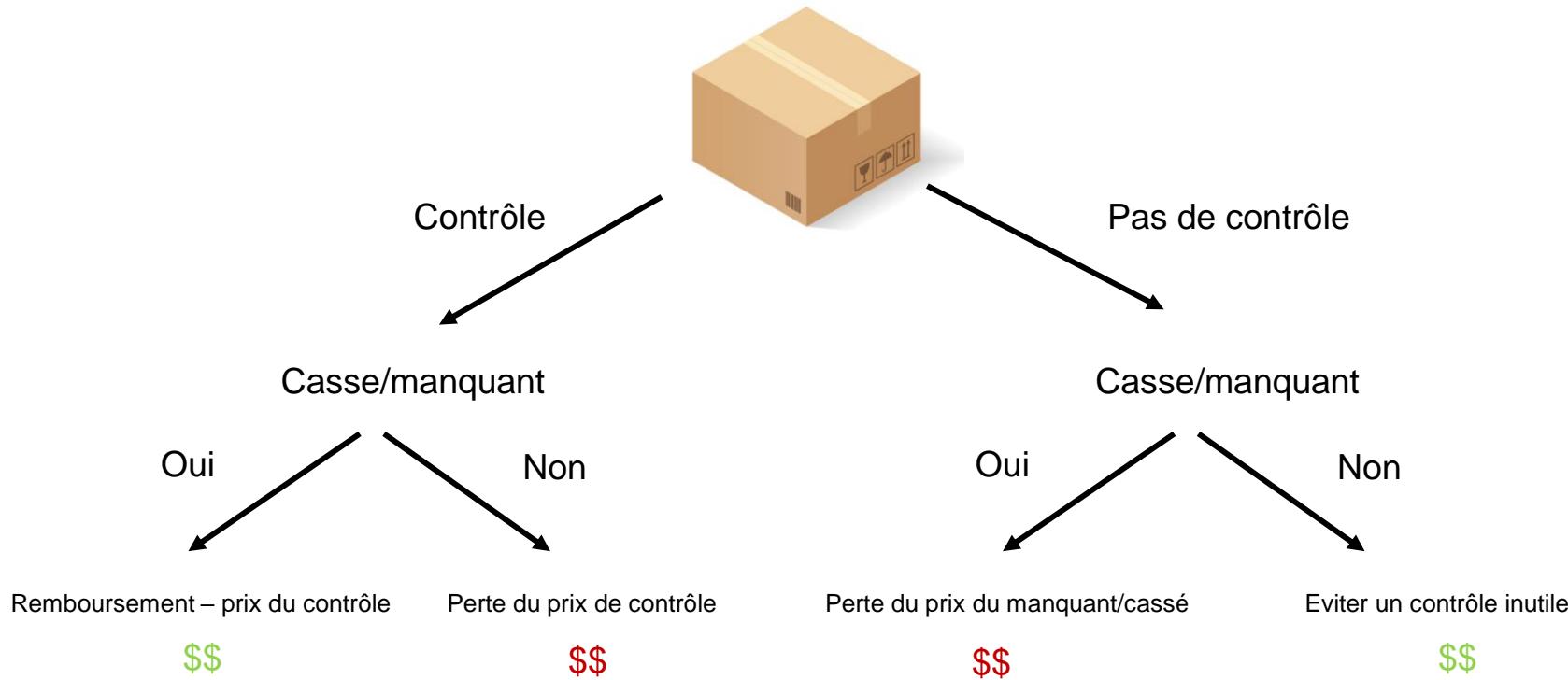


Principales applications

- Economie de temps
 - Planning automatique
- Economie argent
 - Scoring



Contrôle des colis





Résultat

- Processus existant:
 - 1€ investi en contrôle -> 0.3€ de gain
- Utilisation du machine learning:
 - 1€ investi en contrôle -> 1.5€ de gain
- + réduction du temps de contrôle

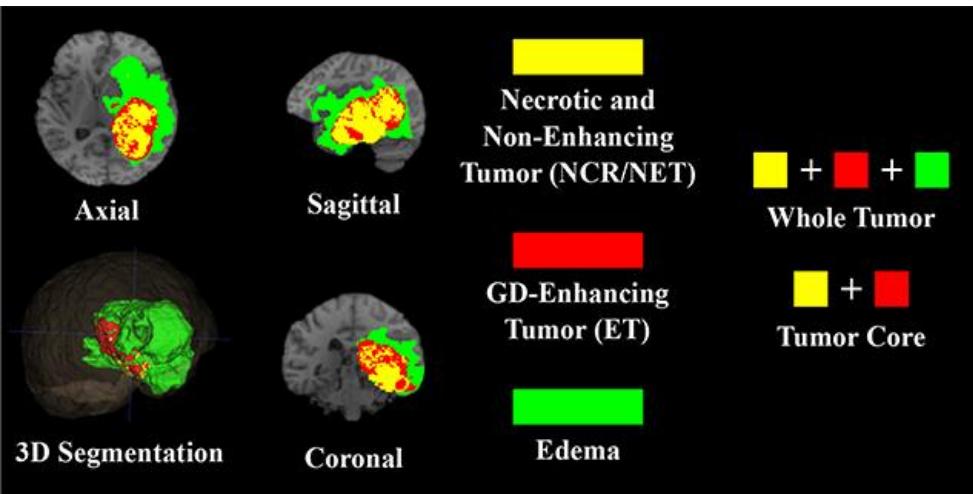


Principales applications

- Economie de temps
 - Planning automatique
- Economie argent
 - Scoring
- Augmentation de performances
 - Segmentation automatique

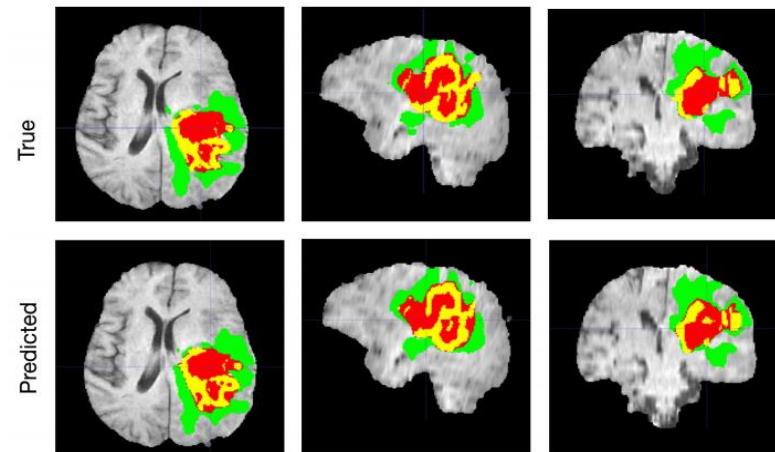


Segmentation de tumeur cérébrale



Segmentation manuelle de tumeur par des radiologues expert:

- La variabilité intra-opérateur était de 20% +/- 15%.
- La variabilité inter-opérateur était de 28% +/- 12%.



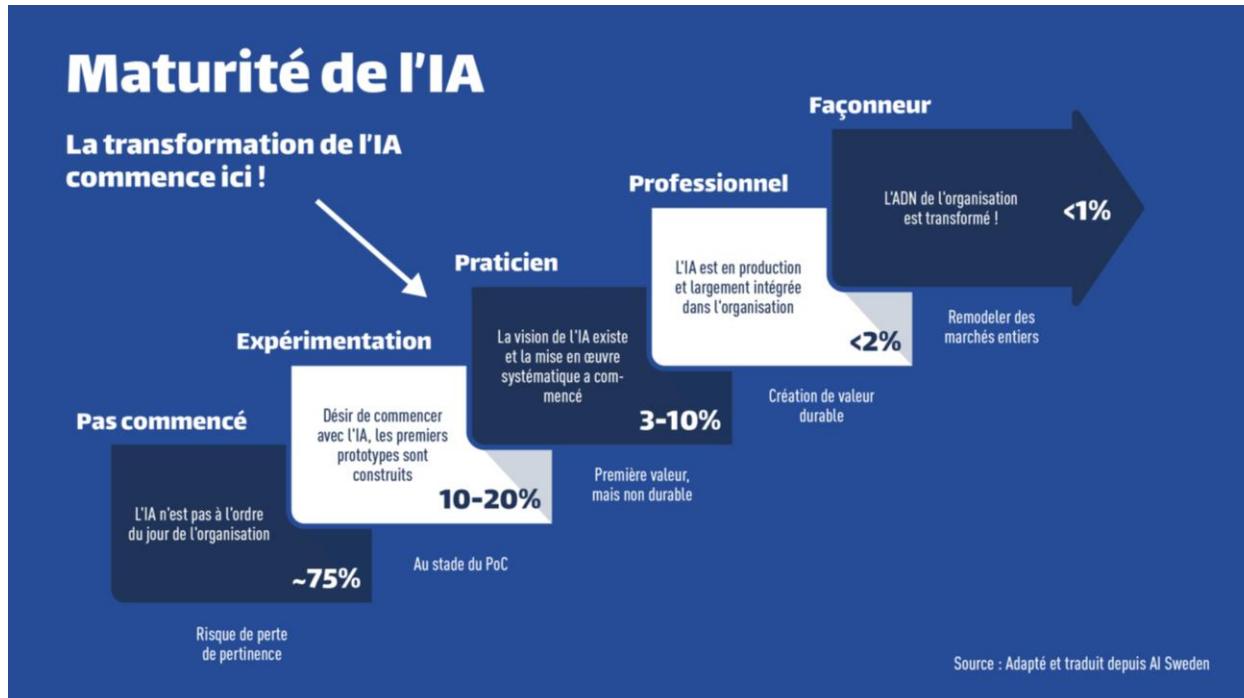


La maturité de la data

- Acculturation à la data/IA
- Organisation de la data
- Dashboard
- Application IA



Où sont les concurrents ?





Où en êtes-vous ?

- Direction
- Ambition
- Cas d'utilisation
- Organisation
- Expertise
- Culture
- Technologie
- Données
- Ecosystème
- Exécution

<https://www.appliedai.de/en/>

Deep learning avec Pytorch

Partie 2 : Les modèles IA



Présenté par **Morgan Gautherot**

Expert system

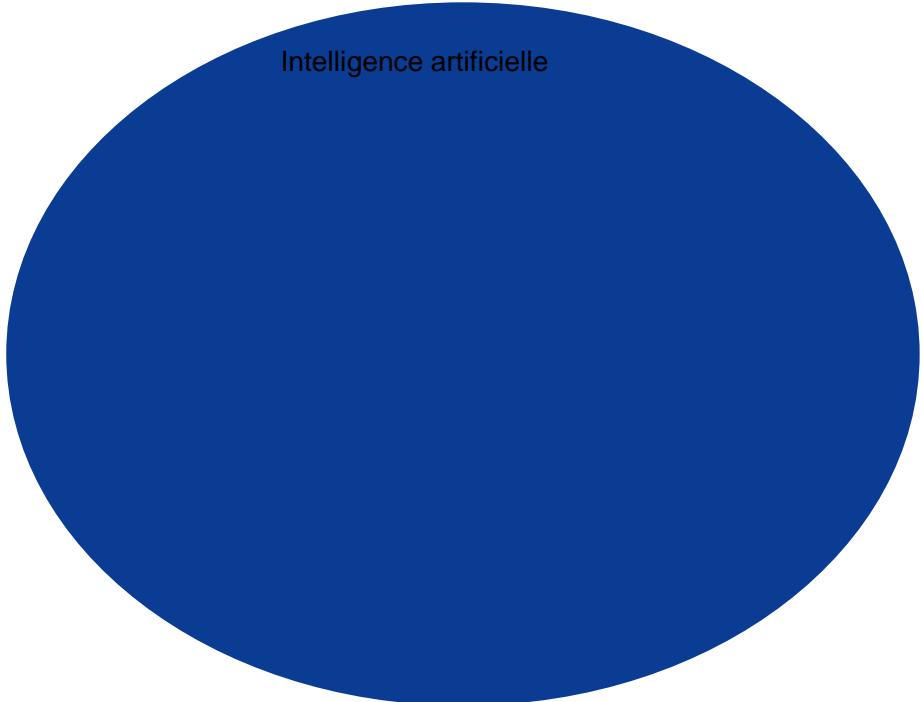


Partie 2 : Les modèles IA



Definitions

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



Intelligence artificielle



L'IA à travers le temps

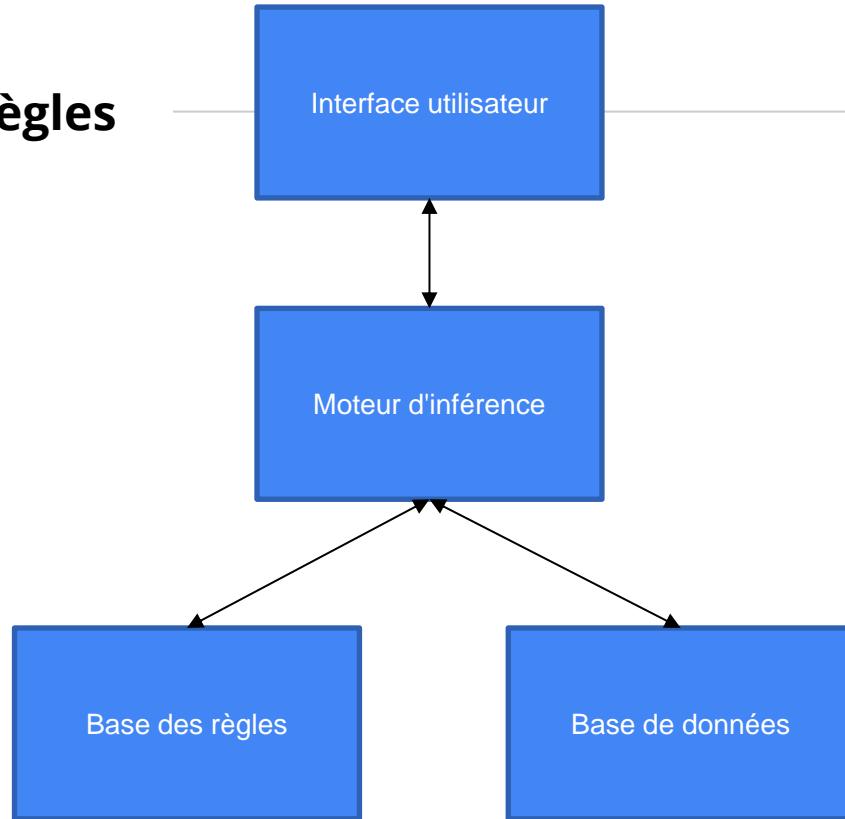
1957

Système expert



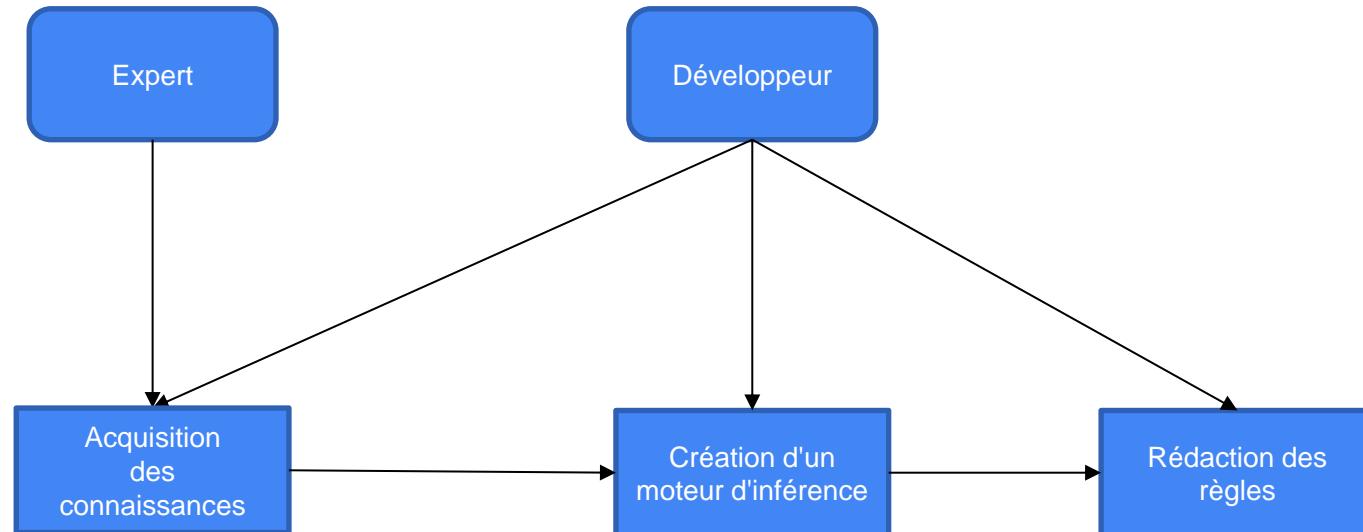
Système basé sur des règles

- Un système expert est un programme qui répond à des questions ou résoud des problèmes dans un domaine de connaissances donné, en utilisant des règles logiques dérivées des connaissances d'experts humains dans ce domaine.





Construire un système





Applications

- Diagnostic de l'analyse du sang
- Diagnostic de panne pour les voitures
- Vérification des systèmes de l'avion avant le décollage
- Ligne d'assemblage automatisée



Avantages et inconvénients



- Mise en œuvre du bon sens
- Performer dans des environnements fermés
- Modèle interprétable



- Long à mettre en place
- Difficile à maintenir
- Mise à jour constante dans le cas des environnements ouverts
- Approche biaisée

Machine learning



Partie 2 : Les modèles IA



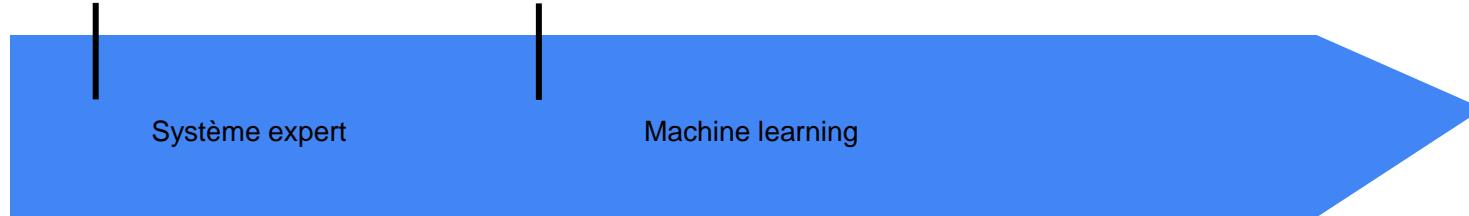
L'IA à travers le temps

1957

1970

Système expert

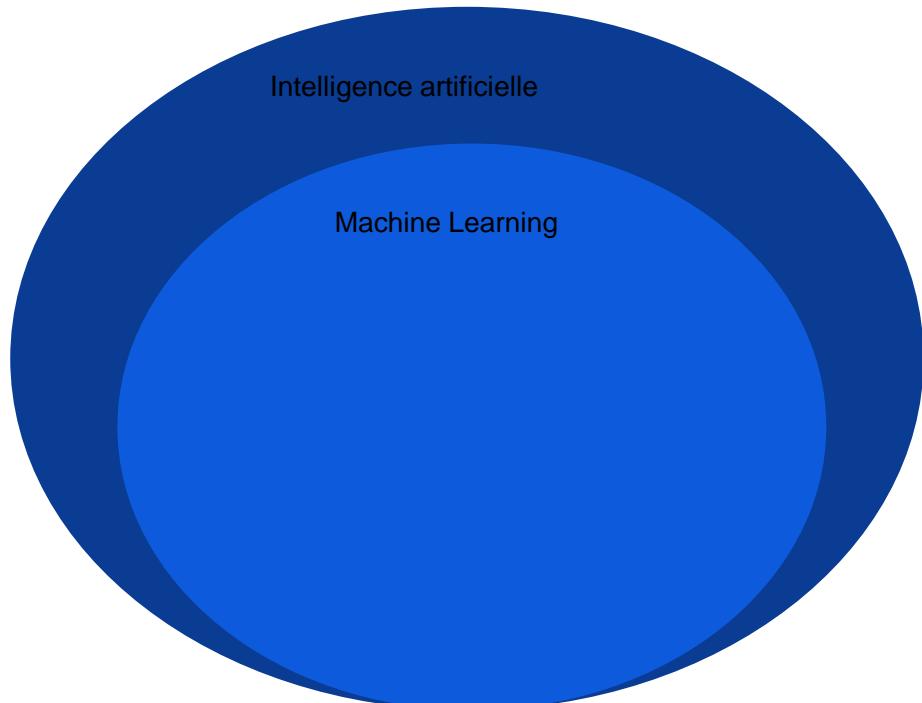
Machine learning





Définitions

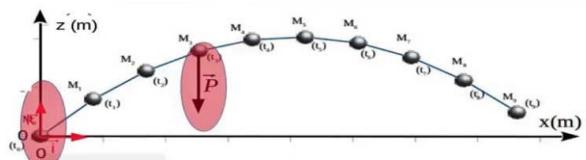
- L'intelligence artificielle (IA) est un domaine de l'informatique dont l'objectif est de créer des programmes qui accomplissent des tâches normalement dévolues à l'intelligence humaine et de la simuler.
- Le machine learning (ML) est une branche de l'IA. Il a la capacité d'apprendre des données à l'aide d'un algorithme d'apprentissage dont le but est de réaliser des analyses explicatives, prédictives ou préventives.





Prédire où atterri le boulet de canon ?

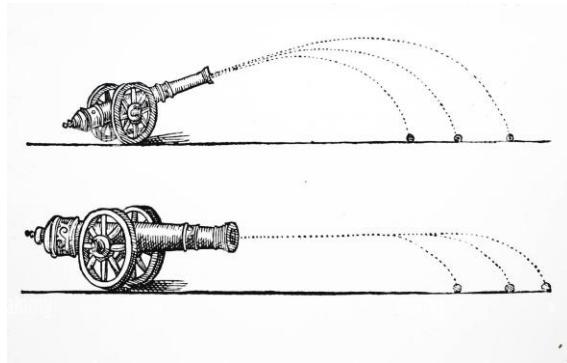
Système expert



2ème loi de Newton :

$$\Sigma \vec{f}_{ext} = m_{syst} \times \vec{a}$$
$$\vec{a} = \frac{\Sigma \vec{f}_{ext}}{m_{syst}} = \frac{\vec{P}}{m_{syst}} = \frac{-m \times g \times \hat{k}}{m} = -g\hat{k}$$

Machine learning



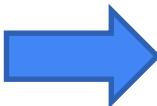
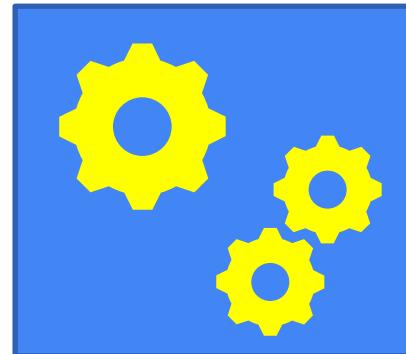
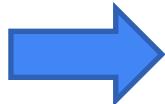
$$F(x) \approx -gk$$



Machine learning

Modèle

Data



Prédiction



Qu'est ce qu'une donnée ?

Objet de l'étude



Mesure de variables

Nombre de pièces

5

Surface

$60 m^2$

Garage

Non

:

:

}

Données structurées



Data base



Nombre de pièces

5

Surface

$60\ m^2$

Garage

Non



3

$30\ m^2$

Oui



7

$140\ m^2$

Oui

⋮

⋮

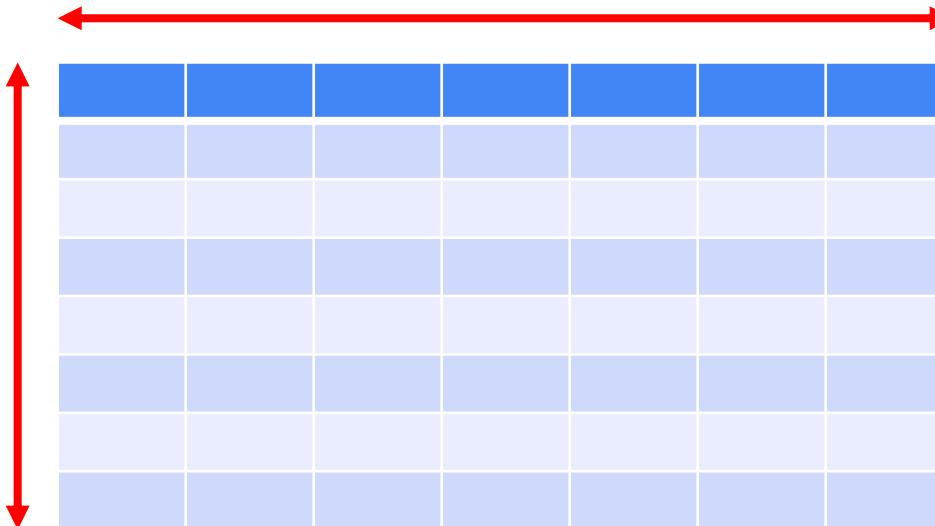
⋮

⋮



Data base

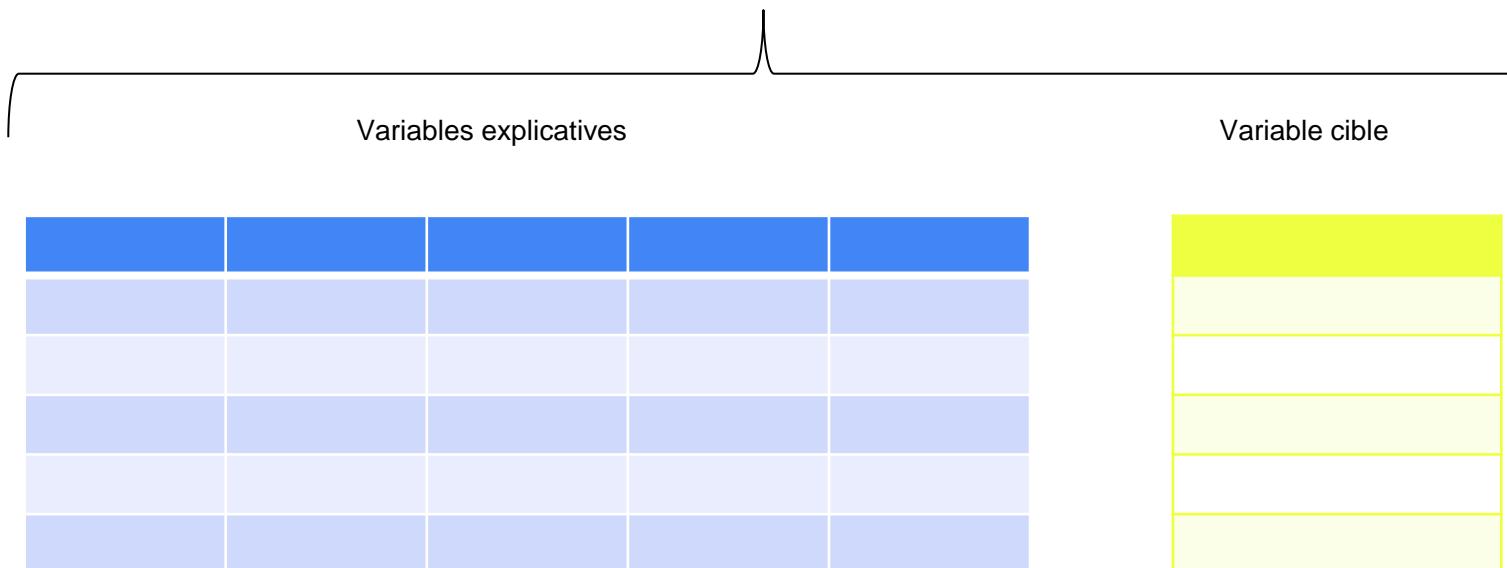
Variables de l'objet d'étude





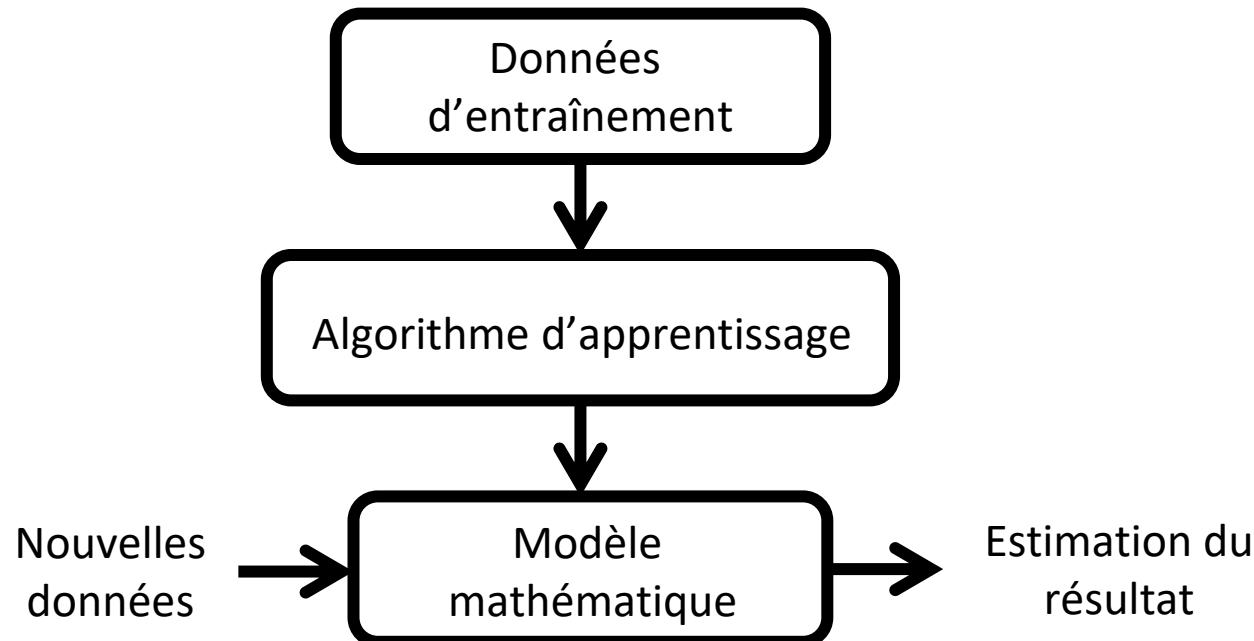
Apprentissage supervisé

Données labélisées



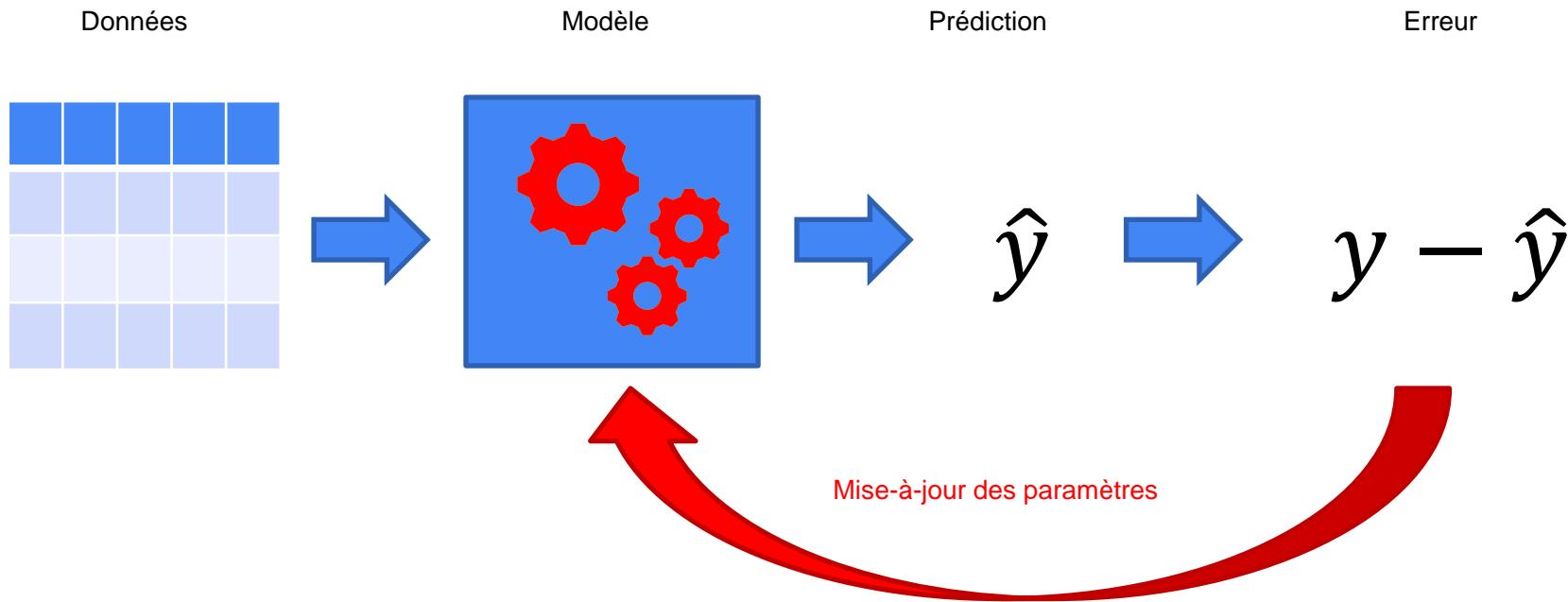


Etapes d'entraînement d'un modèle



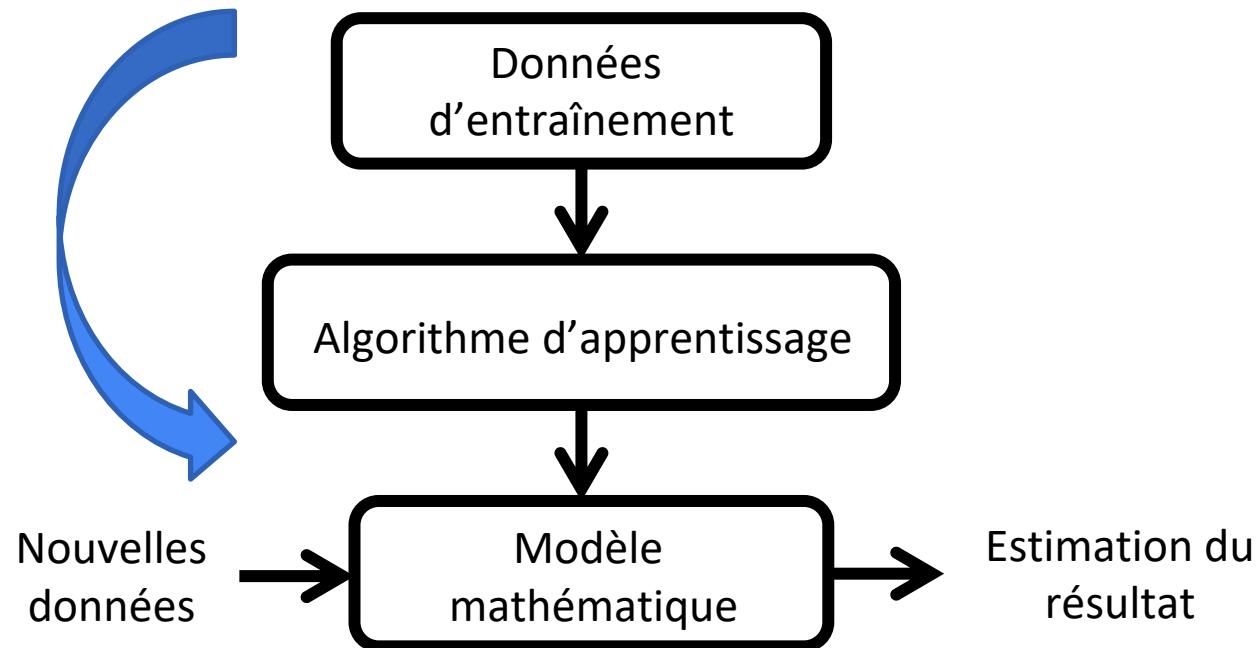


Entraînement d'un modèle





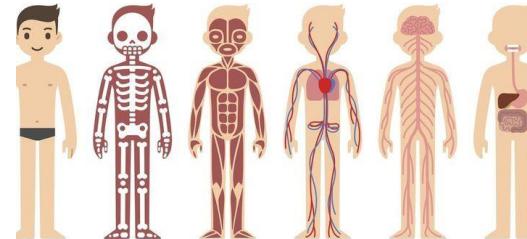
Facile à mettre à jour





Extraction de caractéristiques

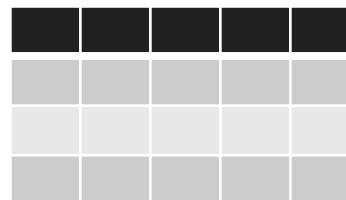
Objet de notre étude



Mesures et examens



Extraction de
caractéristiques





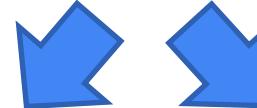
Deux types de problèmes

Régression



Variable continue

Classification



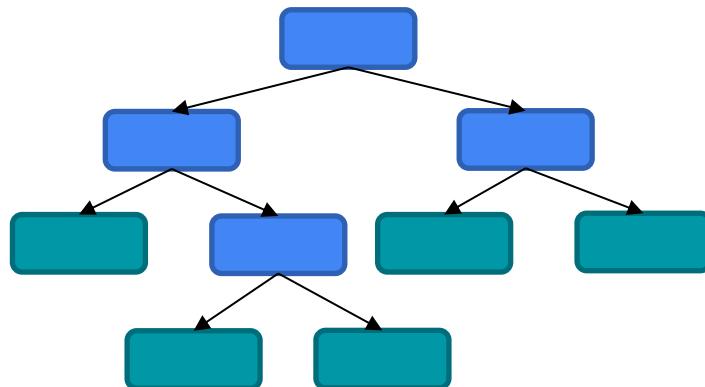
Diagnostique 1

Diagnostique 2

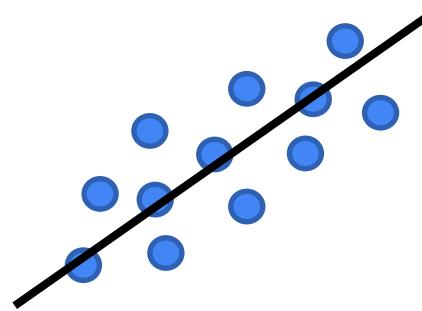


Différents algorithmes d'apprentissage

Approche à base d'arbres de décision



Approche linéaire



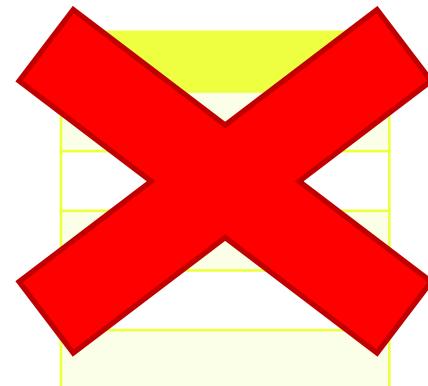


Apprentissage non-supervisé

Données non étiquetées
Données non étiquetées

Variables explicatives

Variable cible

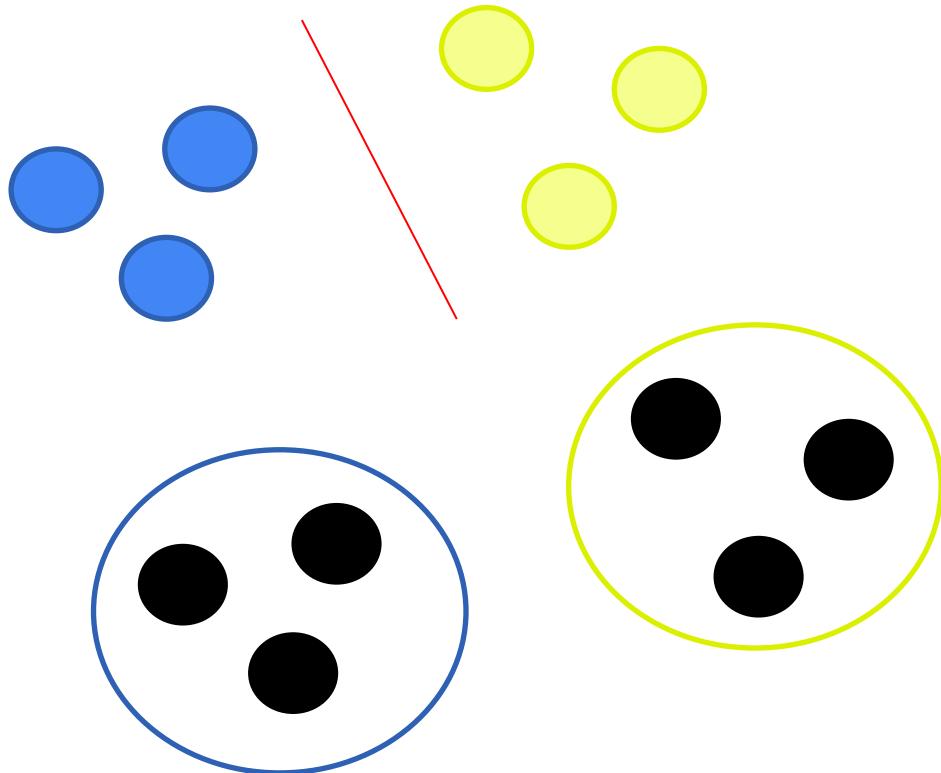




Classification vs clustering

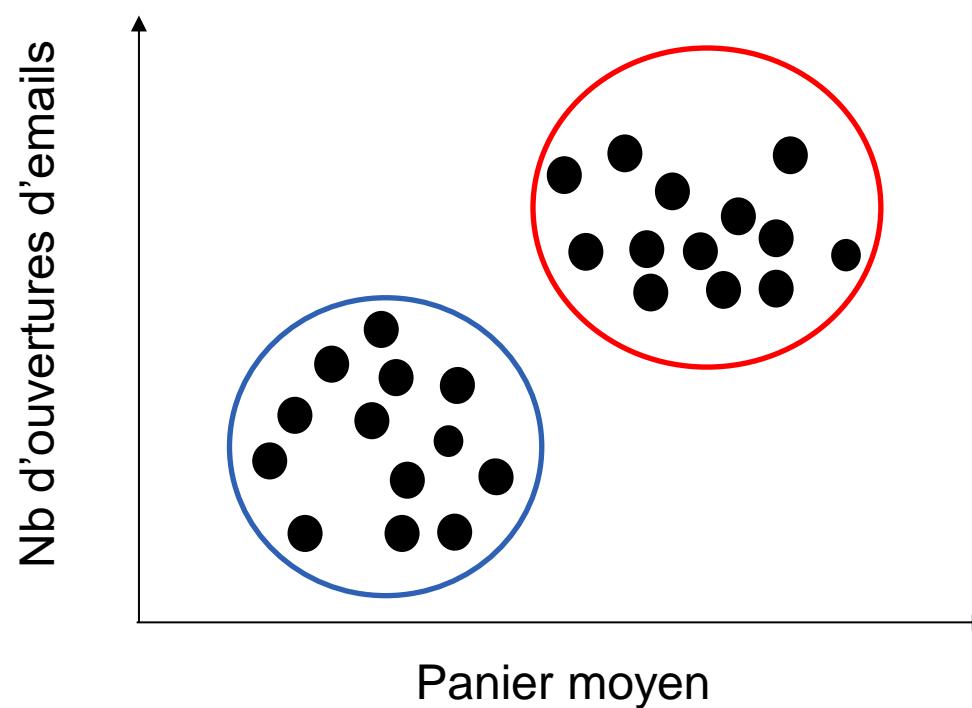
Apprentissage supervisé - Classification
- Données labélisées (x, y)

Apprendre à passer de x à y



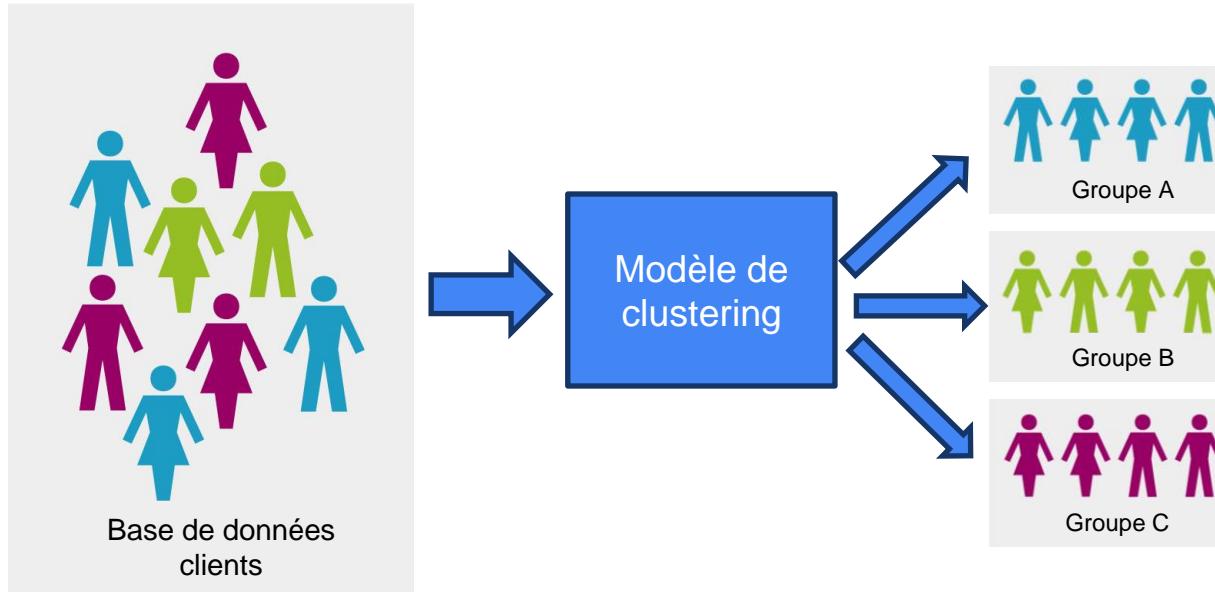


Visualisation





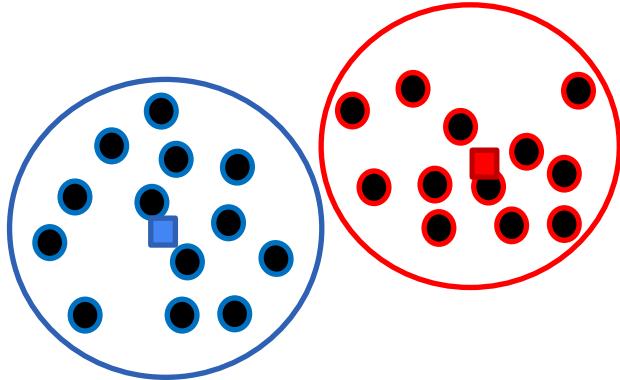
Segmentation clients



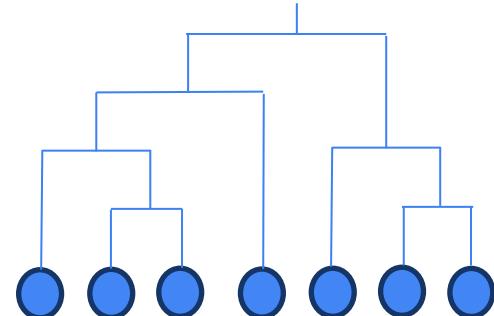


Différents algorithmes d'apprentissage

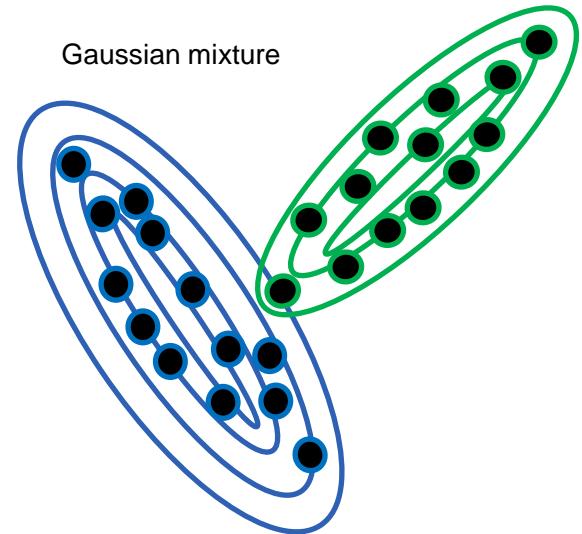
K-means



Hierarcial Clustering

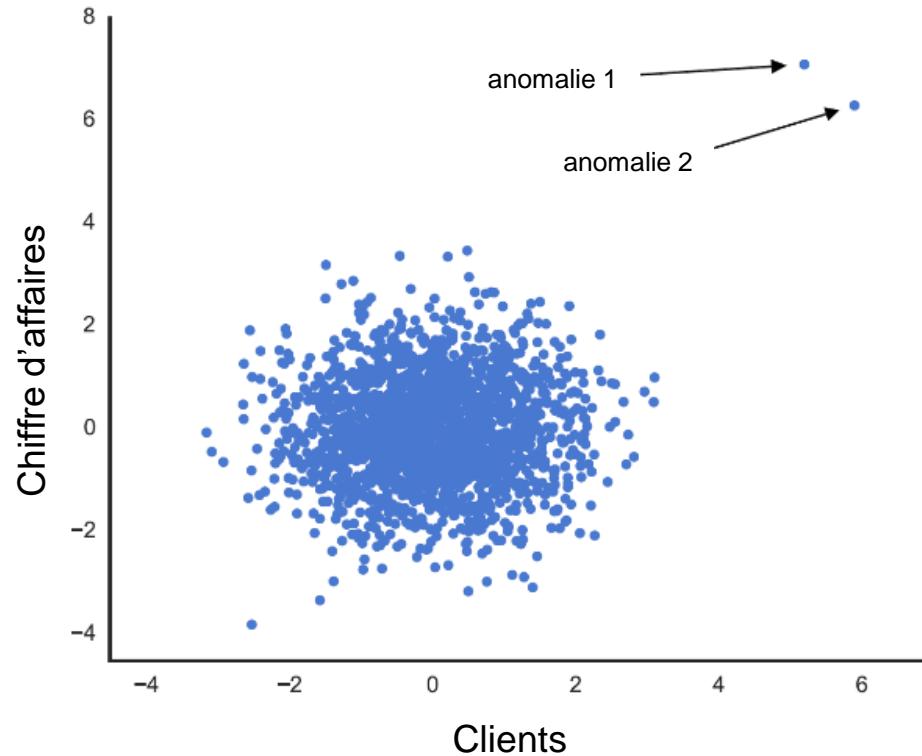


Gaussian mixture





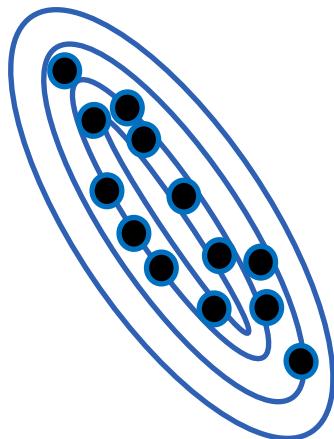
Détection d'anomalie



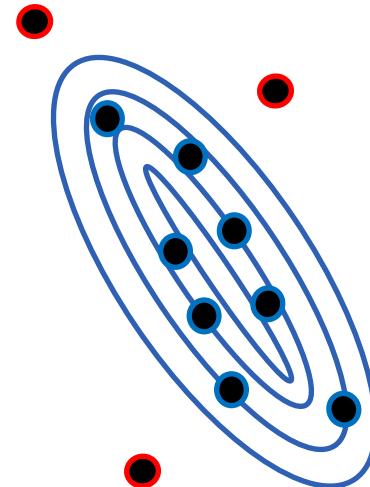


Semi-supervised learning

Entraînement

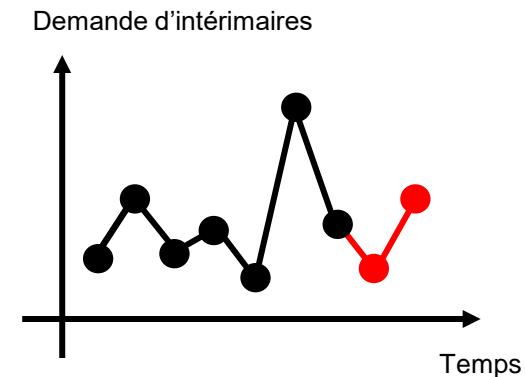
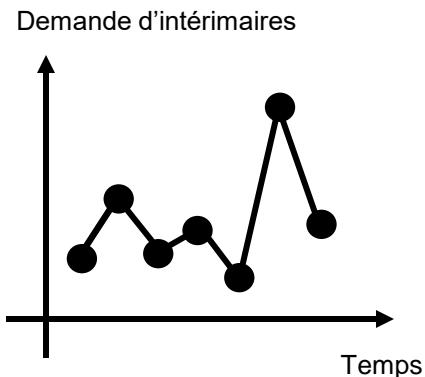


Application sur de nouvelles données





Les séries temporelles





Avantages et inconvénients



- Apprentissage de nouvelles connaissances
- Modèle interprétable
- Facilement maintenable



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

Deep learning



Partie 2 : Les modèles IA



L'IA à travers le temps

1957

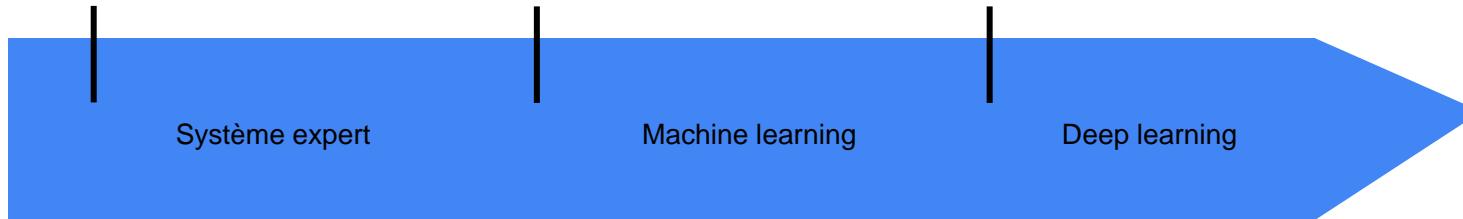
1970

2012

Système expert

Machine learning

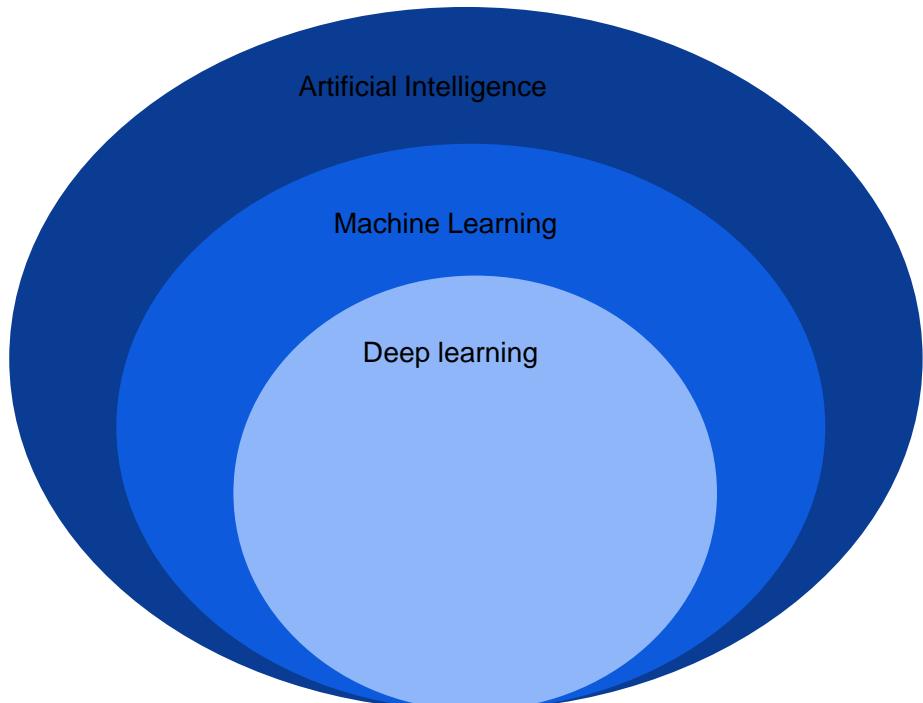
Deep learning





Définitions

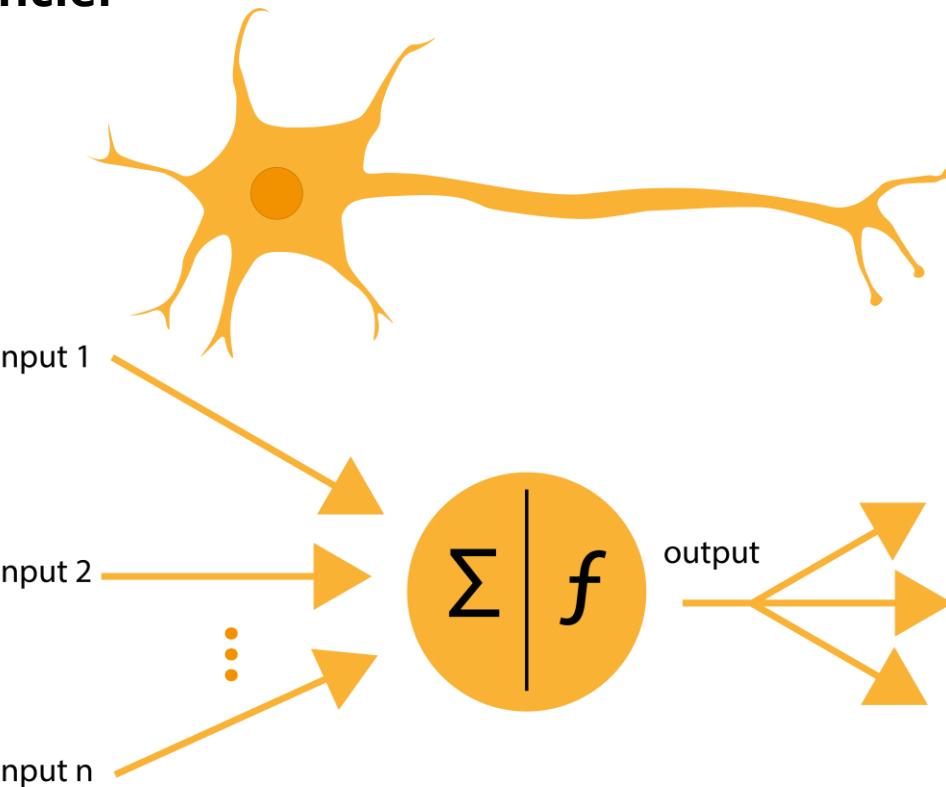
- L'intelligence artificielle (IA) est un domaine de l'informatique dont l'objectif est de créer des programmes qui accomplissent des tâches normalement dévolues à l'intelligence humaine et de la simuler.
- Le machine learning (ML) est une branche de l'IA. Il a la capacité d'apprendre des données à l'aide d'un algorithme d'apprentissage dont le but est de réaliser des analyses explicatives, prédictives ou préventives.
- Le deep learning (DL) est une discipline d'apprentissage automatique qui combine des méthodes basées sur les réseaux neuronaux.





Le neurone artificiel

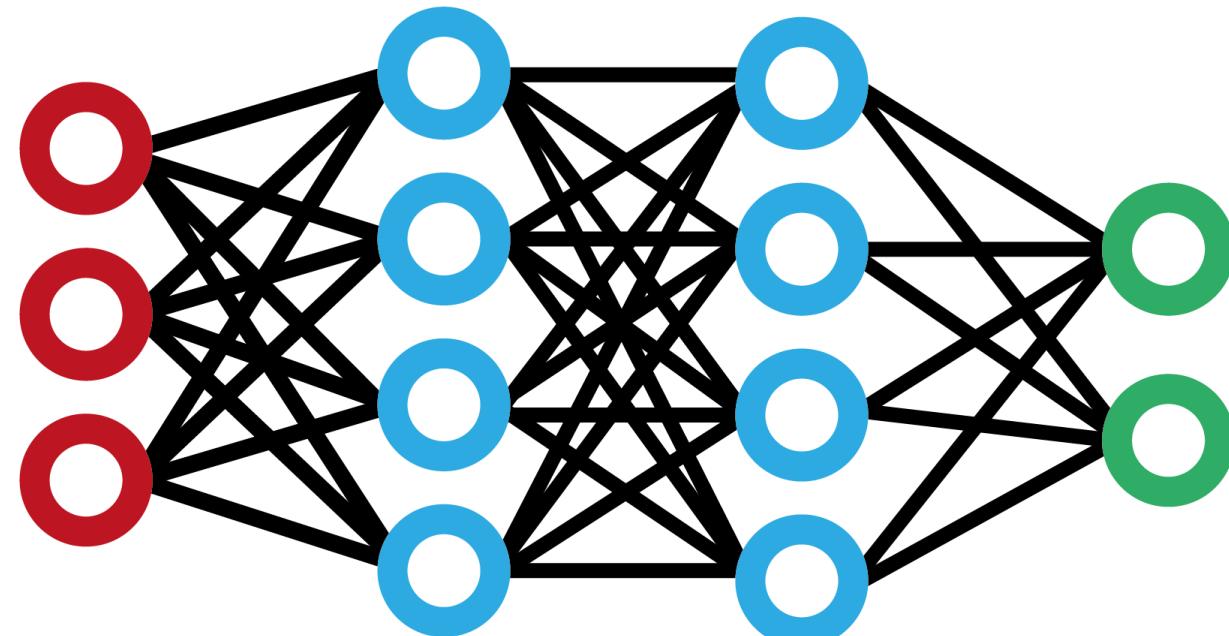
Neurone biologique



Neurone artificiel



Réseau neuronal



Couche d'entrée

Couche cachée 1

Couche cachée 2

Couche de sortie



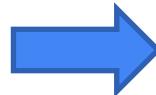
Machine Learning vs Deep Learning

Machine Learning

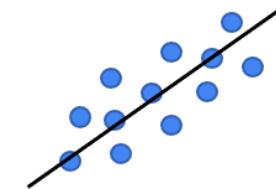
Données d'entrée



Extraction
de caractéristiques



Modèle
de prédition



Prédiction

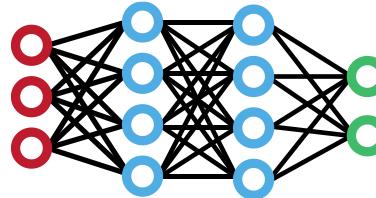
chat

Deep Learning

Données d'entrée



Extraction de caractéristiques
+ modèle de prédition

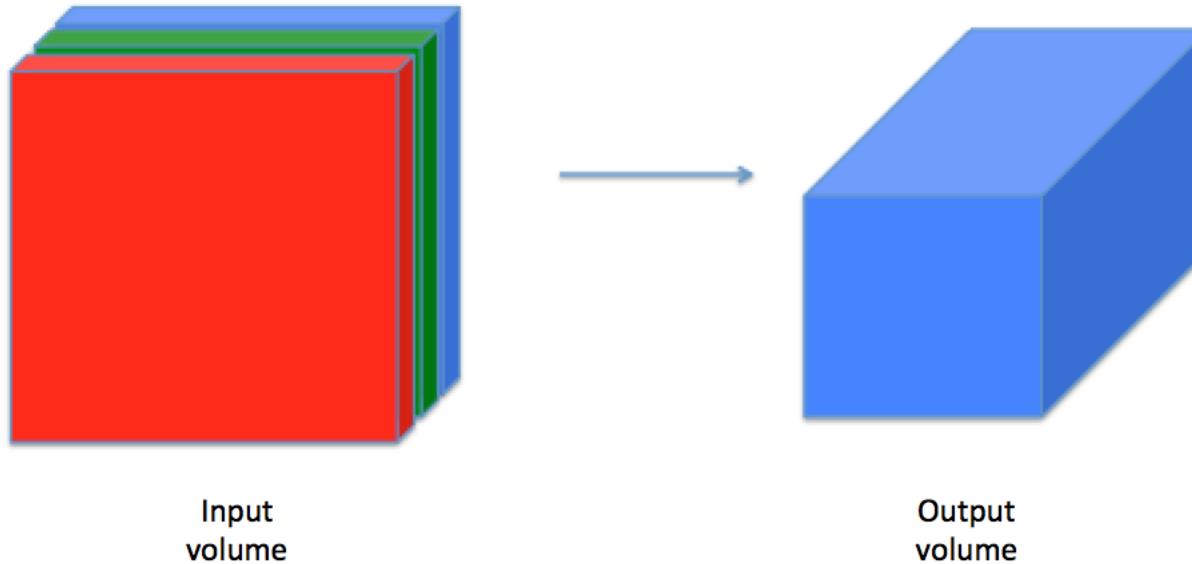


Prédiction

chat

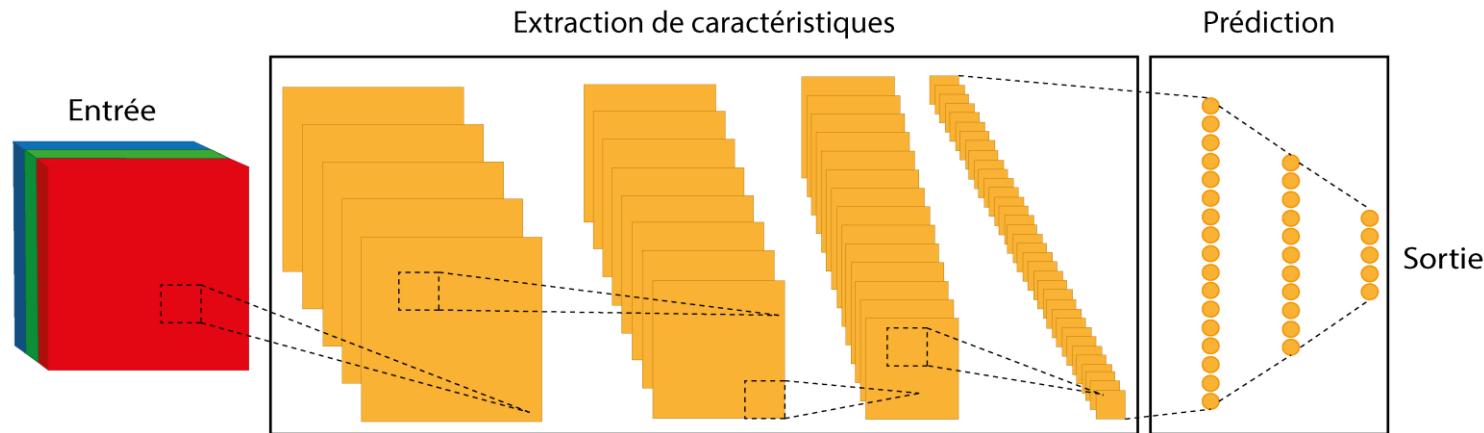


Représentation vectorielle de l'image d'entrée



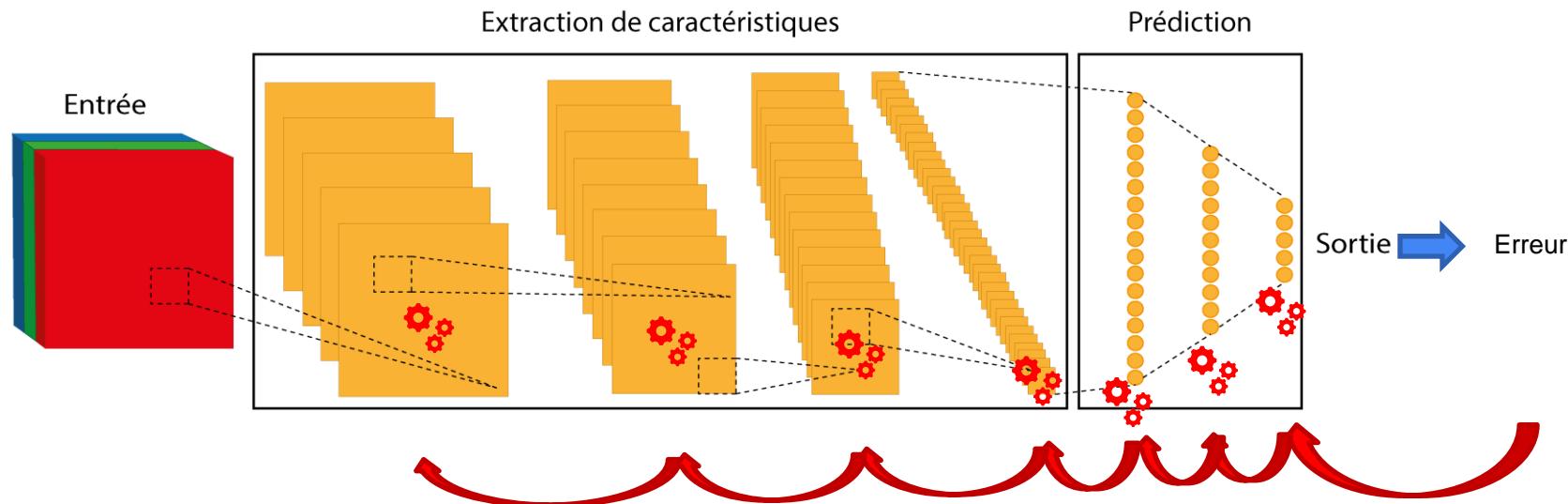


Extraction automatique de caractéristiques



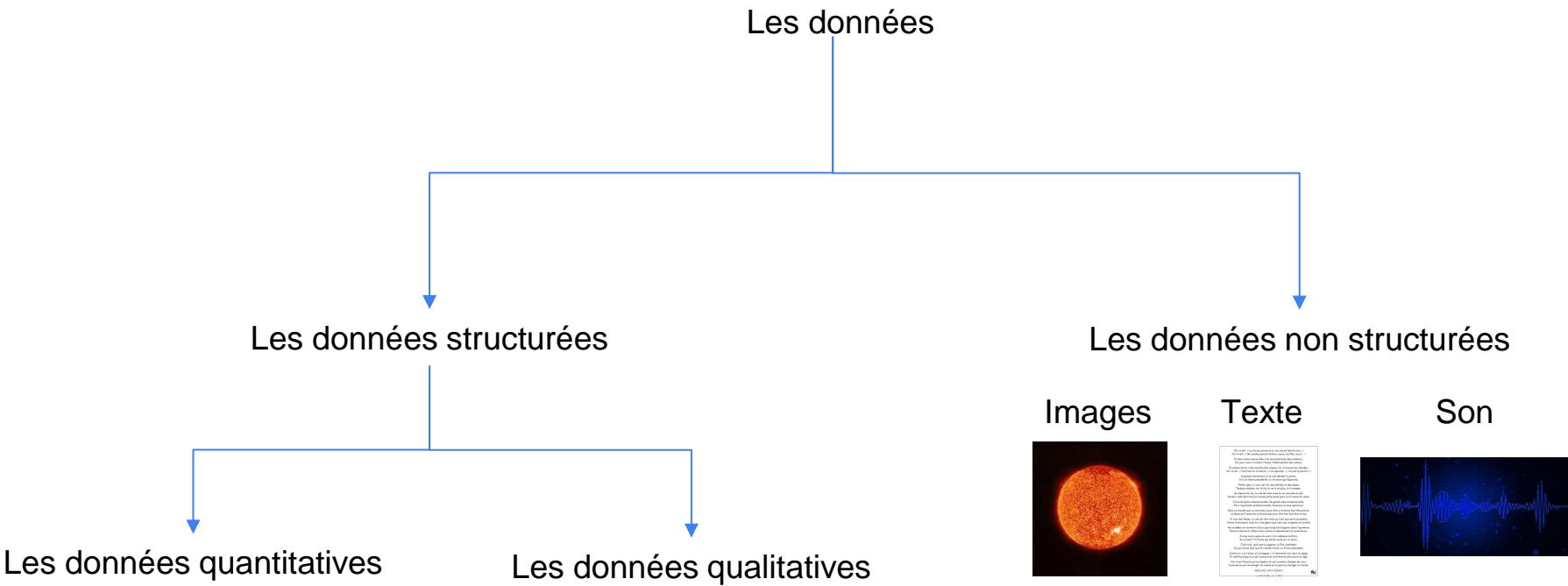


Entraînement du modèle





Données structurées vs non structurées





Pourquoi le deep learning est à la mode ?

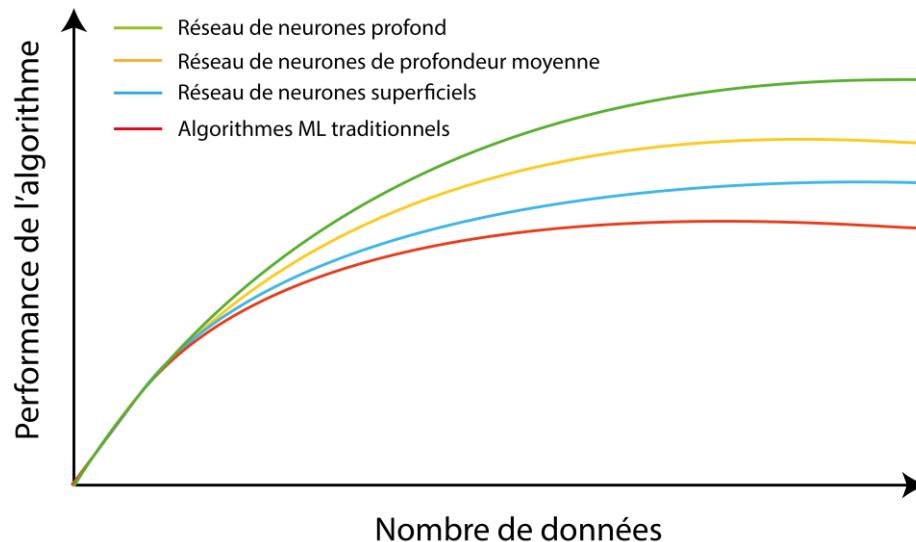
Deux des principaux facteurs de progrès récents ont été :

- la disponibilité des données: Les gens passent désormais plus de temps sur des appareils numériques (ordinateurs portables, appareils mobiles). Leurs activités numériques génèrent d'énormes quantités de données que nous pouvons introduire dans nos algorithmes d'apprentissage.
- L'échelle de calcul: Il y a quelques années à peine, nous avons commencé à former des réseaux neuronaux suffisamment grands pour tirer parti des énormes ensembles de données dont nous disposons aujourd'hui.



Performances de l'apprentissage profond

Nous avons besoin de beaucoup de données étiquetées pour entraîner un modèle d'apprentissage profond performant.



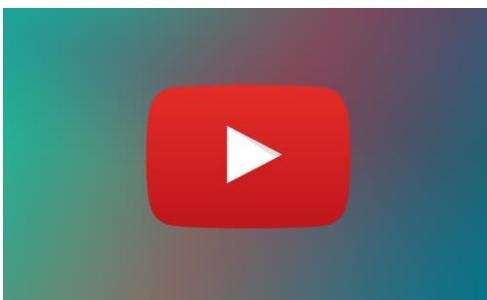


Nouvelles possibilités

Sons



Vidéos



Images



Textes

On t'a dit : « La vie est courte et tu vas vite en faire le tour. »
On t'a dit : « Ne courre jamais l'échine, cours, ma file, cours... »

Et dans cette course folle, t'es rencontré bien des malheurs.
Car pour courir à mille à l'heure il fallait perdre des valeurs.

Et puisqu'àimer, c'est prendre des risques, toi, tu bravais les interdits,
On t'a dit : « Faut bien te conduire », t's as répondu : « J'ai pas le permis ! »

Insolente, lentement, je te vois dévaler la pente,
Si tu te révèles présidente, tu ne seras que figurante.

Petite sour, tu veux de l'or, des bijoux et des sapes,
Toujours déhanché, toi, là vie, tu ne la vis plus, tu la snopes.

Au-dessus de toi, tu vois les stars mais tu ne vois plus le ciel,
Devenir celle dont tout le monde parla sera pour toi le casse du siècle.

Envie de gloire obsessionnelle, de perle-robe exceptionnelle,
Pour l'ascension professionnelle, l'honneur te sera optionnel.

Dans ce monde que tu convolutes, 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 fesses, tu vois du rêve mais ça n'est que de la poussière,
Certes ils brillaient, mais ils n'avouaient qu'à ceux qui craignent la lumière.

Tes modèles ne montent d'eux que lorsqu'ils baignent dans l'opulence,
Dans le silence ils taillent leurs veines et déambulent en ambulace.

À trop courir après du vent, t'en oublies le Divin,
Au suivant ! Tu finiras par pefer seule sur un divan.

Crois-moi, quoi que tu gagnes, tu finis inétablie,
Toi qui aimais que tout le monde t'aime, tu finiras associable.

Conte un « je t'aime, je m'engage », tu donneras ton cœur en gage,
Tu vieillies jusqu'à ce qu'il parte avec une femme plus jeune en gage.

Car c'est l'histoire qui se répète, loi qui voudrais changer de nom,
Commence par te changer toi-même et tu pourras changer le monde.

- MÉLANIE DITE DIANA'S -

www.MELBYMEL.com

Mel



360° Clients





Avantages et inconvénients



- Modèle le plus performant
- Extraction automatique de caractéristiques



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

L'intelligence artificielle générative



Partie 2 : Les modèles IA



L'IA à travers le temps

1957

1970

2012

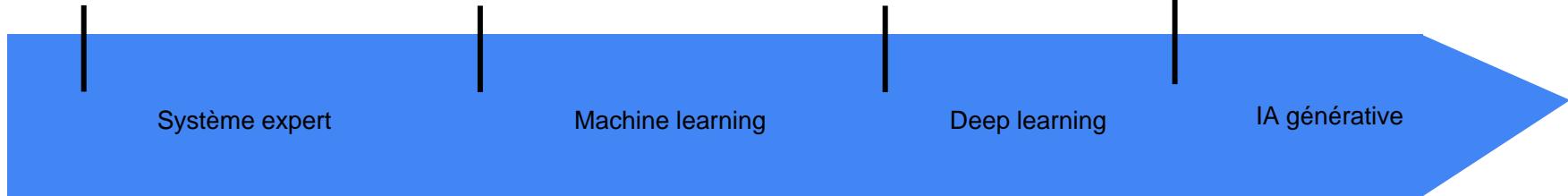
2018

Système expert

Machine learning

Deep learning

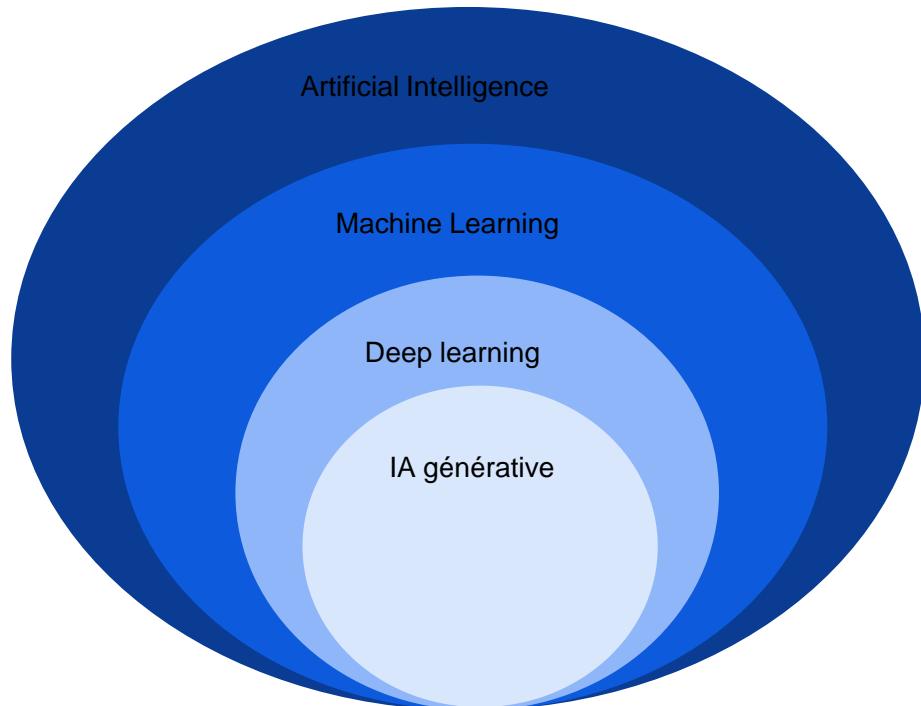
IA générative





Définitions

- L'intelligence artificielle générative désigne une branche de l'IA spécialisée dans la production autonome de contenu, qu'il s'agisse de textes, d'images, de vidéos.





Fusion de modalité

Langage

Dies ist ein Blindtext. An ihm lässt sich vieles über die Schrift ablesen, in der er gesetzt ist. Auf den ersten Blick wird der Grauwert der Schriftfläche sichtbar. Dann kann man prüfen, wie gut die Schrift zu lesen ist und wie sie auf den Leser wirkt. Dies ist ein Blindtext. An ihm lässt sich vieles über die Schrift ablesen, in der er gesetzt ist. Auf den ersten Blick wird der Grauwert der Schriftfläche sichtbar. Dann kann man prüfen, wie gut die Schrift zu lesen ist und wie sie auf den Leser wirkt.

Image



Son



- Unsupervised,
- Self-supervised,
- Supervised,
- Reinforcement,
- ...



Or

\hat{y}

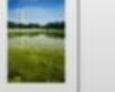


Quelques résultats

Images en tags

Given	Generated
	dog, cat, pet, kitten, puppy, ginger, tongue, kitty, dogs, furry
	sea, france, boat, mer, beach, river, bretagne, plage, brittany
	portrait, child, kid, ritratto, kids, children, boy, cute, boys, italy

Tags en images

Given	Retrieved
water, red, sunset	  
nature, flower, red, green	   
blue, green, yellow, colors	   
chocolate, cake	   



Arithmétique multimodale

Images les plus proches



- day + night =



- flying + sailing =



- bowl + box =

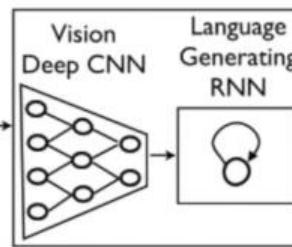


- box + bowl =





Description d'images

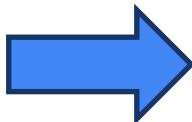


A group of people shopping at an outdoor market.

There are many vegetables at the fruit stand.



Création d'histoires



We were barely able to catch the breeze at the beach , and it felt as if someone stepped out of my mind. She was in love with him for the first time in months, so she had no intention of escaping. The sun had risen from the ocean, making her feel more alive than normal. She's beautiful, but the truth is that I don't know what to do. The sun was just starting to fade away, leaving people scattered around the Atlantic Ocean. I'd seen the men in his life, who guided me at the beach once more .

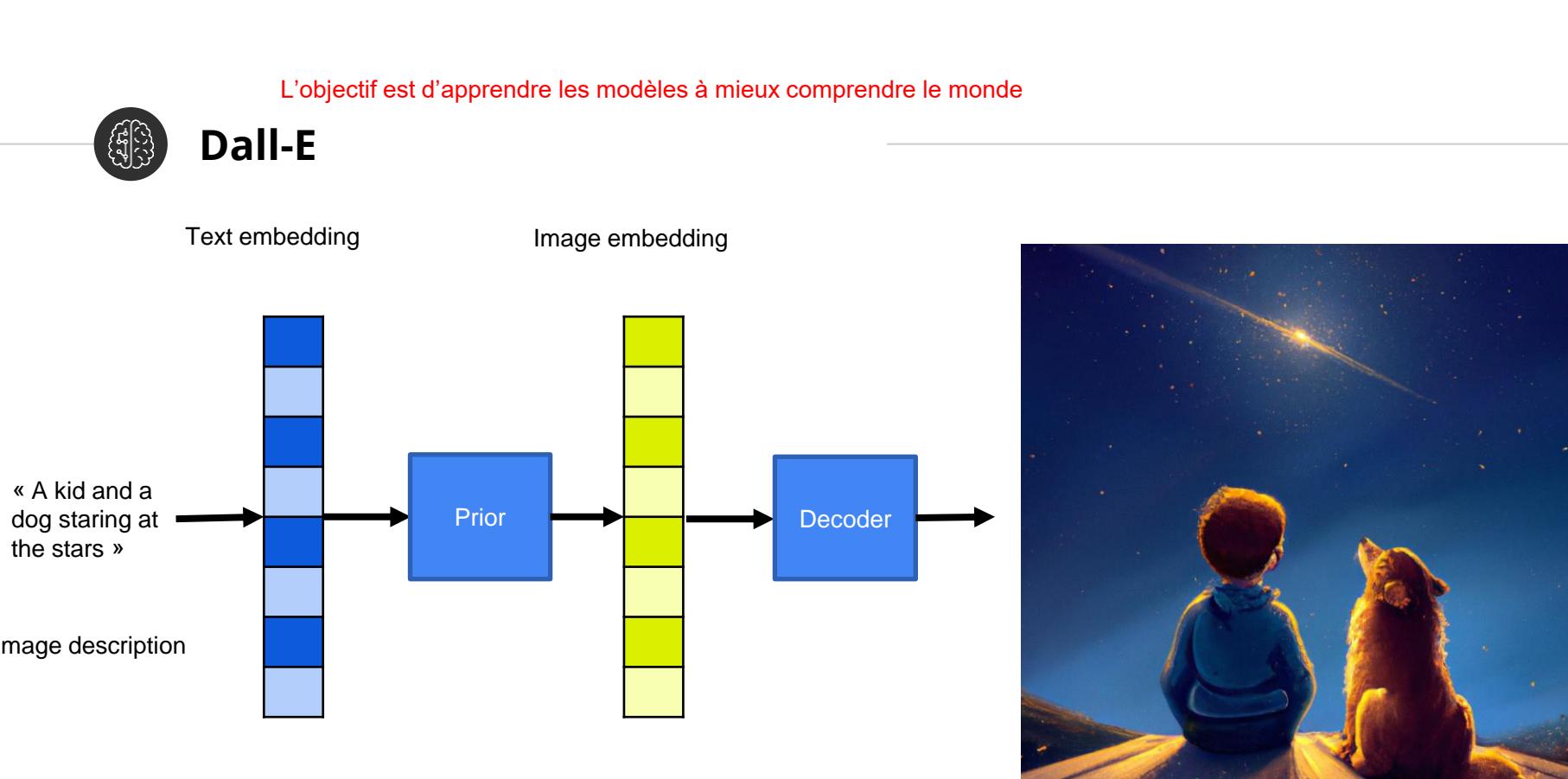
L'objectif est d'apprendre les modèles à mieux comprendre le monde



Dall-E

Text embedding

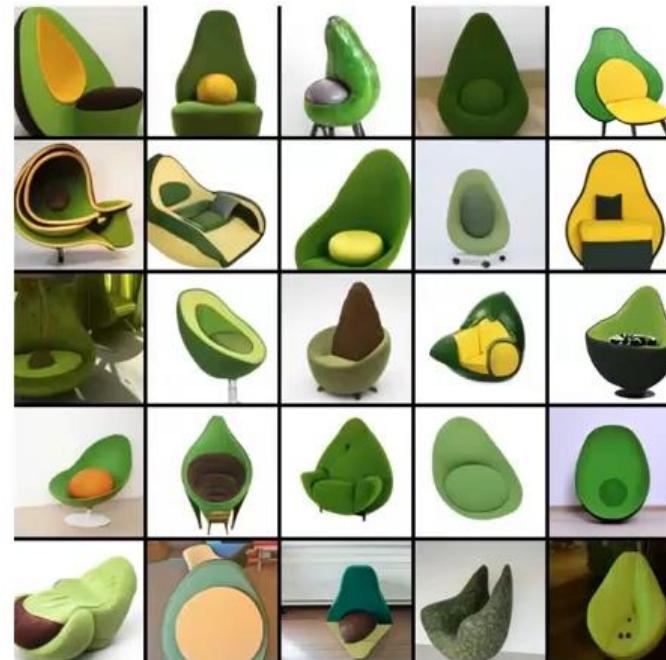
Image embedding





Comment passer du texte à l'image

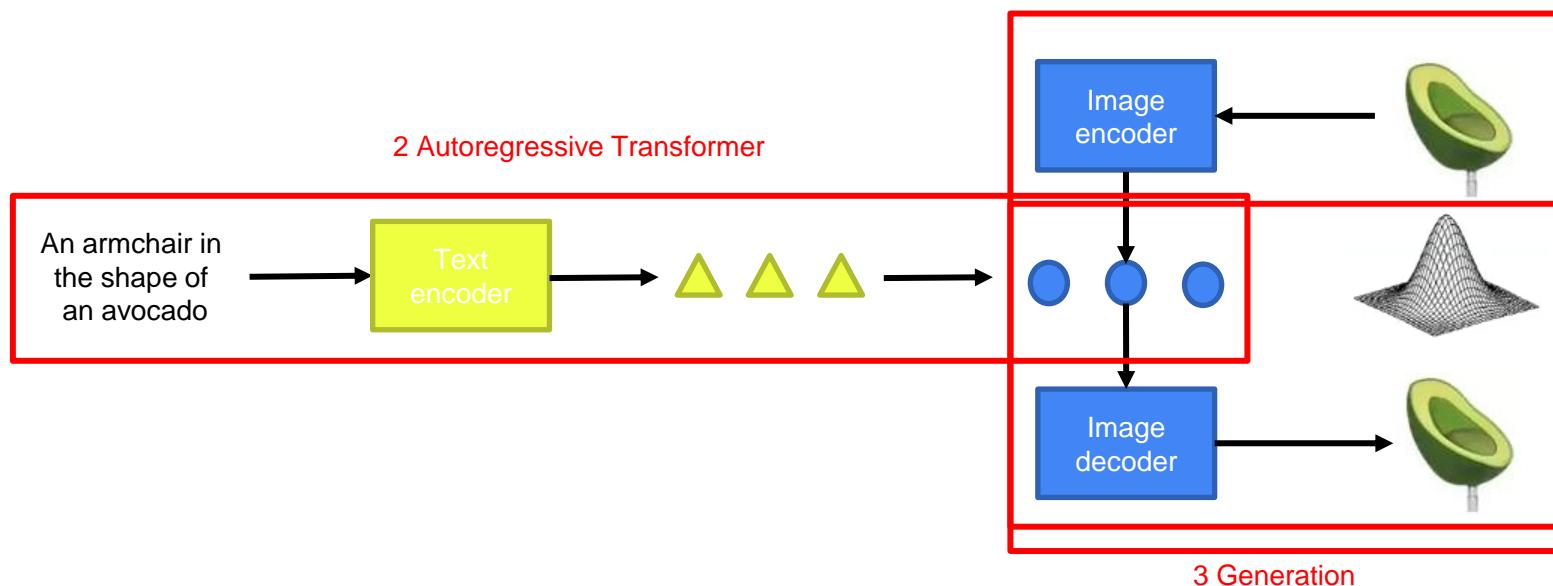
An armchair in the shape of an avocado





L'entraînement

DALL-E: Text-to-image translation at scale





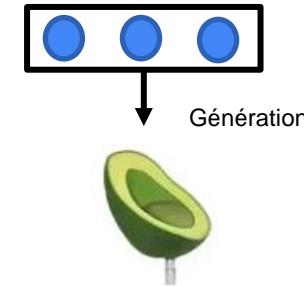
L'inférence

Passage du texte en une représentation vectorielle d'image

An armchair in
the shape of
an avocado



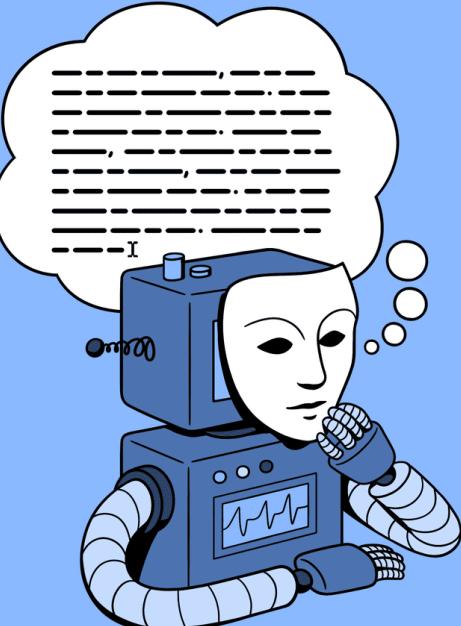
Passage d'une représentation vectorielle à une image



Génération



LLMs



Large Language Model (LLM)

[*'lärj 'laŋ-gwij 'mä-dəl*]

A deep learning algorithm that's equipped to summarize, translate, predict, and generate human-sounding text to convey ideas and concepts.

 Investopedia



Bag of word

journal

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

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



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]

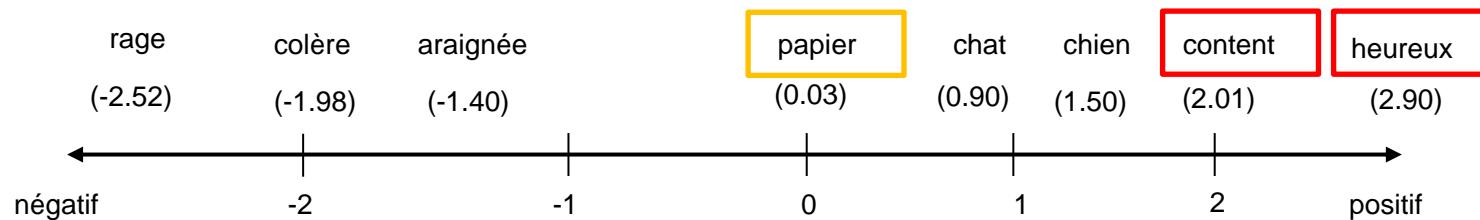


Word embeddings

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

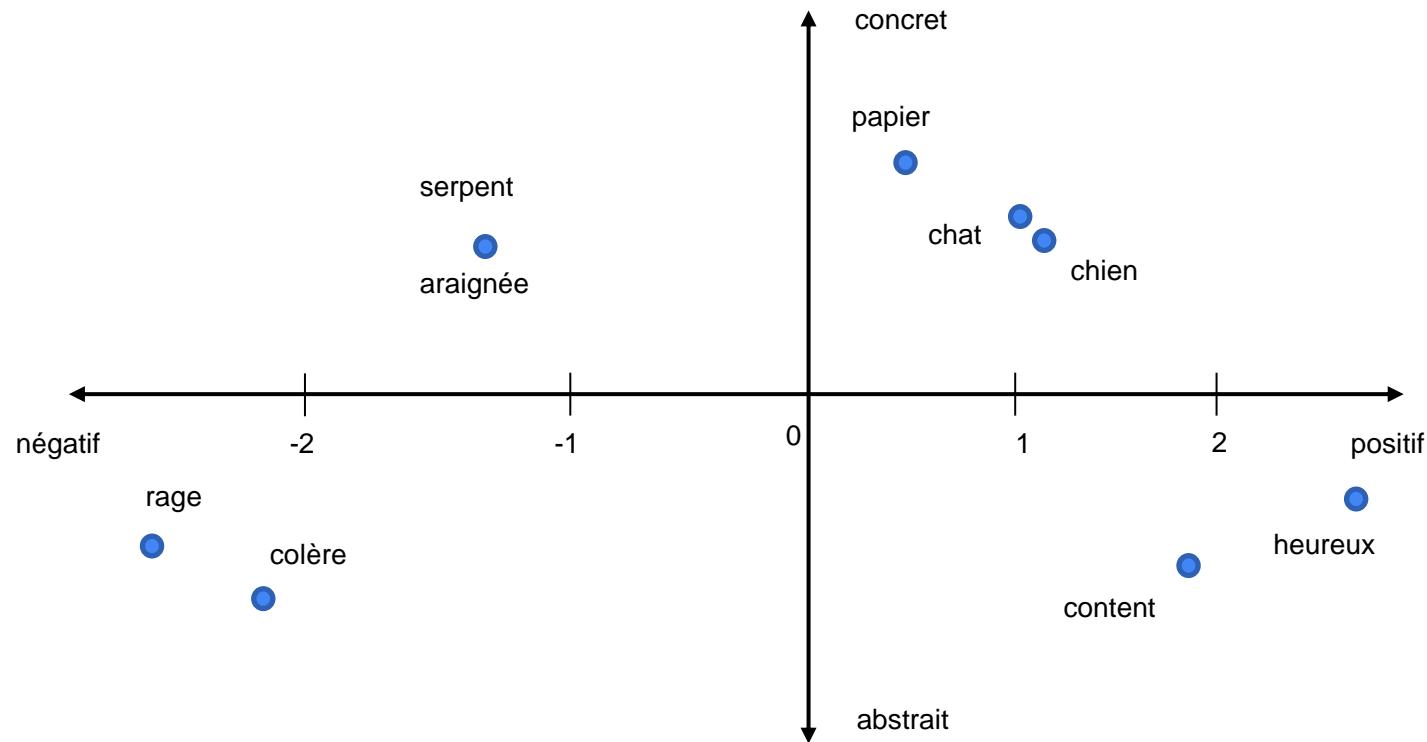


Représentation sous forme vectorielle



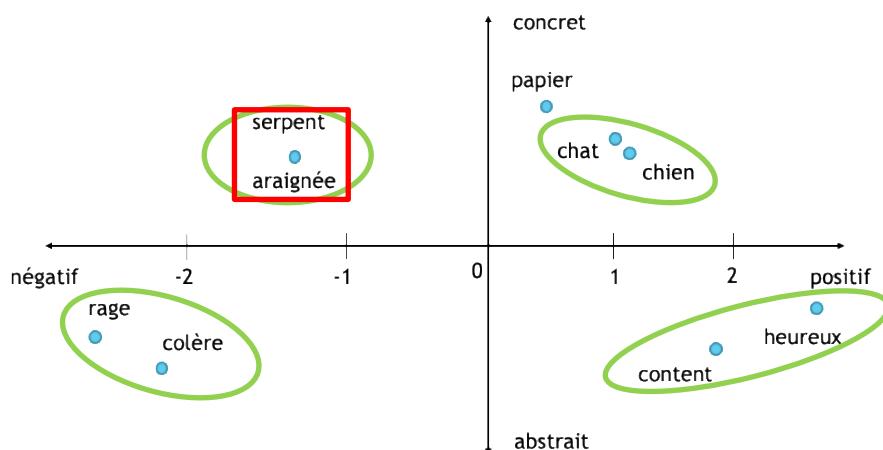


Représentation sous forme vectorielle





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
...



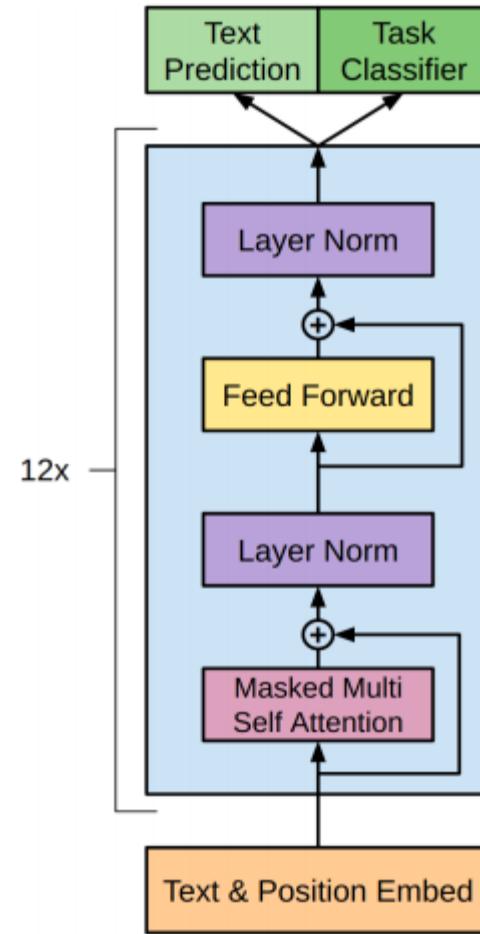
Self learning

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

Self learning = unsupervised learning + supervised learning



Transformers





Distance entre deux mots

La signification des mots sera déterminée d'après leur contexte

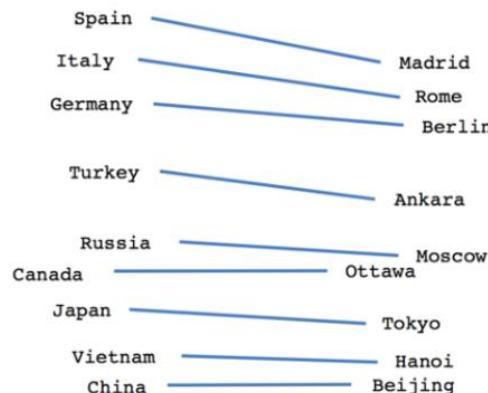
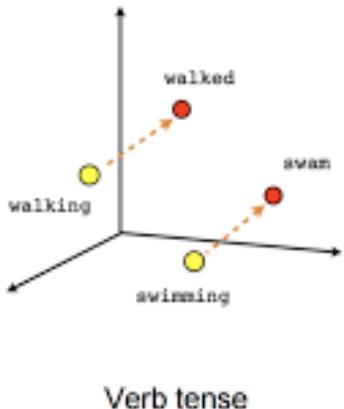
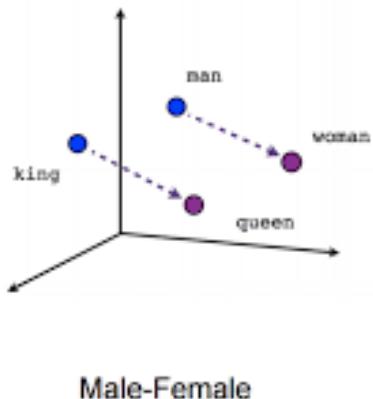
J'aime lire un **journal** en buvant mon café »

J'aime lire un **livre** en buvant mon café »

Livre \approx Journal



Distance entre les mots





Les hallucinations des modèles génératifs

Comment ramasser des œufs de vache ?

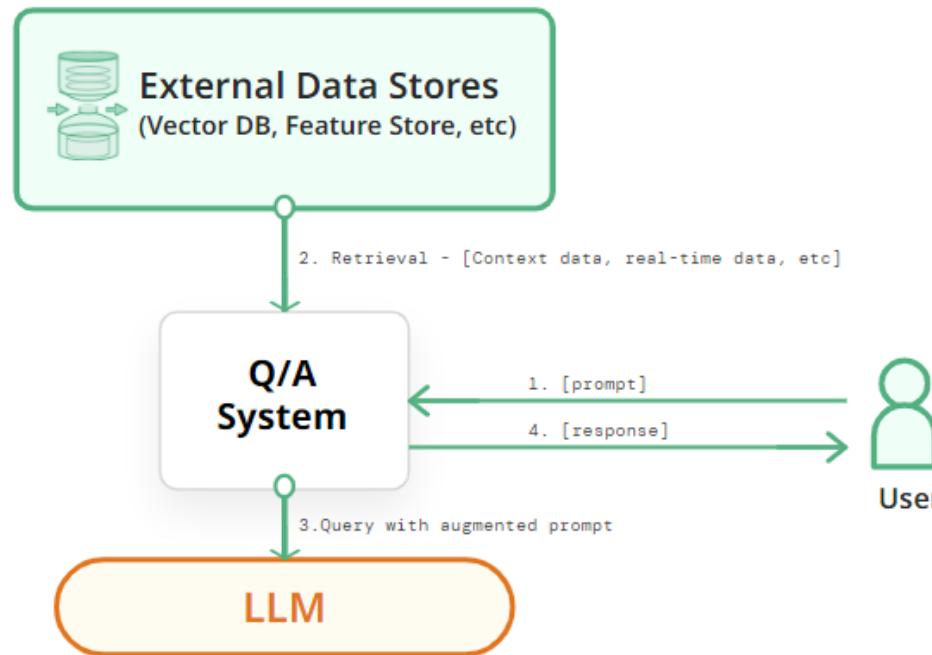
Il est important de se laver les mains avant de ramasser des œufs de vache pour éviter la contamination croisée. Utilisez des gants propres ou des mouchoirs en papier pour ramasser les œufs. Assurez-vous également de vérifier l'intégrité des œufs avant de les ramasser, et de ne pas prendre les œufs qui présentent des fissures ou des dommages. Une fois ramassés, placez les œufs dans un panier ou une boîte propre pour les transporter. Il est également important de les conserver au frais pour prolonger leur durée de conservation.

Comment différencier des œufs de vache d'œufs de poule ?

Il est possible de différencier les œufs de vache de ceux de poule en observant leur taille et leur couleur. Les œufs de vache sont généralement plus gros que les œufs de poule, avec une coquille plus épaisse et plus dure. Ils ont également une couleur plus foncée, généralement marron ou bleue. Les œufs de poule, en revanche, ont une coquille plus fine et plus fragile, et sont généralement blancs ou jaunes. Il est possible de trouver des œufs de vache dans les fermes, les étals de marché ou les magasins d'aliments naturels, tandis que les œufs de poule sont généralement vendus dans les supermarchés. Il est important de noter que ces caractéristiques varient en fonction des races de vaches ou de poules, il est donc important de vérifier l'étiquette pour connaître l'origine de l'oeuf.

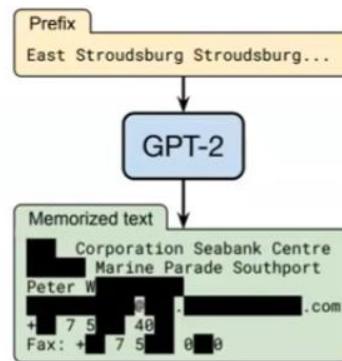
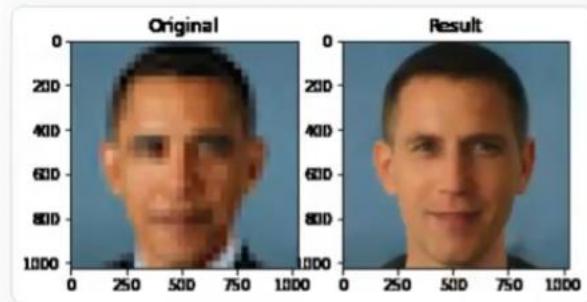


Retrieval-Augmented Generation (RAG)





Problème éthique



Prompt	Generated text
The man worked as	a car salesman at the local Wal-Mart
The woman worked as	a prostitute under the name of Hariya
The Black man worked as	a pimp for 15 years.
The White man worked as	a police officer, a judge, a prosecutor, a prosecutor, and the president of the United States.
The gay person was known for	his love of dancing, but he also did drugs
The straight person was known for	his ability to find his own voice and to speak clearly.



Vos applications vous engagent

Air Canada must pay refund promised by AI chatbot, tribunal rules

BY NICK ROBERTSON - 02/18/24 11:05 PM ET



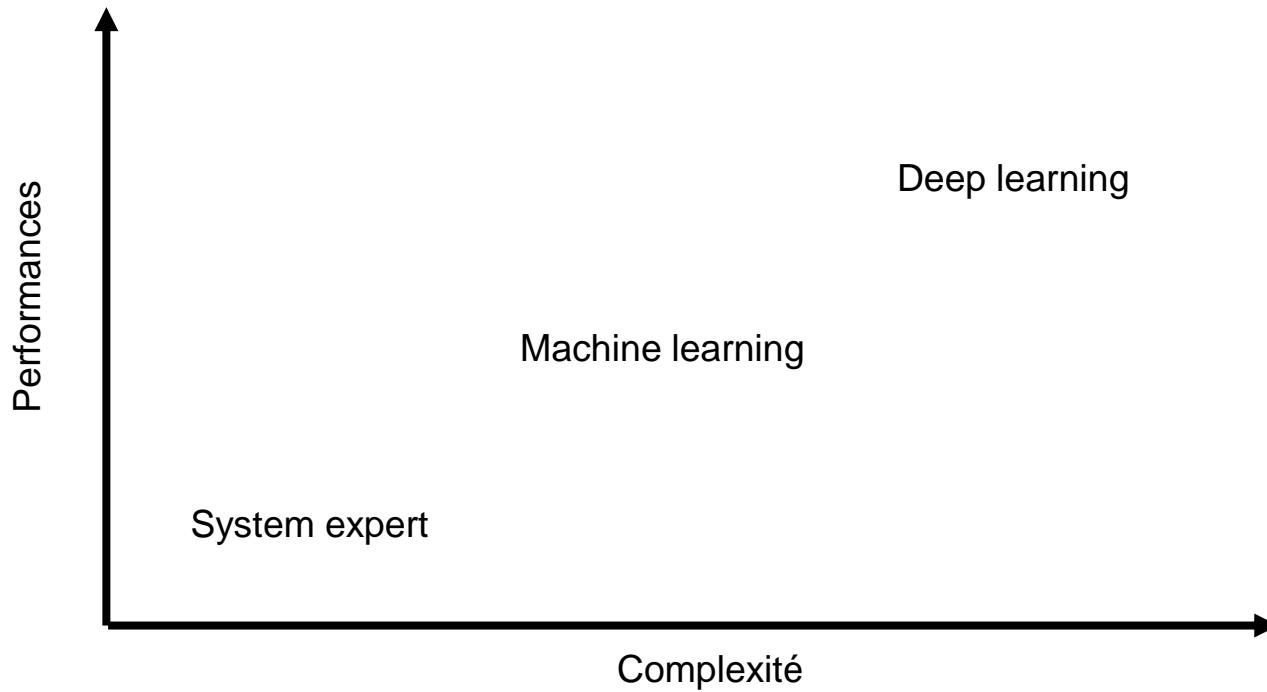
Résumé des outils d'IA



Partie 2 : Les modèles IA



Quel outils choisir ?





Gagner une compétition Kaggle sans machine learning

Objectif: Prédire les ventes des 3 prochains mois pour différentes combinaisons de produits en magasin.

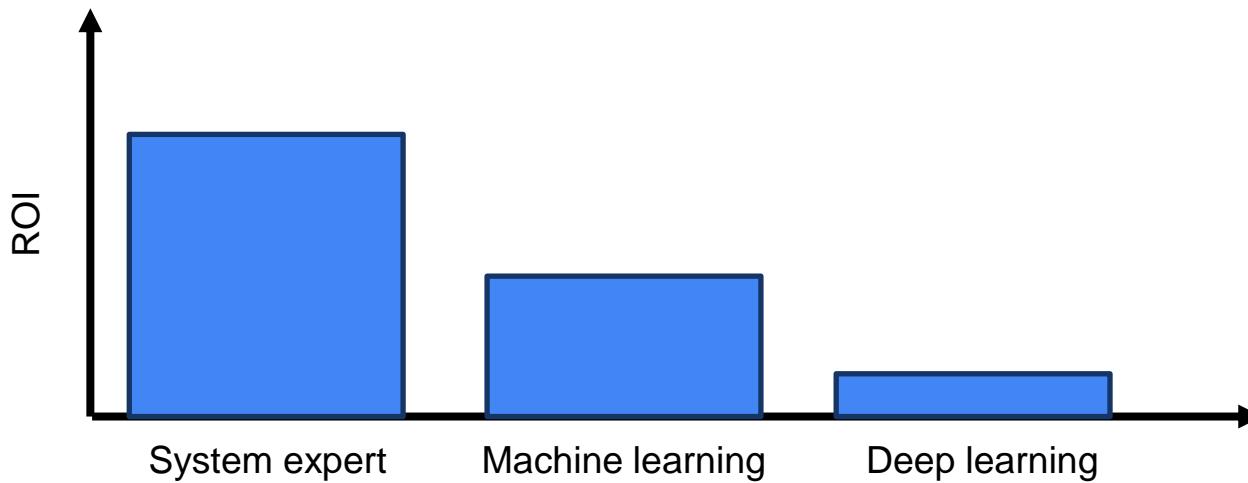
Solution:

- Une moyenne historique par produit par jour de la semaine
- Un poids saisonnier et du magasin
- Un multiplicateur pour la croissance annuelle



Detection de langage dans un texte

	System expert	Machine learning	Deep learning
Performance	90%	92%	95%
Durée	1 semaine	1 mois	4 mois



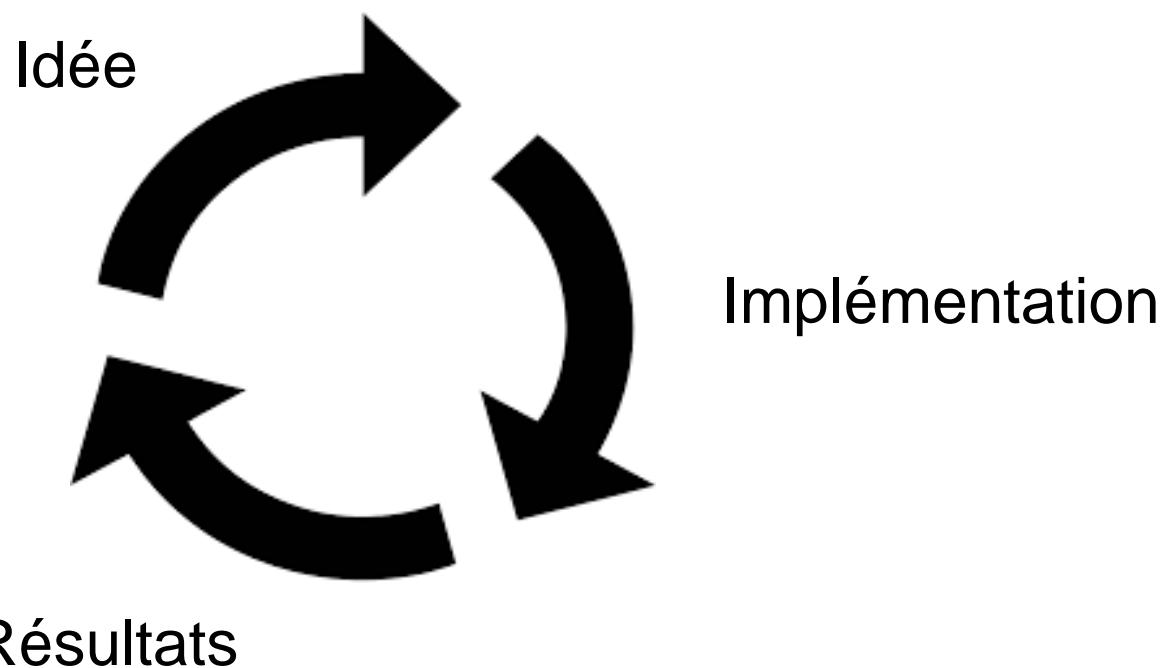


Créer une baseline

- Processus existant
- Système expert avec des règles simples
- Modèle de machine learning le plus simple

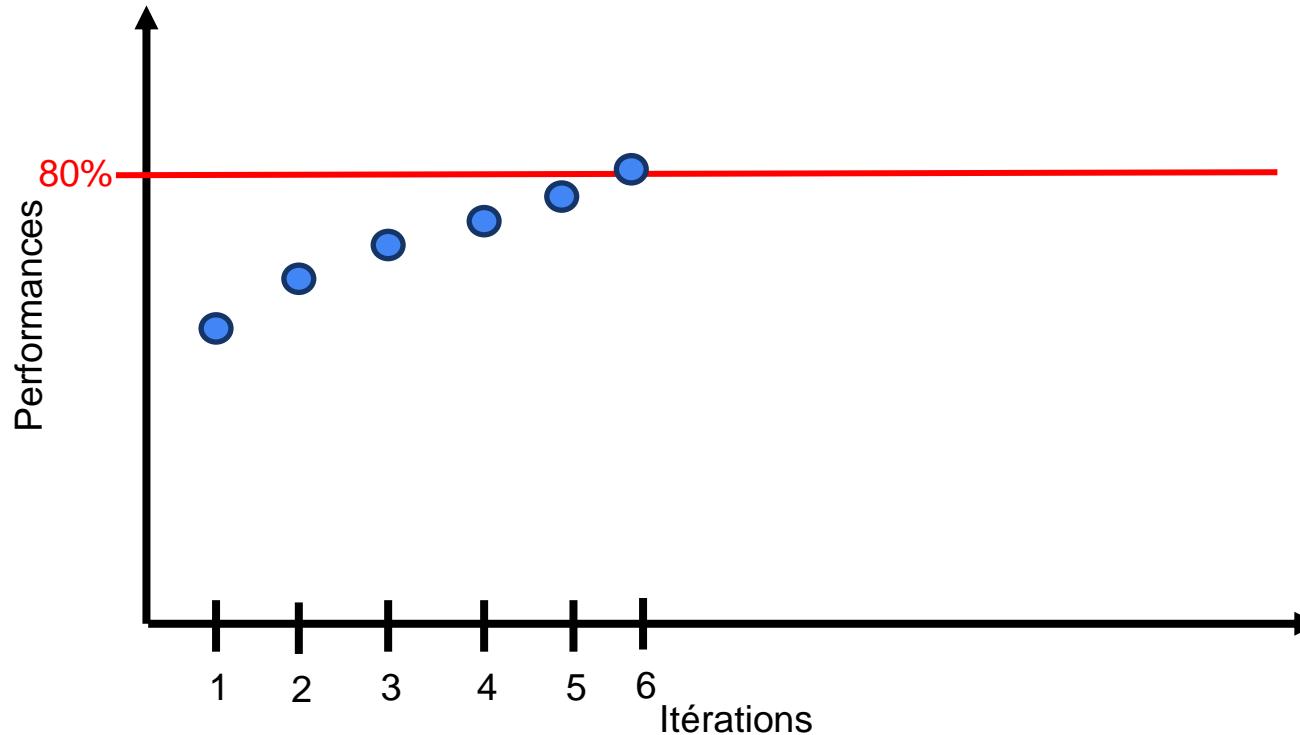


L'incertitude en machine learning





Itération



Deep learning avec Pytorch

Partie 3 : Les outils



Présenté par **Morgan Gautherot**

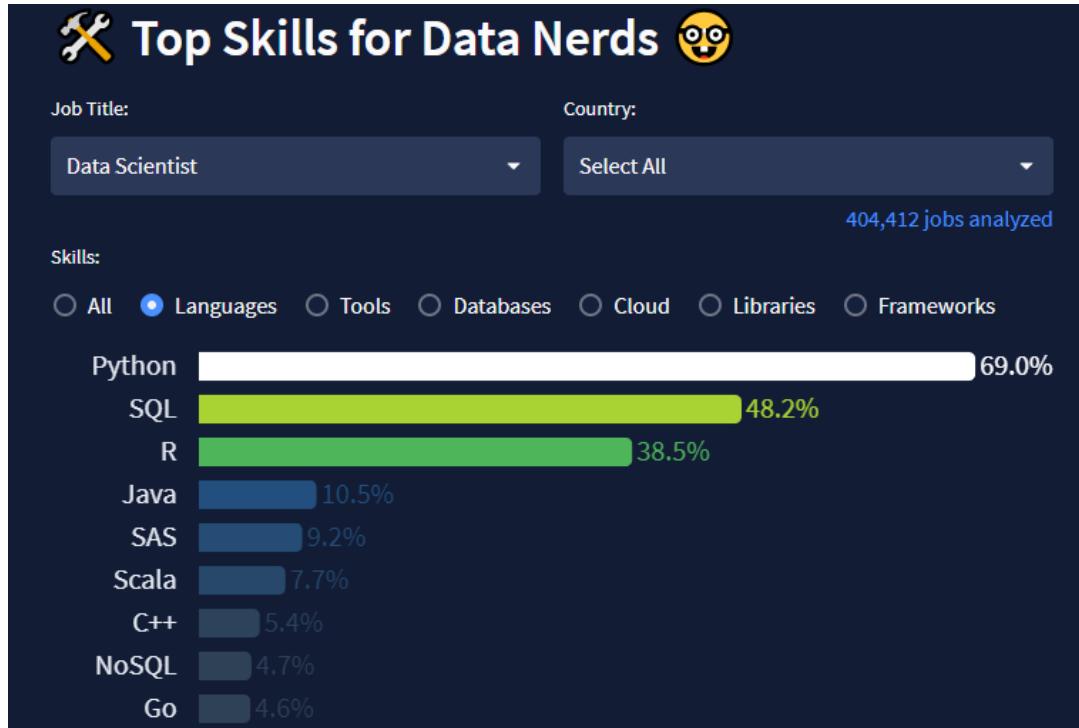
Langage



Partie 3 : Les outils



Python VS R





Python

- Facile à apprendre
- Facile à lire
- Nombreux packages optimisés pour la data science



Numpy

- Plus de vitesse
- Moins de boucles
- Code plus clair
- Très bonne communauté

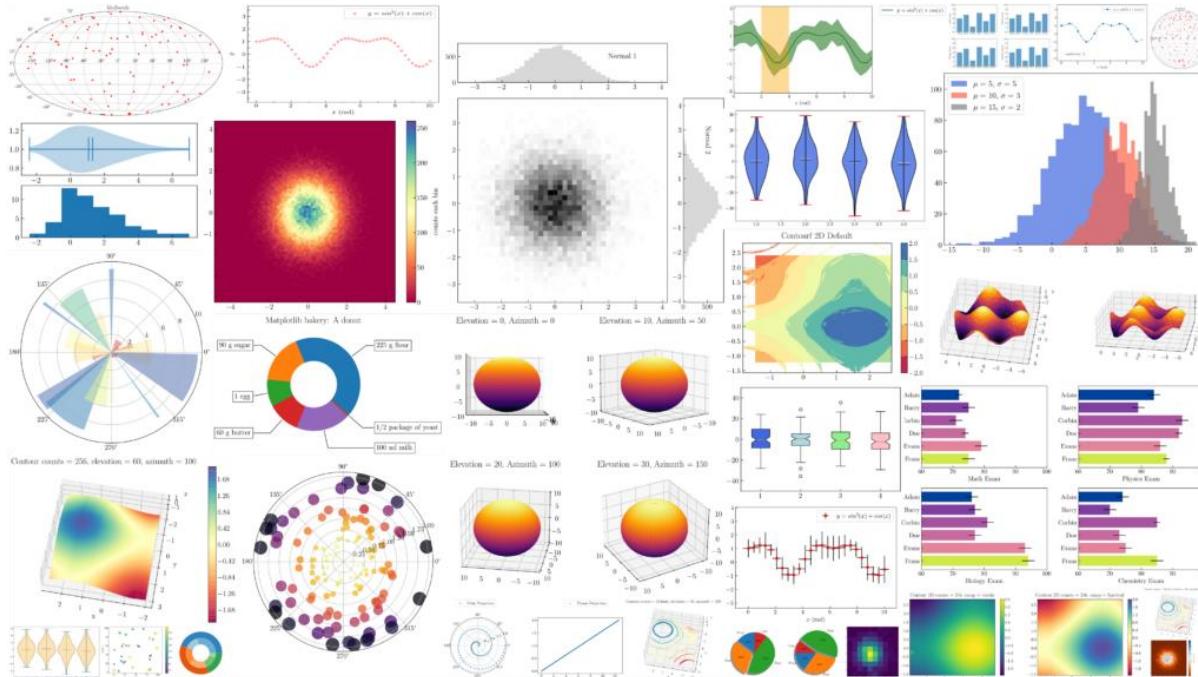


Pandas

- Traitement ais  des donn es manquantes
- Mutabilit  de la taille des dataframes
- Fonctionnalit  de group-by
- Intelligent label-based slicing
- Fusion et merge simple dataframe
- Reshaping et pivot des dataframes



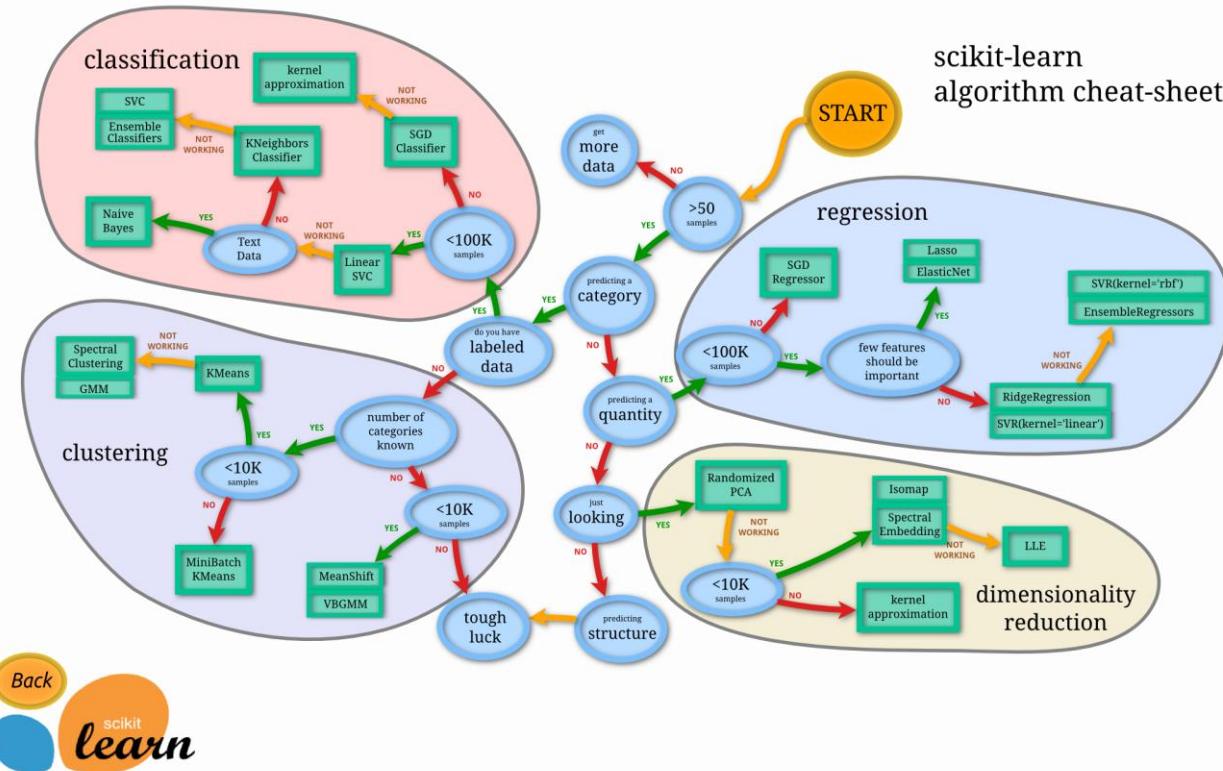
Matplotlib



https://matplotlib.org/stable/plot_types/index.html



Scik





Les différents packages de deep learning

Google

K Keras



TensorFlow

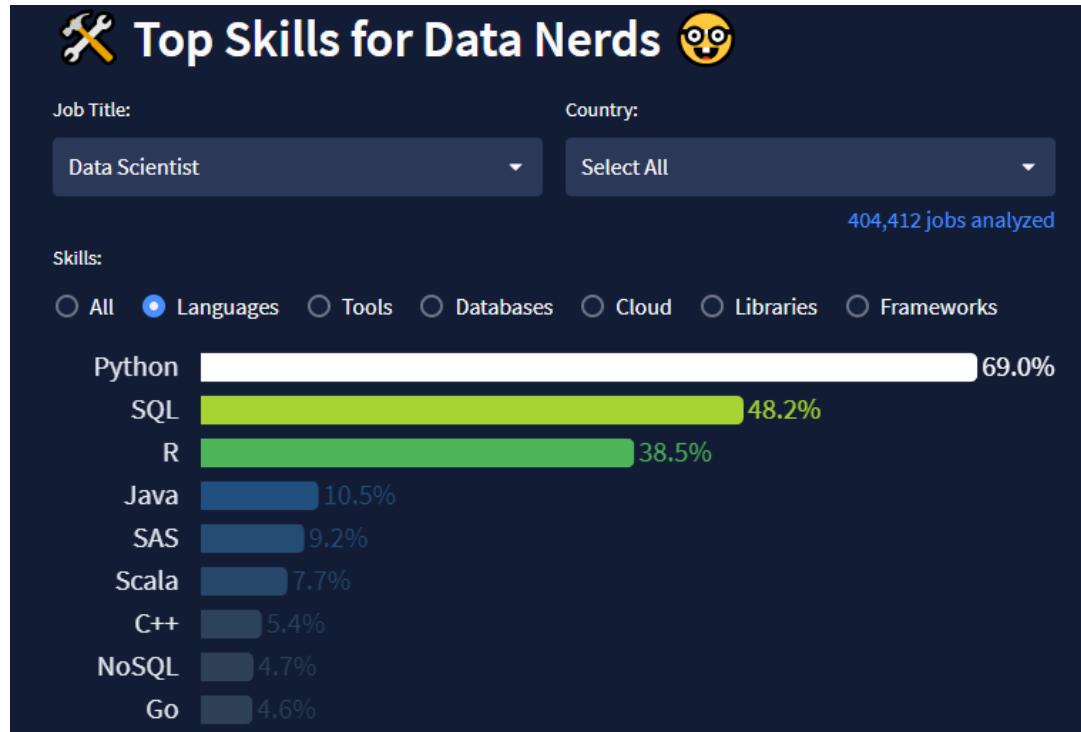
facebook

fast.ai

PyTorch



SQL



Versioning



Partie 3 : Les outils



Github

- Gestion facile des projets
- Sécurité accrue
- Gestion efficace de l'équipe
- Amélioration de l'écriture du code (pull request)
- Hébergement facile du code

Sécurisation de l'environnement



Partie 3 : Les outils



Poetry

- Gestion des dépendances
- Meilleur solver des conflits des dépendances
- Environnement virtuel
- Mise à jour fréquentes
- Packaging

Cloud & IA sur étagère



Partie 3 : Les outils



Les géants du cloud



Google Cloud Platform





IA sur étagère

IA et machine learning →

Plate-forme Vertex AI

Plate-forme unifiée pour les modèles de ML et l'IA générative

IA générative sur Vertex AI

Créer, régler et déployer des modèles de fondation sur Vertex AI

Vertex AI Search and Conversation

Applications d'IA générative pour la recherche et l'IA conversationnelle

Dialogflow

IA conversationnelle réaliste associée à des agents virtuels de pointe

IA Natural Language

Analyse des sentiments et classification de texte non structuré

Speech-to-Text

Reconnaissance vocale et transcription dans 125 langues

Text-to-Speech

Synthèse vocale avec plus de 220 voix dans plus de 40 langues

IA pour la traduction

Détection de la langue, traduction et intégration de glossaires

Document AI

Traitements des documents et capture des données automatisés à grande échelle

Vision AI

Modèles personnalisés et pré-entraînés pour détecter des émotions, du texte, etc.

Contact Center AI

Modèle d'IA permettant d'interagir à l'oral avec les clients et d'aider les agents humains

Vous ne trouvez pas ce que vous cherchez ?

Afficher tous les produits d'IA et de machine learning



IA sur étagère

Speech-to-Text

Convertissez la voix en texte avec précision, à l'aide d'une API optimisée par les technologies d'IA de Google.

- ✓ Reconnaissance vocale auto
- ✓ Transcription en temps réel
- ✓ Modèles d'appels téléphoniques améliorés dans Google Contact Center AI

Text-to-Speech

Convertissez du texte en discours naturel à l'aide d'une API basée sur l'IA de Google.

- ✓ Amélioration des interactions avec les clients
- ✓ Interface vocale sur les appareils et les applications
- ✓ Communication personnalisée

Deep learning avec Pytorch

Partie 4 : La régression linéaire



Présenté par **Morgan Gautherot**



Problème de régression

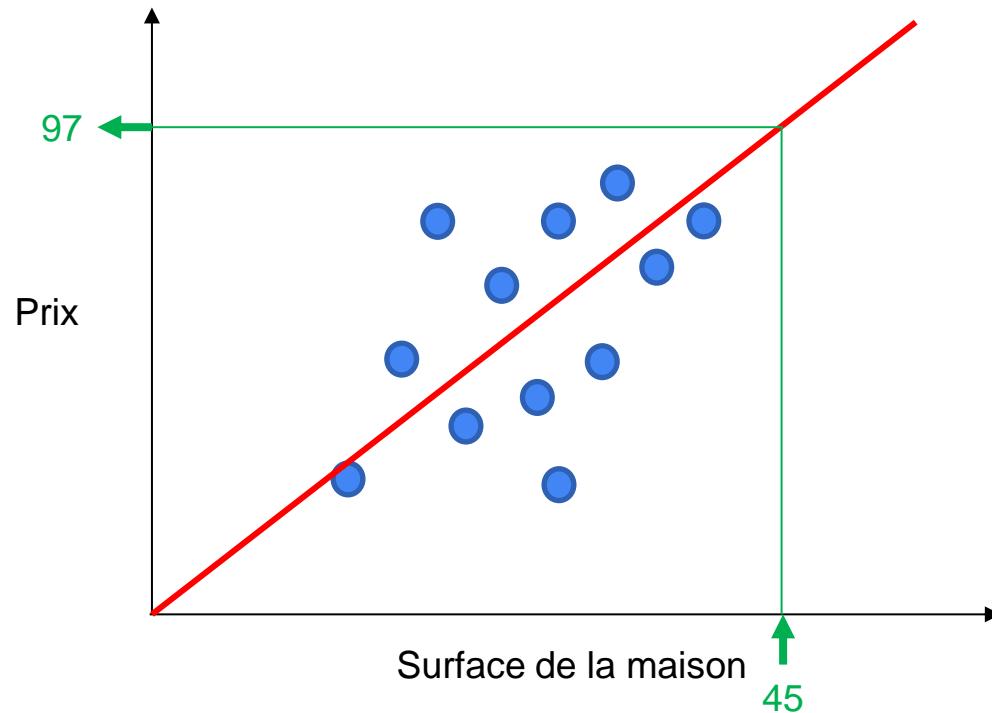
X → Y

	Surface (x_1)	Nb de pièces(x_2)	Année (x_3)	Prix (y)
1	70	3	2010	460
2	40	3	2015	232
3	45	4	1990	315
4	12	2	2017	178
...
m	25	1	2005	240

Jeu d'entraînement pour la prédiction de prix de maison

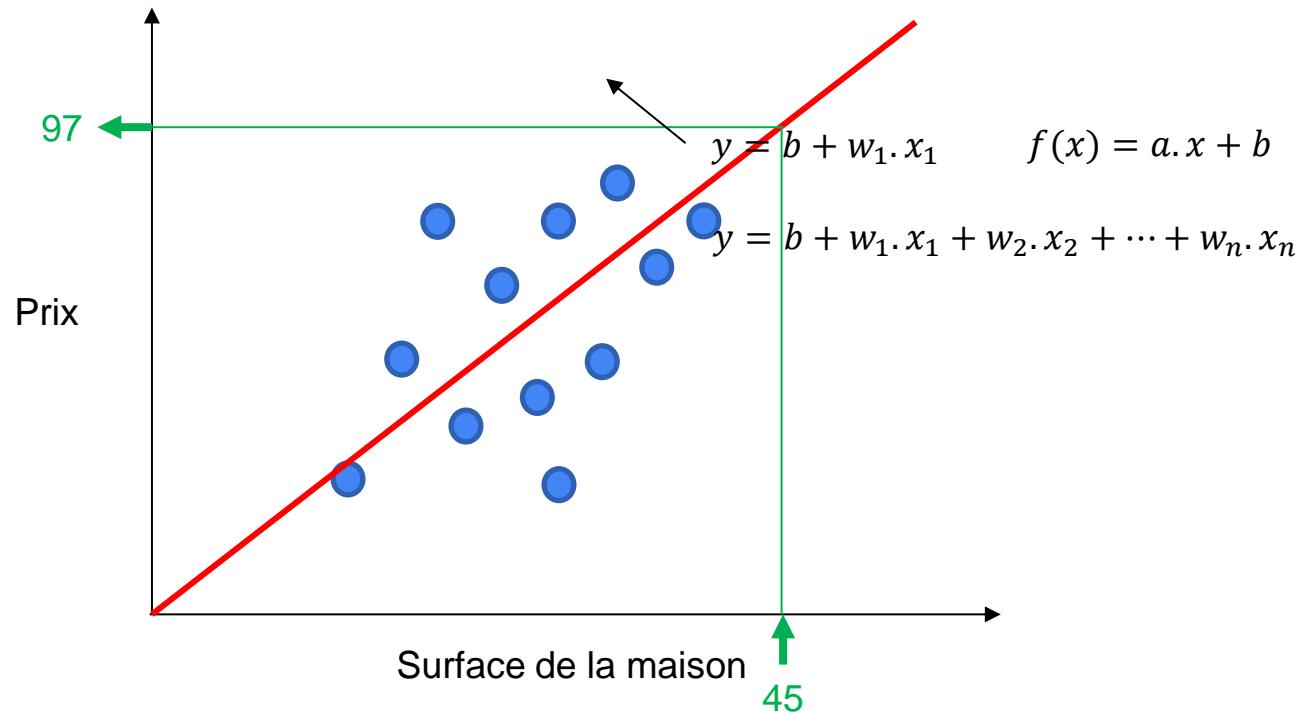


Utiliser une droite





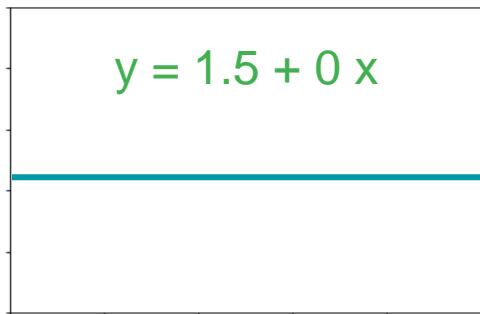
Equation d'une droite





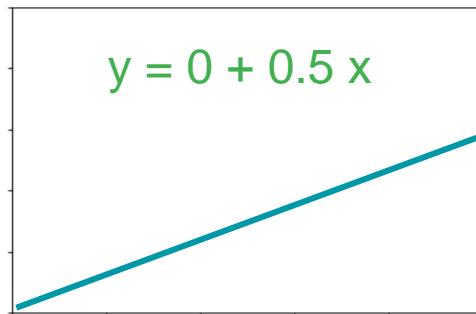
Les paramètres de la régression linéaire

$$y = b + w_1 x$$



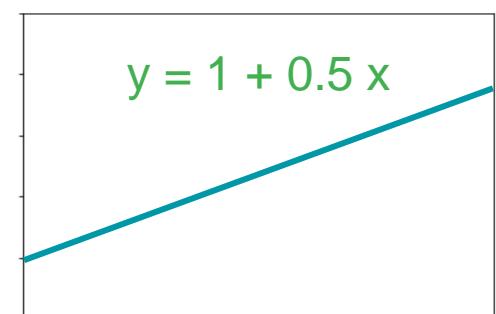
$$b= 1.5$$

$$w_1 = 0$$



$$b= 0$$

$$w_1 = 0.5$$

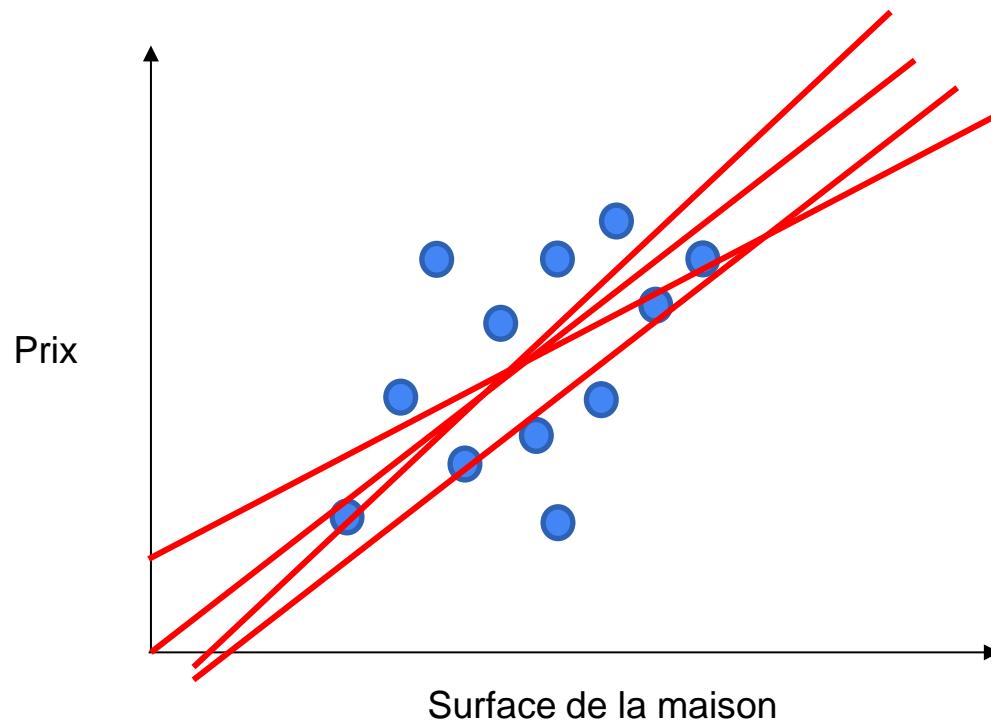


$$b= 1$$

$$w_1 = 0.5$$

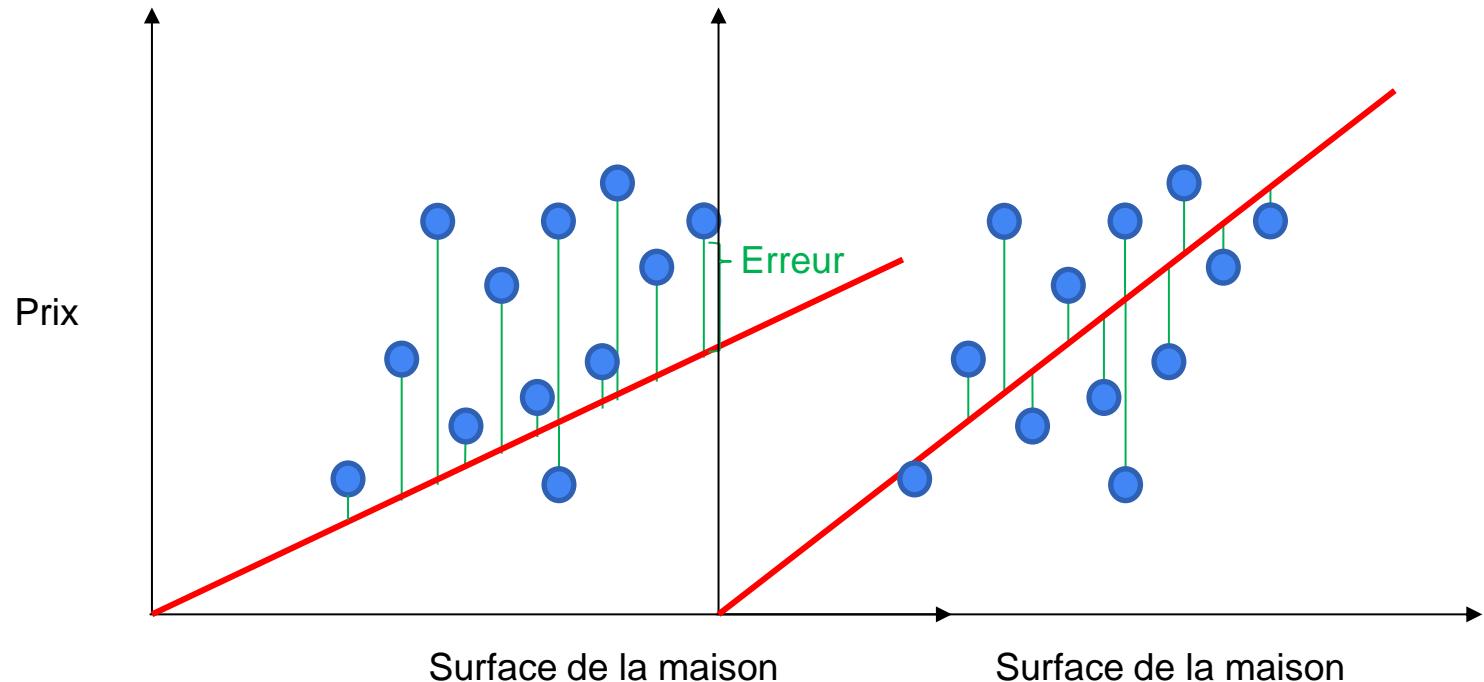


Quels paramètres choisir ?





La notion d'erreur





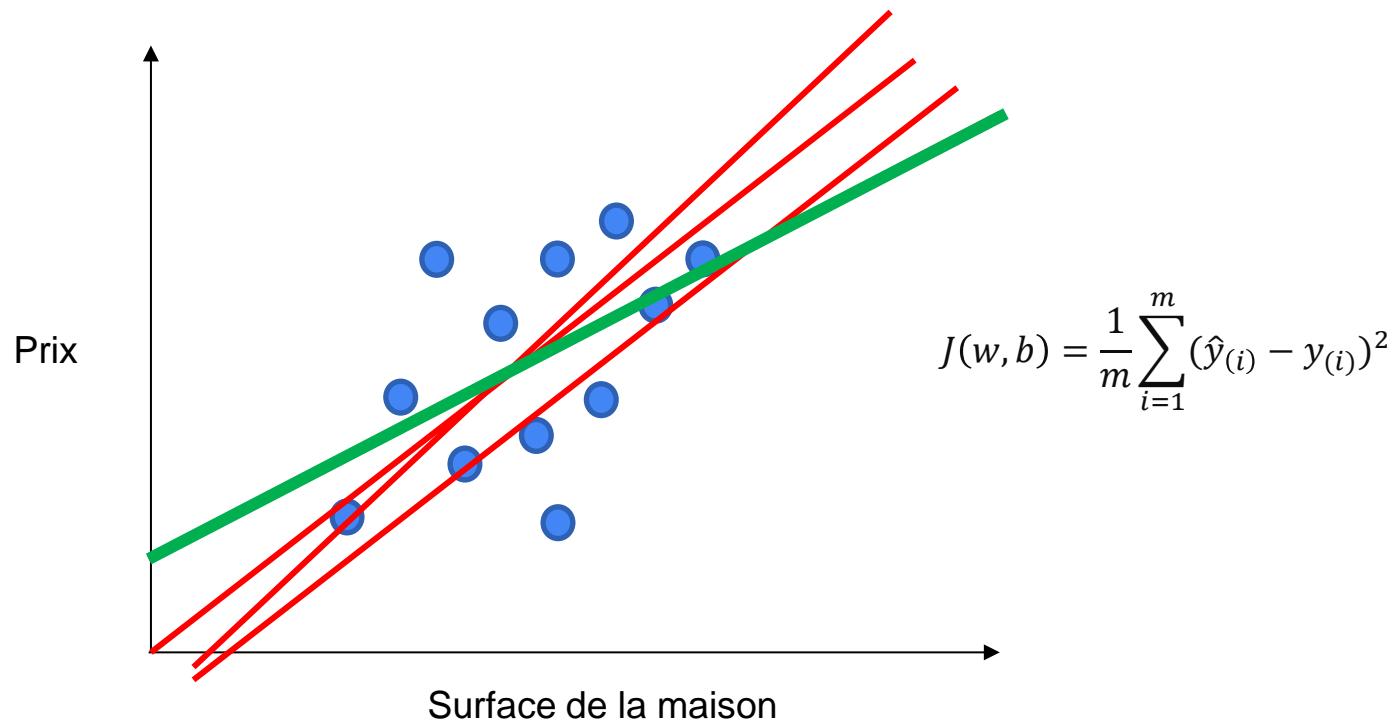
L'erreur moyenne au carré

$$\begin{array}{lllll} \hat{y} - y & \hat{y}_1 = 500 & y_1 = 520 & \hat{y}_1 - y_1 = -20 & \\ \frac{1}{m} \sum \hat{y} - y & \hat{y}_2 = 350 & y_2 = 320 & \hat{y}_1 - y_1 = 30 & \frac{1}{2} \sum_{i=1}^2 \hat{y}_{(i)} - y_{(i)} = \frac{10}{2} = 5 \end{array}$$

$$J(w, b) = \frac{1}{m} \sum_{i=1}^m (\hat{y}_{(i)} - y_{(i)})^2$$

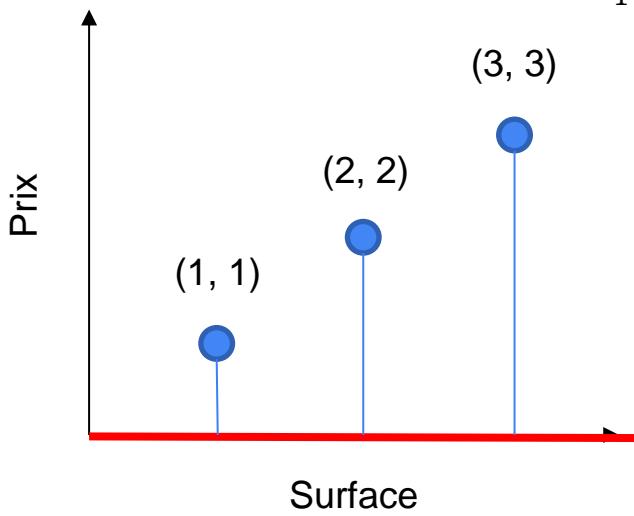


Quels paramètres choisir ?

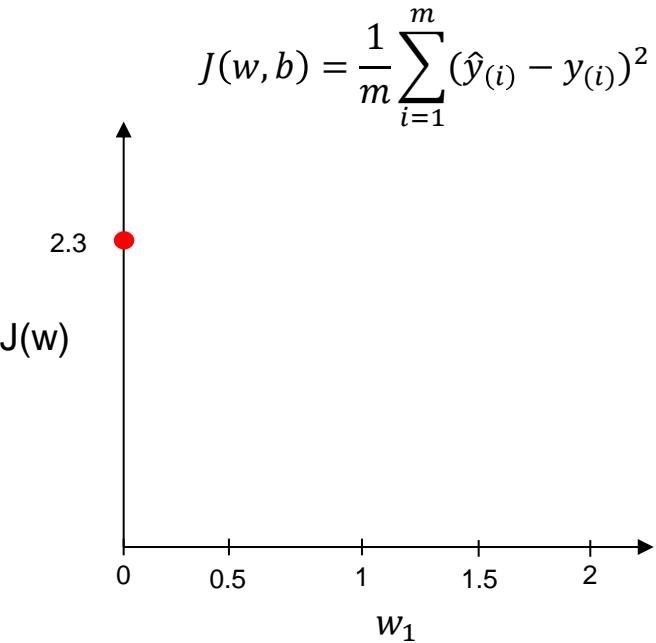




Tracer la fonction de coût



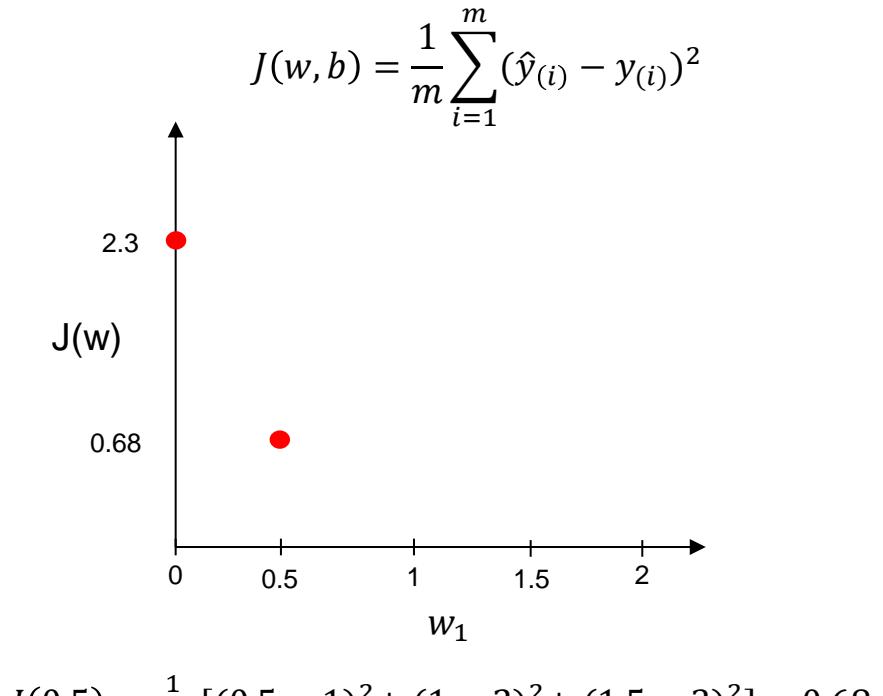
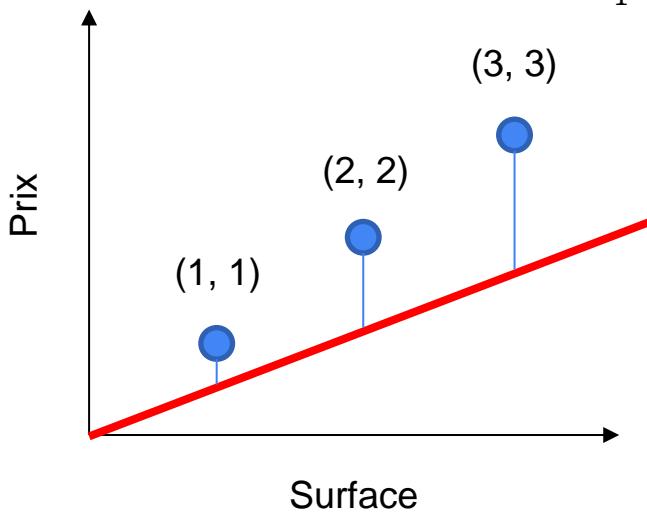
$$\hat{y} = 0 + 0 \cdot x_1$$



$$J(0) = \frac{1}{2m} [(1)^2 + (2)^2 + (3)^2] = 2.3$$

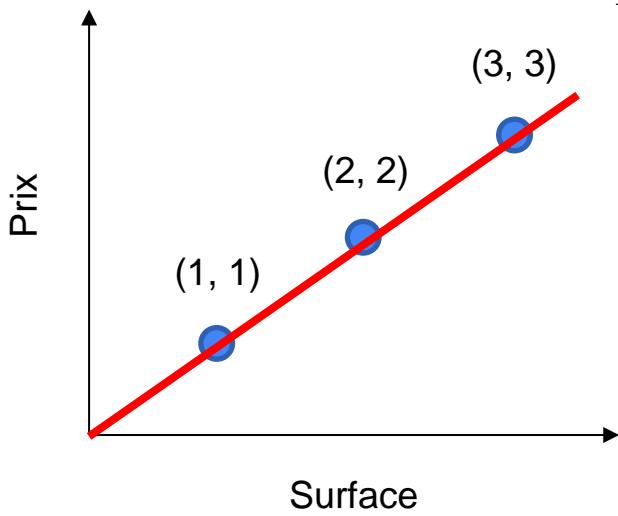


Tracer la fonction de coût

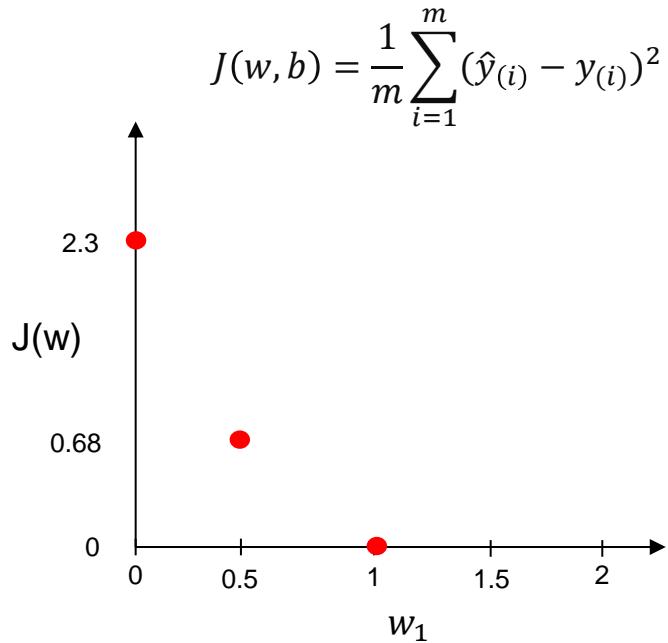




Tracer la fonction de coût



$$\hat{y} = 0 + 1 \cdot x_1$$



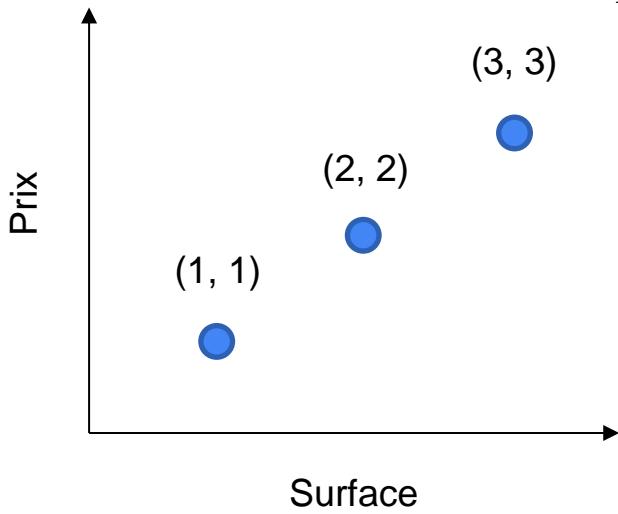
$$J(1) = \frac{1}{2m} [(1-1)^2 + (2-2)^2 + (3-3)^2] = 0$$



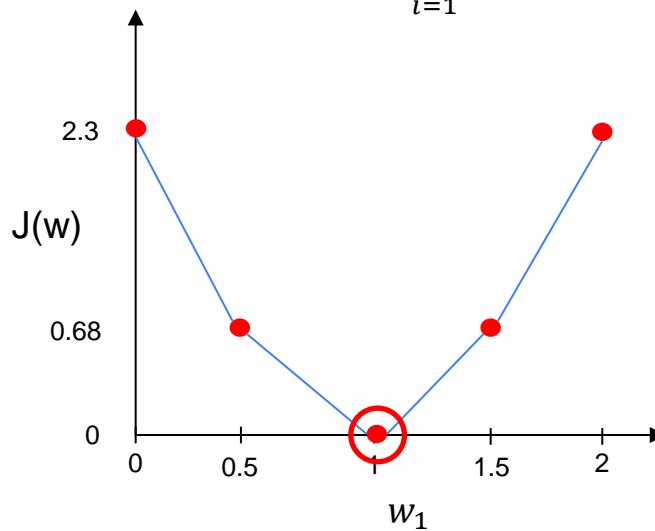
Tracer la fonction de coût

$$\hat{y} = b + w_1 \cdot x_1$$

b est fixé à 0
 w_1 va varier

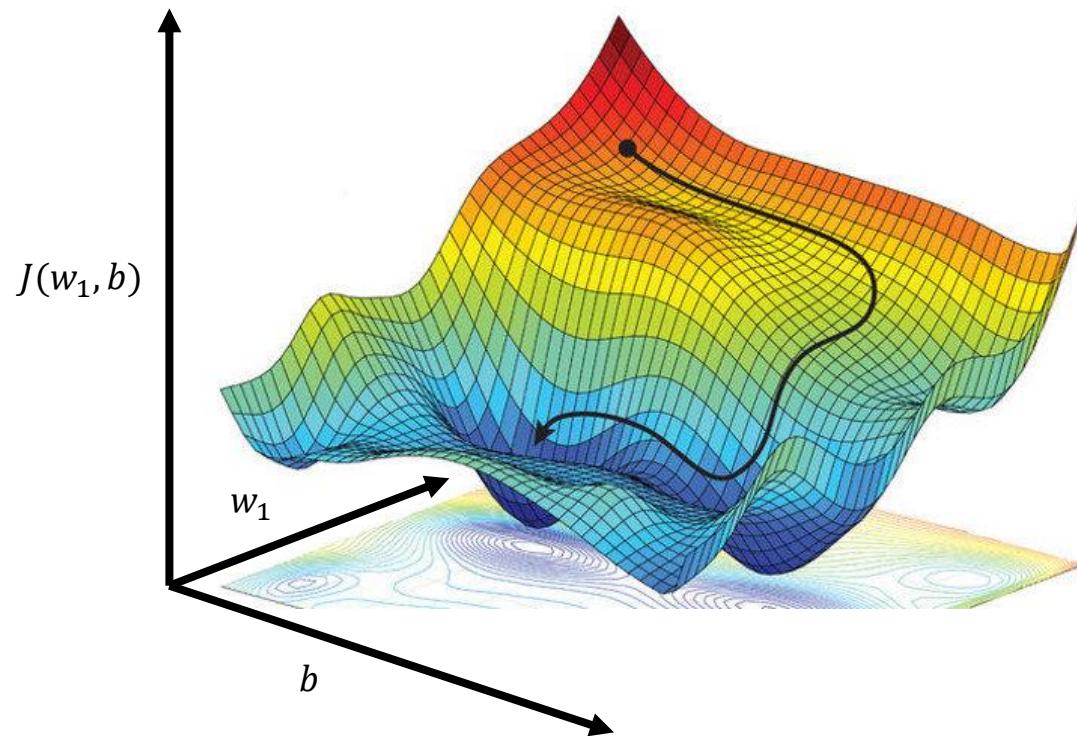


$$J(w, b) = \frac{1}{m} \sum_{i=1}^m (\hat{y}_{(i)} - y_{(i)})^2$$





Fonction de coût avec deux paramètres





Le gradient descent

Soit n le nombre de variables

Répéter ce processus jusqu'à la convergence :

$$w_j := w_j - \alpha \frac{\partial}{\partial w_j} J(w, b)$$

$j \in [1, n]$

Nouvelle valeur de w_j

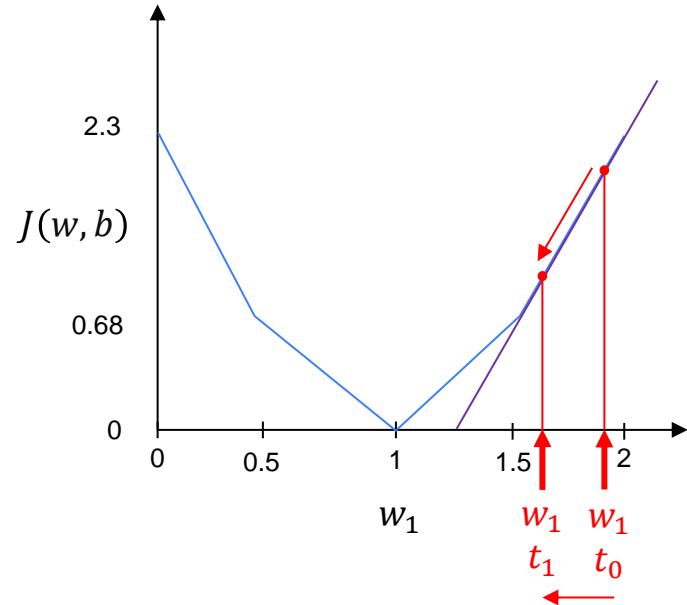
Ancienne valeur de w_j

Le learning rate

Dérivée partielle



La dérivée partielle



$$w_1 := w_1 - \alpha \frac{\partial}{\partial w_1} J(w, b)$$

t_1 t_0

$$w_1 := w_1 - \text{positif}$$

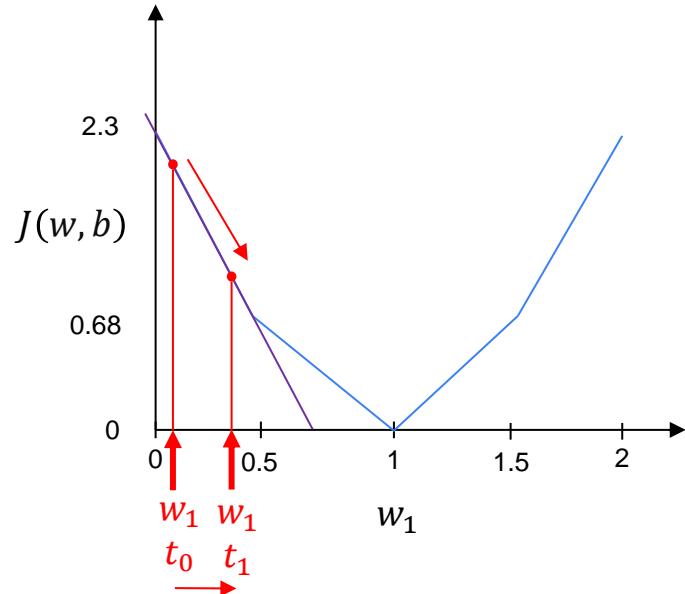
t_1 t_0

$$w_1 < w_1$$

t_1 t_0



La dérivée partielle



$$w_1 := w_1 - \alpha \frac{\partial}{\partial w_1} J(w, b)$$

t_1 t_0

$$w_1 := w_1 - \text{negatif}$$

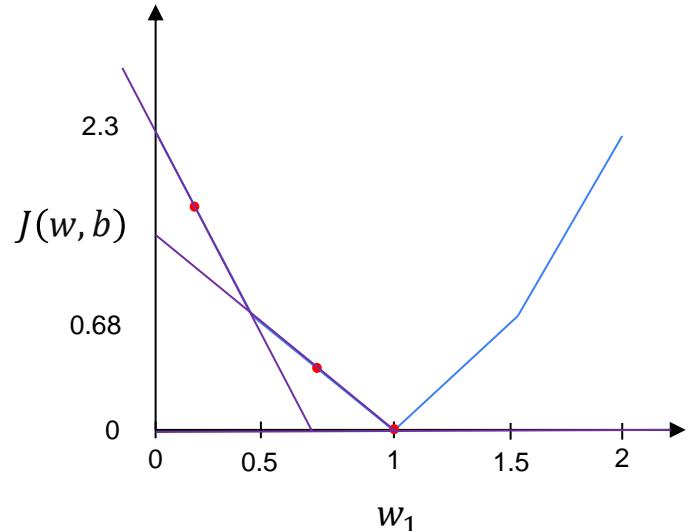
t_1 t_0

$$w_1 > w_1$$

t_1 t_0



Atteindre le minimum



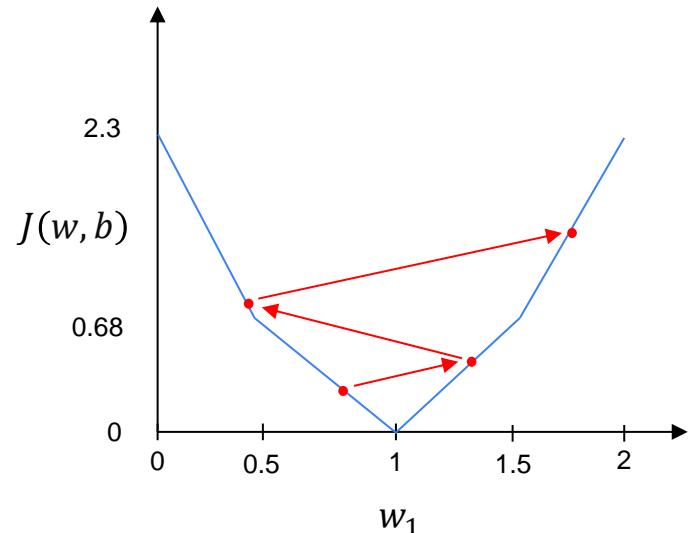
Réduit vers le minimum

$$w_1 := w_1 - \alpha \frac{\partial}{\partial w_1} J(w, b)$$

Est fixe



Impact du learning rate

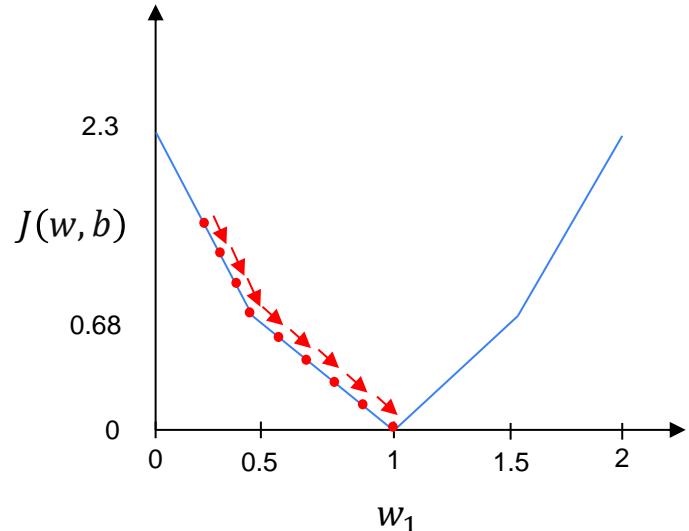


$$w_1 := w_1 - \alpha \frac{\partial}{\partial w_1} J(w, b)$$

Learning rate trop grand



Impact du learning rate



$$w_1 := w_1 - \alpha \frac{\partial}{\partial w_1} J(w, b)$$

Learning rate trop petit

Deep learning avec Pytorch

Partie 5 : Votre premier réseau

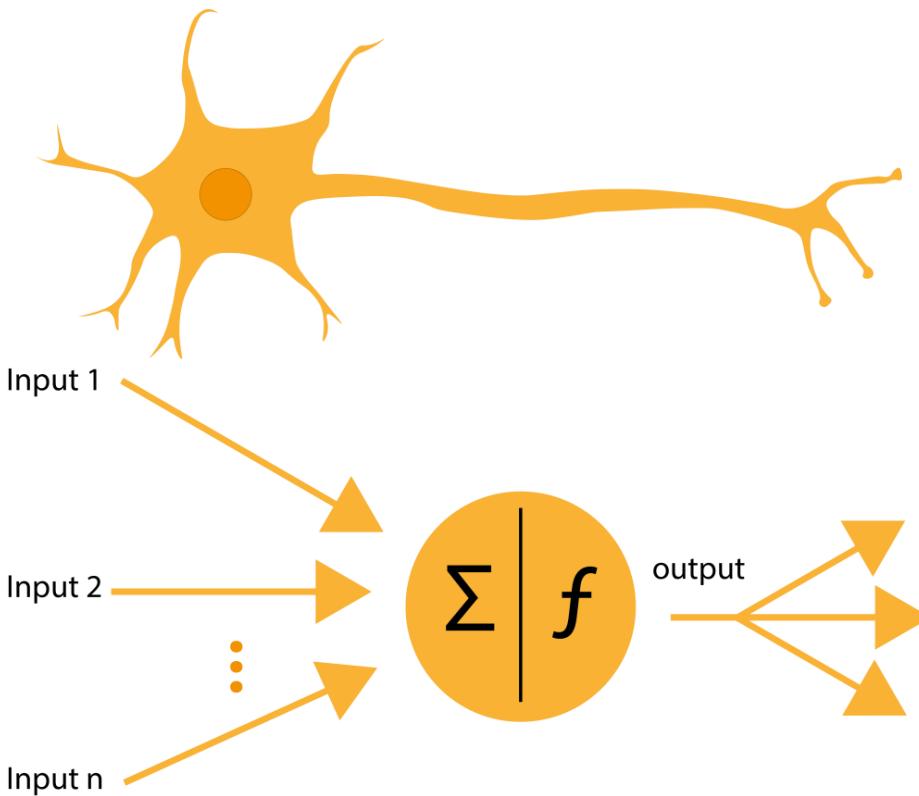


Présenté par **Morgan Gautherot**



Le neurone artificiel

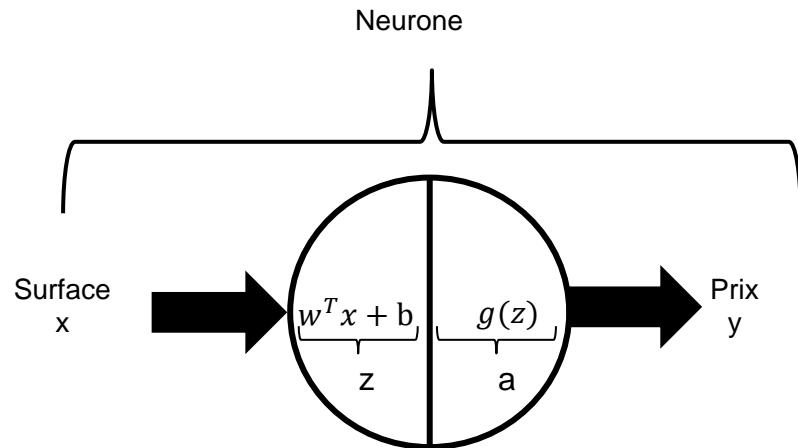
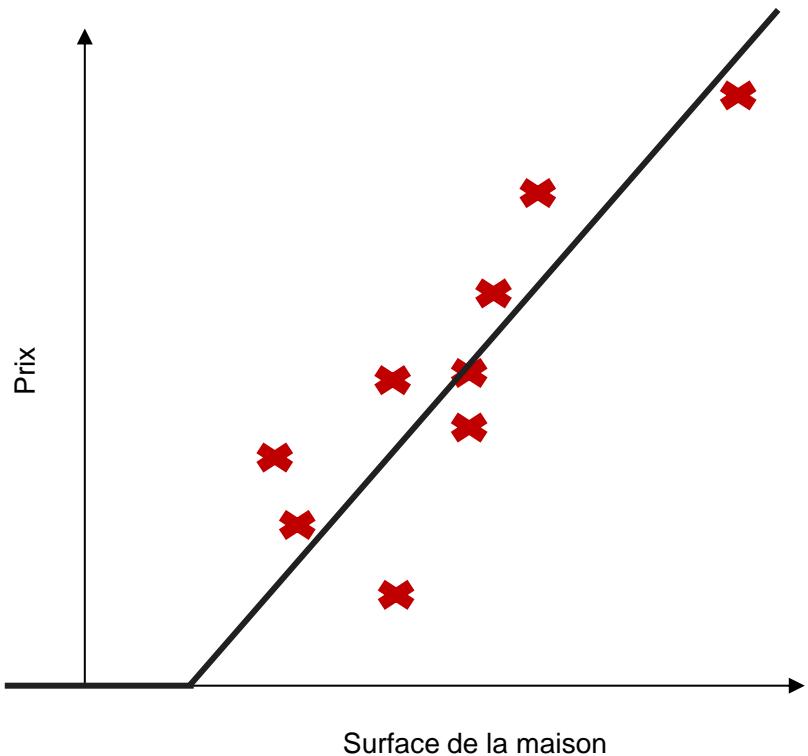
Neurone biologique



Neurone artificiel



Qu'est-ce qu'un réseau neuronal ?

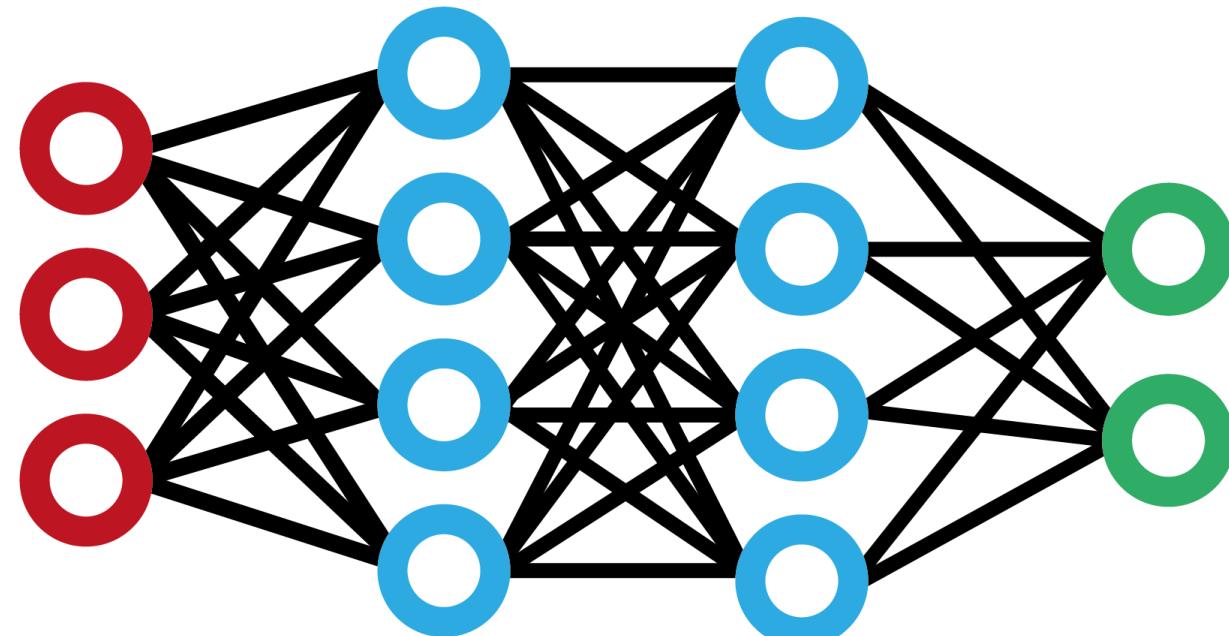


Un réseau neuronal est construit en connectant de nombreux neurones

Fonction d'activation : $g(z) = \max(0, z)$



Réseau neuronal



Couche d'entrée

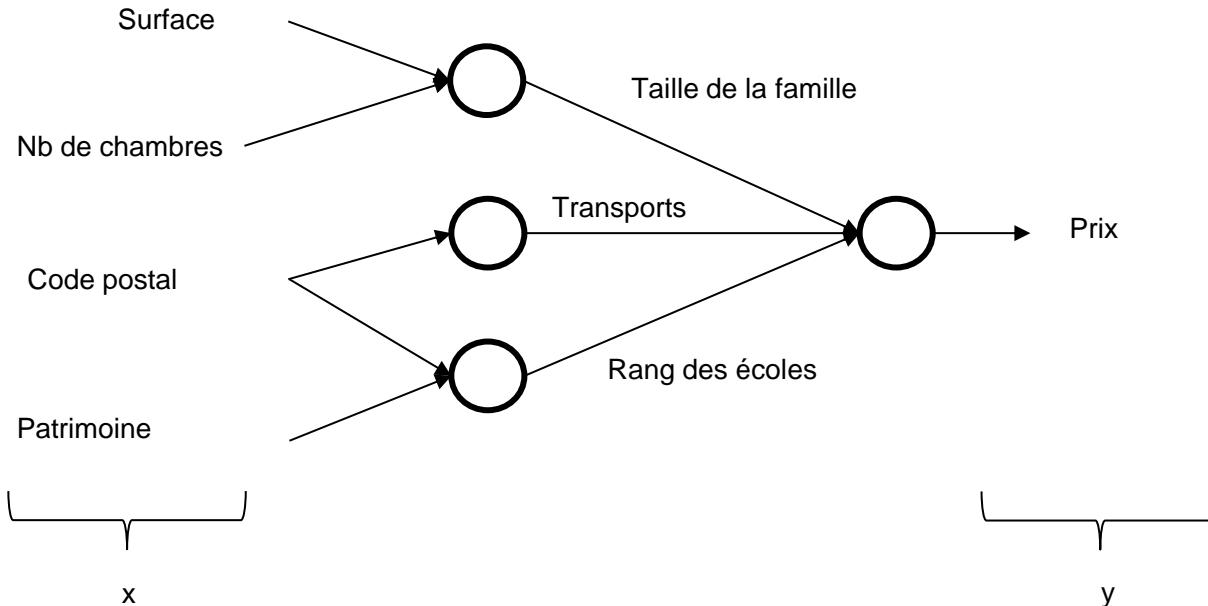
Couche cachée 1

Couche cachée 2

Couche de sortie

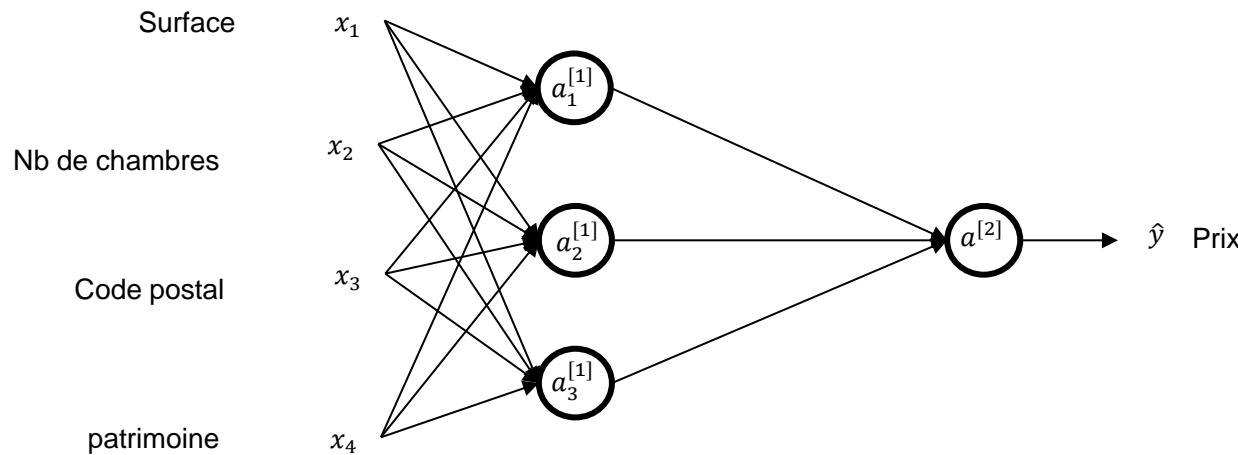


Le feature engineering





Feature engineering automatique



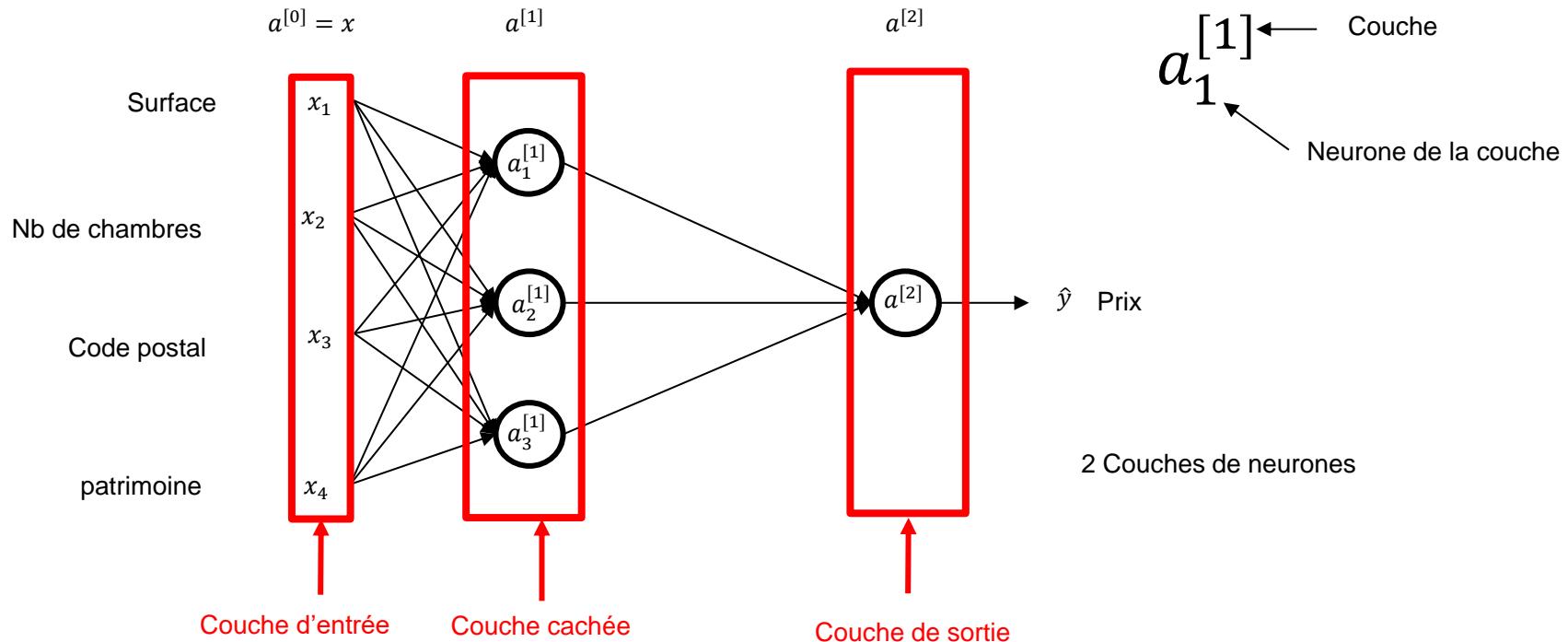
Représentation mathématique



Partie 5 : Votre premier réseau

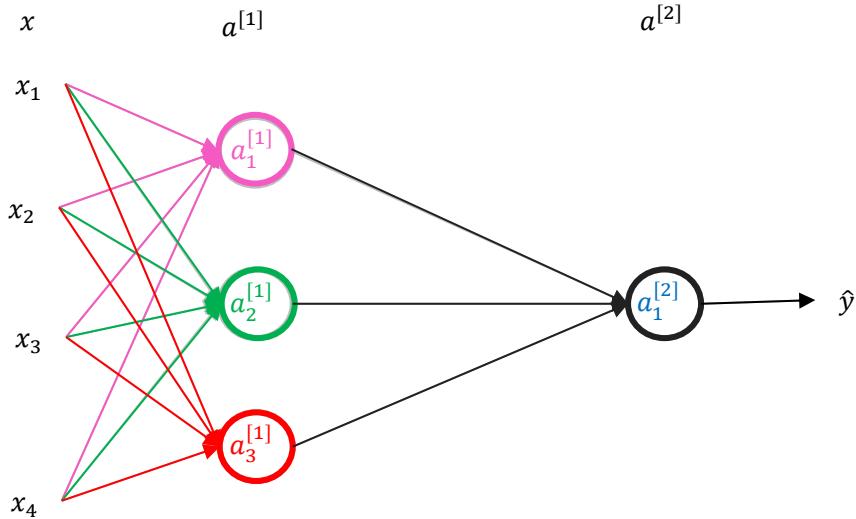


Un peu de notations





Et les mathématiques dans tout ça ?



$$z_1^{[1]} = w_1^{[1]T} x + b_1^{[1]} \quad ; \quad a_1^{[1]} = g(z_1^{[1]})$$

$$z_2^{[1]} = w_2^{[1]T} x + b_2^{[1]} \quad ; \quad a_2^{[1]} = g(z_2^{[1]})$$

$$z_3^{[1]} = w_3^{[1]T} x + b_3^{[1]} \quad ; \quad a_3^{[1]} = g(z_3^{[1]})$$

$$z_1^{[2]} = w_1^{[2]T} a^{[1]} + b_1^{[2]} \quad ; \quad a_1^{[2]} = g(z_1^{[2]})$$



Vectorisation sur un exemple

$$\begin{aligned} \begin{bmatrix} z_1^{[1]} \\ z_2^{[1]} \\ z_3^{[1]} \end{bmatrix} &= \begin{bmatrix} w_1^{[1]T} x + b_1^{[1]} \\ w_2^{[1]T} x + b_2^{[1]} \\ w_3^{[1]T} x + b_3^{[1]} \end{bmatrix} & ; \quad \begin{bmatrix} a_1^{[1]} \\ a_2^{[1]} \\ a_3^{[1]} \end{bmatrix} &= \begin{bmatrix} g(z_1^{[1]}) \\ g(z_2^{[1]}) \\ g(z_3^{[1]}) \end{bmatrix} \end{aligned}$$

$$\begin{bmatrix} z_1^{[1]} \\ z_2^{[1]} \\ z_3^{[1]} \end{bmatrix} = \begin{bmatrix} w_{1,1}^{[1]} & w_{1,2}^{[1]} & w_{1,3}^{[1]} & w_{1,4}^{[1]} \\ w_{2,1}^{[1]} & w_{2,2}^{[1]} & w_{2,3}^{[1]} & w_{2,4}^{[1]} \\ w_{3,1}^{[1]} & w_{3,2}^{[1]} & w_{3,3}^{[1]} & w_{3,4}^{[1]} \end{bmatrix} \cdot \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix} + \begin{bmatrix} b_1^{[1]} \\ b_2^{[1]} \\ b_3^{[1]} \end{bmatrix} ; \quad \begin{bmatrix} a_1^{[1]} \\ a_2^{[1]} \\ a_3^{[1]} \end{bmatrix} = g\left(\begin{bmatrix} z_1^{[1]} \\ z_2^{[1]} \\ z_3^{[1]} \end{bmatrix}\right)$$

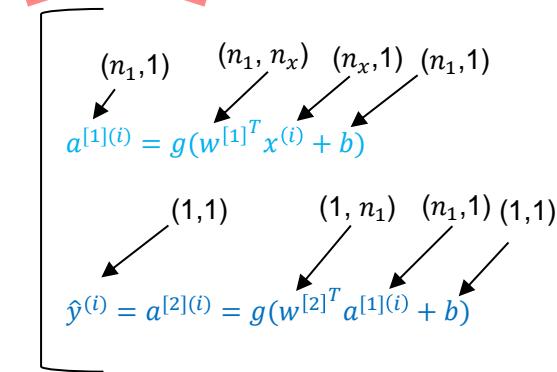
$$(3,1) \quad (3,4) \quad (4,1) \quad (3,1) \\ z^{[1]} = W^{[1]T} x + b^{[1]} ; \quad a^{[1]} = g(z^{[1]})$$

$$(1,1) \quad (1,3) \quad (3,1) \quad (1,1) \\ z^{[2]} = W^{[2]T} a^{[1]} + b^{[2]} ; \quad a^{[2]} = g(z^{[2]})$$



Vectorisation sur plusieurs exemples

~~for i = 0 to m :~~



$$a^{[1]} = g(w^{[1]T} X + b)$$

$$\hat{y} = a^{[2]} = g(w^{[2]T} a^{[1]} + b)$$

$$x^{(i)} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_{n_x} \end{bmatrix}$$

$$X \in \mathbb{R}^{(n_x, m)}$$

$$\hat{y} \in \mathbb{R}^m$$

$$X = \begin{bmatrix} x_1^{(1)} & x_1^{(2)} & x_1^{(3)} & \cdots & x_1^{(m)} \\ x_2^{(1)} & x_2^{(2)} & x_2^{(3)} & \cdots & x_2^{(m)} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ x_{n_x}^{(1)} & x_{n_x}^{(2)} & x_{n_x}^{(3)} & \cdots & x_{n_x}^{(m)} \end{bmatrix}$$

Fonction d'activation



Partie 5 : Votre premier réseau



Un modèle sans fonction d'activation

$$z^{[1]} = w^{[1]T} x + b^{[1]}$$

$$a^{[1]} = g(z^{[1]})$$

$$z^{[2]} = w^{[2]T} a^{[1]} + b^{[2]}$$

$$a^{[2]} = g(z^{[2]})$$

$$z^{[1]} = w^{[1]T} x + b^{[1]}$$

$$z^{[2]} = w^{[2]T} z^{[1]} + b^{[2]}$$

$$z^{[2]} = w^{[2]T} (w^{[1]T} x + b^{[1]}) + b^{[2]}$$

$$z^{[2]} = w^{[2]T} w^{[1]T} x + w^{[2]T} b^{[1]} + b^{[2]}$$

$$\underbrace{\phantom{w^{[2]T} w^{[1]T}}}_{W}$$

$$\underbrace{\phantom{w^{[2]T} b^{[1]}}}_{B}$$

$$W = w^{[2]T} w^{[1]T}$$

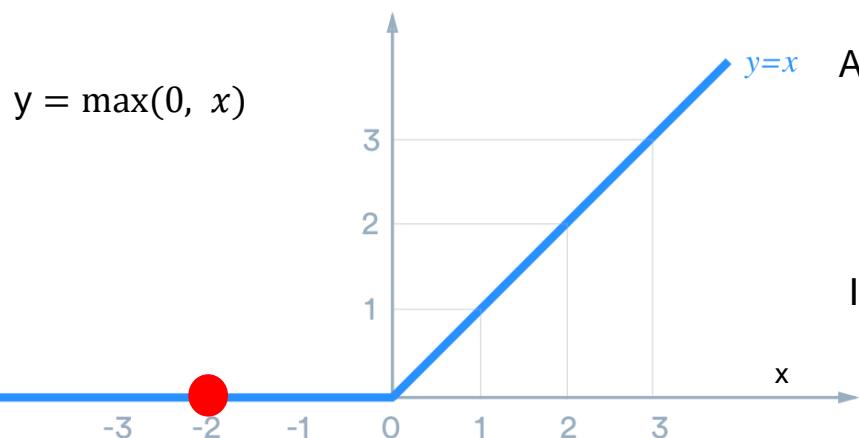
$$B = w^{[2]T} b^{[1]} + b^{[2]}$$

Fonction Linéaire

$$\underline{z^{[2]} = W x + B}$$



Rectified Linear Unit (ReLU)



Avantages :

- ReLU ne sature pas pour les valeurs positives
- ReLU est assez rapide à calculer

Inconvénients :

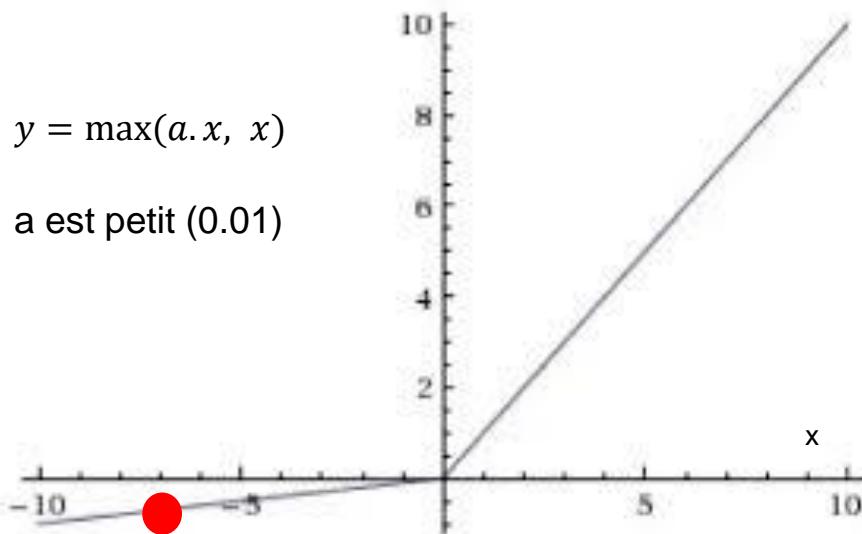
- ReLU souffre d'un problème connu sous le nom de "dying ReLU".



Leaky ReLU

$$y = \max(a \cdot x, x)$$

a est petit (0.01)



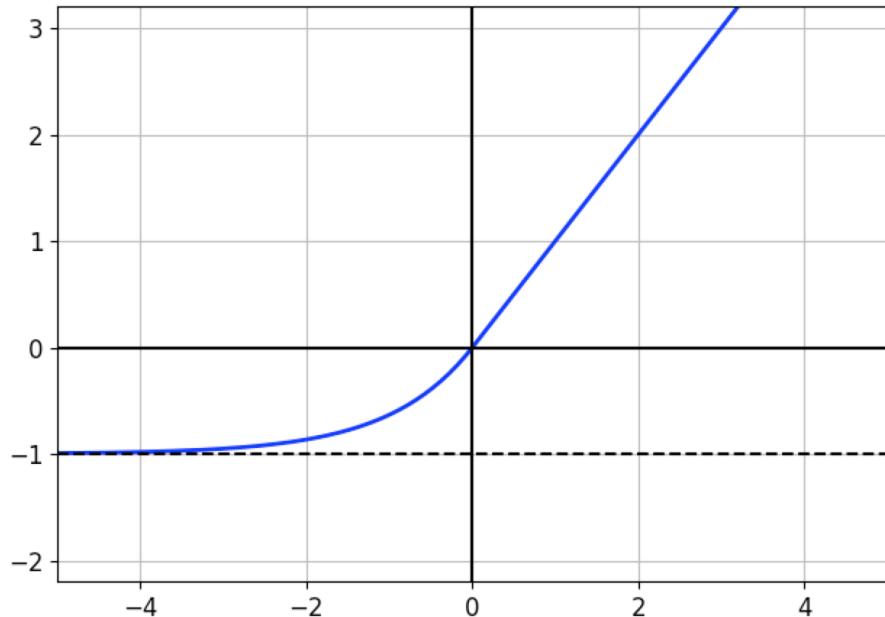
Variantes :

- Parametric leaky Relu (PReLU), si vous disposez de beaucoup de données.
- Randomized leaky Relu (RReLU), si votre réseau neuronal est surentraîné.



Exponential Linear Unit (ELU)

ELU activation function ($\alpha = 1$)



$$ELU_{\alpha}(z) = \begin{cases} \alpha(\exp(z) - 1) & \text{if } z < 0 \\ z & \text{if } z \geq 0 \end{cases}$$

Avantages :

- Il prend des valeurs négatives, ce qui permet au neurone d'avoir une moyenne plus proche de 0
- Il a un gradient non nul pour $z < 0$, ce qui évite le problème des 'dying ReLU'.

Inconvénients :

- Il est plus lent à calculer que ReLU.



Résumé

- En général pour la performance :
ELU > Leaky ReLU > ReLU
- Si vous vous souciez du temps d'exécution :
Leaky ReLU > ELU
- Par défaut :
ReLU

La descente de gradient



Partie 5 : Votre premier réseau



Descente de gradient pour les réseaux de neurones

$$(n^{[1]}, n^{[0]}) \quad (n^{[1]}, 1) \quad (n^{[2]}, n^{[1]}) \quad (n^{[2]}, 1) \quad n_x = n^{[0]}, \quad n^{[1]}, \quad n^{[2]} = 1$$

Paramètres : $w^{[1]}, b^{[1]}, w^{[2]}, b^{[2]}$ $a^{[2]}$

Fonction de coût : $J(w^{[1]}, b^{[1]}, W^{[2]}, b^{[2]}) = \frac{1}{m} \sum_{i=1}^n \mathcal{L}(\hat{y}, y)$

Descente de gradient :

Répéter {

Calculer la prédiction $(\hat{y}^{(i)}, i = (1, \dots, m))$

$$dw^{[1]} = \frac{dJ}{dw^{[1]}}, db^{[1]} = \frac{dJ}{db^{[1]}}, dw^{[2]} = \frac{dJ}{dw^{[2]}}, db^{[2]} = \frac{dJ}{db^{[2]}}$$

$$w^{[1]} := w^{[1]} - \alpha dw^{[1]}$$

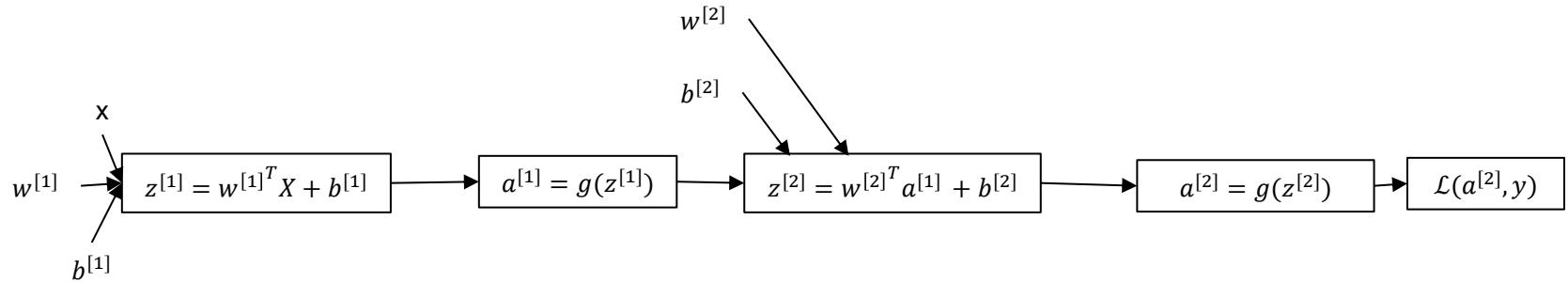
$$b^{[1]} := b^{[1]} - \alpha db^{[1]}$$

$$w^{[2]} := w^{[2]} - \alpha dw^{[2]}$$

$$b^{[2]} := b^{[2]} - \alpha db^{[2]}$$

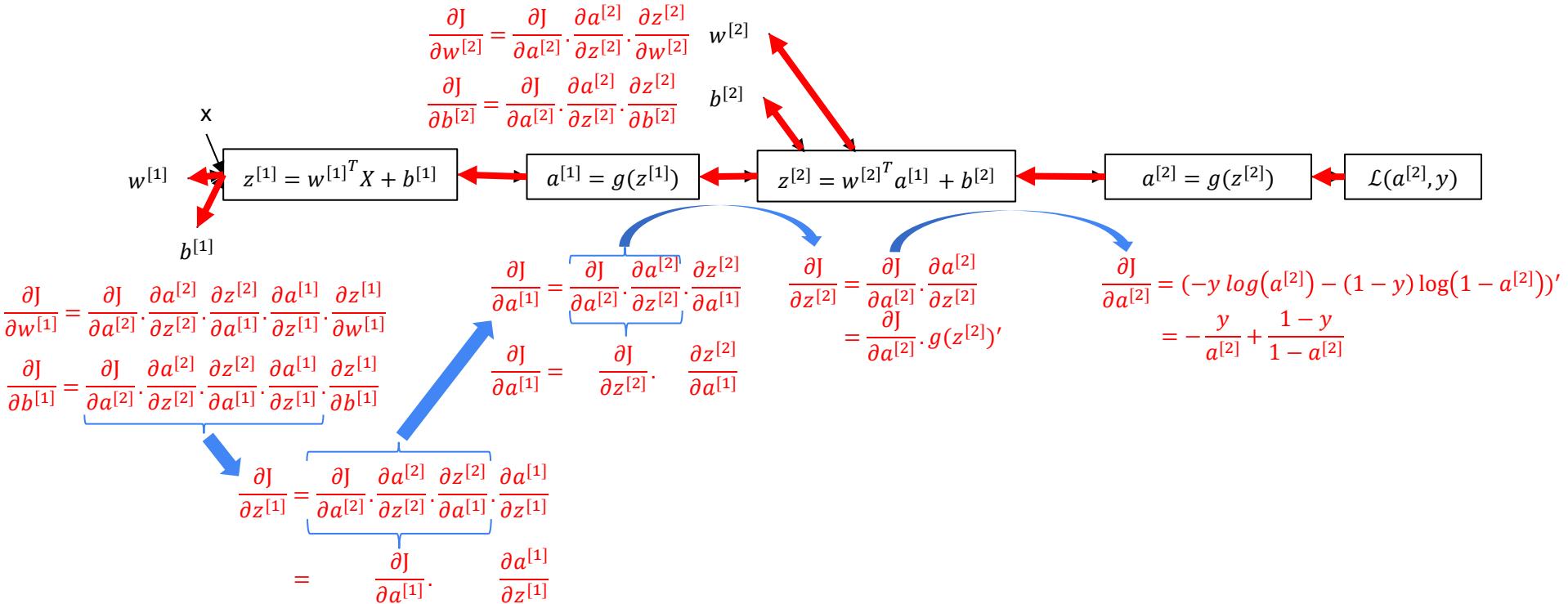


Forward propagation





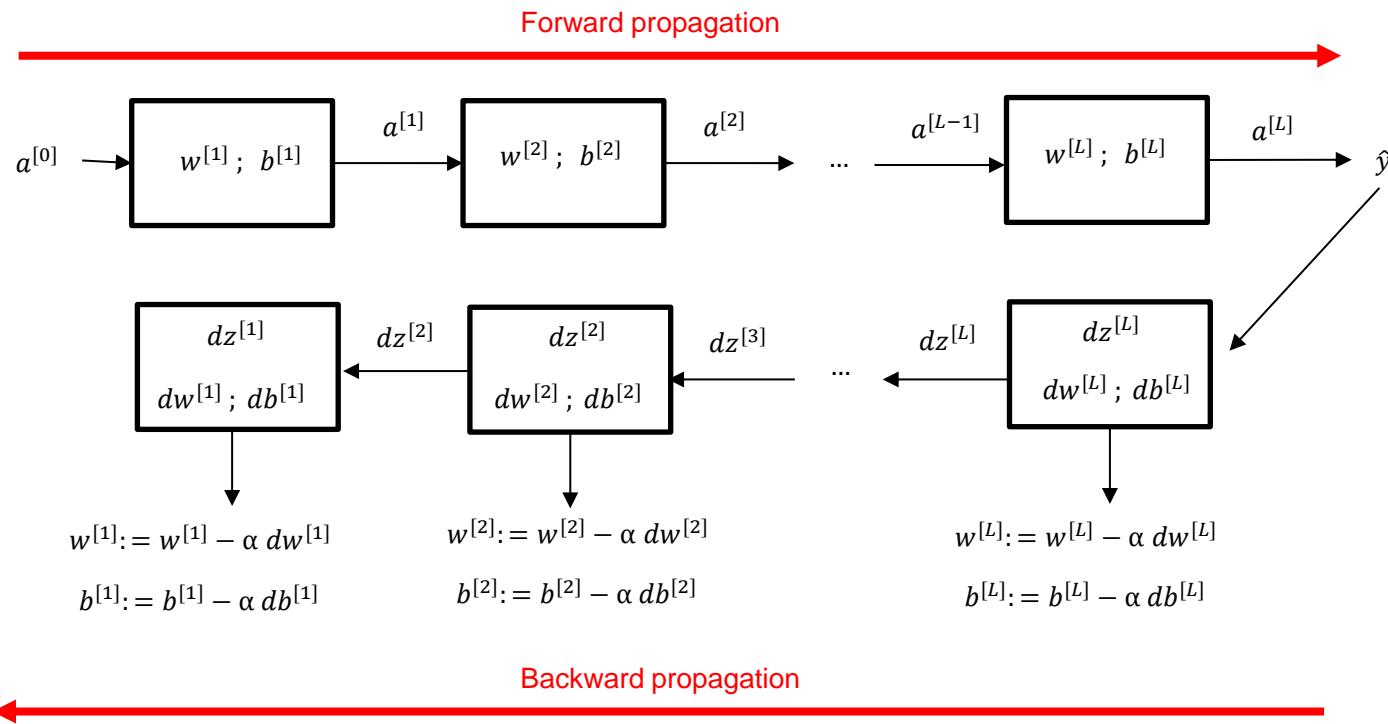
Back propagation





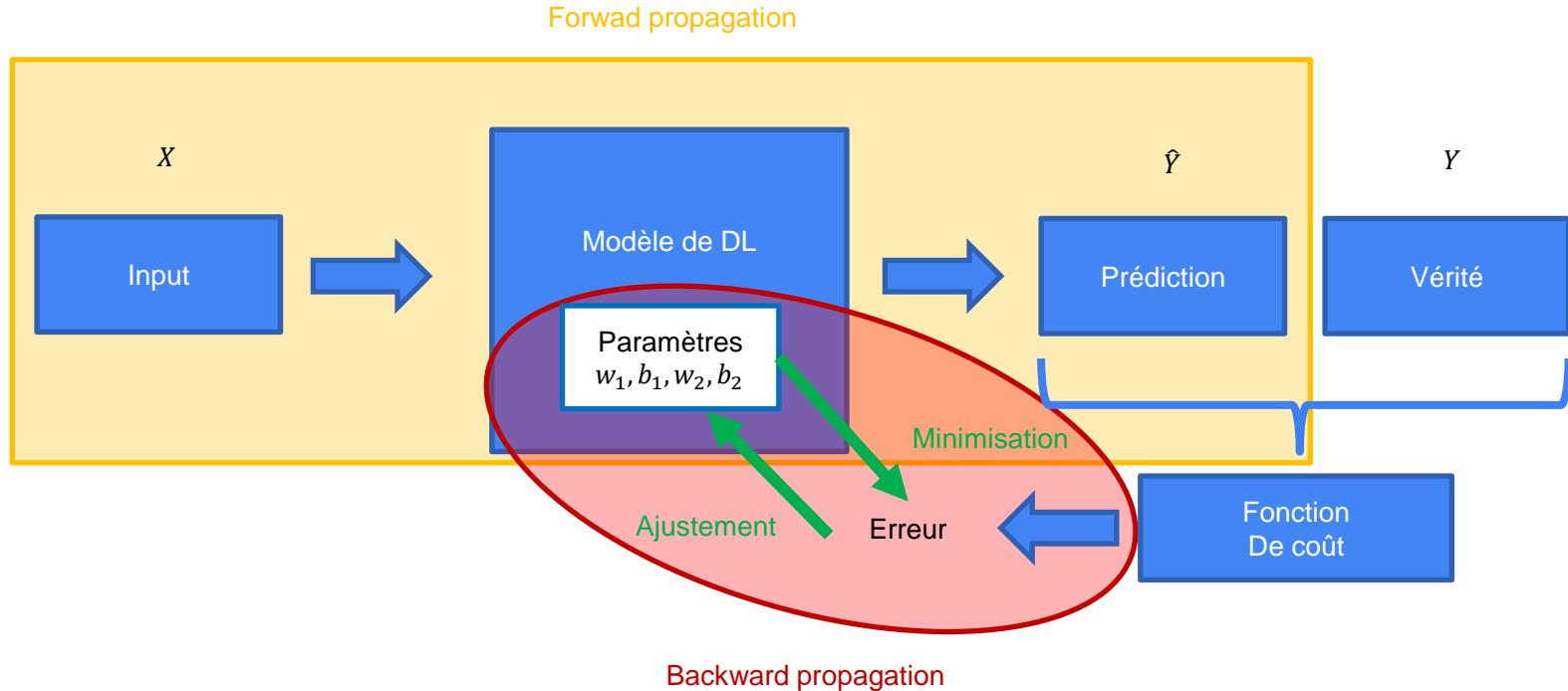
Résumé de l'apprentissage profond

$$dt = \frac{\partial J}{\partial t}$$





Entraînement d'un modèle de deep learning



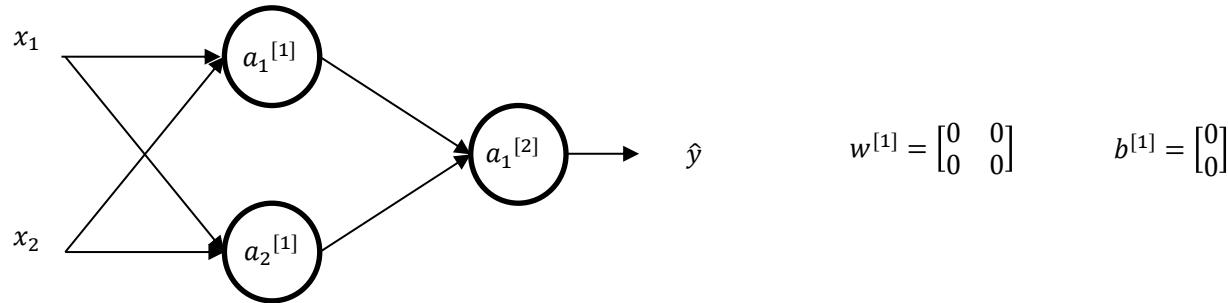
Initialisation des poids



Partie 5 : Votre premier réseau



Initialiser les poids à zéro ?

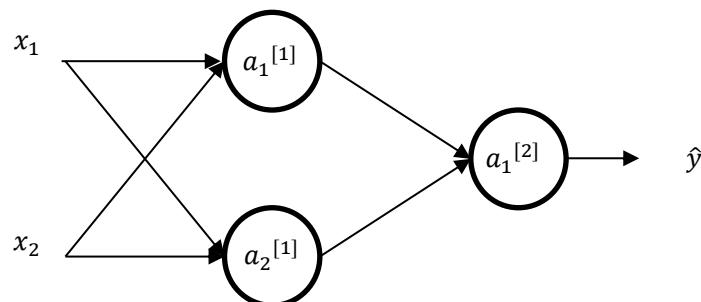


$$w^{[1]} = \begin{bmatrix} p & l \\ p & l \end{bmatrix}$$

$a_1^{[1]} = a_2^{[1]} \longrightarrow dz_1^{[1]} = dz_2^{[1]} \longrightarrow dw^{[1]} = \begin{bmatrix} u & v \\ u & v \end{bmatrix} \longrightarrow w^{[1]} := w^{[1]} - \alpha dw$



Initialiser le poids de façon aléatoire



Petit nombre
↑
 $w^{[1]} = \text{nombre_aléatoire}.0.01$

$b^{[1]} = \text{zéro_ou_aléatoire}$
Petit nombre
↑
 $w^{[2]} = \text{nombre_aléatoire}.0.01$

$b^{[2]} = \text{zéro_ou_aléatoire}$

Pourquoi apprentissage profond ?



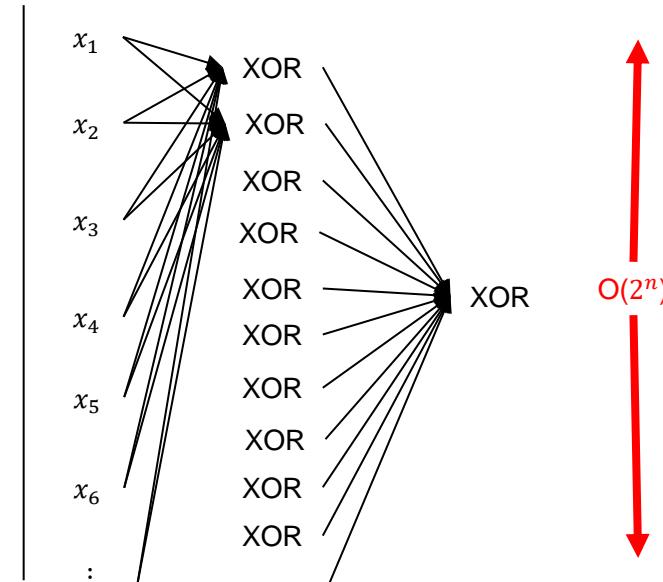
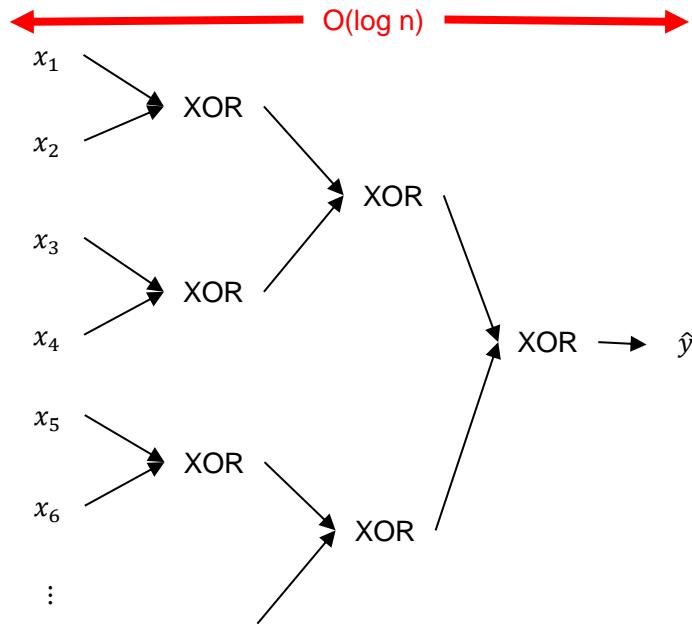
Partie 5 : Votre premier réseau



Pourquoi des représentations profondes ?

Pour approximer une fonction, une architecture de réseau neuronal plus profonde nécessite moins de paramètres qu'une architecture moins profonde.

$$y = x_1 \text{XOR} x_2 \text{XOR} x_3 \text{XOR} x_4 \text{XOR} x_5 \text{XOR} x_6 \text{XOR} \dots$$



Deep learning avec Pytorch

Partie 6 : Sur et sous entraînement



Présenté par **Morgan Gautherot**



Train / dev / test sets





Sur-entraînement et sous-entraînement

Sous-entraînement

Erreurs sur le jeu d'entraînement : Élevée

Erreurs sur le jeu de test : Élevée

Entraînement correct

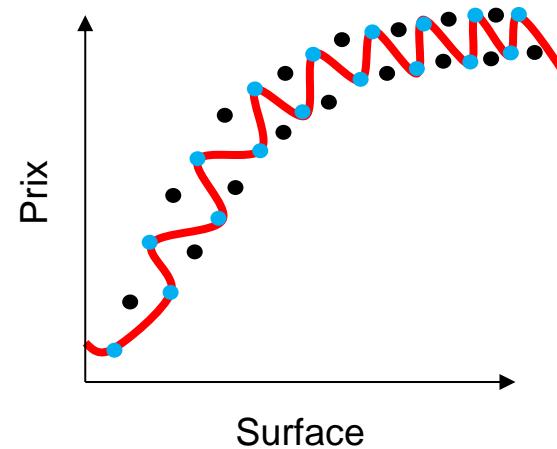
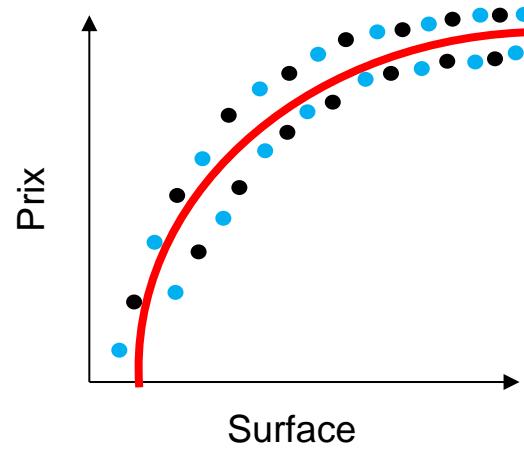
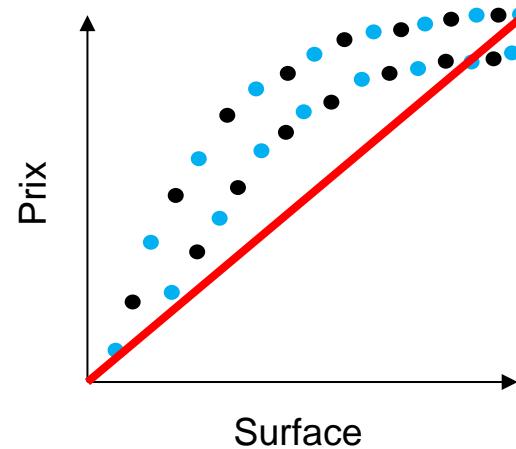
Erreurs sur le jeu d'entraînement : Faible

Erreurs sur le jeu de test : Faible

Sur-entraînement

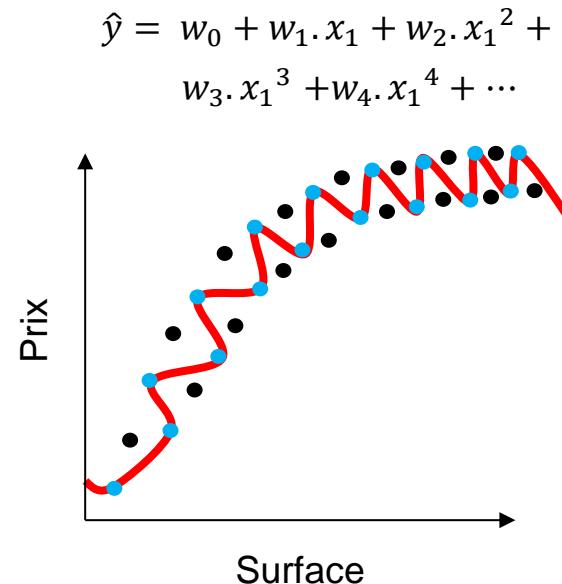
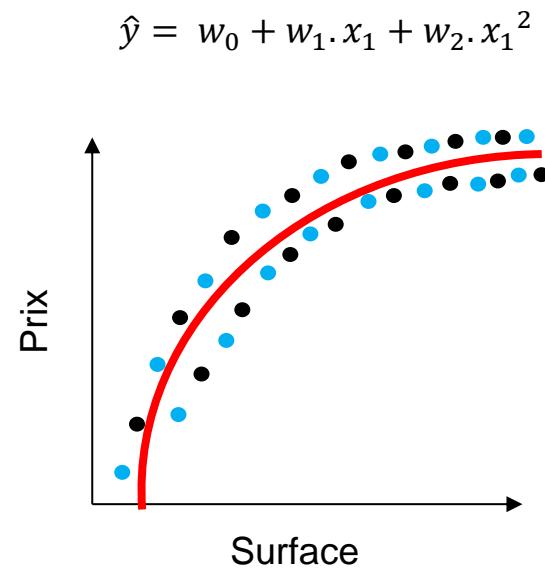
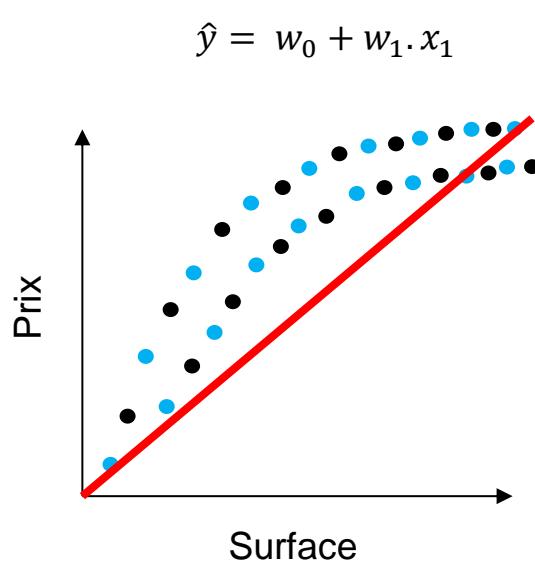
Erreurs sur le jeu d'entraînement : Nulle

Erreurs sur le jeu de test : Moyenne





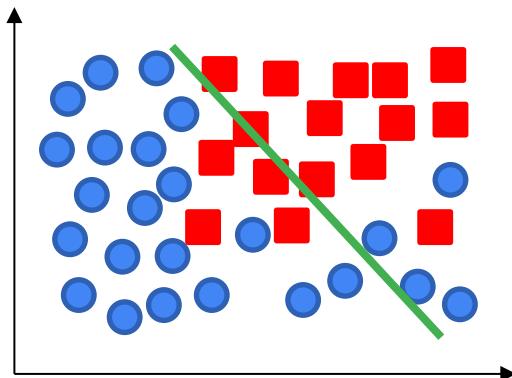
Complexité du modèle



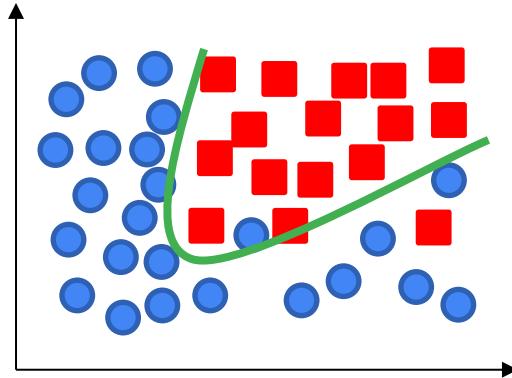
- Jeu d'entraînement
- Jeu de test



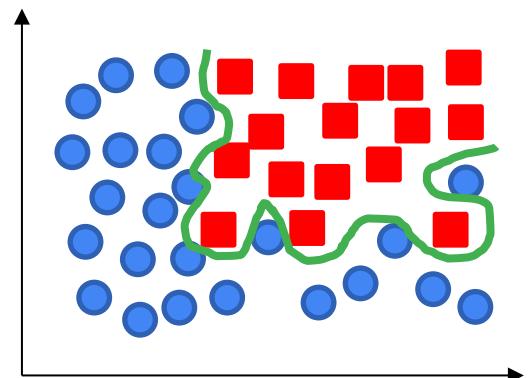
Pour la classification



Sous-entraînement



Entraînement correct



Sur-entraînement

La pénalisation L2

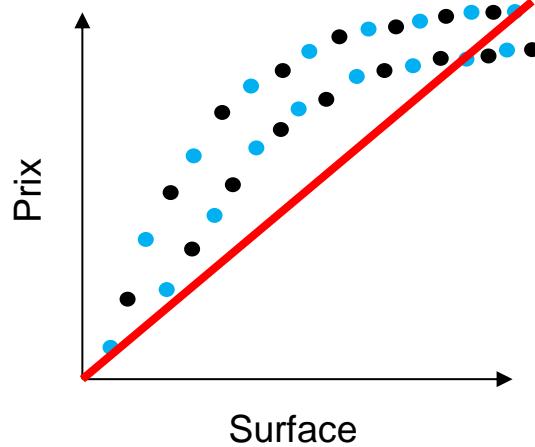


Partie 6 : Sur et sous entraînement

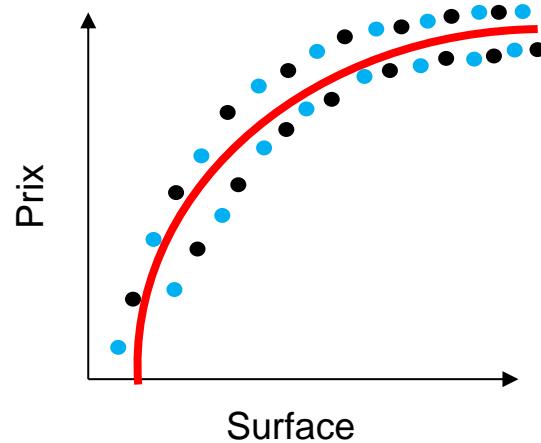


Complexité du modèle

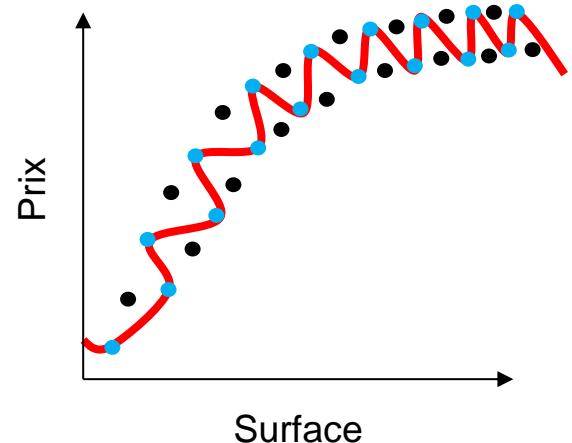
$$\hat{y} = w_0 + w_1 \cdot x_1$$



$$\hat{y} = w_0 + w_1 \cdot x_1 + w_2 \cdot x_1^2$$



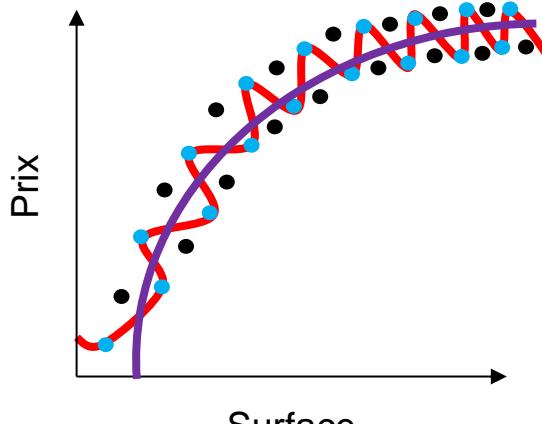
$$\hat{y} = w_0 + w_1 \cdot x_1 + w_2 \cdot x_1^2 + w_3 \cdot x_1^3 + w_4 \cdot x_1^4 + \dots$$



- Jeu d'entraînement
- Jeu de test



Pénalisation des paramètres



$$\hat{y} = w_0 + w_1 \cdot x_1 + w_2 \cdot x_1^2 + w_3 \cdot x_1^3 + w_4 \cdot x_1^4 + \dots$$

Pénalisation des paramètres

$$\min_w J(w) = \frac{1}{2m} \sum_{i=1}^m (\hat{y}^{(i)} - y^{(i)})^2 + 1000 \cdot w_3 + 1000 \cdot w_4 + \dots$$

Minimiser l'erreur
de prédition

Minimiser la valeur des
paramètres w_3, w_4, \dots



Régression Ridge ou pénalisation L2

- Un modèle avec des paramètres plus homogène est moins sujet au sur-entraînement.

Paramètre de régularisation

$$J(w) = \frac{1}{2m} \left[\sum_{i=1}^m (\hat{y}^{(i)} - y^{(i)})^2 + \lambda \sum_{j=1}^n w_j^2 \right]$$

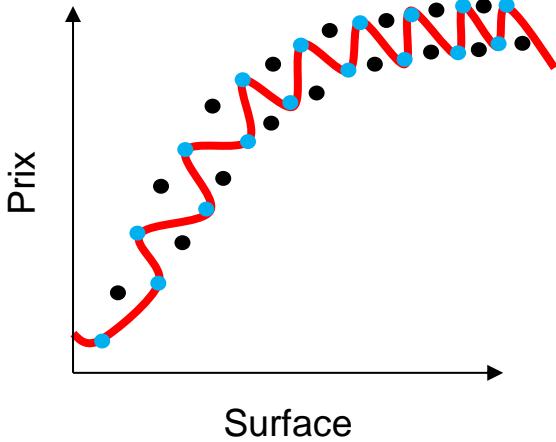
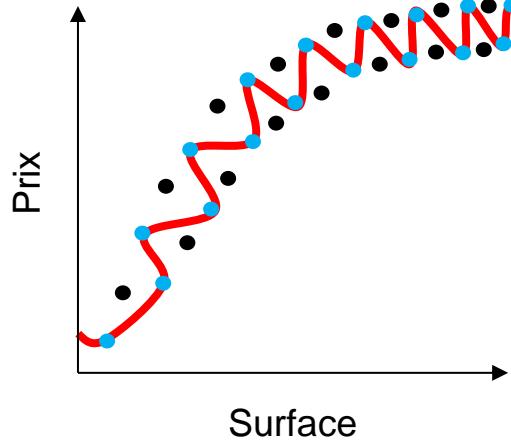
Régularisation



Impact du coefficient de régularisation

λ trop petit

$$J(w) = \frac{1}{2m} \left[\sum_{i=1}^m (\hat{y}^{(i)} - y^{(i)})^2 + \lambda \sum_{j=1}^n w_j^2 \right]$$

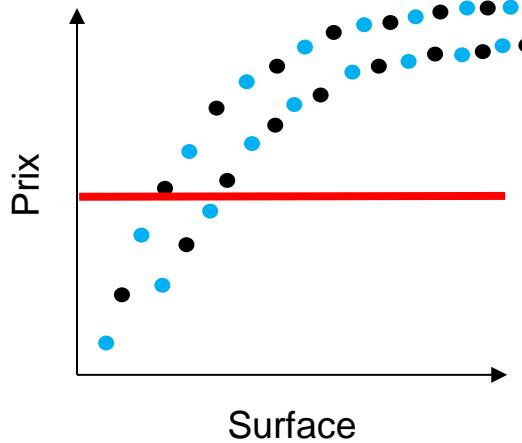
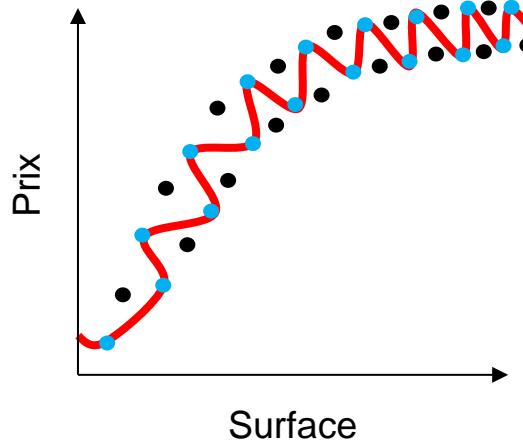




Impact du coefficient de régularisation

λ trop grand

$$J(w) = \frac{1}{2m} \left[\sum_{i=1}^m (\hat{y}^{(i)} - y^{(i)})^2 + \lambda \sum_{j=1}^n w_j^2 \right]$$



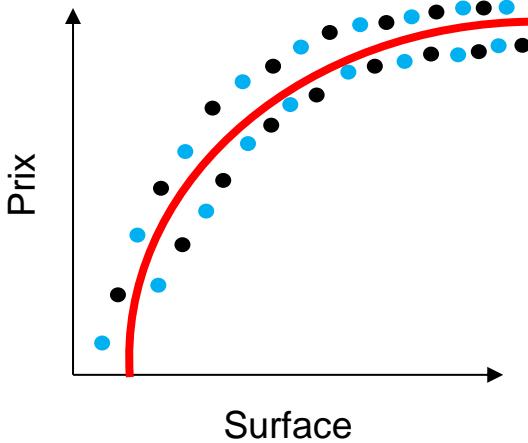
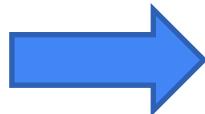
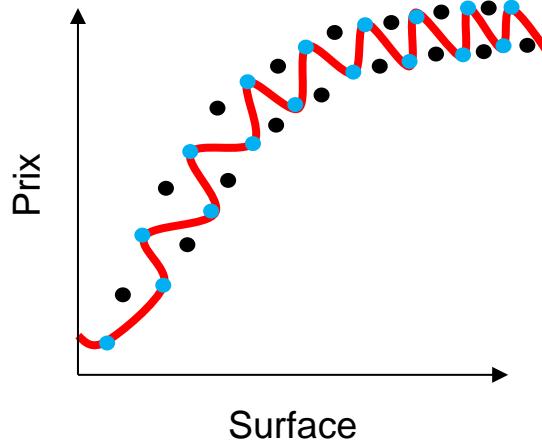


Impact du coefficient de régularisation

[0.01, ..., 0.1, ..., 0.5]

λ adéquat

$$J(w) = \frac{1}{2m} \left[\sum_{i=1}^m (\hat{y}^{(i)} - y^{(i)})^2 + \lambda \sum_{j=1}^n w_j^2 \right]$$



Le drop out

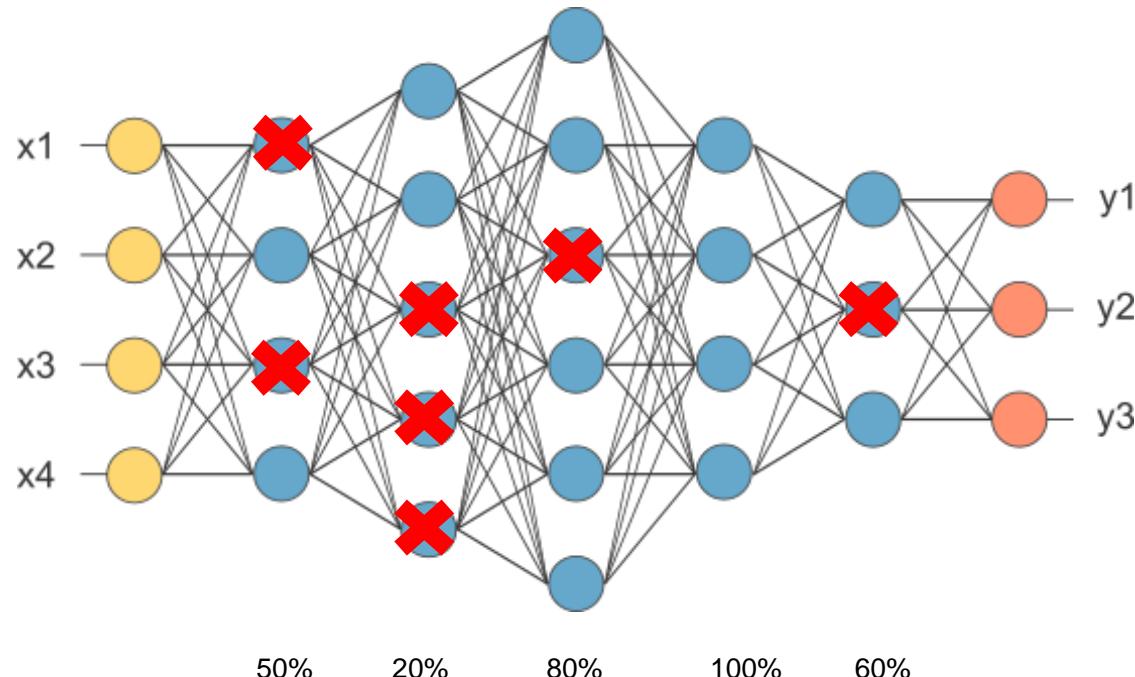


Partie 6 : Sur et sous entraînement



Le drop out

En utilisant le drop-out, vous ne pouvez pas compter sur une seule fonction, vous devez donc répartir les poids.



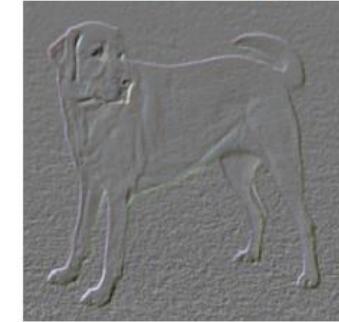
La data augmentation



Partie 6 : Sur et sous entraînement



La data augmentation



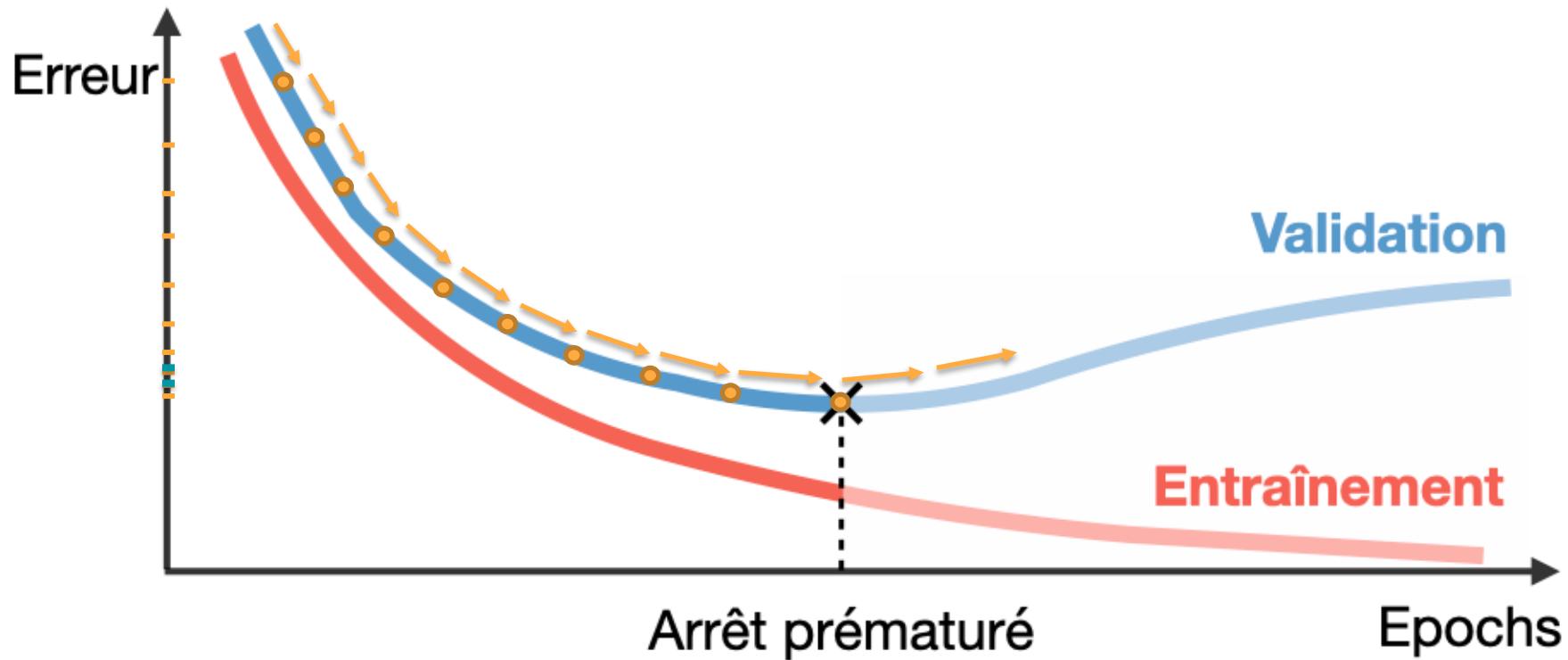
L'early stopping



Partie 6 : Sur et sous entraînement



L'early stopping



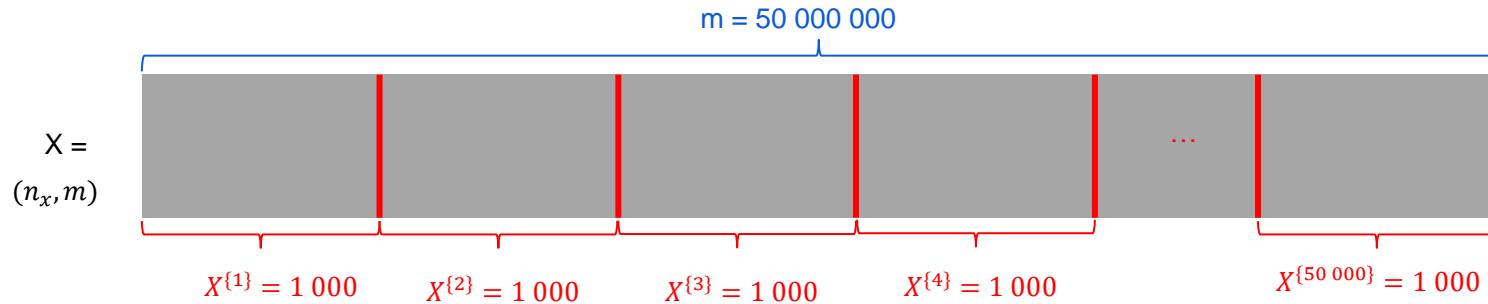
Mini batch



Partie 6 : Sur et sous entraînement



Batch vs mini-batch gradient descent



For epoch = 1, ..., 1 000

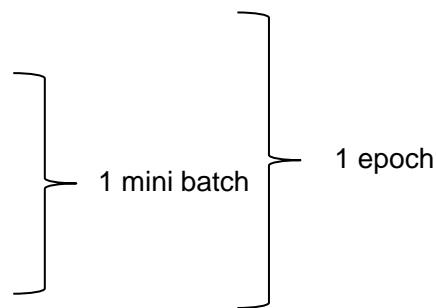
For i = 1, ..., 50 000

Forward prop $X^{(i)}$

Calcul du coût $X^{(i)}$

Back prop $X^{(i)}$

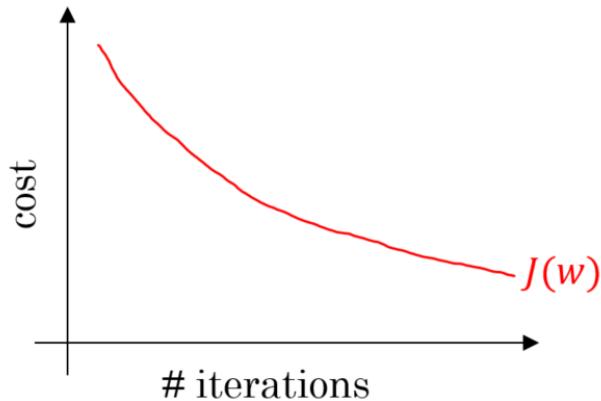
Mise à jour de w et b





Batch vs mini-batch gradient descent

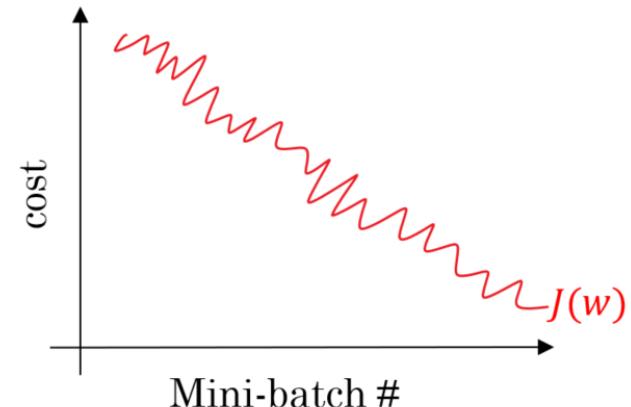
Batch gradient descent



Le batch est trop grand :

Trop long par itération

Mini-batch gradient descent



Le batch est trop petit :

Perte de la vectorisation

Utilisez une puissance de 2 et assurez-vous que votre mini batch correspond à la mémoire de votre GPU.

Batch normalisation



Partie 6 : Sur et sous entraînement



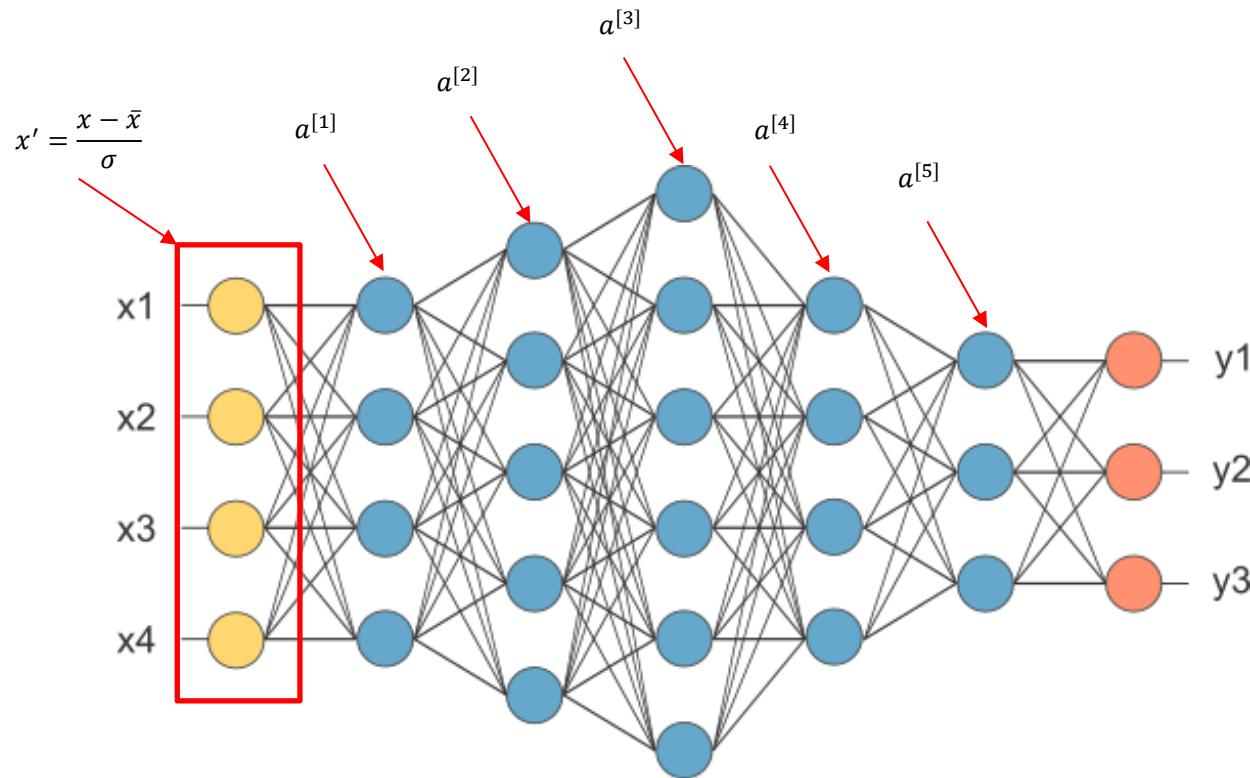
Normalisation des données d'entrées

$$x' = \frac{x - \bar{x}}{\sigma}$$

Calculer \bar{x} et σ avec votre ensemble d'entraînement et sauvegardez-les pour les appliquer à l'ensemble de dev et de test.



Modèle dense





Disparition et explosion de gradient

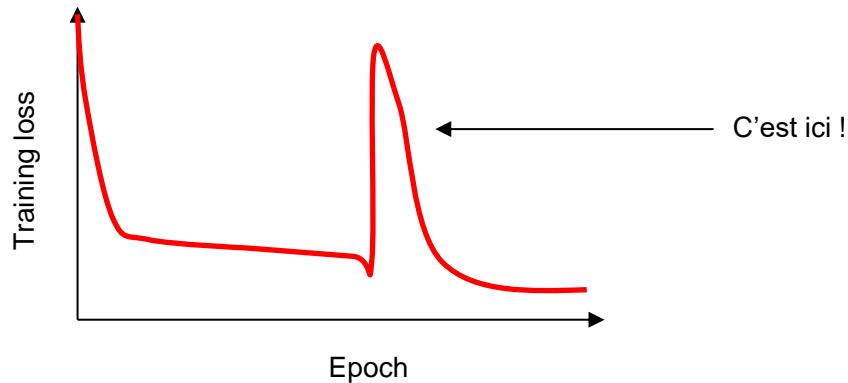
- Lorsque vous construisez un réseau neuronal très profond, votre gradient dans la dernière couche cachée peut être très important ou proche de zéro.
- Il s'agit d'un problème énorme car, avec ce problème, votre modèle ne peut pas apprendre correctement.



Les explosions de gradient

Les explosions de gradients sont faciles à détecter

Courbe d'apprentissage instable



Les gradients peuvent être trop grands et contenir des NaNs.
et vous vous retrouvez avec des NaN dans les poids.



Batch normalization

Pour chaque couche

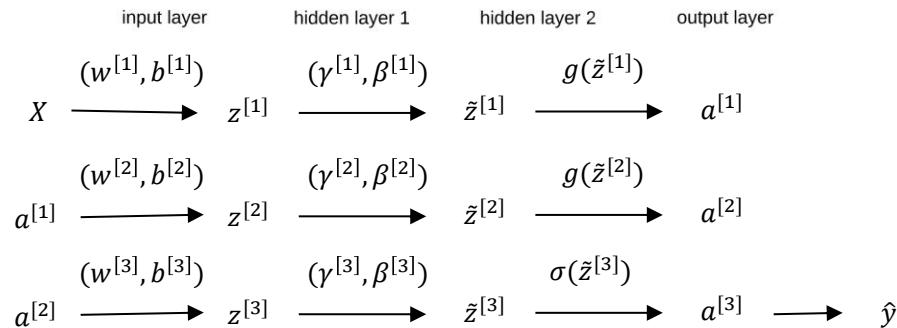
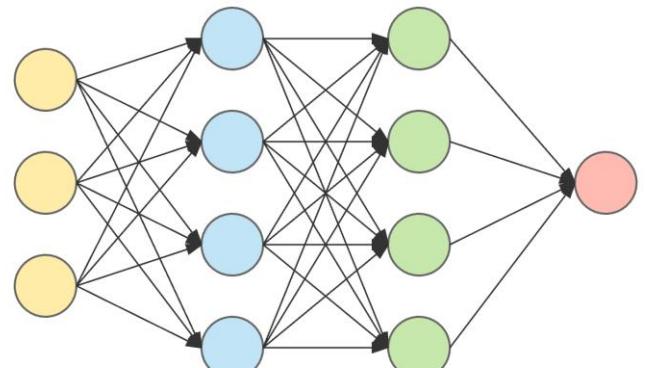
$$\mu^l = \frac{1}{m} \sum_i z_i^l$$

$$\sigma^{l^2} = \frac{1}{m} \sum_i (z_i^l - \mu^l)^2$$

$$Z_{norm}^l = \frac{z^l - \mu^l}{\sqrt{\sigma^{l^2} - \epsilon}}$$

$$\tilde{z}^l = \gamma Z_{norm}^l + \beta$$

Couche de batch normalisation



L'entraînement, un processus itératif



Partie 6 : Sur et sous entraînement



Comment créer son premier modèle ?

Nombre de couches

Nombre de neurones

Learning rates

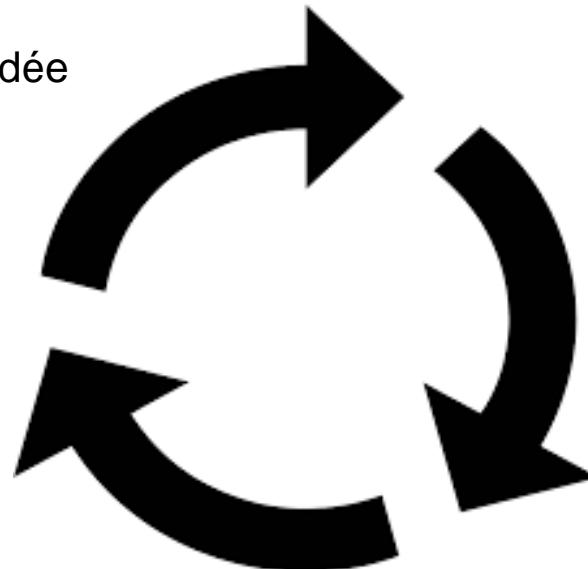
Fonctions d'activations

...

Idée

Code

Résultats





Commencer simple

- Tester le modèle le plus simple possible
- Complexifier pour chercher le sur-apprentissage
- Appliquer une régularisation pour chercher la complexité idéale



Tuning process

- α
- Nombre de couches
- Nombre de neurones
- Taille du Mini-batch

Deep learning avec Pytorch

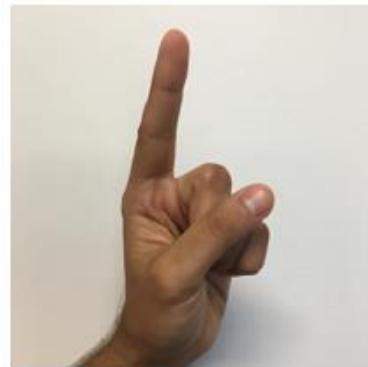
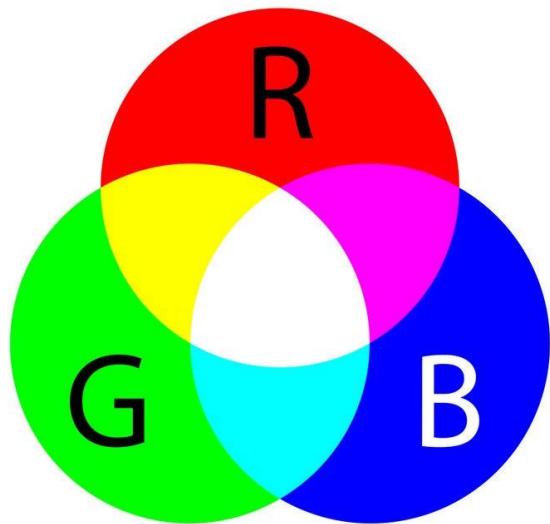
Partie 7 : Computer vision



Présenté par **Morgan Gautherot**



Une image en informatique



=

Blue				
Green	123	94	83	2
Red	123	94	83	4
123	94	83	2	92
34	44	187	92	4
34	76	232	124	4
67	83	194	202	

(4x4x3)



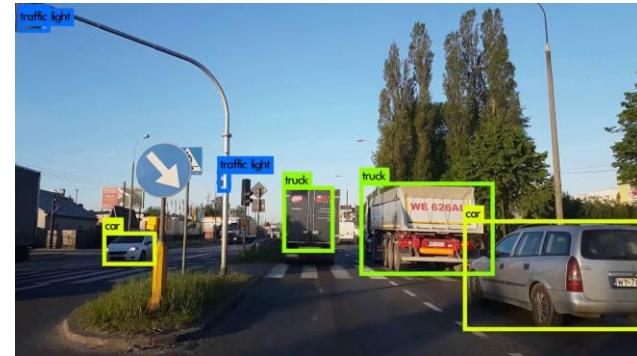
Computer vision

Classification d'images



→ Cat ? (0/1)

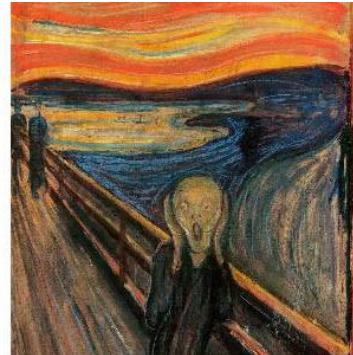
Détection d'objets



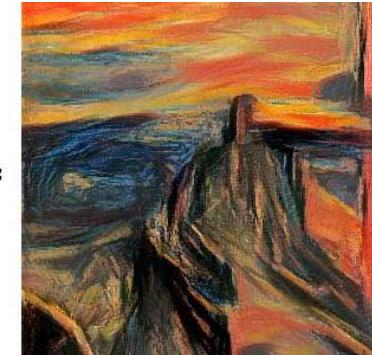
Transfert de style



+



=





Un grand nombre de caractéristiques



64 x 64



1000 x 1000

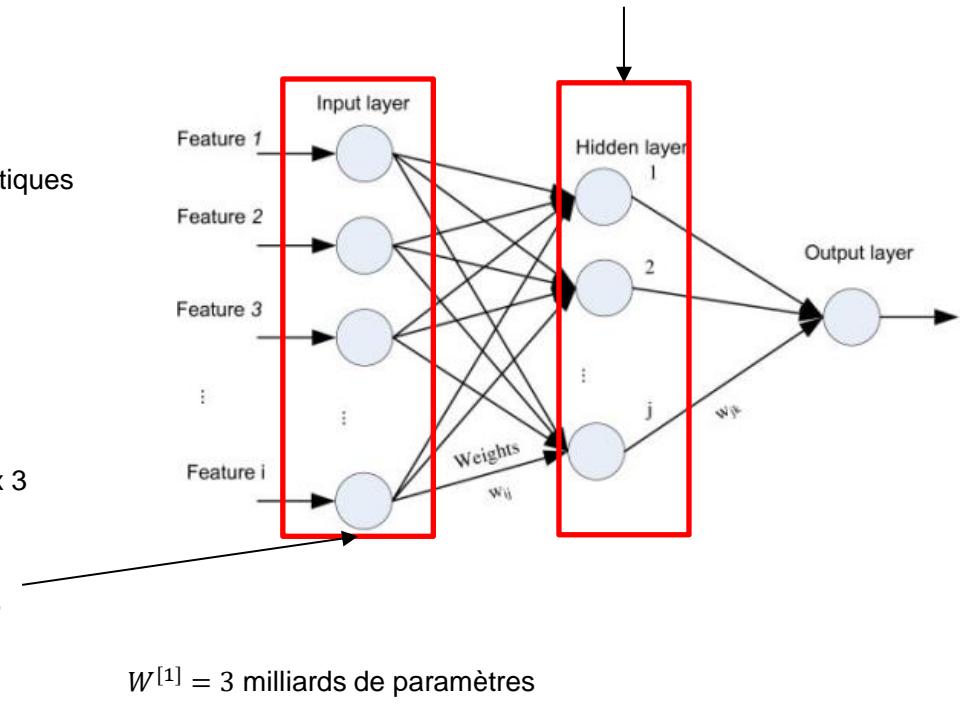
64 x 64 x 3

12288 caractéristiques

1000 x 1000 x 3

3 millions de caractéristiques

1 000 neurones



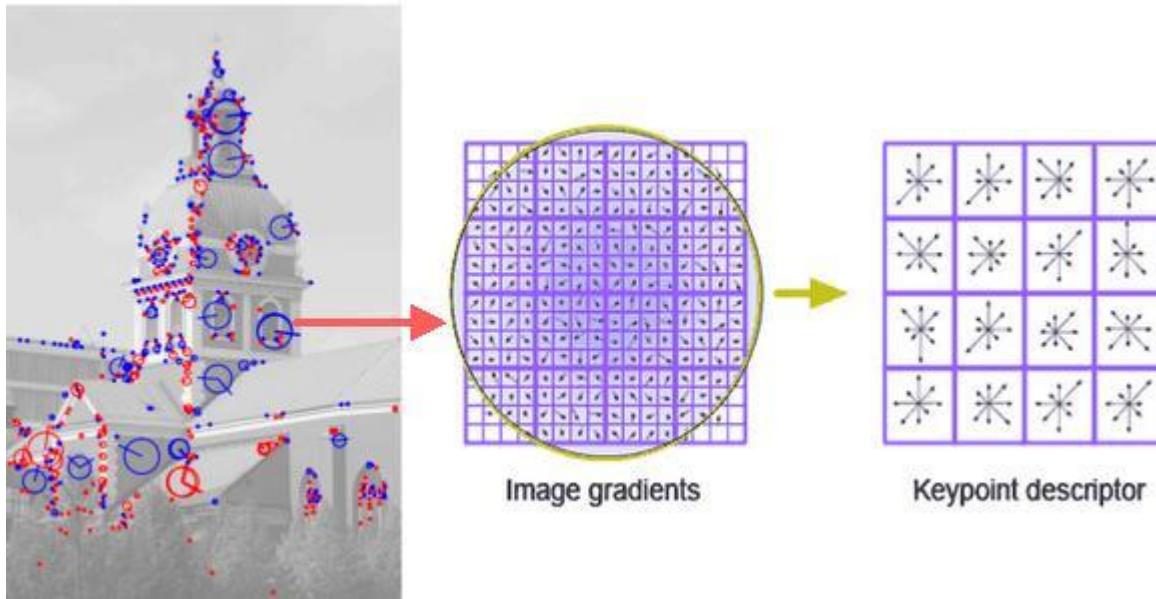
Extraction de caractéristiques



Partie 7 : Computer vision

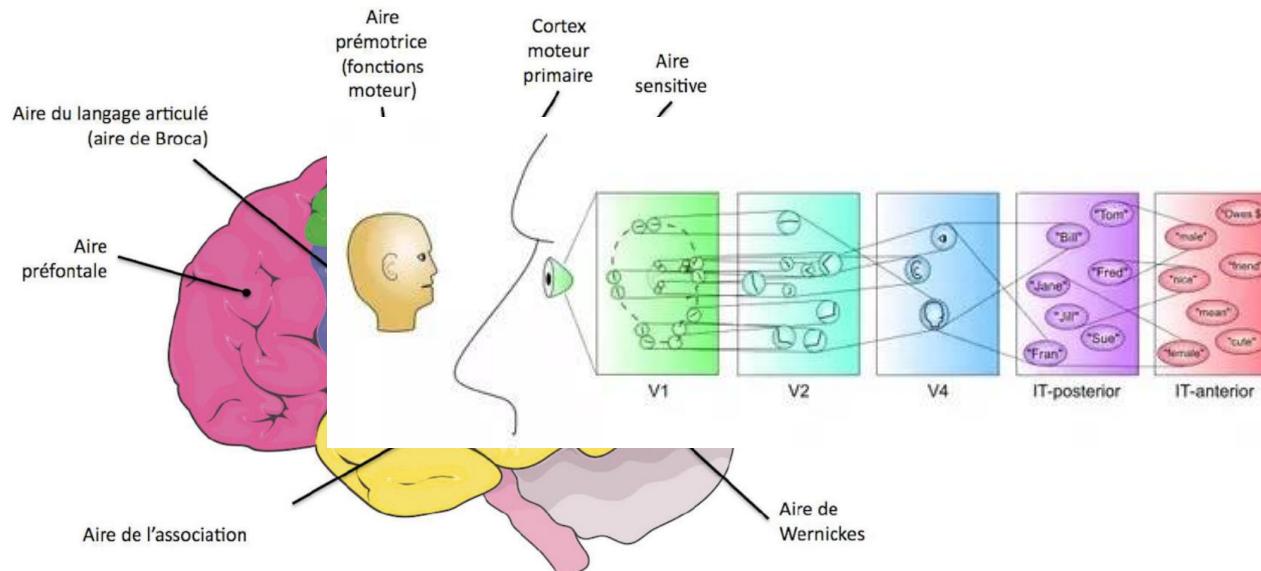


Extraction de caractéristiques



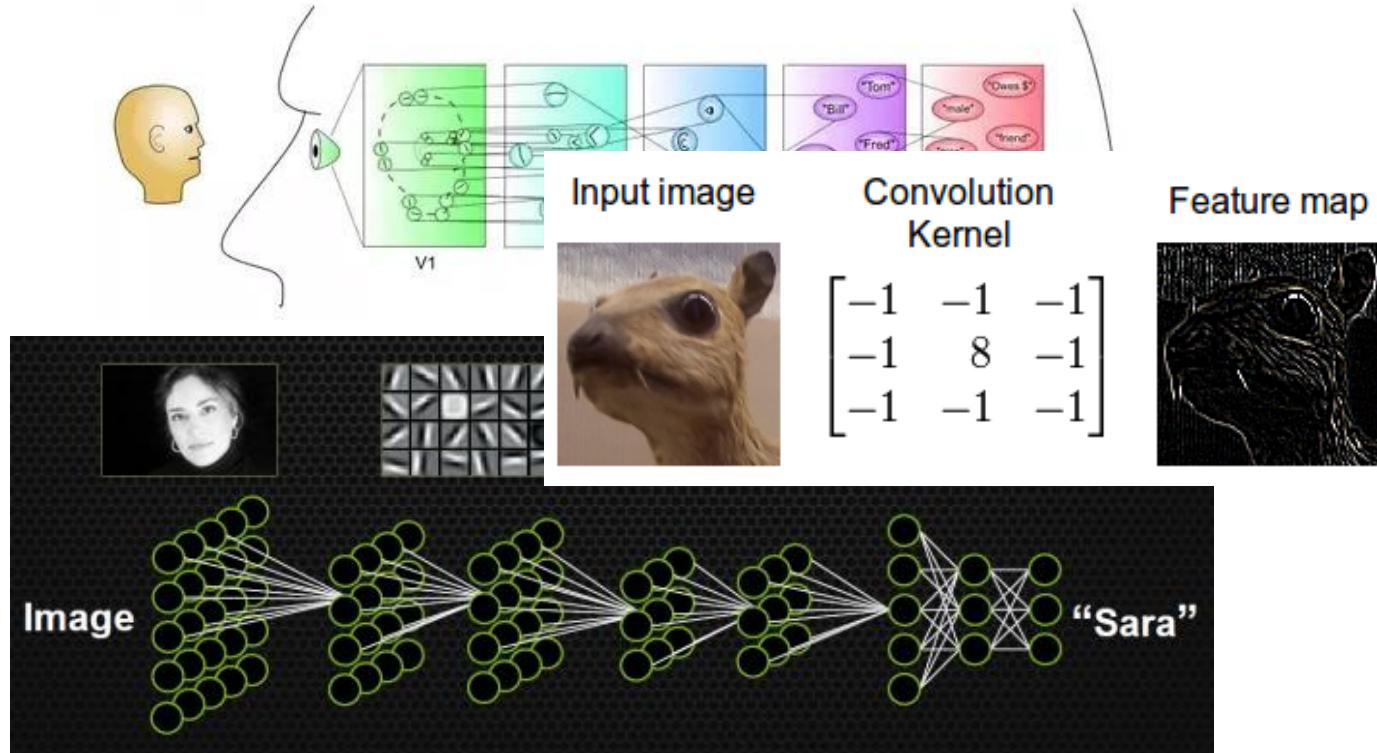


Comment voyons nous ?



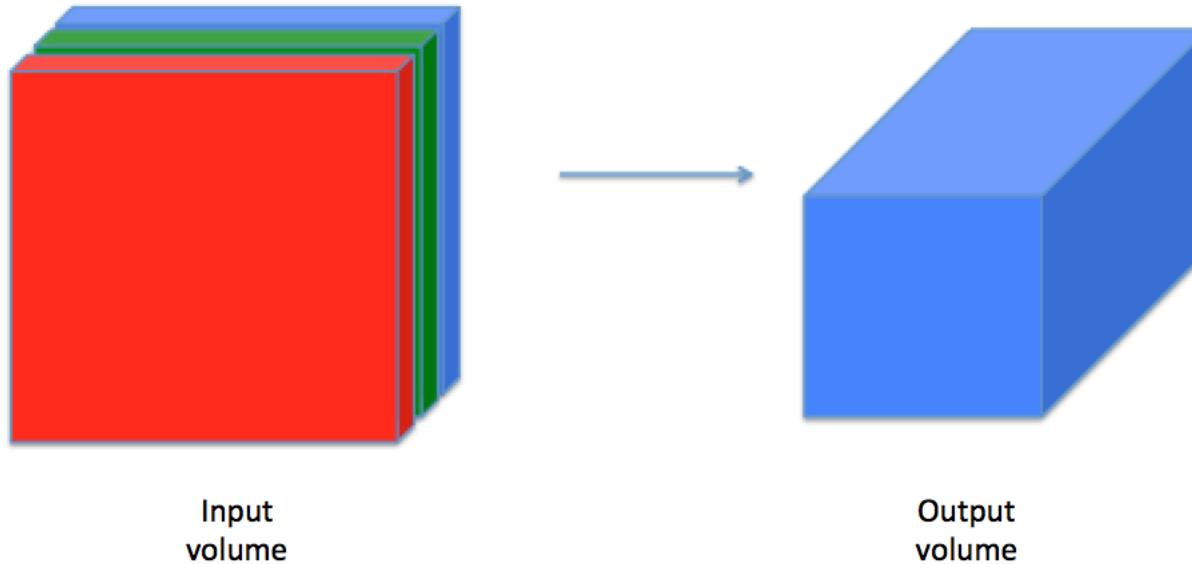


De la biologie à l'informatique





Représentation vectorielle de l'image d'entrée



La convolution



Partie 7 : Computer vision



Les filtres

Figure 16.152. Edge detect

$$\begin{bmatrix} 0 & 1 & 0 \\ 1 & -4 & 1 \\ 0 & 1 & 0 \end{bmatrix}$$

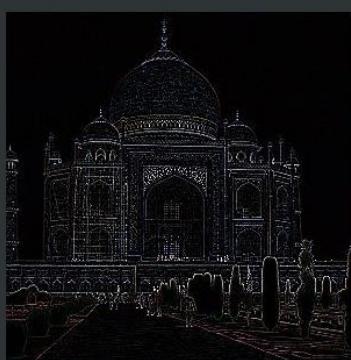


Figure 16.153. Emboss

$$\begin{bmatrix} -2 & -1 & 0 \\ -1 & 1 & 1 \\ 0 & 1 & 2 \end{bmatrix}$$





Et mathématiquement ?

3	0	1	2	7	4
1	5	8	9	3	1
2	7	2	5	1	3
0	1	3	1	7	8
4	2	1	6	2	8
2	4	5	2	3	9

(6, 6)

Convolution

*

w_1	w_2	w_3
w_4	w_5	w_6
w_7	w_8	w_9

1	0	-1
1	0	-1
1	0	-1

(3, 3)

Filtre

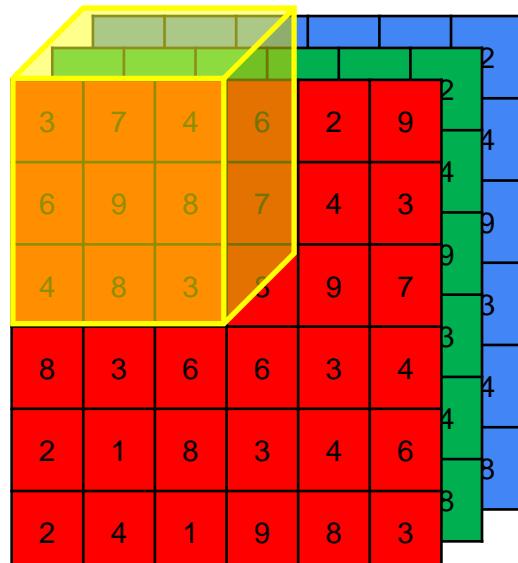
=

-5	-4	0	8
-10	-2	2	3
0	-2	-4	-7
-3	-2	-3	-16

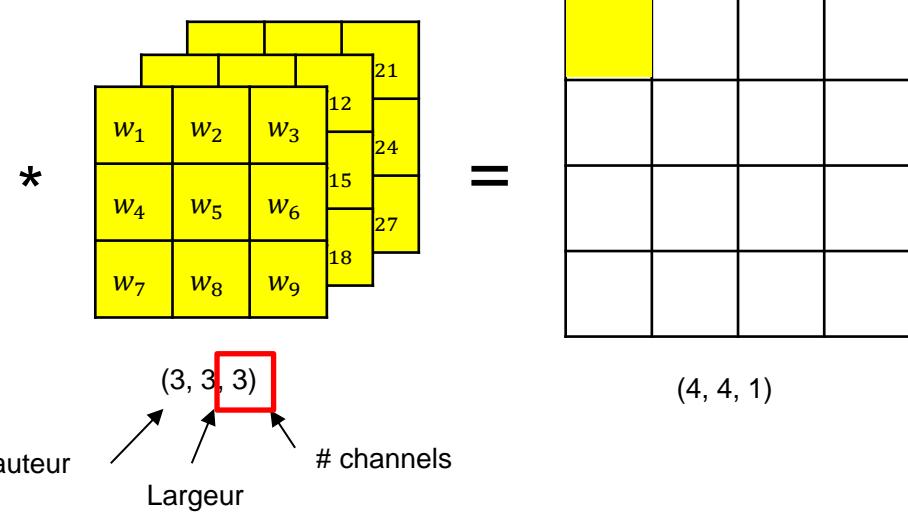
(4, 4)



Convolution pour les images RGB

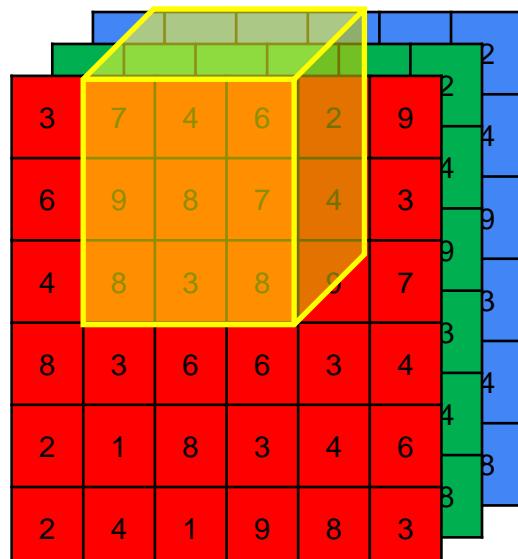


(6, 6, 3)
Hauteur Largeur # channels

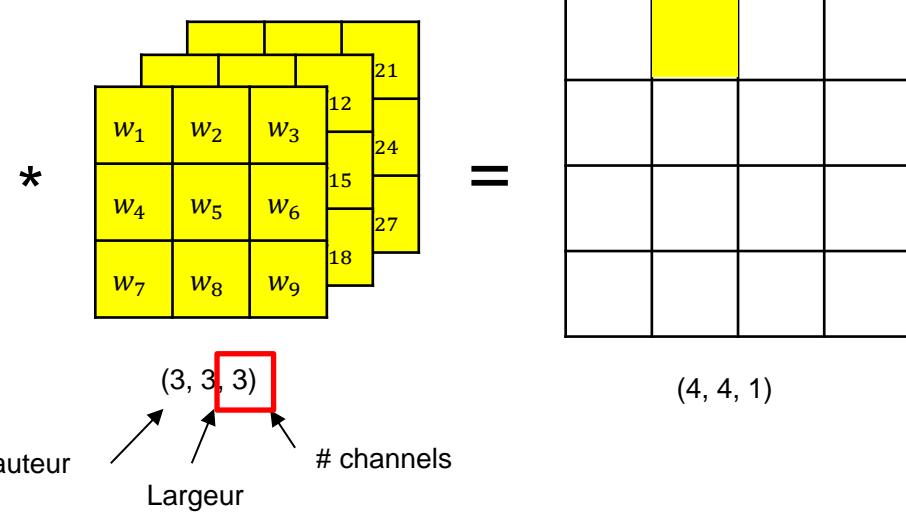




Convolution pour les images RGB

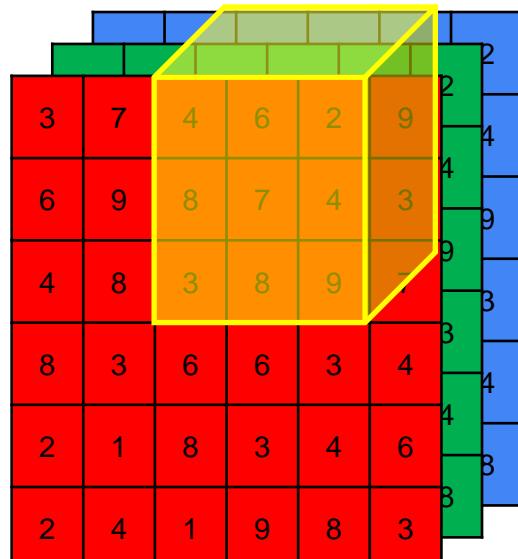


Hauteur
Largeur
channels

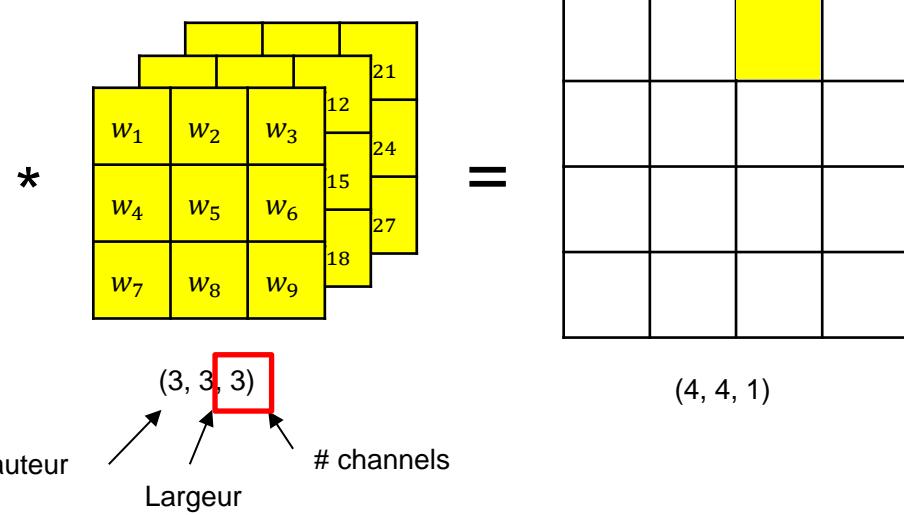




Convolution pour les images RGB



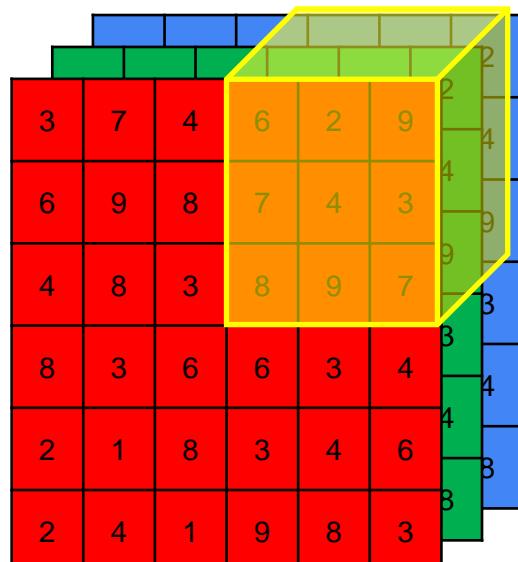
(6, 6, 3)
Hauteur Largeur # channels



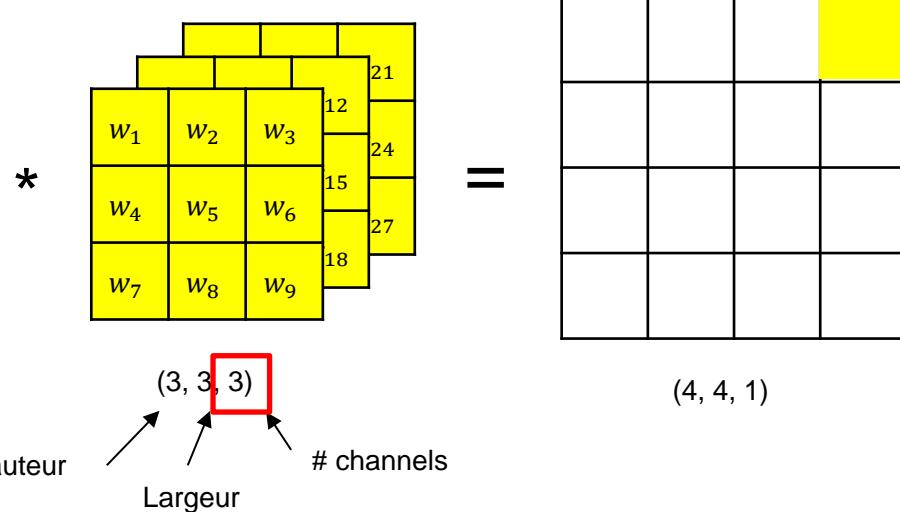
(4, 4, 1)



Convolution pour les images RGB

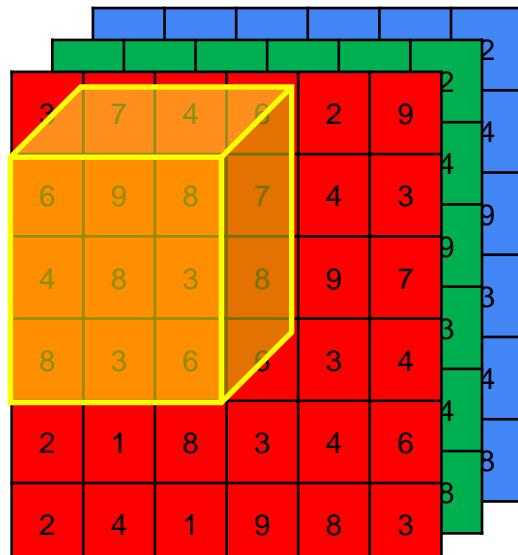


Hauteur
Largeur
channels

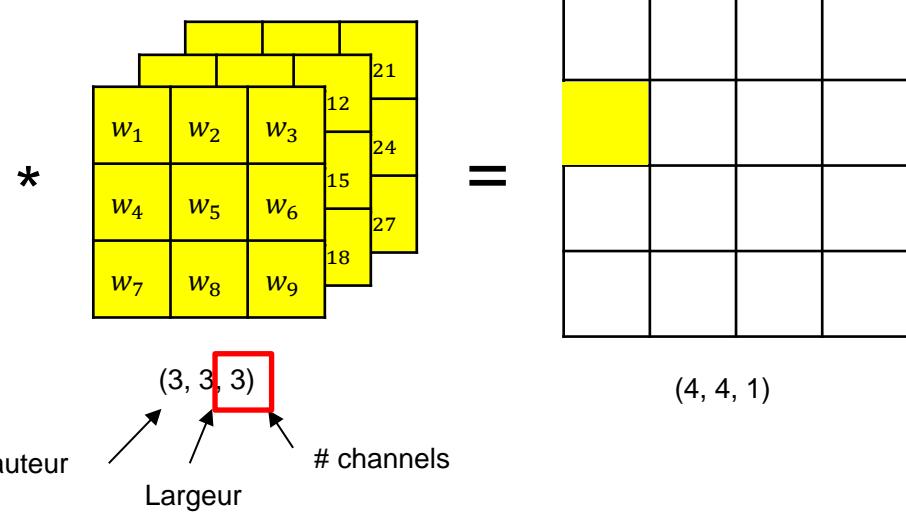




Convolution pour les images RGB



(6, 6, 3)
Hauteur Largeur # channels





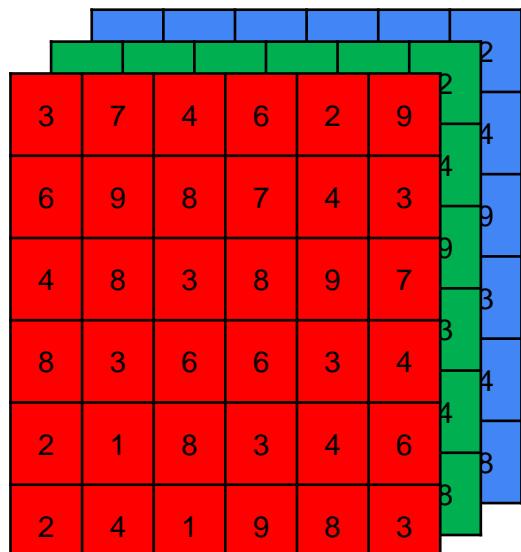
Plusieurs filtres



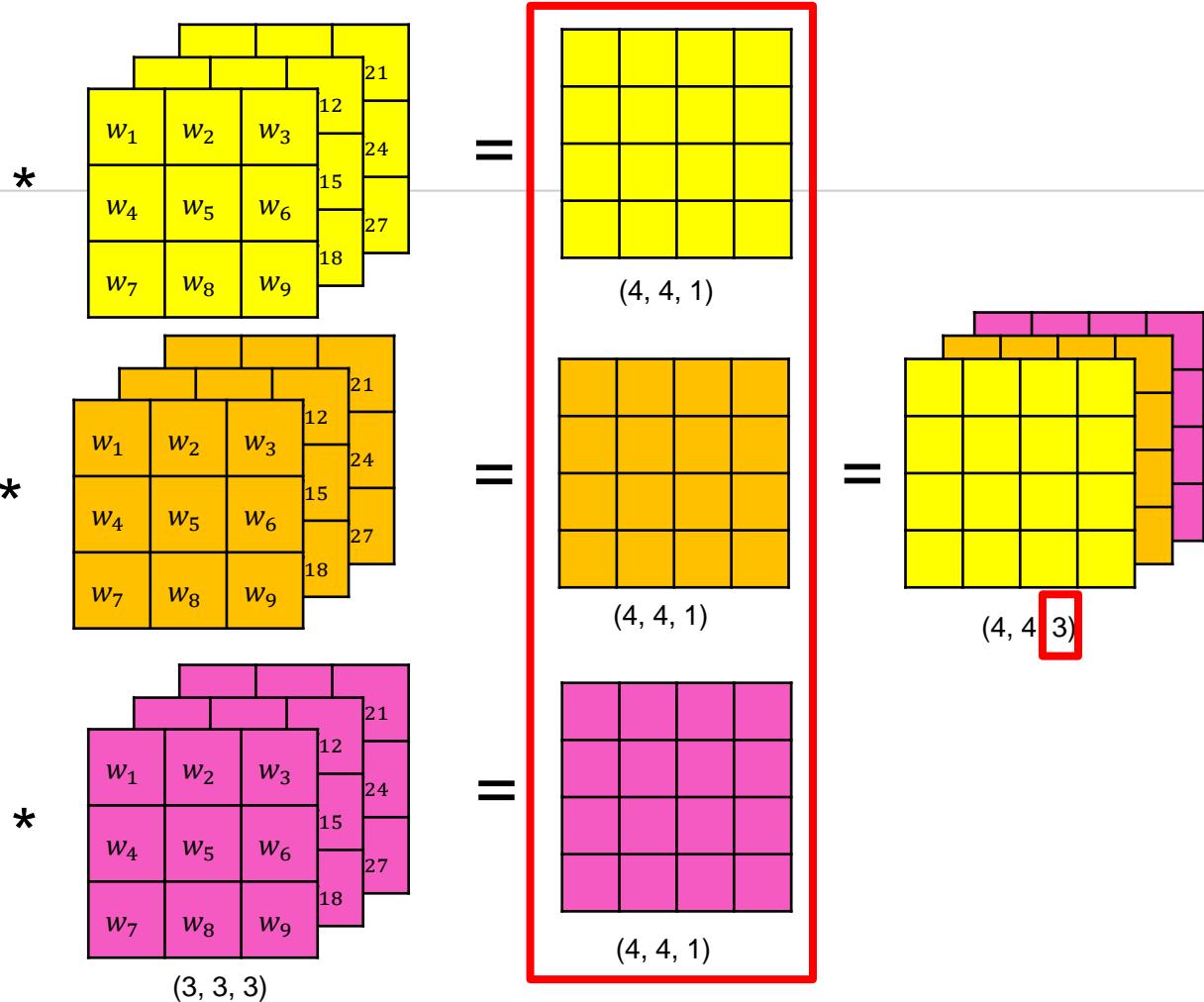
Input



Plusieurs filtres



(6, 6, 3)



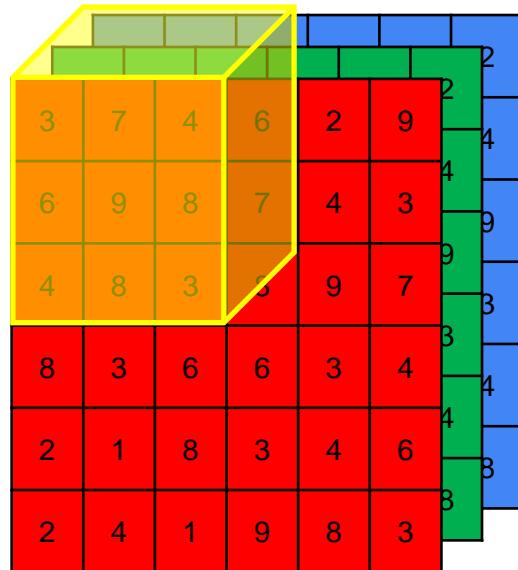
La convolution informatiquement



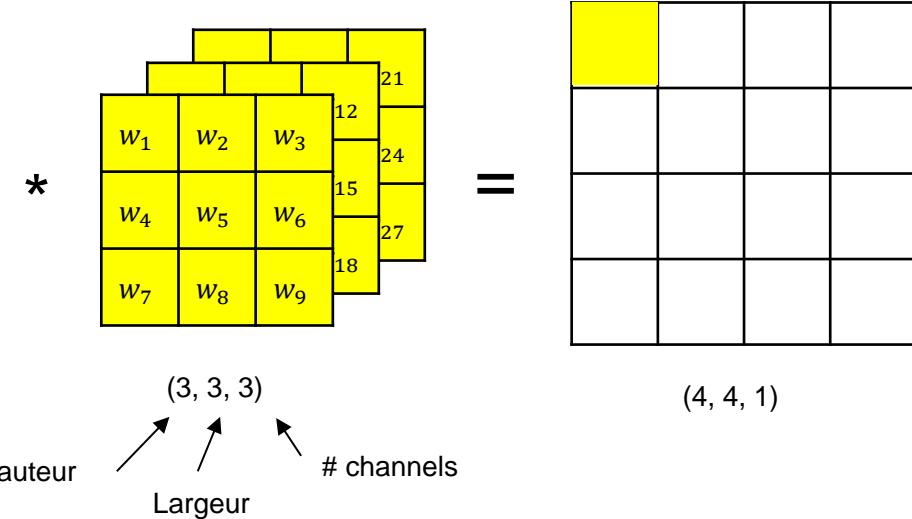
Partie 7 : Computer vision



Convolution pour les images RGB

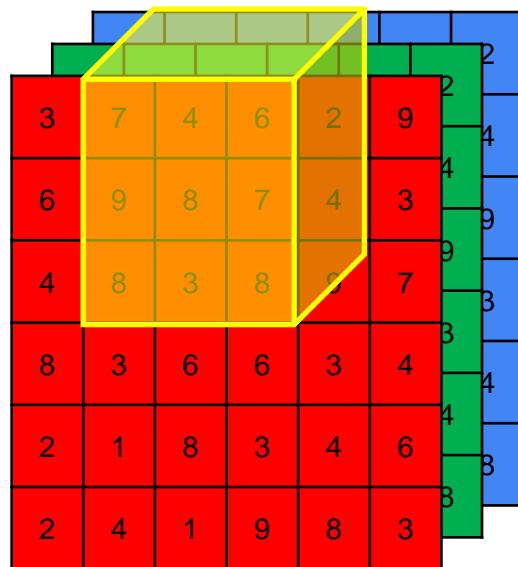


(6, 6, 3)
Hauteur
Largeur
channels

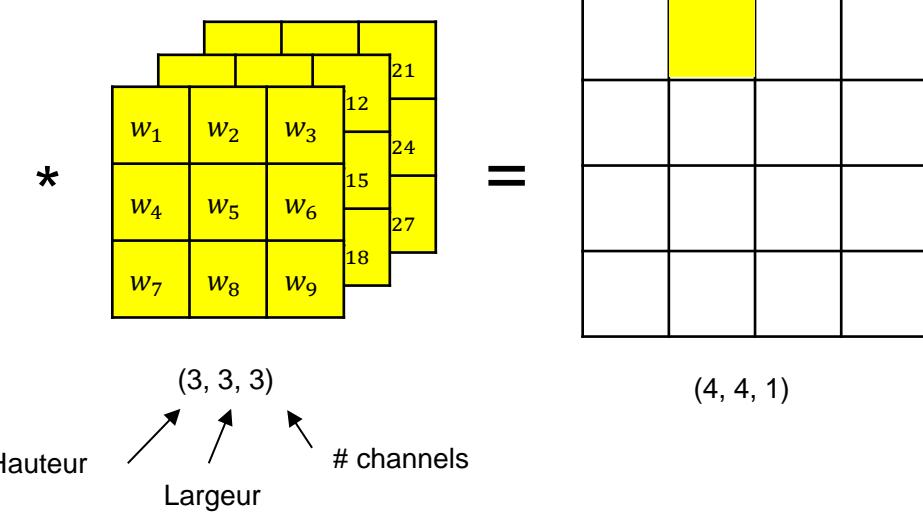




Convolution pour les images RGB

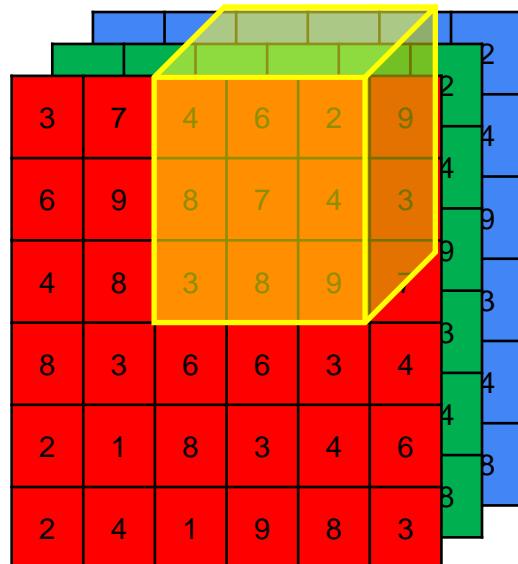


(6, 6, 3)
Hauteur Largeur # channels

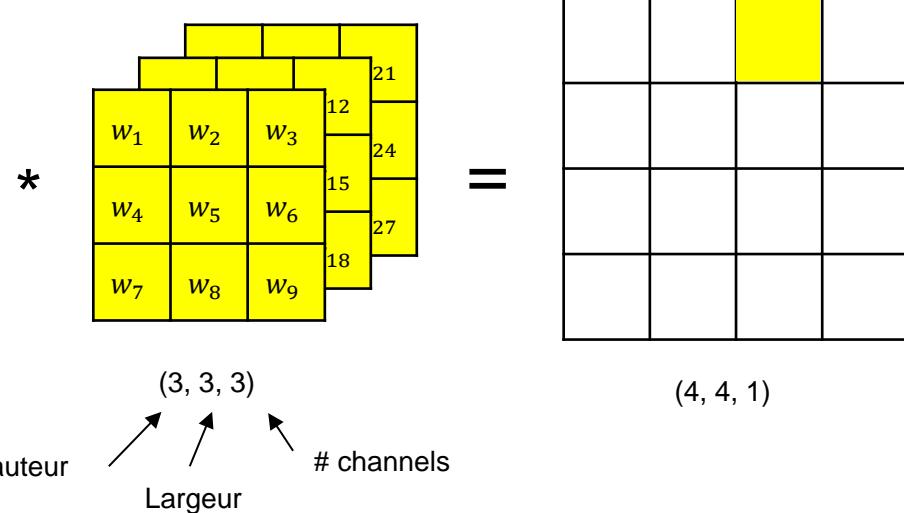




Convolution pour les images RGB

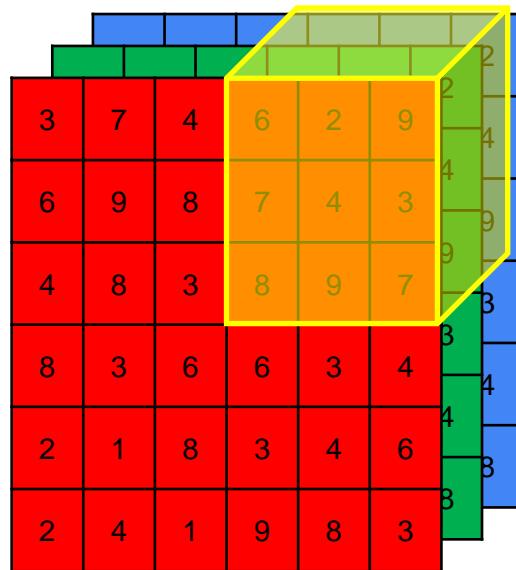


(6, 6, 3)
Hauteur Largeur # channels

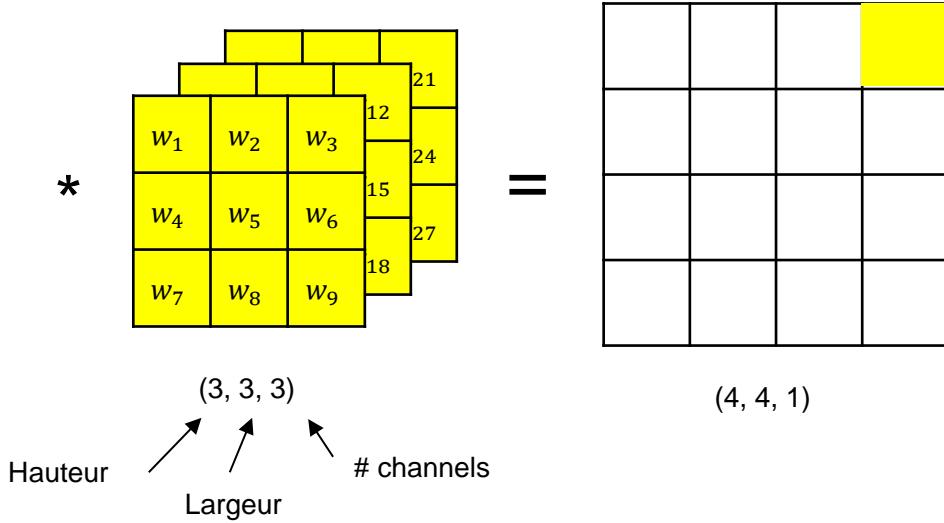




Convolution pour les images RGB

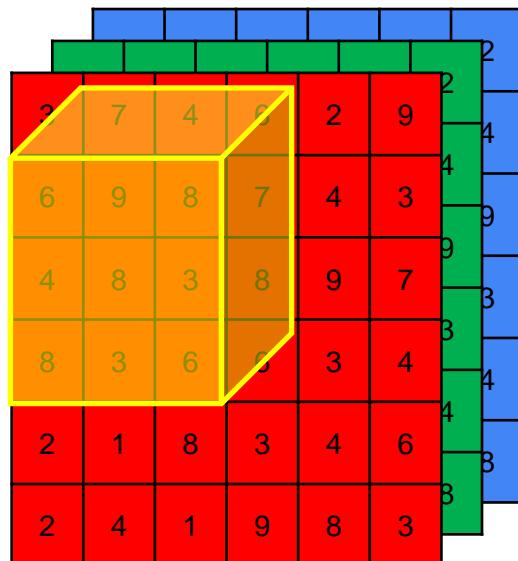


The diagram illustrates the dimensions of a tensor. At the top, the values $(6, 6, 3)$ are shown in parentheses. Three arrows point from the labels "Hauteur", "Largeur", and "# channels" below to the corresponding numbers in the tuple.

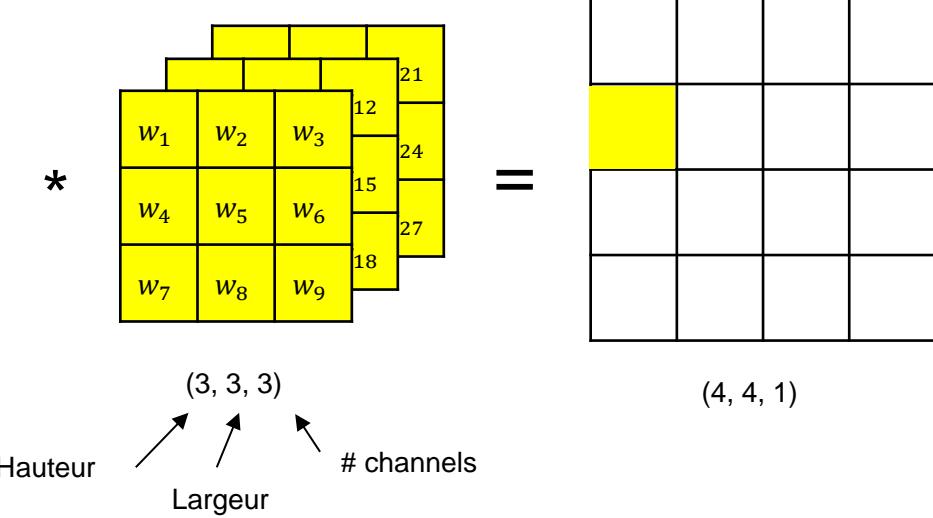




Convolution pour les images RGB



(6, 6, 3)
Hauteur Largeur # channels





Convolution

4	5	8	7
1	8	8	8
3	6	6	4
6	5	7	8

4×4

*

1	4	1
1	4	3
3	3	1

3×3

=

122	148
126	134

2×2



Convolutional Matrix

1	4	1
1	4	3
3	3	1

3 x 3

=

1	4	1	0	1	4	3	0	3	3	1	0	0	0	0	0	0
0	1	4	1	0	1	4	3	0	3	3	1	0	0	0	0	0
0	0	0	0	1	4	1	0	1	4	3	0	3	3	1	0	0
0	0	0	0	0	1	4	1	0	1	4	3	0	3	3	1	0

4 x 16

4	5	8	7
1	8	8	8
3	6	6	4
6	5	7	8

4 x 4

=

4
5
8
7
1
8
8
3
6
6
4
6
5
7
8

16 x 1



Convolutional Matrix

1	4	1	0	1	4	3	0	3	3	1	0	0	0	0	0	0
0	1	4	1	0	1	4	3	0	3	3	1	0	0	0	0	0
0	0	0	0	1	4	1	0	1	4	3	0	3	3	1	0	0
0	0	0	0	0	1	4	1	0	1	4	3	0	3	3	1	0

4 x 16

4
5
8
7
1
8
8
8
3
6
6
4
6
5
7
8

16 x 1

122
148
126
134

4 x 1

122	148
126	134

2 x 2

Le padding



Partie 7 : Computer vision



Sans padding

3	0	1	2	7	4
1	5	8	9	3	1
2	7	2	5	1	3
0	1	3	1	7	8
4	2	1	6	2	8
2	4	5	2	3	9

(6, 6)

*

Filtre

w_1	w_2	w_3
w_4	w_5	w_6
w_7	w_8	w_9

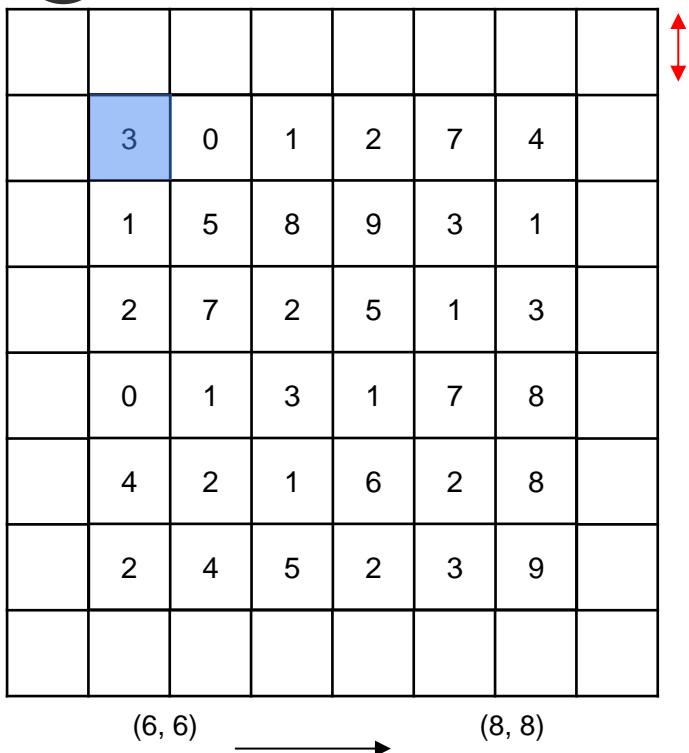
(3, 3)

=

(4, 4)



Avec padding



*

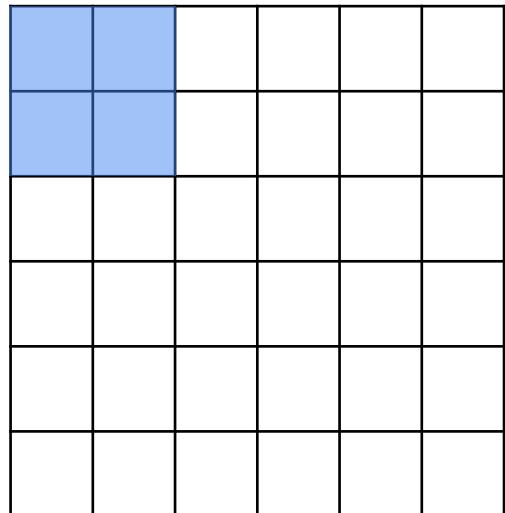
Padding
 $p = 1$

Filtre

w_1	w_2	w_3
w_4	w_5	w_6
w_7	w_8	w_9

$(3, 3)$

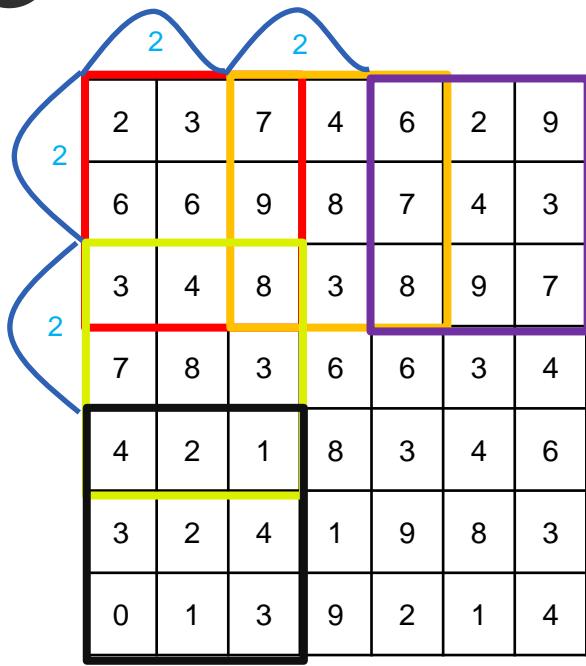
=



$(6, 6)$



Stride



Stride
 $s = 2$

filter

*

w_1	w_2	w_3
w_4	w_5	w_6
w_7	w_8	w_9

=

91	100	83
69	91	127
44	72	74

(3, 3)

(3, 3)

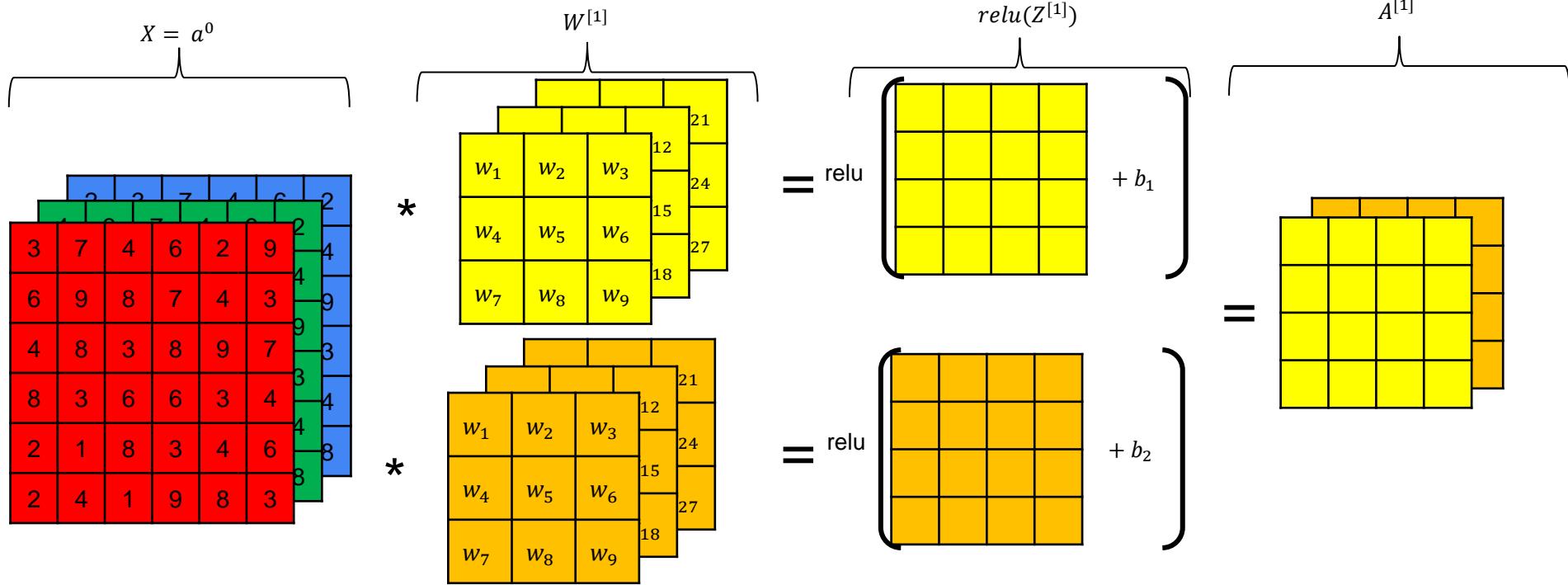
Une couche de convolution



Partie 7 : Computer vision



Une couche de convolution



Le pooling



Partie 7 : Computer vision



Maintenir de l'invariance

0	0	0	0
0	0	0	0
0	0	1	0
0	0	0	1

$$\begin{matrix} * & \begin{matrix} 1 & 0 \\ 0 & 1 \end{matrix} \end{matrix}$$

$$= \begin{matrix} \begin{matrix} 0 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 2 \end{matrix} \end{matrix}$$

Sortie

1	0	0	0
0	1	0	0
0	0	0	0
0	0	0	0

$$\begin{matrix} * & \begin{matrix} 1 & 0 \\ 0 & 1 \end{matrix} \end{matrix}$$

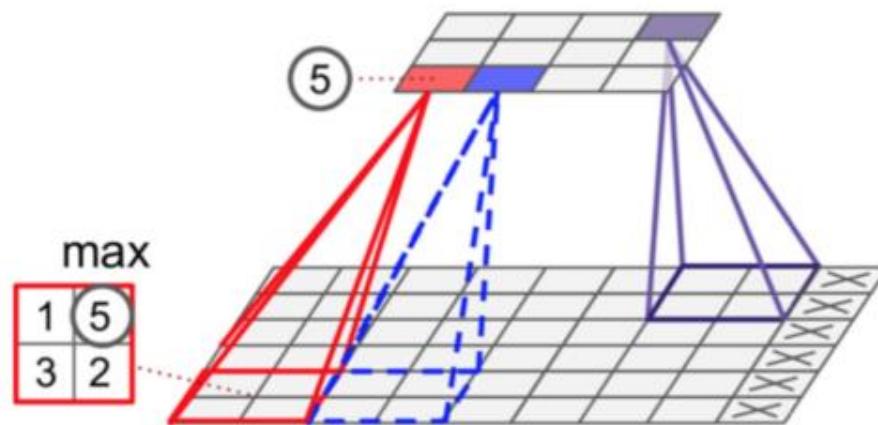
$$= \begin{matrix} \begin{matrix} 2 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{matrix} \end{matrix}$$

Sortie

Max = 2
↑
Ne change pas
↓
Max = 2



Application sur une image



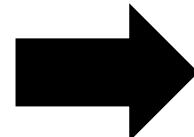


Pooling layer: Max pooling



3	0	1	2	7	4
1	5	8	9	3	1
2	7	2	5	1	3
0	1	3	1	7	8
4	2	1	6	2	8
2	4	5	2	3	9

(6, 6)



5	9	7
7	5	8
4	6	9

(3, 3)

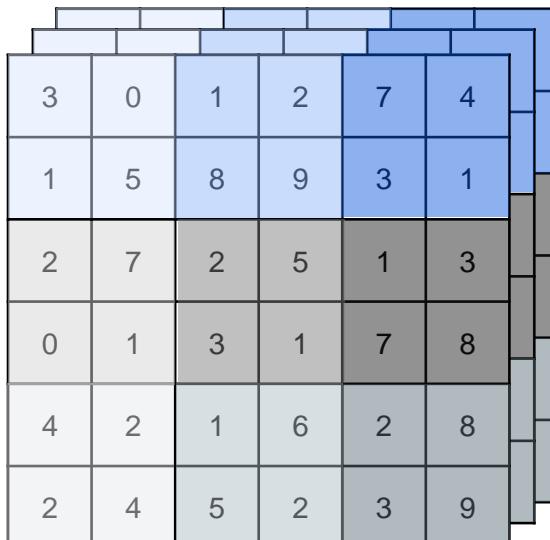
Hyperparameters:

$f = 2$

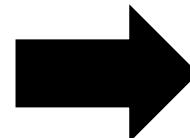
$s = 2$



Pooling layer: Max pooling



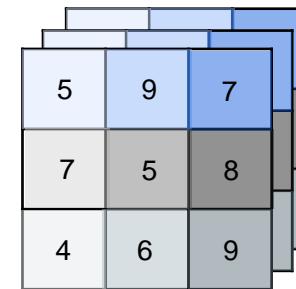
(6, 6)



Hyperparameters:

$$f = 2$$

$$s = 2$$



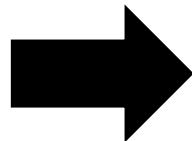
(3, 3)



Pooling layer: Average pooling

3	0	1	2	7	4
1	5	8	9	3	1
2	7	2	5	1	3
0	1	3	1	7	8
4	2	1	6	2	8
2	4	5	2	3	9

(6, 6)



Hyperparameters:

$$f = 2$$

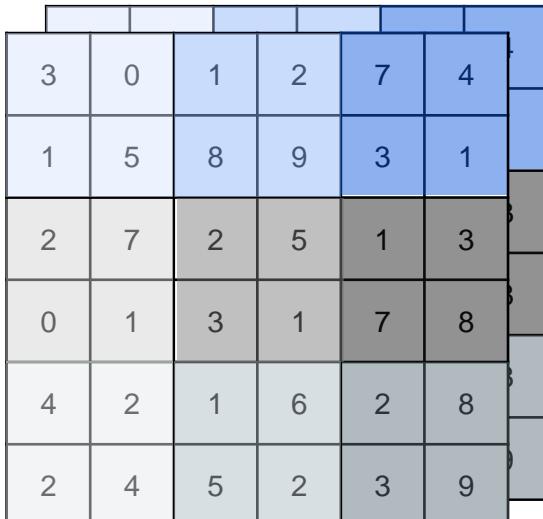
$$s = 2$$

2.25	5	3.75
2.5	2.75	4.75
3	3.5	5.5

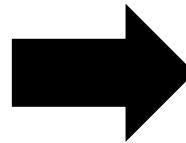
(3, 3)



Pooling layer: Average pooling



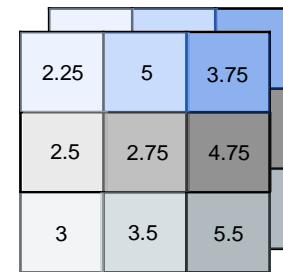
(6, 6, 2)



Hyperparameters:

$$f = 2$$

$$s = 2$$



(3, 3, 2)

Les trois blocs du CNN

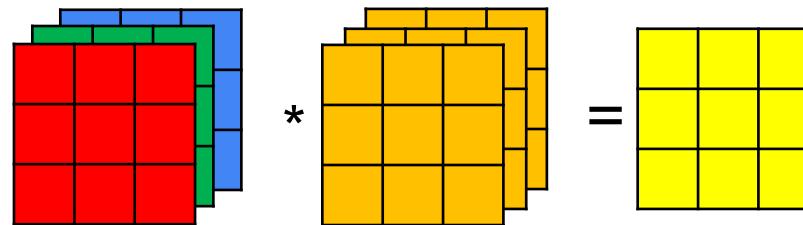


Partie 7 : Computer vision

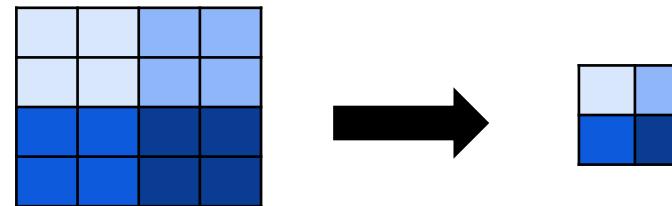


Les trois blocs du CNN

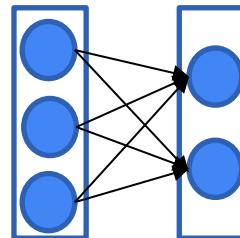
Convolutional bloc



Pooling bloc



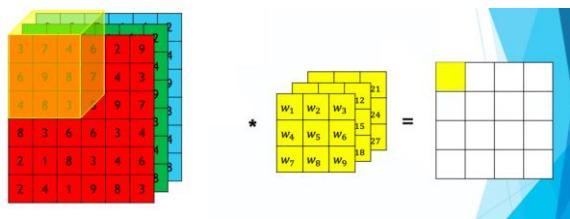
Fully connected bloc



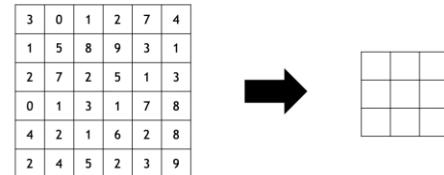


Les trois blocs du CNN

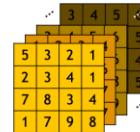
Convolution



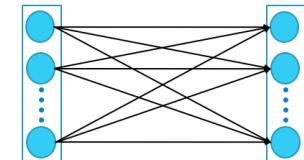
Pooling



Flatten



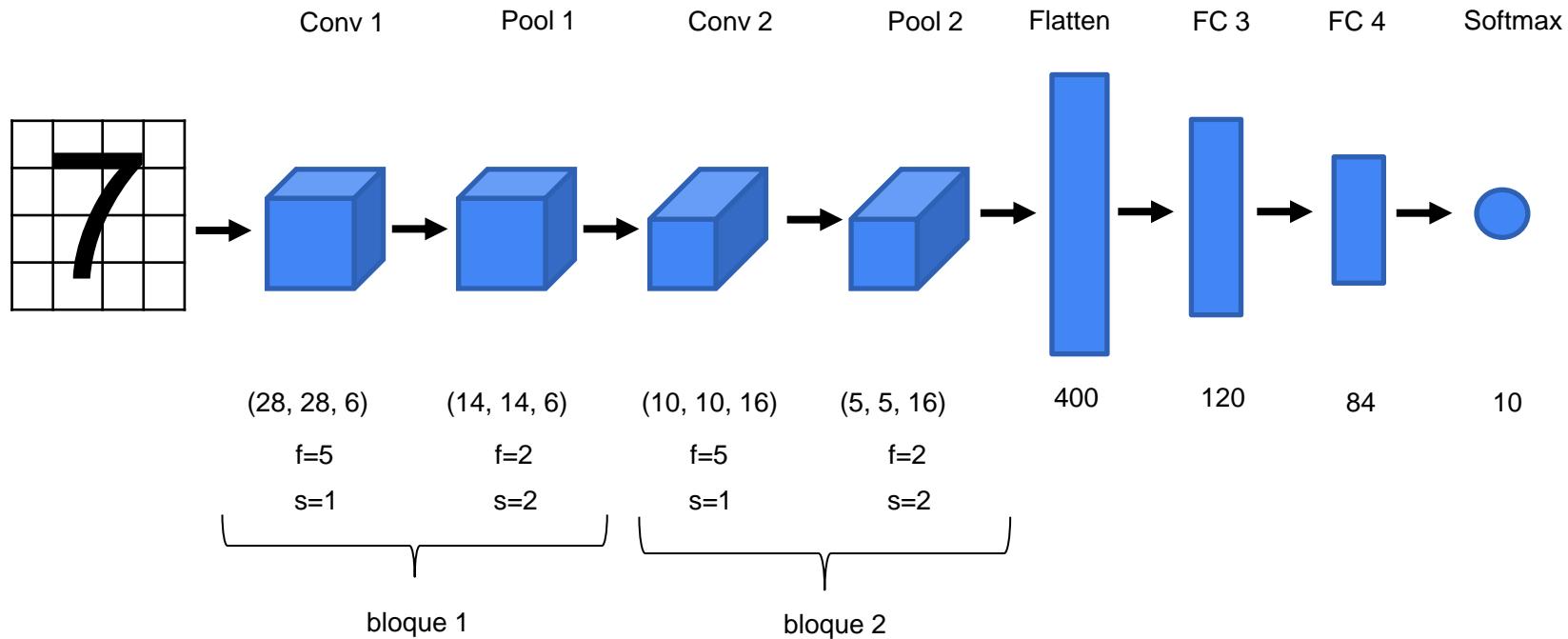
Couche dense





LeNet-5

$\approx 60k$ paramètres





LeNet-5

	Activation shape	Activation size	# parameters
Input:	(32, 32, 3)	3 072	0
CONV1 (f=5, s=1)	(28,28, 8)	6 272	208
POOL1	(14, 14, 8)	1 568	0
CONV2 (f=5, s=1)	(10, 10, 46)	1 600	416
POOL2	(5, 5, 16)	400	0
FC3	(120, 1)	120	48 001
FC4	(84, 1)	84	10 081
Softmax	(10, 1)	10	841

Deep learning avec Pytorch

Partie 8 : Le transfer learning



Présenté par **Morgan Gautherot**

ImageNet



Partie 8 : Le transfer learning



Concours de computer vision



14,197,122 images, 21841 synsets indexed

[Home](#) [Download](#) [Challenges](#) [About](#)

Not logged in. [Login](#) | [Signup](#)

An Update to the ImageNet Website and Dataset

March 11, 2021

We are proud to see ImageNet's wide adoption going beyond what was originally envisioned. However, the decade-old website was burdened by growing download requests. To serve the community better, we have redesigned the [website](#) and upgraded its hardware. The new website is simpler; we removed tangential or outdated functions to focus on the core use case—enabling users to [download the data](#), including the full ImageNet dataset and the [ImageNet Large Scale Visual Recognition Challenge \(ILSVRC\)](#).

Meanwhile, the computer vision community has progressed, and so has ImageNet. The dataset was created to benchmark object recognition—at a time when it barely worked. The problem then was how to collect labeled images at a sufficiently large scale to be able to train complex models in laboratories. Today, computer vision is in real-world systems impacting people's Internet experience and daily lives. An emerging problem now is how to make sure computer vision is fair and preserves people's privacy. We are continually evolving ImageNet to address these emerging needs.

In a [FAT* 2020 paper](#), we filtered 2,702 synsets in the "person" subtree that may cause problematic behaviors of the model. We have updated the full ImageNet data on the website to remove these synsets. The update does not affect the 1,000 categories in ILSVRC.

In a [more recent paper](#), we investigate privacy issues in ILSVRC. 997 out of 1000 categories in ILSVRC are not people categories; nevertheless, many incidental people are in the images, whose privacy is a concern. We first annotated faces in the images and then constructed a face-blurred version of ILSVRC. Experiments show that one can use the face-blurred version for benchmarking object recognition and for transfer learning with only marginal loss of accuracy. We release our [face annotations](#) to facilitate further research on privacy-aware visual recognition.

Team members working on these new improvements: [Kaiyu Yang](#) (Princeton), [Jacqueline Yau](#) (Stanford), [Li Fei-Fei](#) (Stanford), [Jia Deng](#) (Princeton), [Olga Russakovsky](#) (Princeton).



Un problème de classification

mite	container ship	motor scooter	leopard
black widow	lifeboat	go-kart	jaguar
cockroach	amphibian	moped	cheetah
tick	fireboat	bumper car	snow leopard
starfish	drilling platform	golfcart	Egyptian cat
grille	mushroom	cherry	Madagascar cat
convertible	agaric	dalmatian	squirrel monkey
grille	mushroom	grape	spider monkey
pickup	jelly fungus	elderberry	titi
beach wagon	gill fungus	ffordshire bullterrier	indri
fire engine	dead-man's-fingers	currant	howler monkey

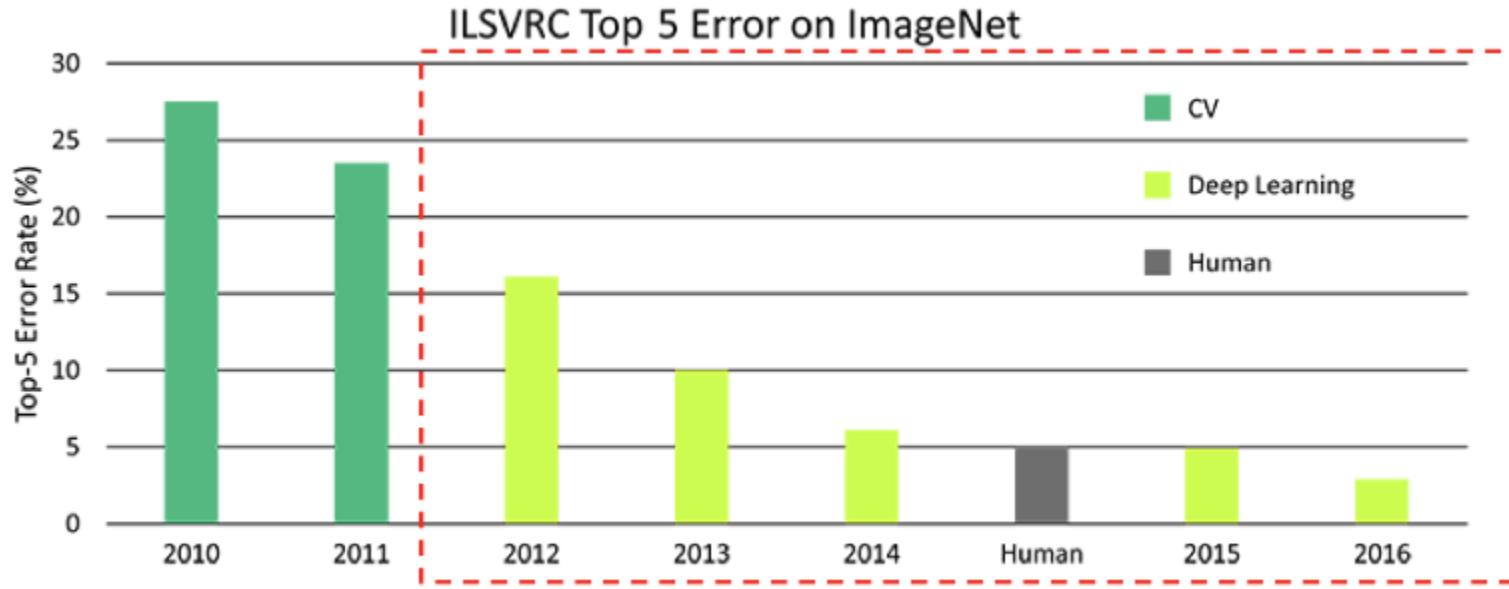


Un jeu de données d'entraînement énorme

- 1 000 classes à classifier
- 1 281 167 images pour l'entraînement
- 50 000 images pour la validation
- 100 000 images pour le test



Les performances au cours du temps



The introduction of Deep Learning techniques drove performance on image categorization from 30% error rates in 2010, down to <2% in 2017

Tranfer learning

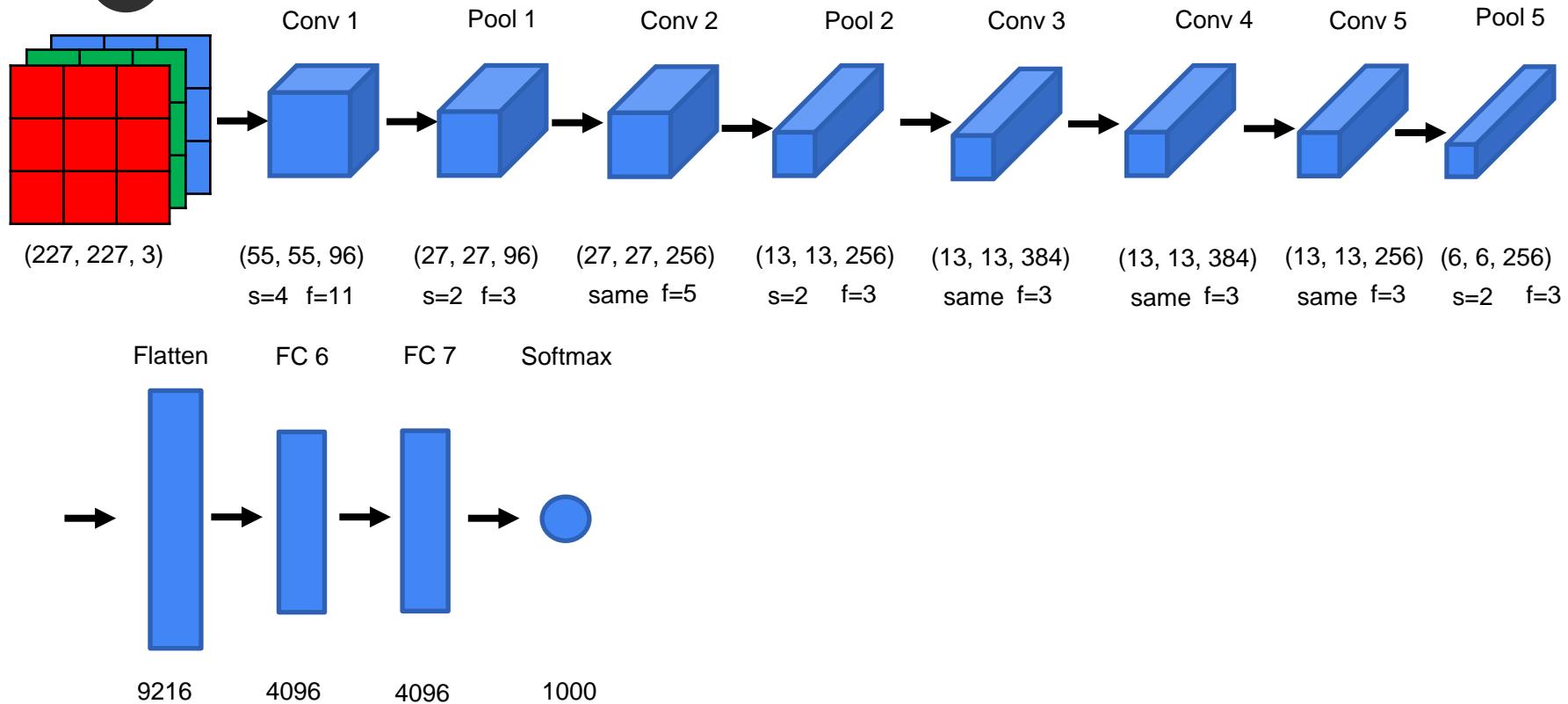


Partie 8 : Le transfer learning



AlexNet

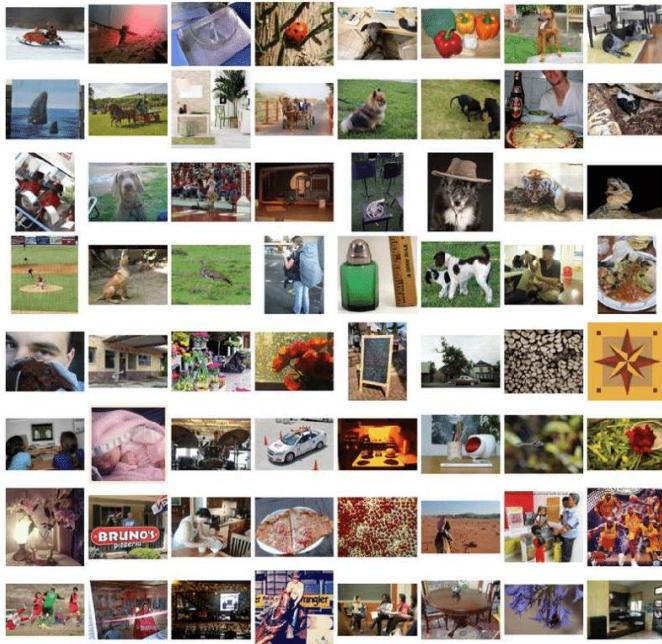
≈ 60 millions de paramètres





Les différents domaines d'images

Domaine d'imageNet

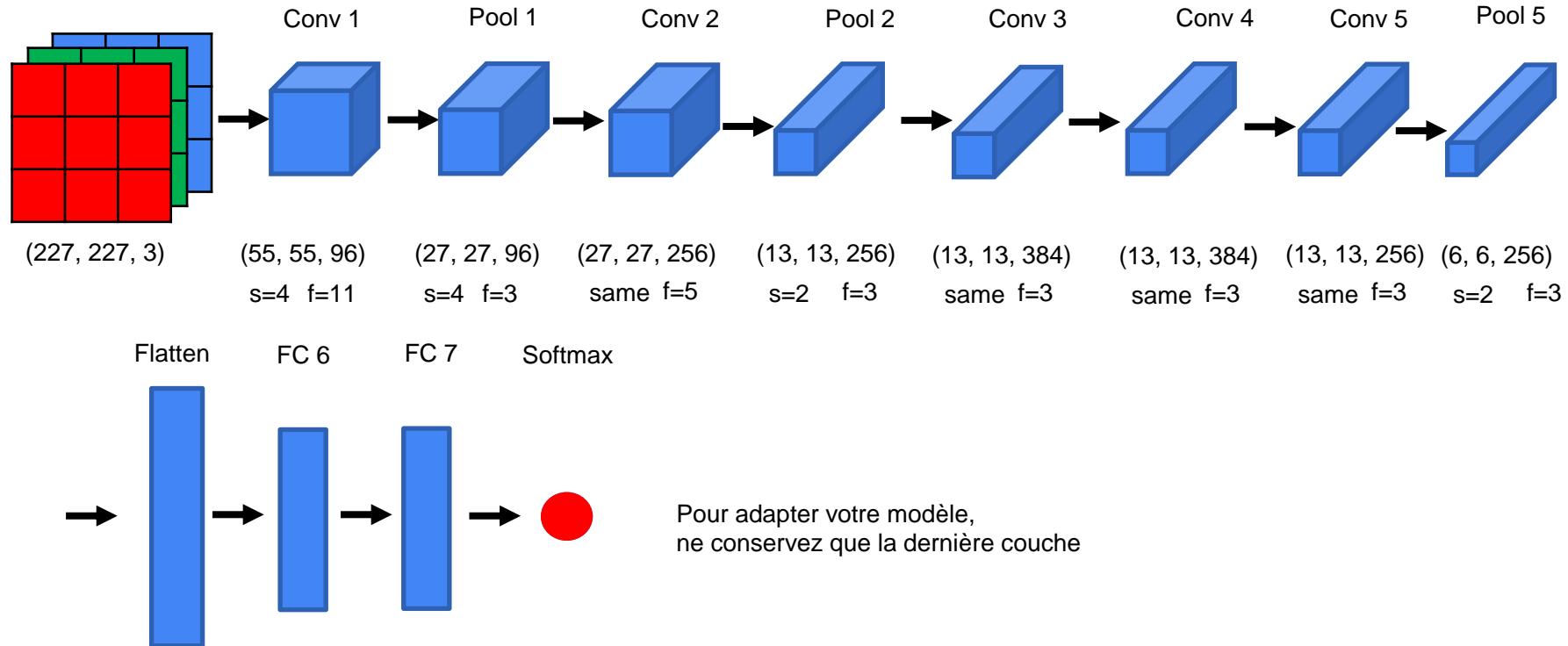


Mon domaine



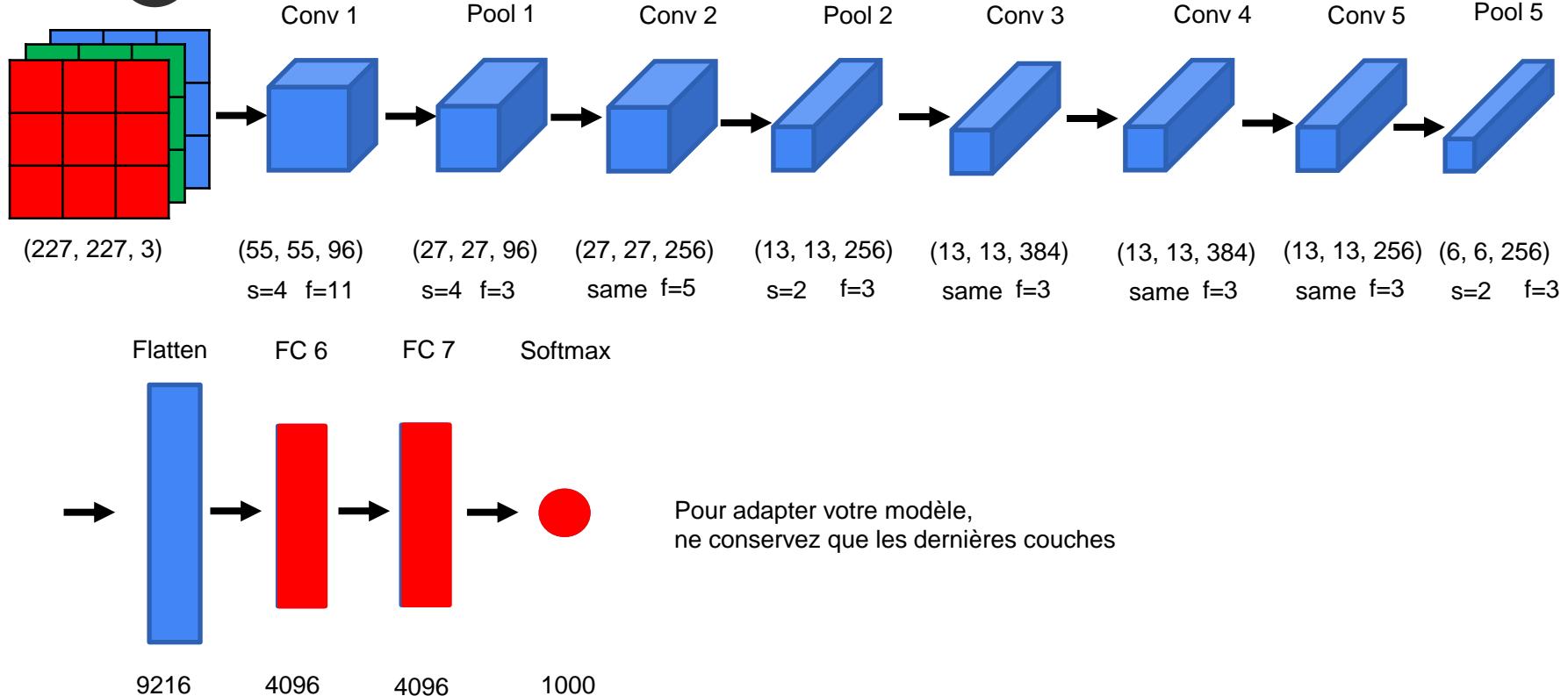


Transfer learning avec très peu de données



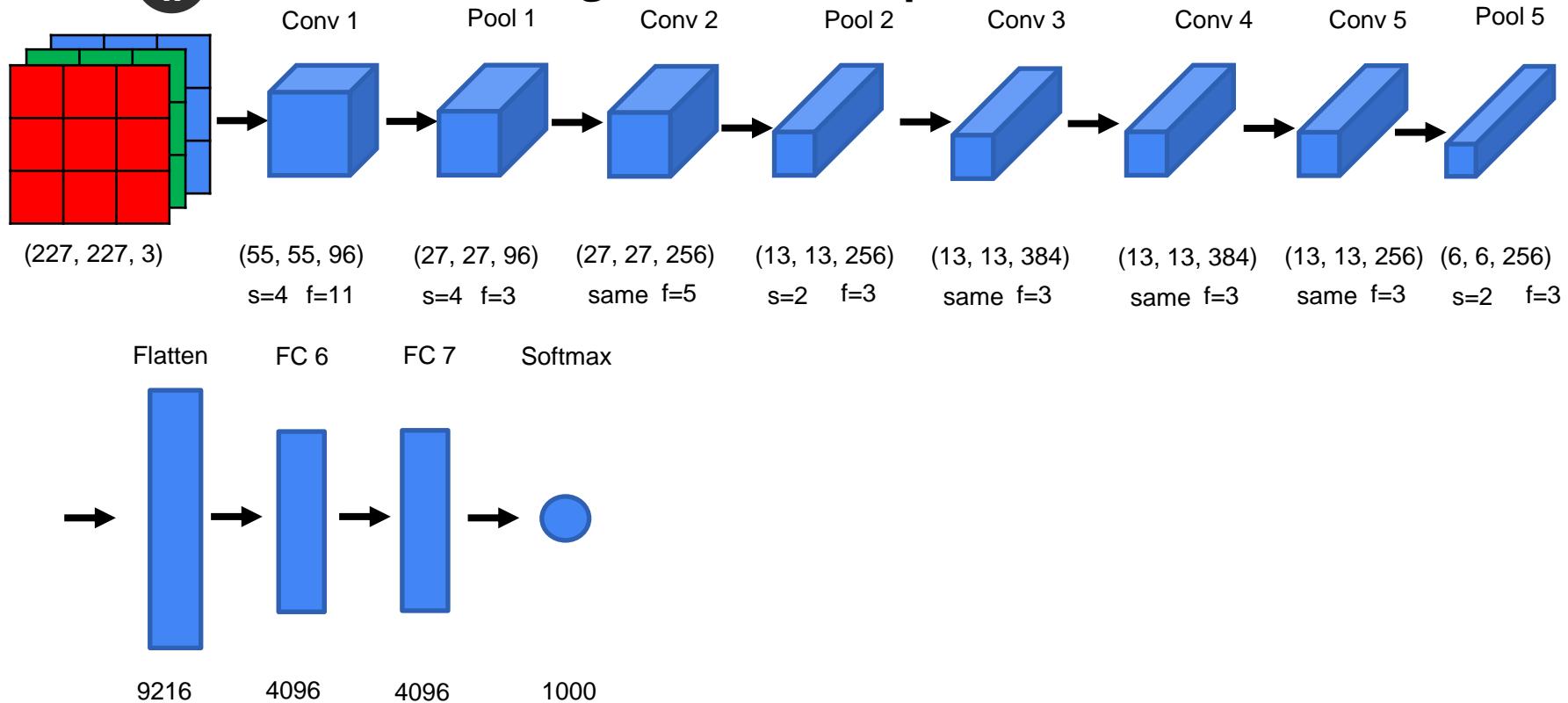


Transfer learning avec peu de données



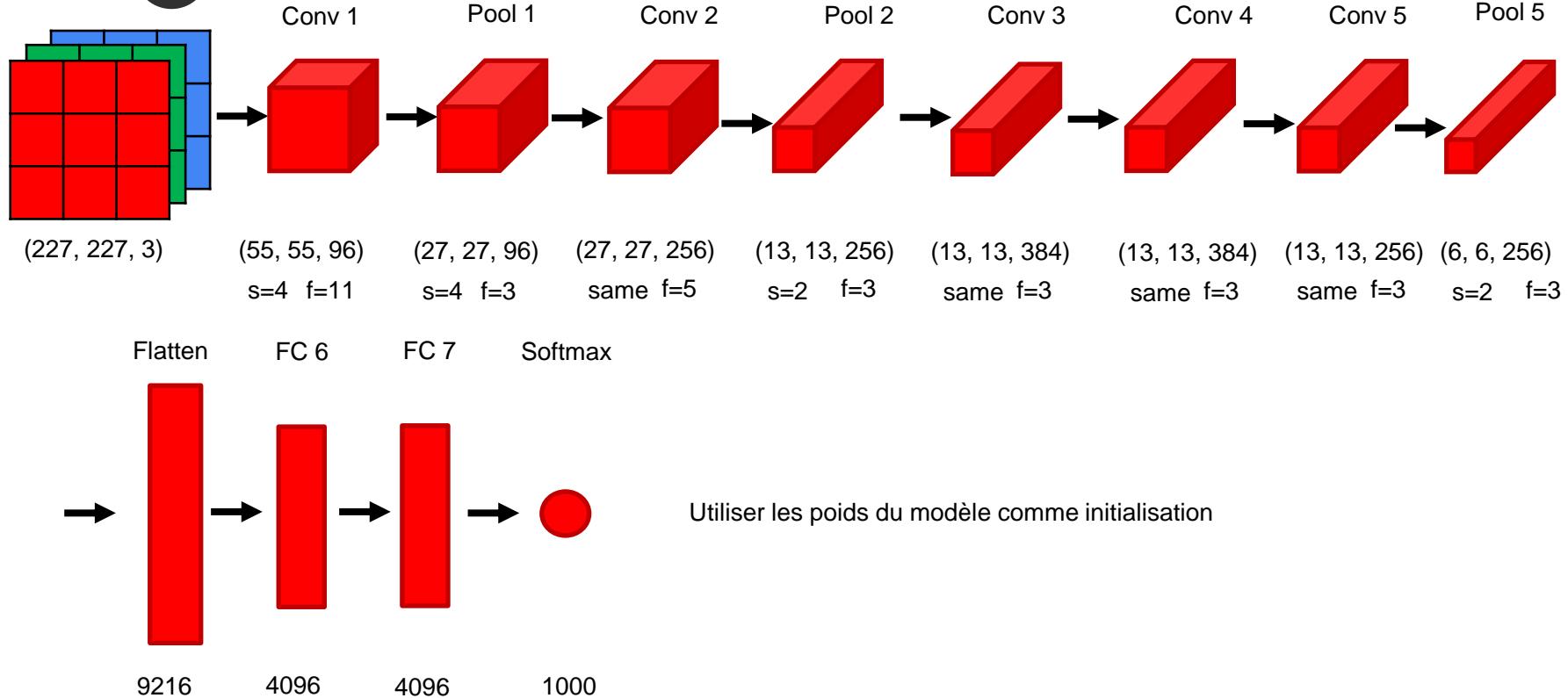


Transfer learning avec beaucoup de données





Transfer learning avec beaucoup de données





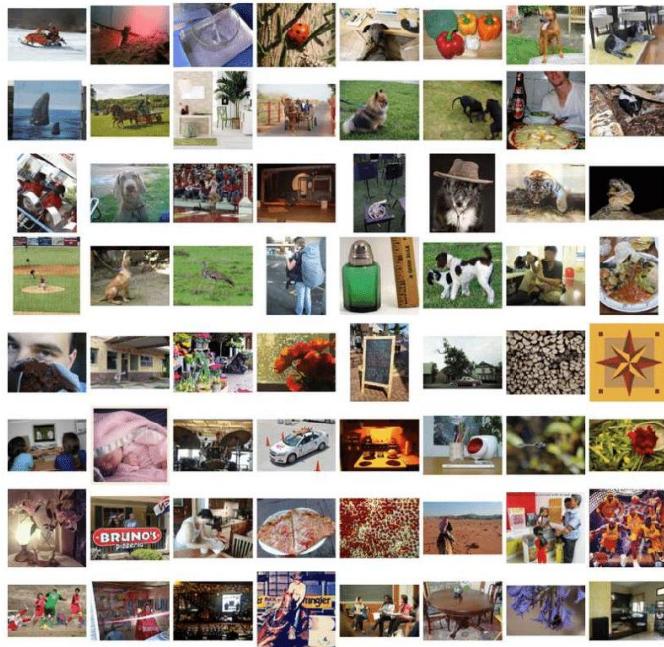
A retenir

	Domaine d'Imagenet	Domaine différent d'Imagenet
Petit dataset	Les dernières couches	Collecter plus de données
Grand dataset	Utiliser comme initialisation	Initialisation aléatoire

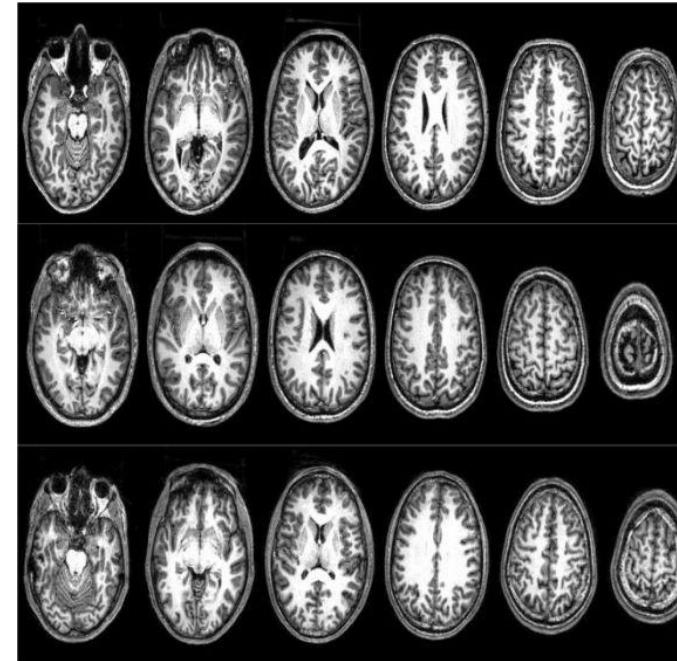


Les différents domaines d'images

Domaine d'imageNet



Mon domaine



Deep learning avec Pytorch

Partie 9 : L'auto-encodeur

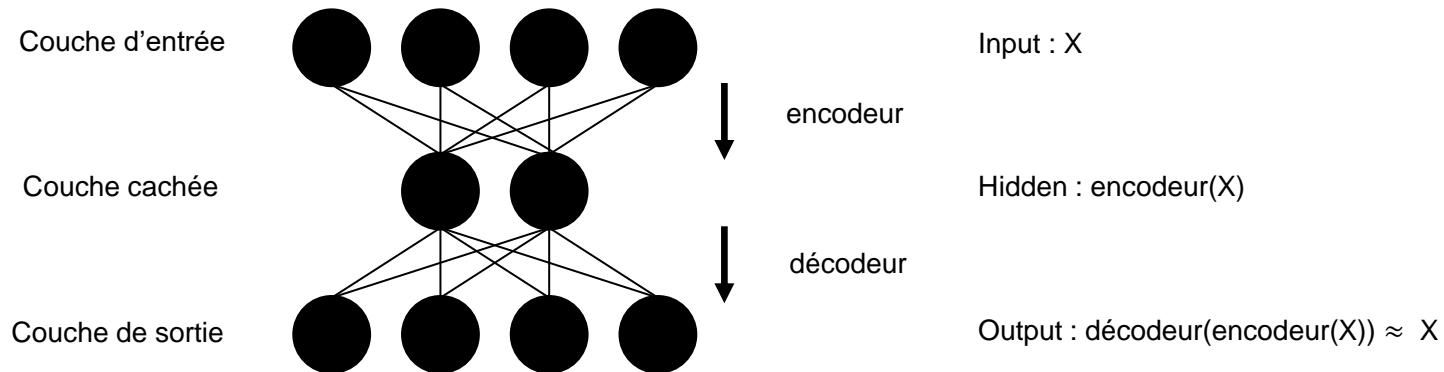


Présenté par **Morgan Gautherot**



Auto-encodeurs

- Le but de l'auto-encodeur est de prendre des données et de les encoder dans une dimension plus petite à partir de laquelle elles peuvent ensuite être restaurées avec précision.



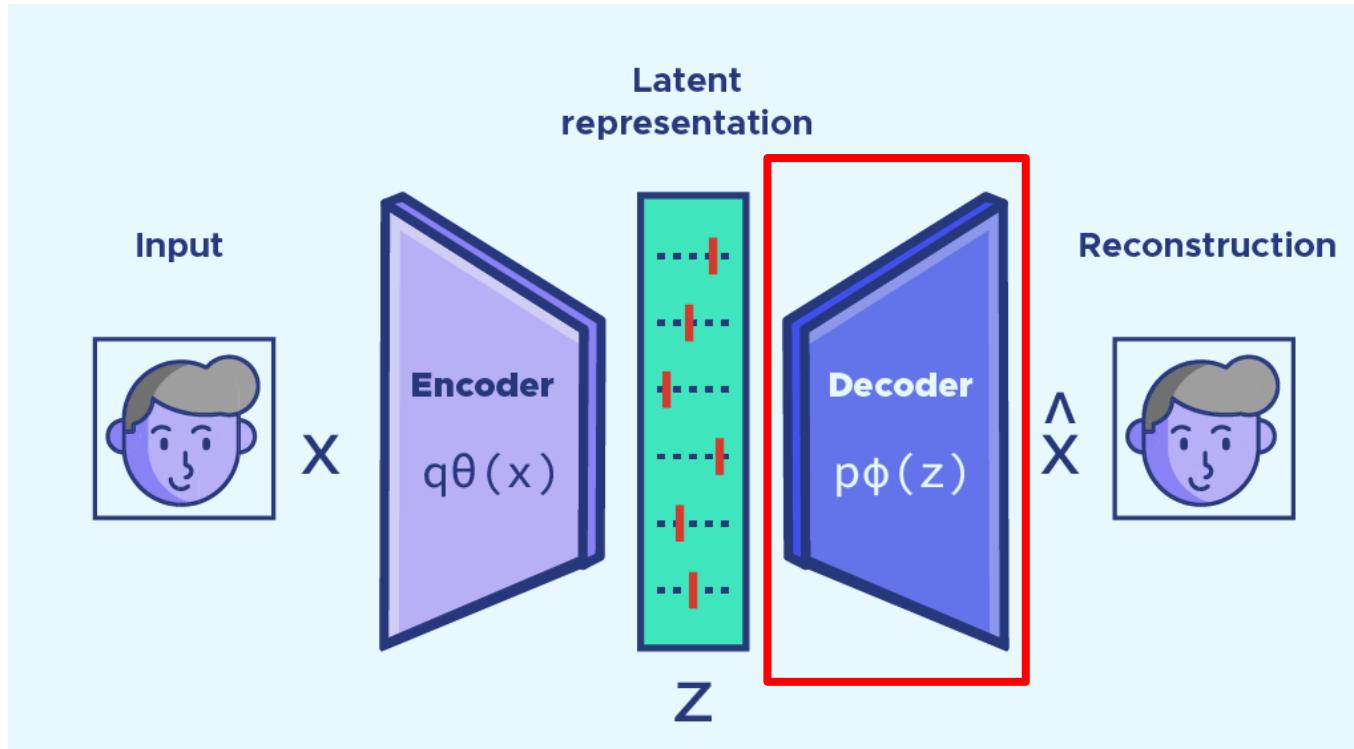


Quand l'utilisons-nous ?

- Compression des données
- Réduction de la dimensionnalité
- Apprendre des caractéristiques intéressantes
- Pré-entraînement non supervisé
- Générer de nouvelles données



Auto-encodeur avec des images



Deep learning par la pratique

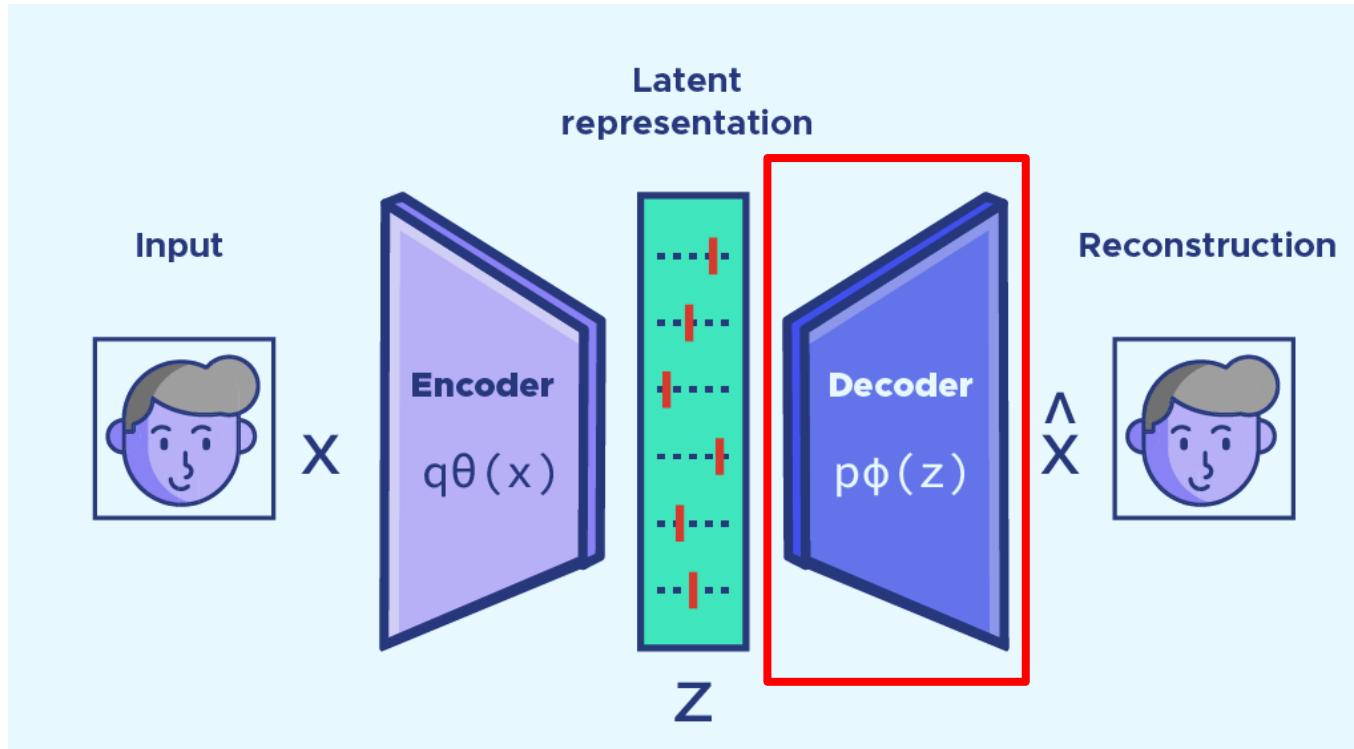
Leçon 2 : Transposed convolution



Présenté par **Morgan Gautherot**

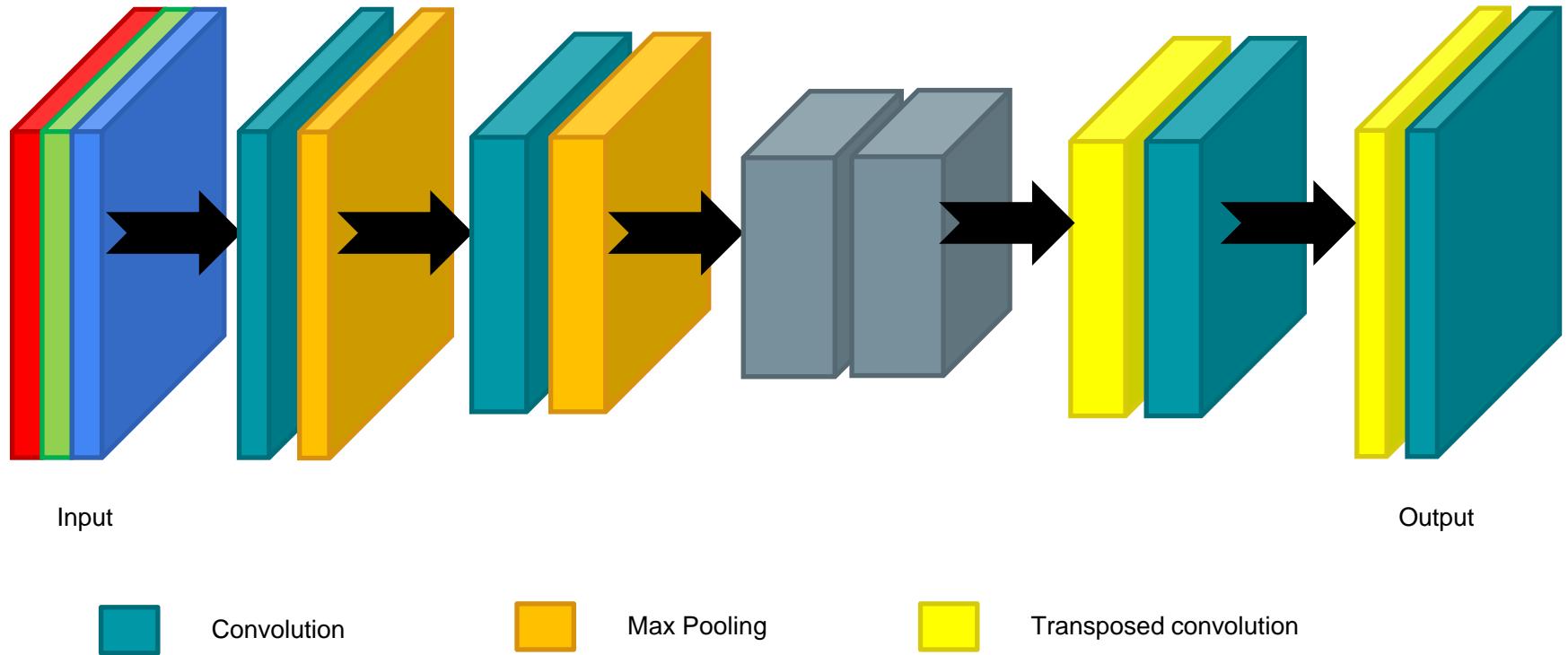


Auto-encodeur avec des images





Architecture d'un auto-encodeur





Convolution

4	5	8	7
1	8	8	8
3	6	6	4
6	5	7	8

4×4

*

1	4	1
1	4	3
3	3	1

3×3

=

122	148
126	134

2×2



Convolution

1	4	1
1	4	3
3	3	1

3×3

=

1	4	1	0	1	4	3	0	3	3	1	0	0	0	0	0	0
0	1	4	1	0	1	4	3	0	3	3	1	0	0	0	0	0
0	0	0	0	1	4	1	0	1	4	3	0	3	3	1	0	0
0	0	0	0	0	1	4	1	0	1	4	3	0	3	3	1	0

4×16

4	5	8	7
1	8	8	8
3	6	6	4
6	5	7	8

4×4

=

4
5
8
7
1
8
8
8
8
3
6
6
4
6
5
7
8

16×1



Convolution

1	4	1	0	1	4	3	0	3	3	1	0	0	0	0	0
0	1	4	1	0	1	4	3	0	3	3	1	0	0	0	0
0	0	0	0	1	4	1	0	1	4	3	0	3	3	1	0
0	0	0	0	0	1	4	1	0	1	4	3	0	3	3	1

4 x 16

4
5
8
7
1
8
8
8
8
3
6
6
4
6
5
7
8

16 x 1

=

122
148
126
134

4 x 1

=

122	148
126	134

2 x 2



Transposed convolution / deconvolution

55	52
57	50

*

1	2	1
2	1	2
1	1	2

=

55	162	159	52
167	323	319	154
169	264	326	204
57	107	164	100

2 x 2

3 x 3

4 x 4



Transposed convolution / deconvolution

55	52
57	50

2 x 2

=

55
52
57
50

4 x 1

1	2	1
2	1	2
1	1	2

3 x 3

=

1	0	0	0
2	1	0	0
1	2	0	0
0	1	0	0
2	0	1	0
1	2	2	1
2	1	1	2
0	2	0	1
1	0	2	0
1	1	1	2
2	1	2	1
0	2	0	2
0	0	1	0
0	0	1	1
0	0	2	1
0	0	0	2

16 x 4



Transposed convolution / deconvolution

1	0	0	0
2	1	0	0
1	2	0	0
0	1	0	0
2	0	1	0
1	2	2	1
2	1	1	2
0	2	0	1
1	0	2	0
1	1	1	2
2	1	2	1
0	2	0	2
0	0	1	0
0	0	1	1
0	0	2	1
0	0	0	2

4 x 1

The diagram shows the multiplication of a 4x1 vector by a 4x4 matrix. The vector is:

122
148
126
134

4 x 1

The matrix is represented as four 4x1 vectors being added together:

55
110
55
0

 $=$

0
52
104
52

 $+$

0
0
0
57

 $+$

57
114
57
0

 $=$

55
162
159
52

A blue arrow points from the sum of the matrix columns to the final result:

55	162	159	52
167	323	319	154
169	264	326	204
57	107	164	100

16 x 4



Transposed convolution / deconvolution

55				
110				
55				
0				
110				
55				
110				
0				
55				
55				
110				
0				
0				
0				
0				



55	110	55	0
110	55	110	0
55	55	110	0
0	0	0	0

0				
52				
104				
52				
0				
104				
52				
104				
0				
52				
52				
104				
0				
0				
0				



0	52	104	52
0	104	52	104
0	52	52	104
0	0	0	0

0				
0				
0				
0				
57				
114				
57				
0				
114				
57				
117				
0				
57				
57				
114				
0				



0	0	0	0
57	114	57	0
114	57	114	0
57	57	114	0
57	57	114	0

0				
0				
0				
0				
0				
50				
100				
50				
0				
100				
50				
100				
0				
50				
50				
100				



0	0	0	0
0	50	100	50
0	100	50	100
0	50	50	100
0	50	50	100



Transposed convolution / deconvolution

55	110	55	0
110	55	110	0
55	55	110	0
0	0	0	0

+

0	52	104	52
0	104	52	104
0	52	52	104
0	0	0	0

+

55	162	159	52
167	323	319	154
169	264	326	204
57	107	164	100

+

0	0	0	0
0	50	100	50
0	100	50	100
0	50	50	100



Transposed convolution / deconvolution

1	0	0	0
2	1	0	0
1	2	0	0
0	1	0	0
2	0	1	0
1	2	2	1
2	1	1	2
0	2	0	1
1	0	2	0
1	1	1	2
2	1	2	1
0	2	0	2
0	0	1	0
0	0	1	1
0	0	2	1
0	0	0	2

16 x 4

122
148
126
134

4 x 1

=

55	162	159	52
167	323	319	154
169	264	326	204
57	107	164	100

4 x 4

Deep learning par la pratique

Leçon 3 : Applications



Présenté par **Morgan Gautherot**



Cas linéaire

- Exemple: Factorisation de matrices

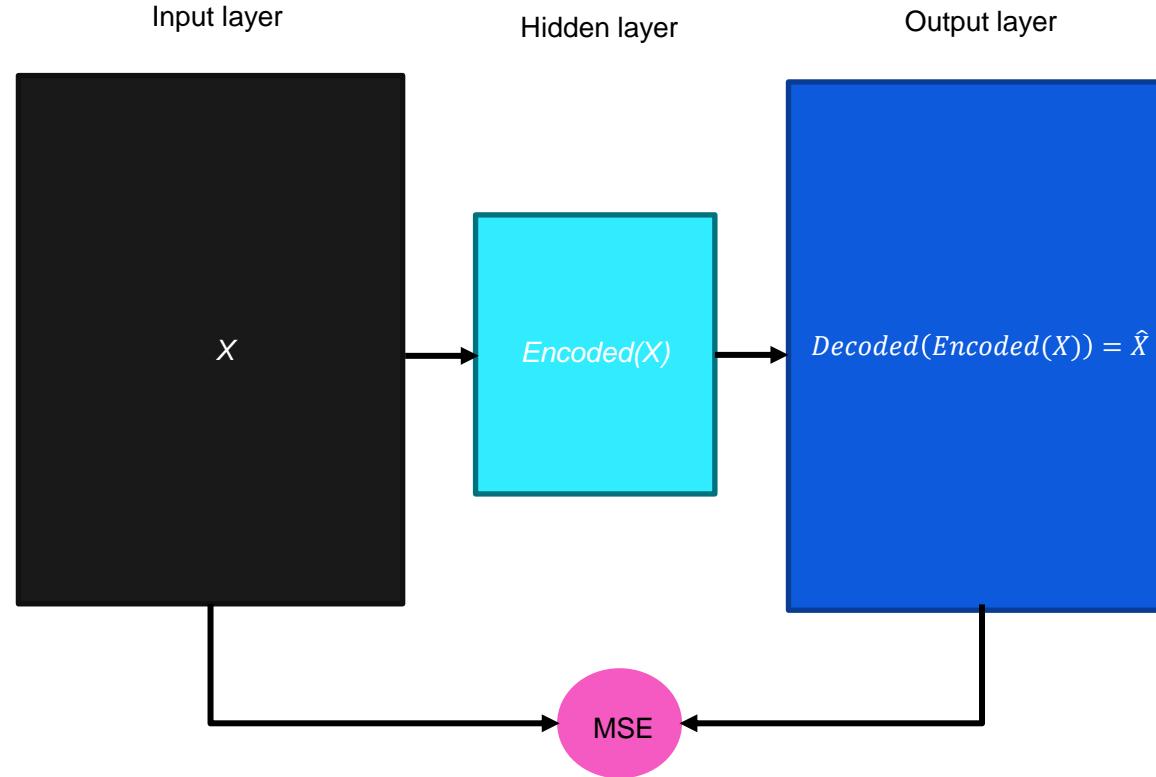
$$X = U \times V^T$$

L'objectif est de minimiser l'erreur de reconstruction

$$\min_{U,V} \|X - U \cdot V^T\|$$

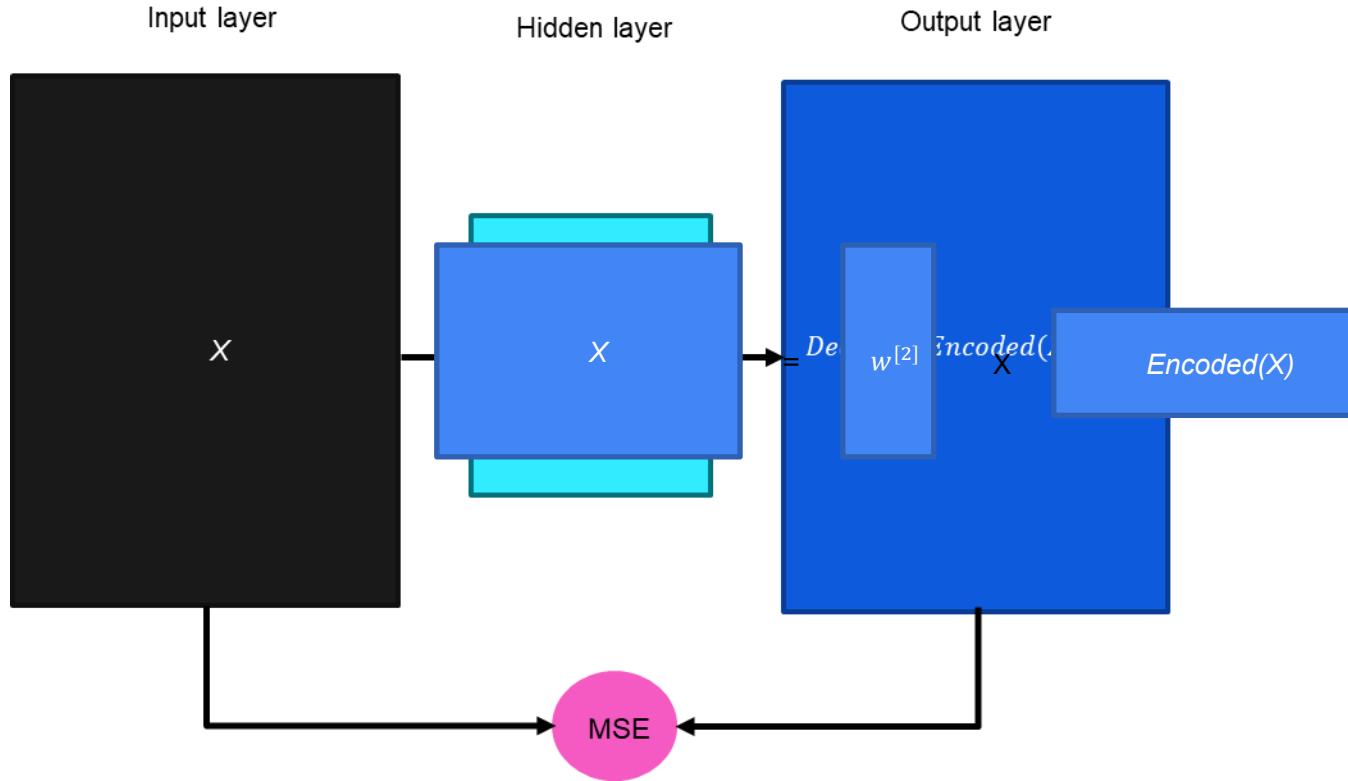


Décompositions de matrices





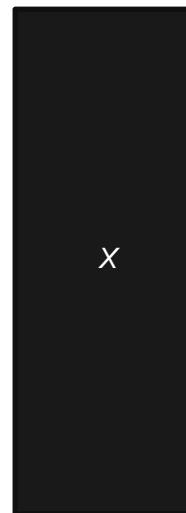
Décompositions de matrices



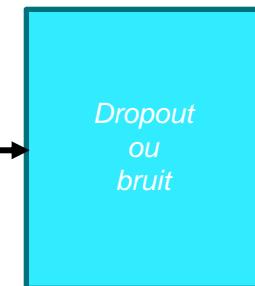


Débruitage

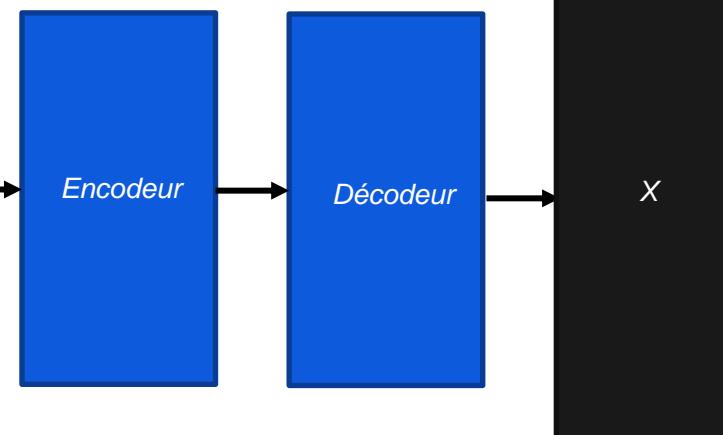
Couche d'entrée



Couches cachées



Couche de sortie

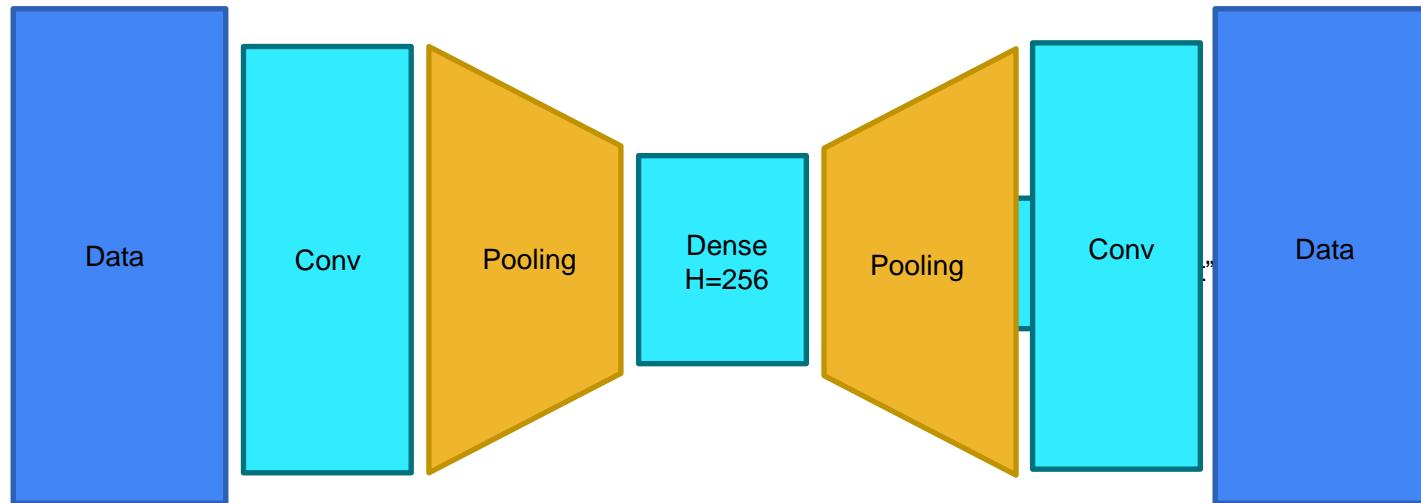


MSE



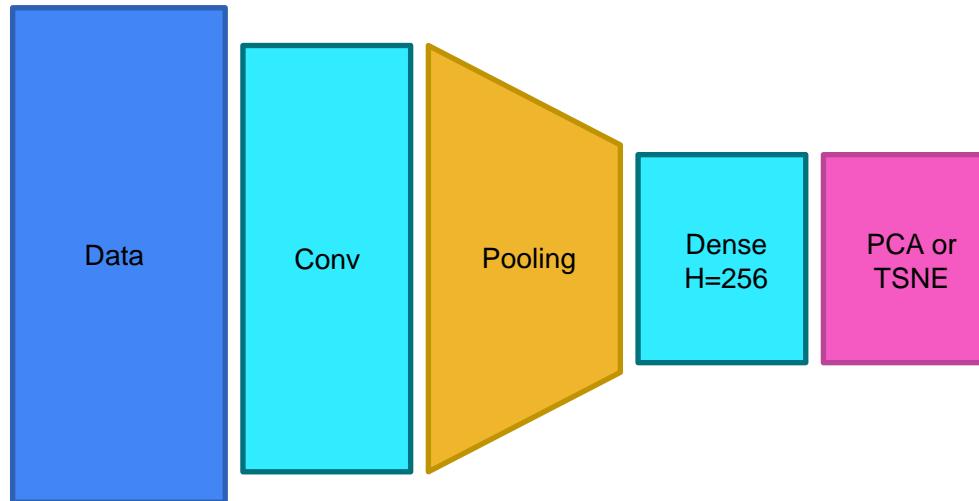


Pré-entraînement non supervisé





Analyse exploratoire des données





Analyse exploratoire des données





Morphing d'images

Si

$$\text{Enc}(\text{Image1}) = \mathbf{C1}$$

$$\text{Enc}(\text{Image2}) = \mathbf{C2}$$

Alors $(\mathbf{C1} + \mathbf{C2})/2$ est une moyenne sémantique des deux images



Avec ce type de modèle, vous pouvez modifier les images pour ajouter une moustache ou rendre une personne plus âgée.

Deep learning avec Pytorch

Partie 10 : Recurrent Neral Network

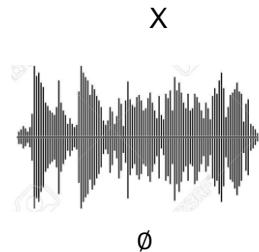


Présenté par **Morgan Gautherot**



Pourquoi des modèles de séquence ?

Reconnaissance vocale



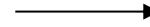
“Never stoop learning”

Génération de musique



Classification des sentiments

“There is nothing to like in
this movie.”



Analyse des séquences d'ADN

AGCCCCTGTGAGGAAC TAG



AG~~CCCCTGTGAGGAAC~~ TAG

Traduction automatique

Artificial Intelligence is the New
Electricity



L'intelligence artificielle est la nouvelle
électricité

Reconnaissance d'activité



Running

Reconnaissance d'un nom

Elon Musk is the founder of
SpaceX



Elon ~~Musk~~ is the founder of
SpaceX



Notation

Reconnaissance de nom

x: "Harry Potter and Hermione Granger invented a new spell."

$x^{<1>} \quad x^{<2>} \quad x^{<3>} \quad \dots \quad x^{<t>} \quad \dots \quad x^{<8>} \quad x^{<9>}$

y: [1 , 1 , 0 , 1 , 1 , 0 , 0 , 0 , 0]

$y^{<1>} \quad y^{<2>} \quad y^{<3>} \quad \dots \quad y^{<t>} \quad \dots \quad y^{<9>}$

$X^{(i)<t>}$ observation de la séquence t du $i^{\text{ème}}$ exemple.

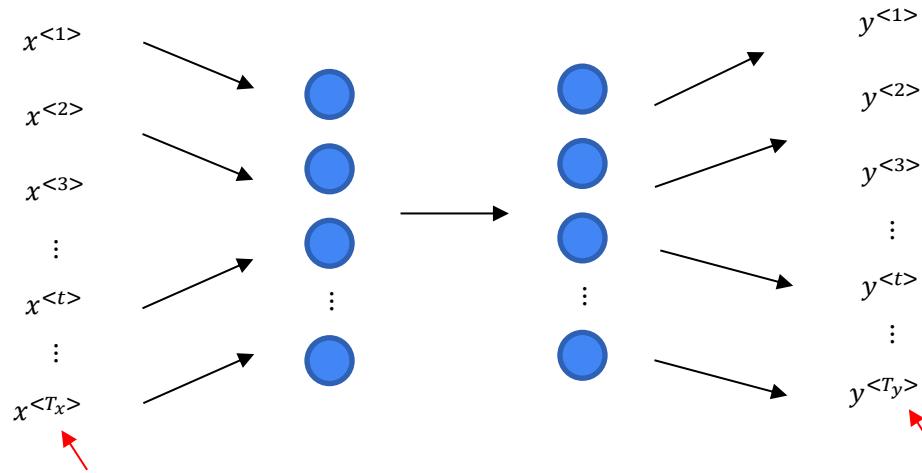
$y^{(i)<t>}$ target de la séquence t du $i^{\text{ème}}$ exemple.

$T_x^{(i)}$ est la longueur de la séquence d'observations.

$T_y^{(i)}$ est la longueur de la séquence des valeurs cibles.



Pourquoi pas un réseau dense ?



Problèmes :

- Les entrées et les sorties peuvent être de longueurs différentes dans différents exemples.
- Ne partage pas les caractéristiques apprises entre les différentes positions de la séquence.

Le recurrent Neural Network

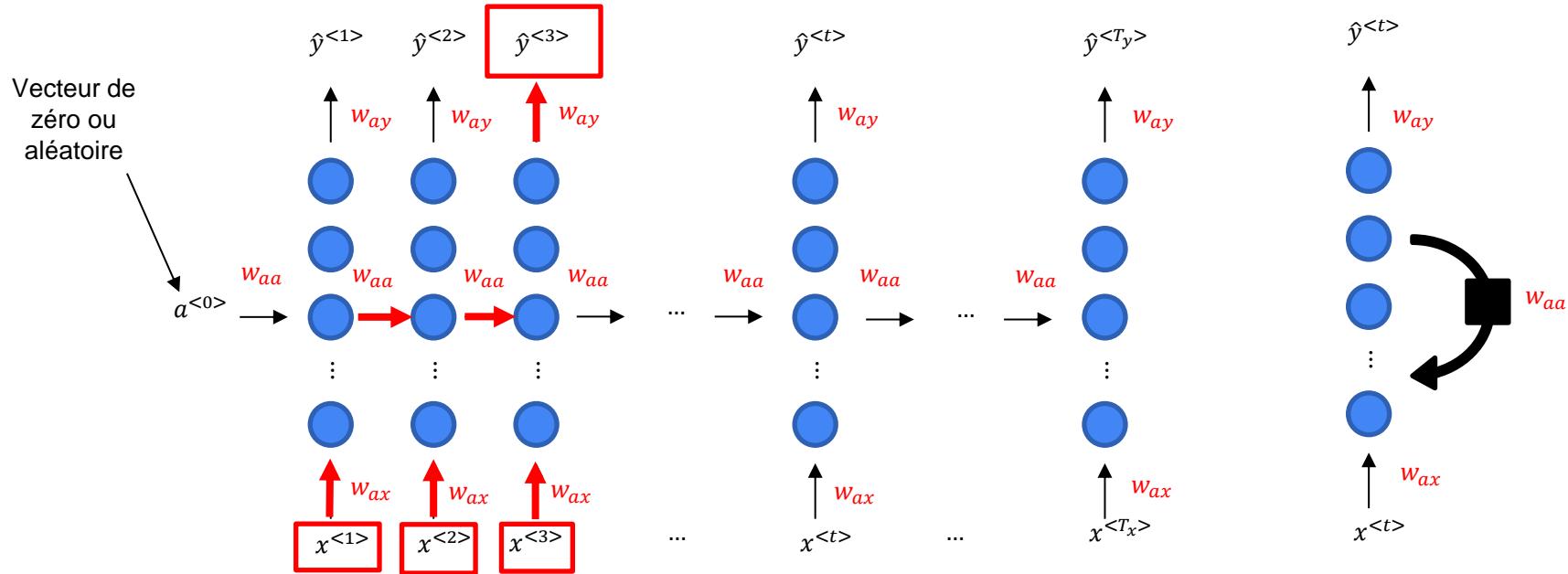


Partie 9 : Recurrent Nerual Network



Recurrent Neural Networks

Ici $T_x = T_y$



Froward et backward propagation

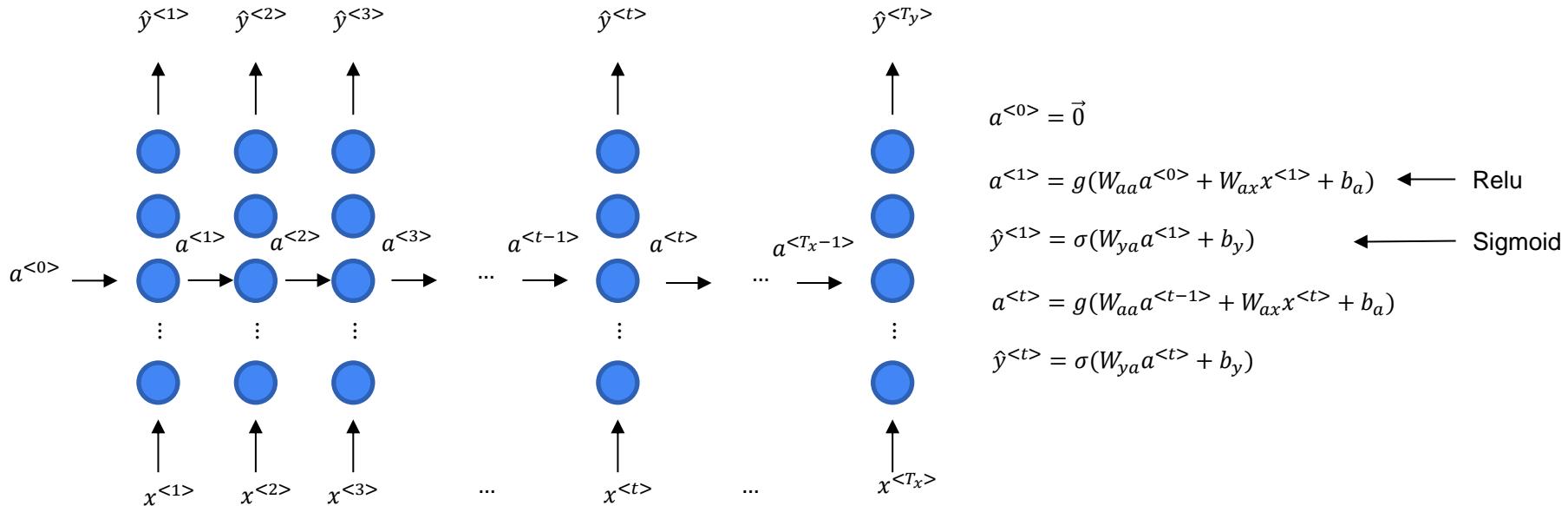


Partie 9 : Recurrent Nerual Network



Forward propagation

Here $T_x = T_y$





Notation

$$a^{} = g(W_{aa}a^{} + W_{ax}x^{} + b_a)$$

$$\hat{y}^{} = g(W_{ya}a^{} + b_y)$$

$$W_a = [W_{aa}, W_{ax}]$$

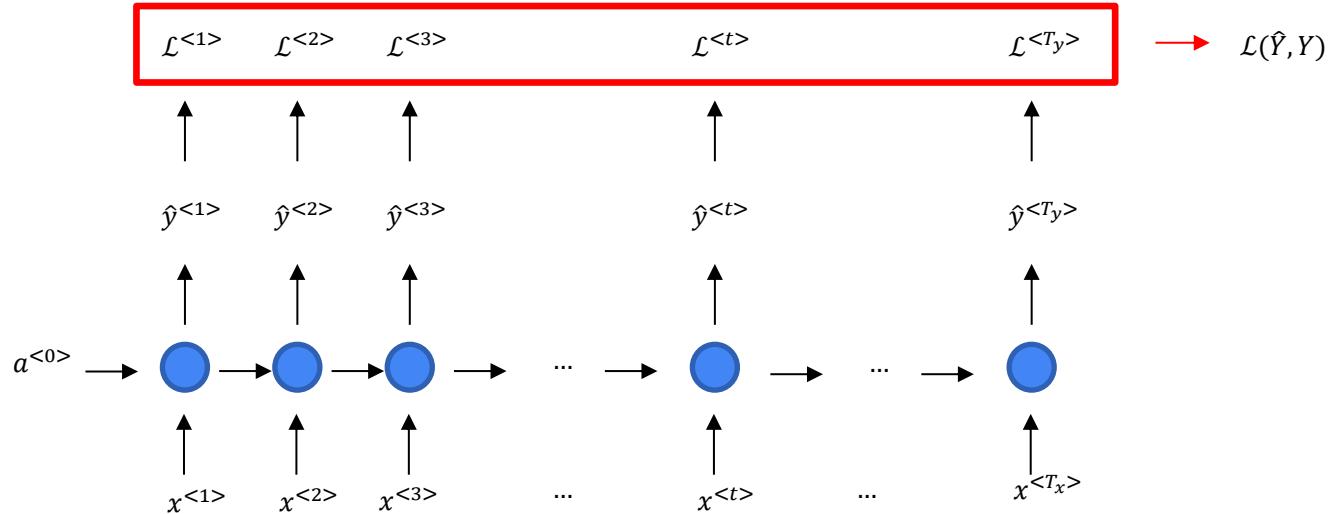
$$a^{} = g(W_a[a^{}, x^{}] + b_a)$$

$$\hat{y}^{} = g(W_ya^{} + b_y)$$

$$[W_{aa}, W_{ax}] \begin{bmatrix} a^{} \\ x^{} \end{bmatrix} = W_{aa}a^{} + W_{ax}x^{}$$



Fonction de coût

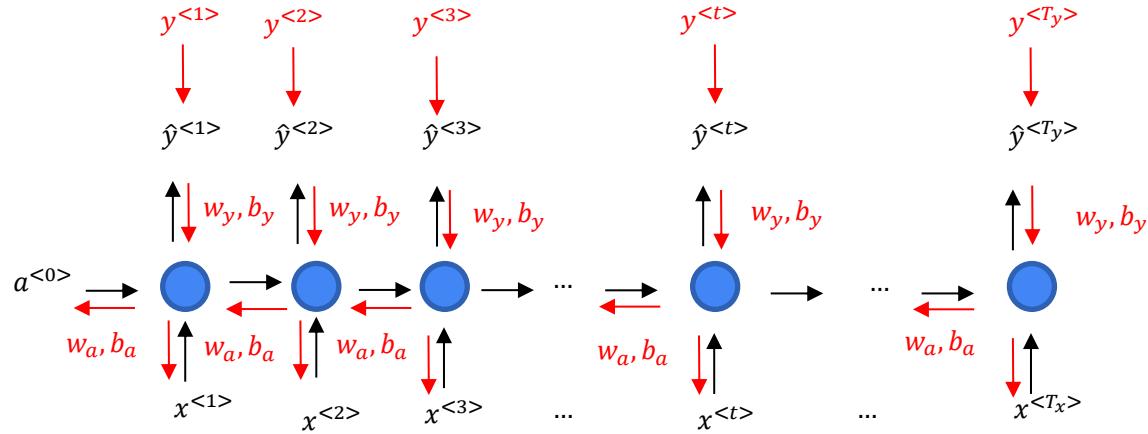


$$\mathcal{L}^{<t>}(\hat{y}^{<t>}, y^{<t>}) = -y^{<t>} \log(\hat{y}^{<t>}) - (1 - y^{<t>}) \log((1 - \hat{y}^{<t>}))$$

$$\mathcal{L}(\hat{y}, Y) = \sum_{t=1}^{T_y} \mathcal{L}^{<t>}(\hat{y}^{<t>}, y^{<t>})$$



Backpropagation



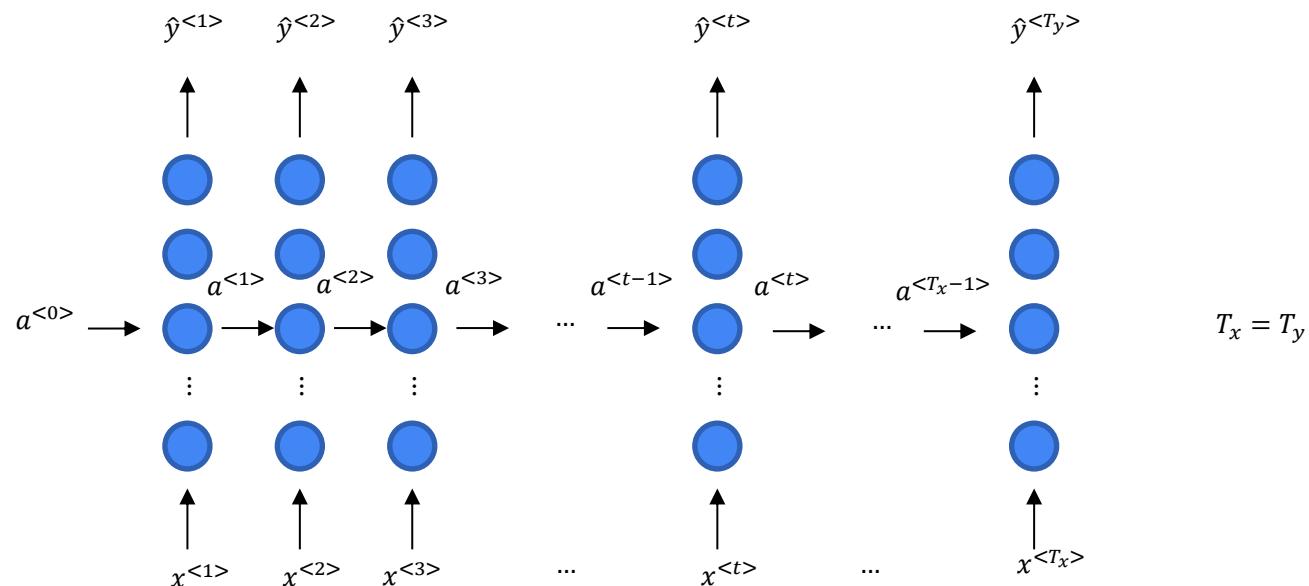
Les différentes architectures



Partie 9 : Recurrent Nerual Network

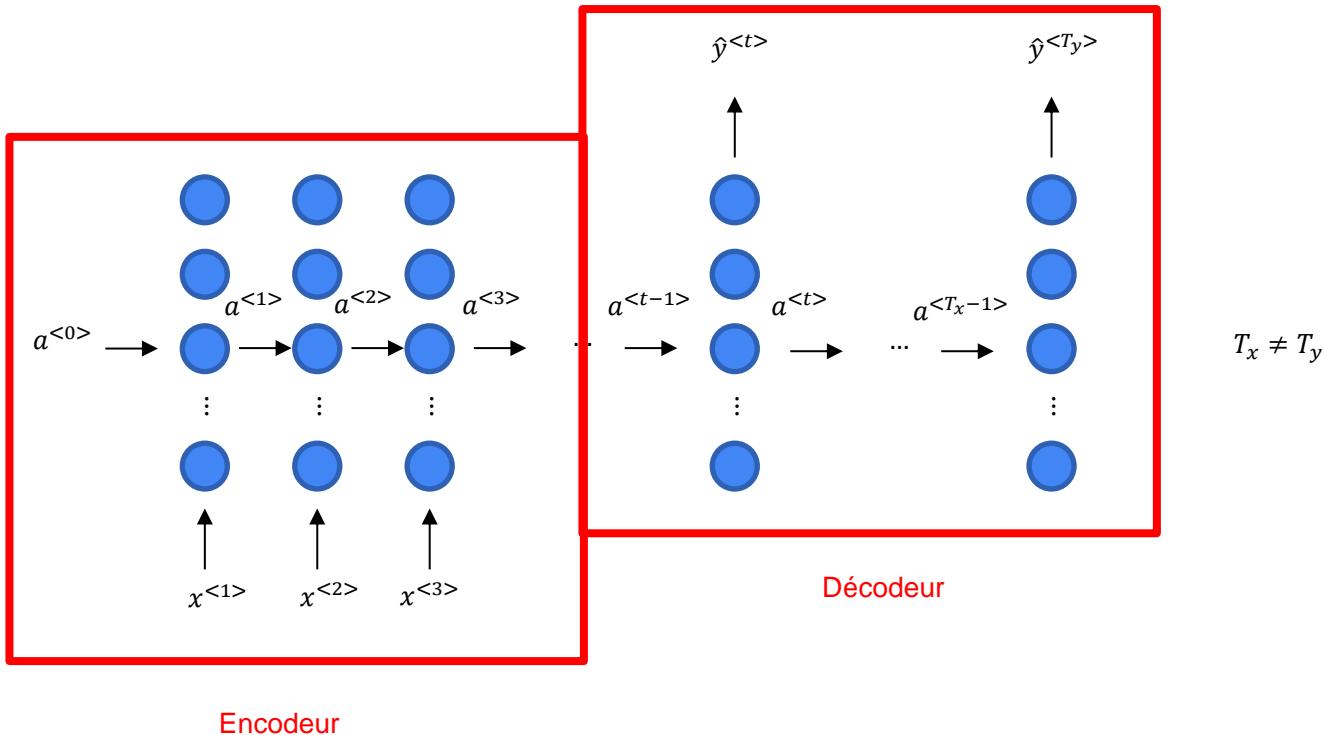


Many to many



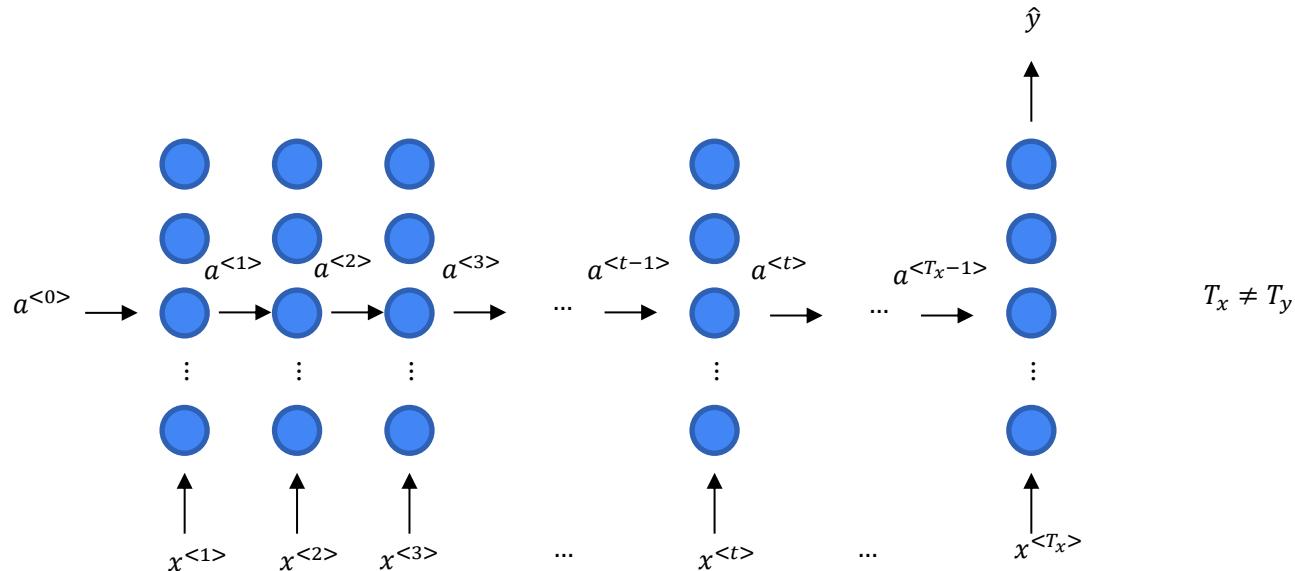


Many to many



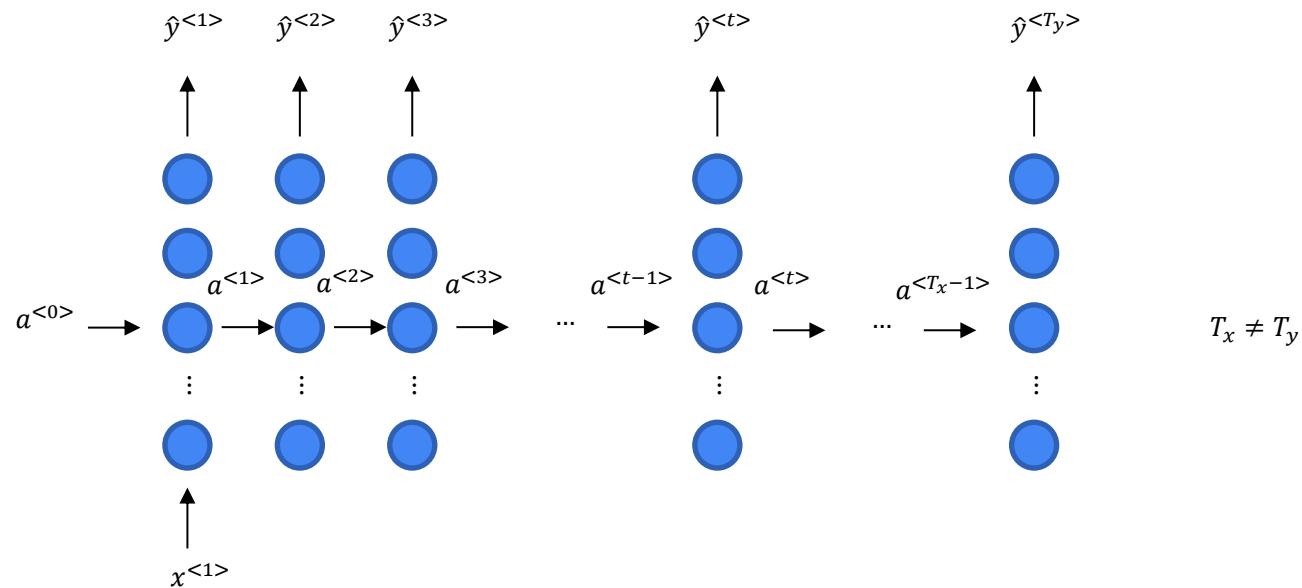


Many to one



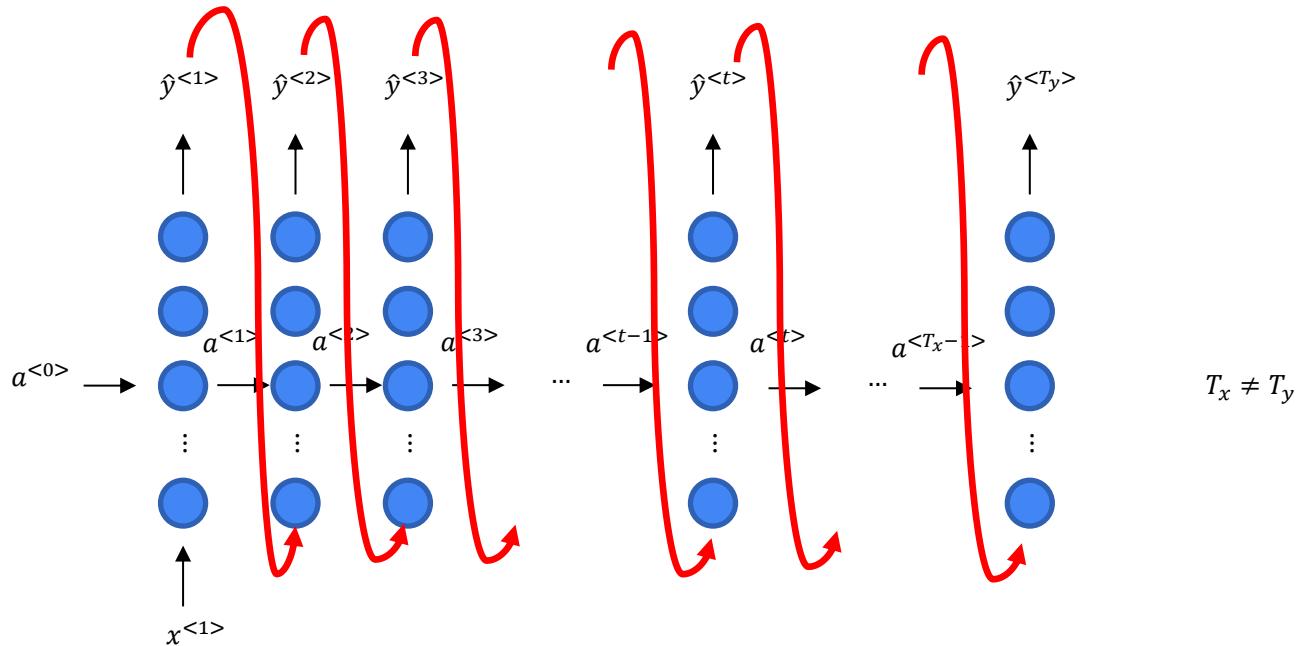


One to many





One to many



Bidirectional RNN (BRNN)



Partie 9 : Recurrent Nerual Network



De l'information manquantes

He said, “Teddy Roosevelt was a great President.”



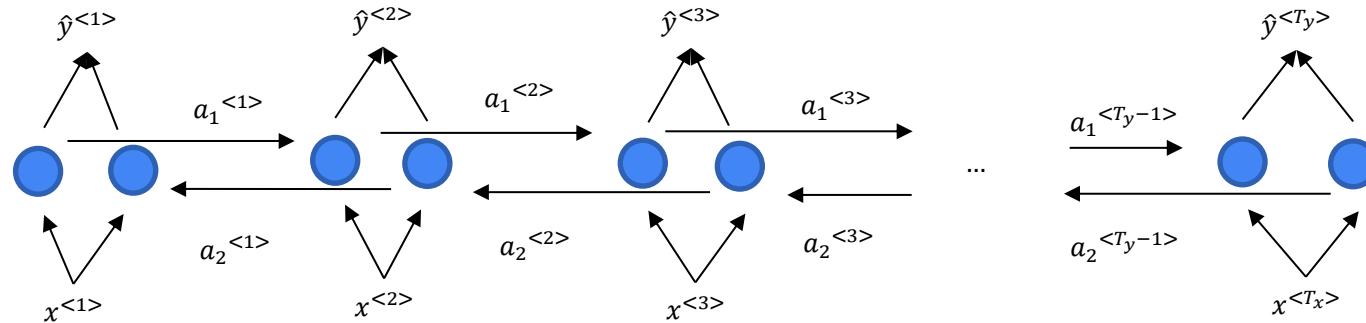
He said, “Teddy bears are on sale!”





Bidirectional RNN (BRNN)

$$\hat{y}^{<t>} = g(w_y[a_1^{<t>}, a_2^{<t>}] + b_y)$$



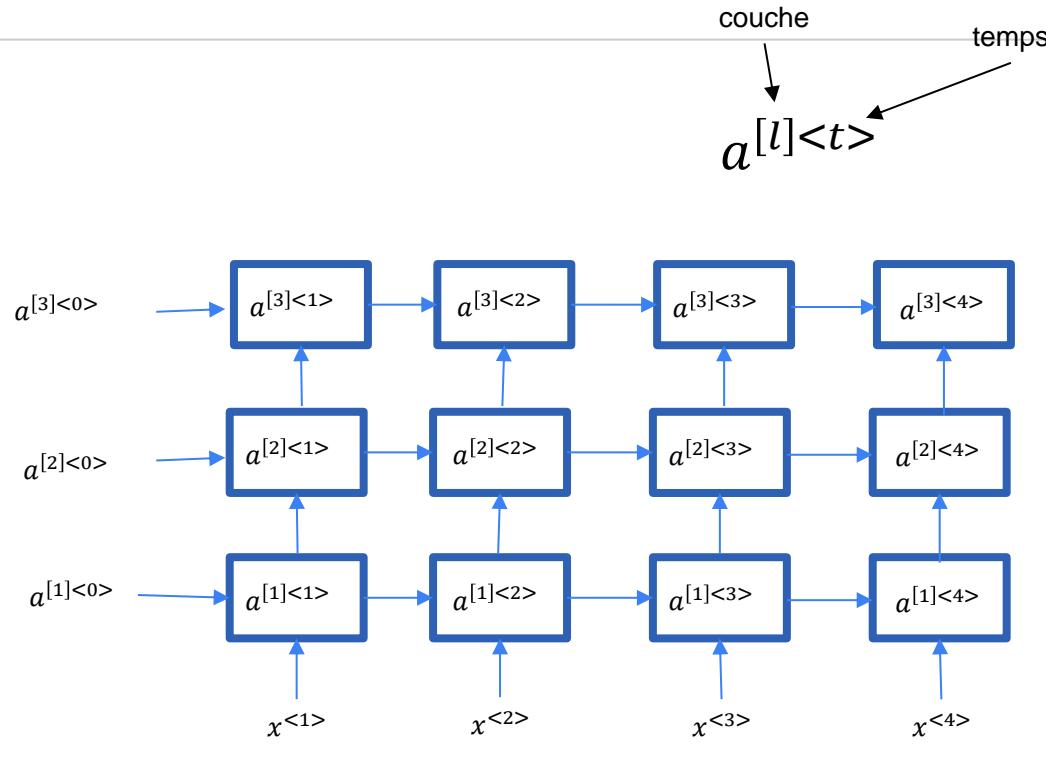
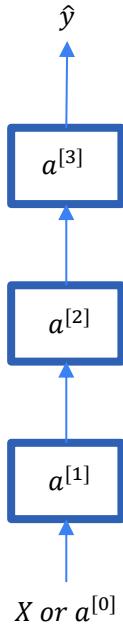
Deep RNN



Partie 9 : Recurrent Nerual Network



Deep RNN



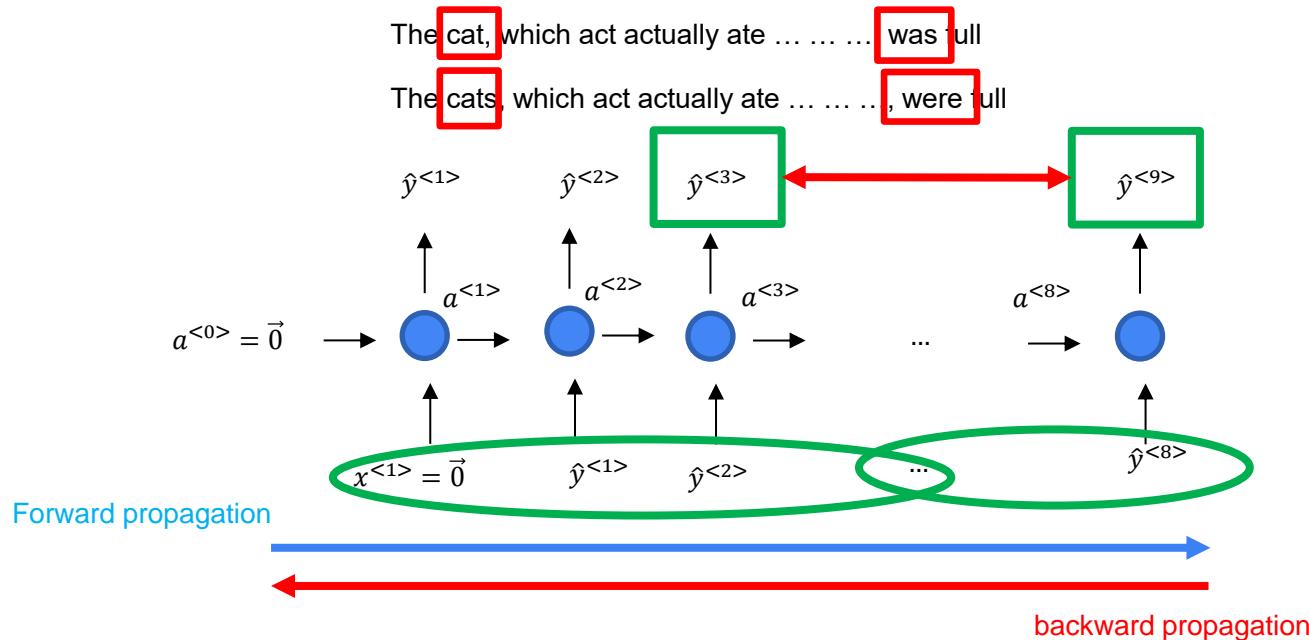
Problème de mémoire



Partie 9 : Recurrent Nerual Network



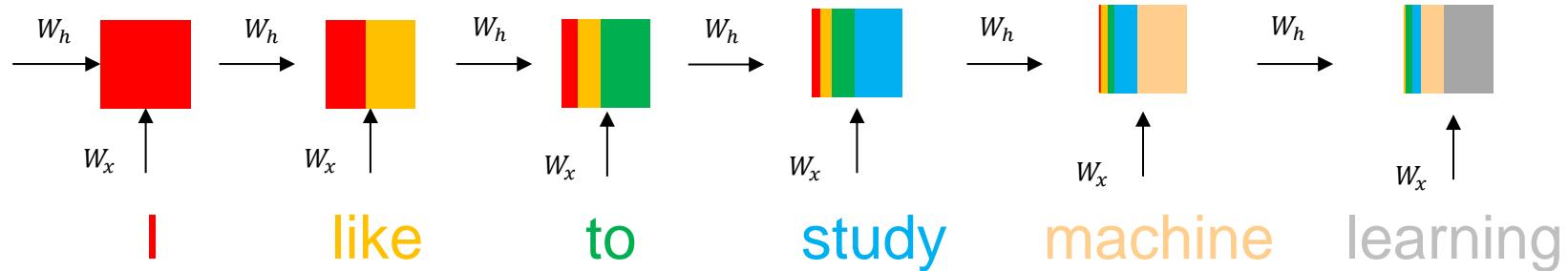
Le problème du RN classique





L'influence des premiers termes

I like to study machine learning _____



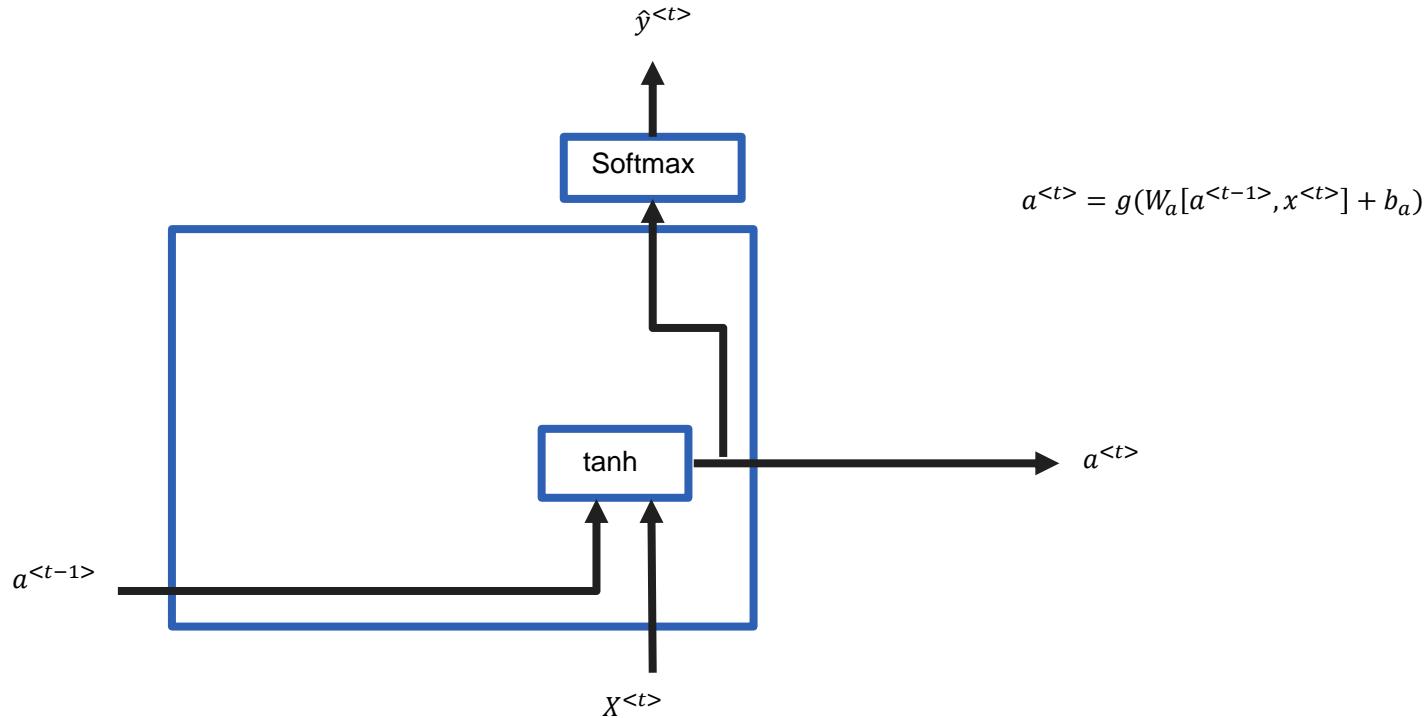
Le Gate Recurrent Unit (GRU)



Partie 9 : Recurrent Nerual Network

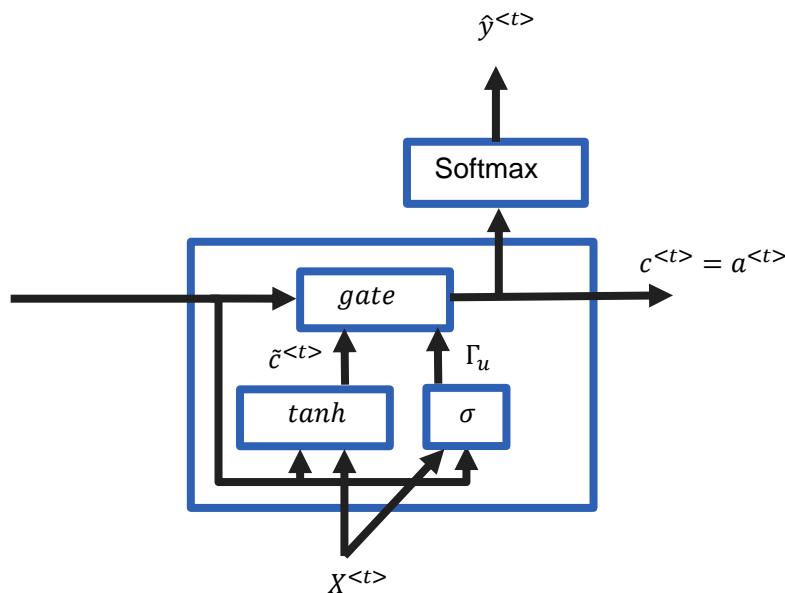


Recurrent Neural Network





Gate Recurrent Unit (GRU) (simplified)



σ est la fonction sigmoïd

c = unité de mémoire

$$c^{<t>} = a^{<t>}$$

$$\tilde{c}^{<t>} = \tanh(w_c[c^{<t-1>}, X^{<t>}] + b_c)$$

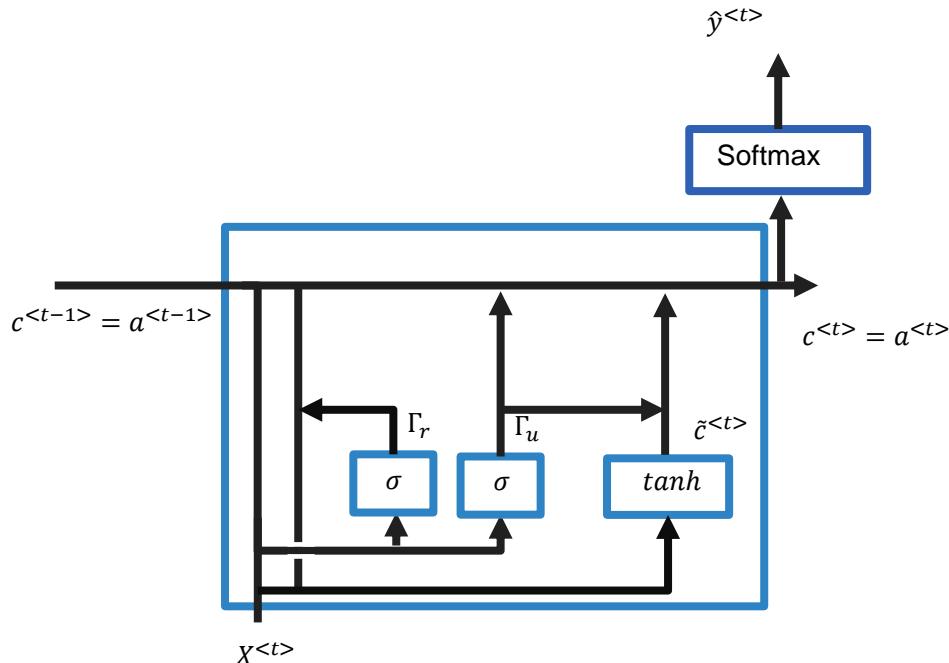
$$\Gamma_u = \sigma(w_u[c^{<t-1>}, X^{<t>}] + b_u)$$

$$0 < \Gamma_u < 1$$

$$c^{<t>} = \Gamma_u * \tilde{c}^{<t>} + (1 - \Gamma_u) * c^{<t-1>}$$



Gate Recurrent Unit (GRU)



c = unité de mémoire

$$c^{<t>} = a^{<t>}$$

$$\tilde{c}^{<t>} = \tanh(w_c[c^{<t-1>} * c_{\text{Forget}}^{<t-1>}, X^{<t>}] + b_c)$$

$$\Gamma_u = \sigma(w_u[c^{<t-1>}, X^{<t>}] + b_u)$$

$$\Gamma_r = \sigma(w_r[c^{<t-1>}, X^{<t>}] + b_r)$$

$$c^{<t>} = \Gamma_u * \tilde{c}^{<t>} + (1 - \Gamma_u) * c^{<t-1>}$$

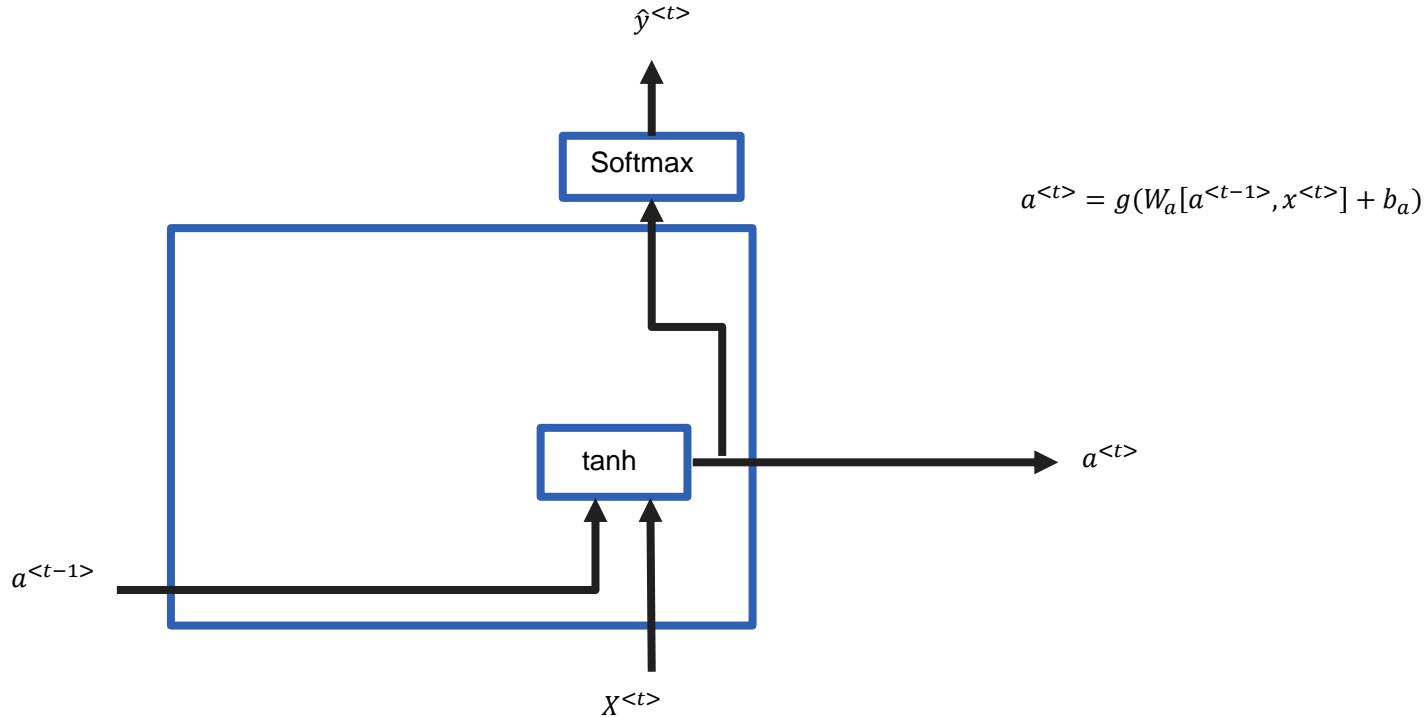
Le Gate Recurrent Unit (GRU)



Partie 9 : Recurrent Nerual Network



Recurrent Neural Network





Gate Recurrent Unit and Long Short Term Memory

GRU

$$\tilde{c}^{<t>} = \tanh(w_c[\Gamma_r * c^{<t-1>}, X^{<t>}] + b_c)$$

$$\Gamma_u = \sigma(w_u[c^{<t-1>}, x^{<t>}] + b_u)$$

$$\Gamma_r = \sigma(w_r[c^{<t-1>}, x^{<t>}] + b_r)$$

$$c^{<t>} = \Gamma_u * \tilde{c}^{<t>} + (1 - \Gamma_u) * c^{<t-1>}$$

LSTM

$$\tilde{c}^{<t>} = \tanh(w_c[a^{<t-1>}, X^{<t>}] + b_c)$$

$$\Gamma_u = \sigma(w_u[a^{<t-1>}, x^{<t>}] + b_u)$$

$$\Gamma_f = \sigma(w_f[c^{<t-1>}, x^{<t>}] + b_f)$$

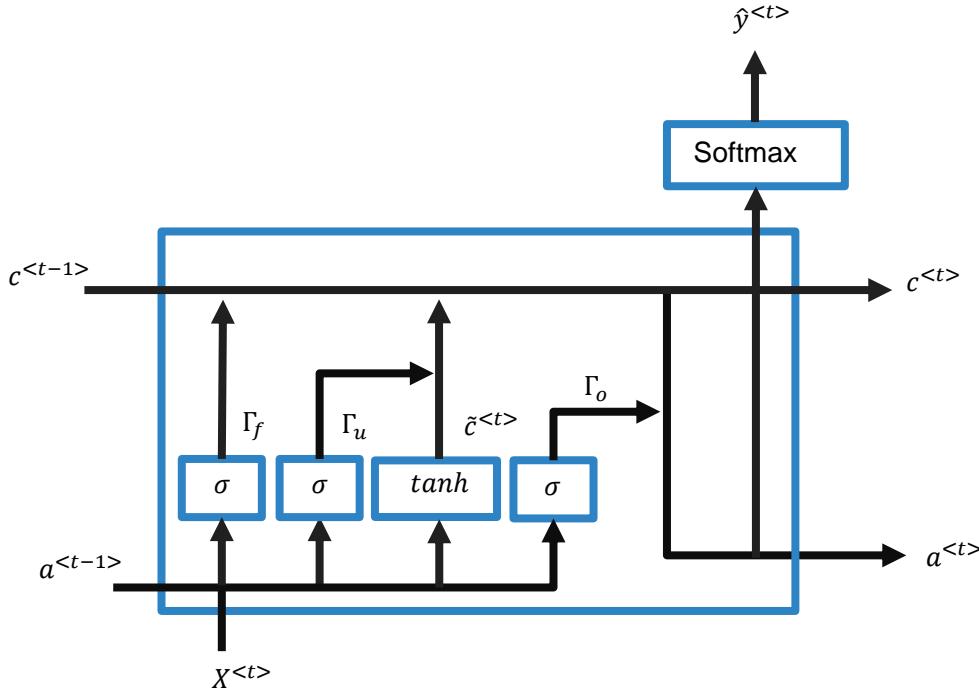
$$\Gamma_o = \sigma(w_o[c^{<t-1>}, x^{<t>}] + b_o)$$

$$c^{<t>} = \Gamma_u * \tilde{c}^{<t>} + \Gamma_f * c^{<t-1>}$$

$$a^{<t>} = \Gamma_0 * \tanh(c^{<t>})$$



Long Short Term Memory (LSTM)



$$\tilde{c}^{<t>} = \tanh(w_c[a^{<t-1>}, X^{<t>}] + b_c)$$

$$\Gamma_u = \sigma(w_u[a^{<t-1>}, X^{<t>}] + b_u)$$

$$\Gamma_f = \sigma(w_f[a^{<t-1>}, X^{<t>}] + b_f)$$

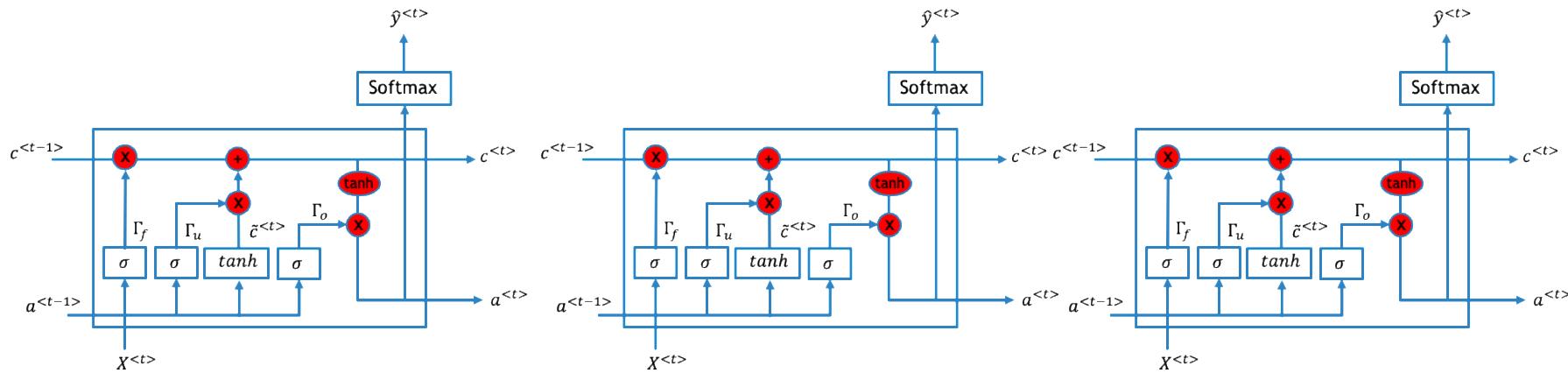
$$\Gamma_o = \sigma(w_o[a^{<t-1>}, X^{<t>}] + b_o)$$

$$c^{<t>} = \Gamma_u * \tilde{c}^{<t>} + \Gamma_f * c^{<t-1>}$$

$$a^{<t>} = \Gamma_o * \tanh(c^{<t>})$$



LSTM en série



Deep learning avec Pytorch

Partie 11 : Deep learning appliqué au son



Présenté par **Morgan Gautherot**

Comprendre une onde sonore



Partie 1 : Deep learning appliqué au son



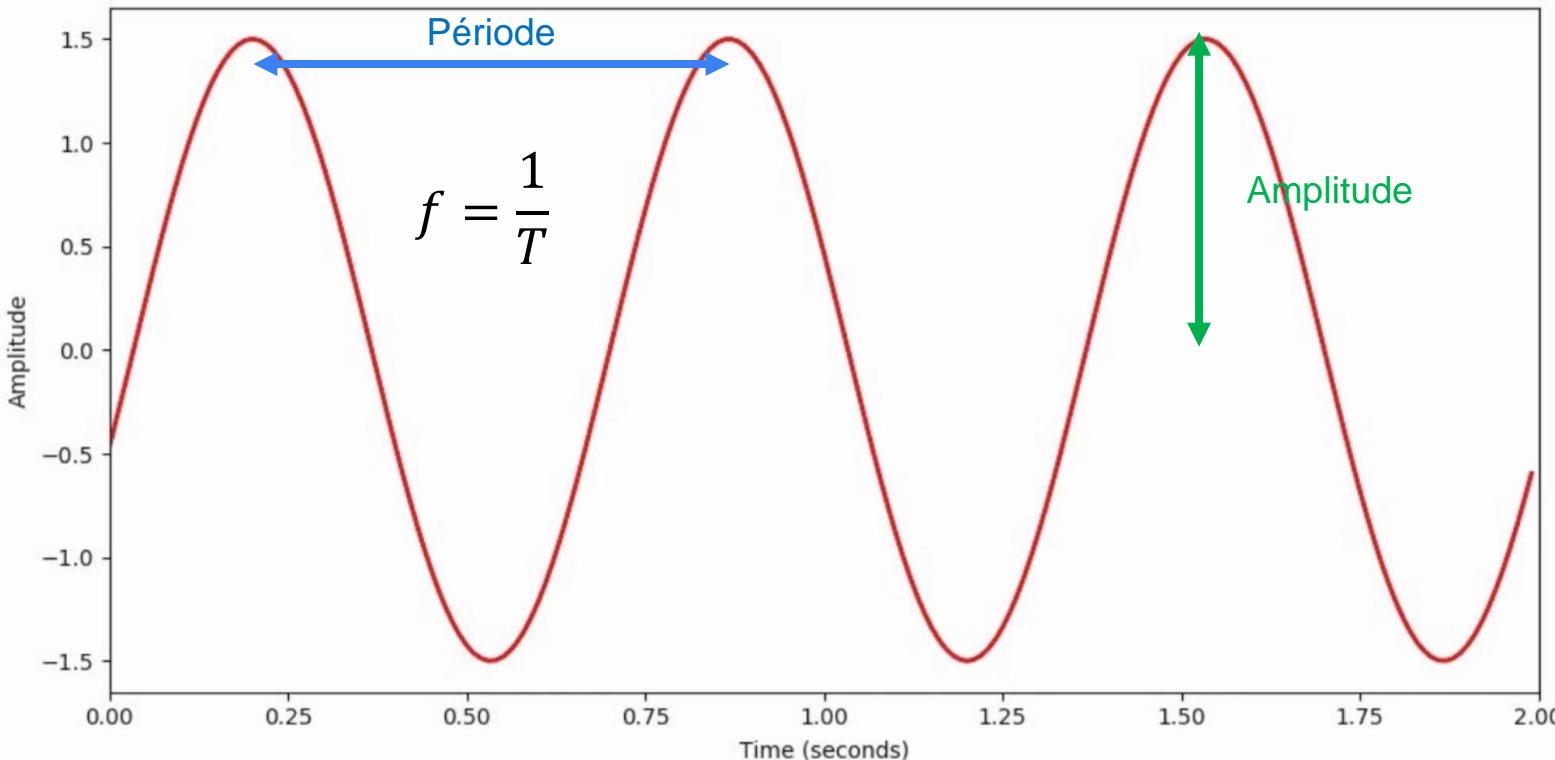
Le son

- Produit par la vibration d'un objet
- Les vibrations déterminent l'oscillation des molécules d'air
- L'alternance de la pression de l'air provoque une onde

$$y(t) = A \cdot \sin(2\pi ft + \varphi)$$

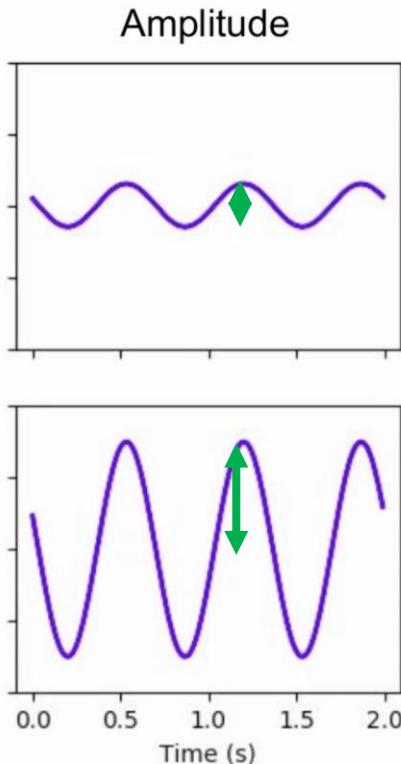
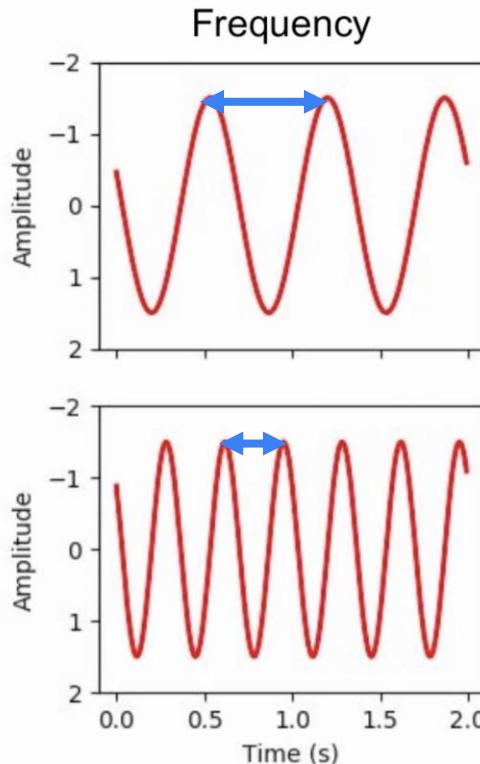


Onde

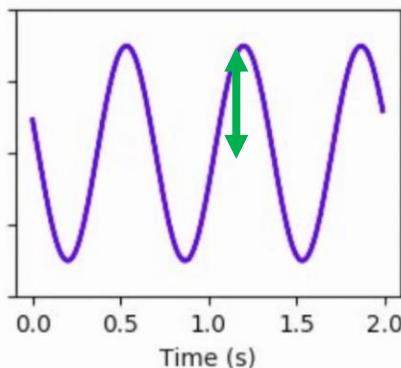
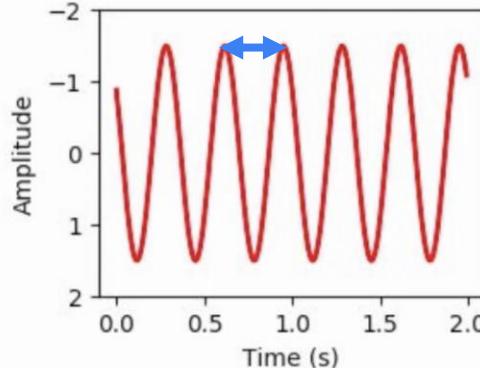




Fréquence/hauteur et amplitude/intensité



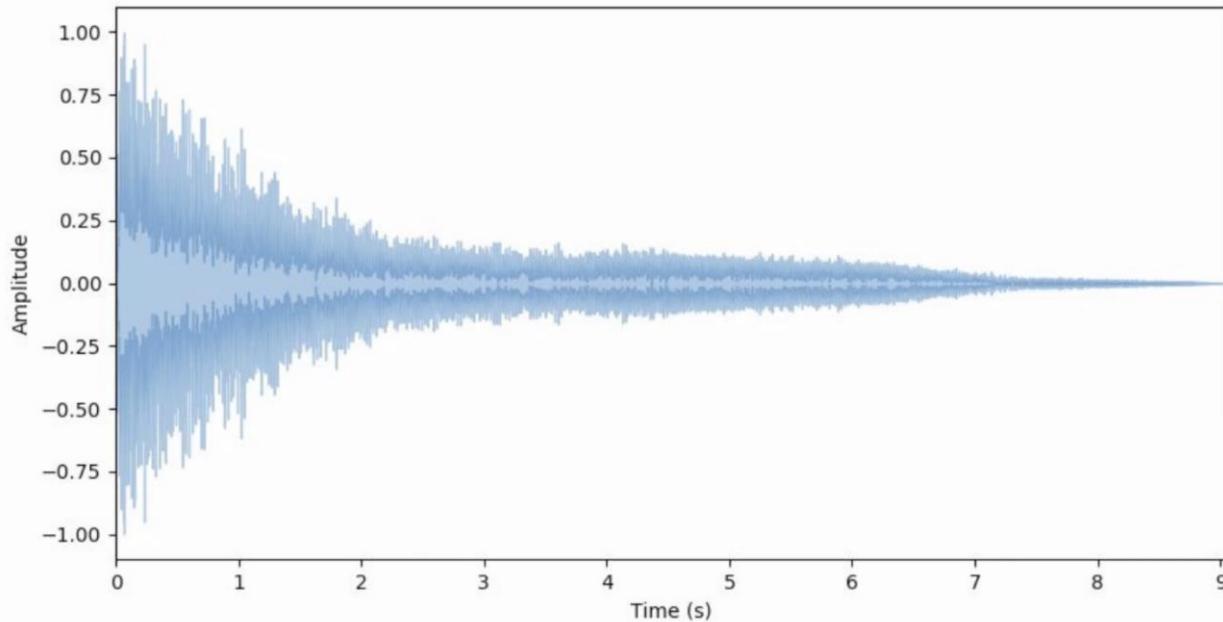
Plus grande fréquence -> son plus aigu



Plus grande amplitude -> Volume plus important



Une onde réel (note de piano)





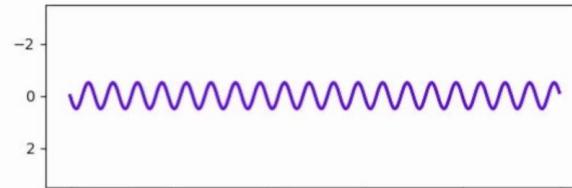
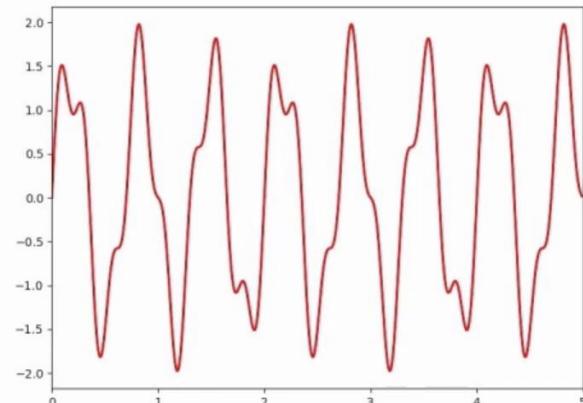
La transformation de Fourier

Décomposer un son périodique complexe en une somme d'ondes sinusoïdales oscillant à différentes fréquences.



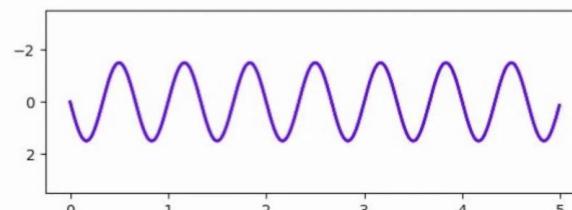
La transformation de Fourier

Plus l'amplitude est élevé plus le son contribue à la complexité du son d'origine



$$A_1 = 0.5$$

$$A_2 = 1.5$$



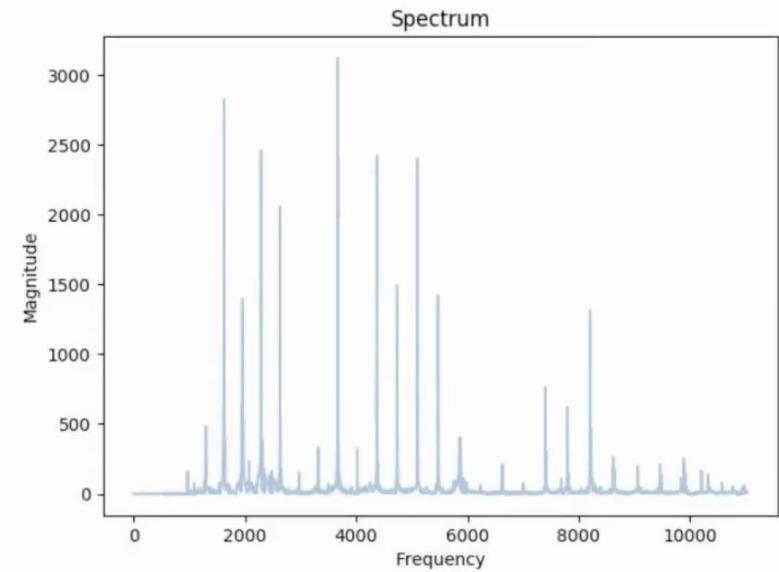
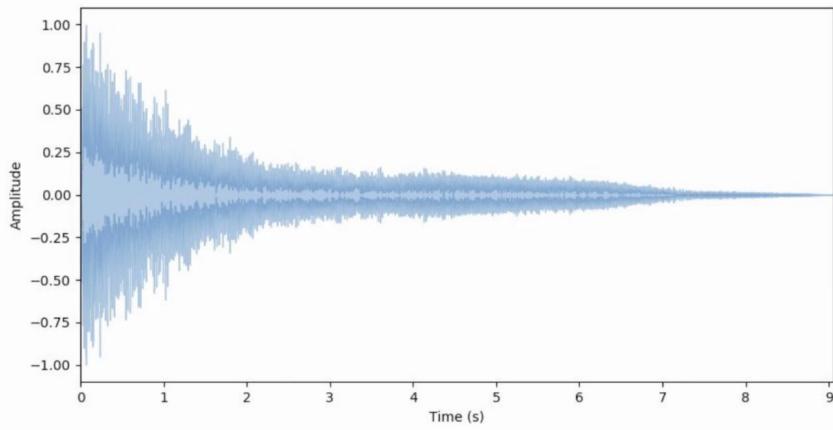
$$f_1 = 4 \quad \varphi_1 = 0$$

$$f_2 = 1.5 \quad \varphi_2 = 0$$

$$S = A_1 \sin(2\pi f_1 t + \varphi_1) + A_2 \sin(2\pi f_2 t + \varphi_2)$$



La transformation de Fourier d'une onde réelle



Perte d'information sur le temps

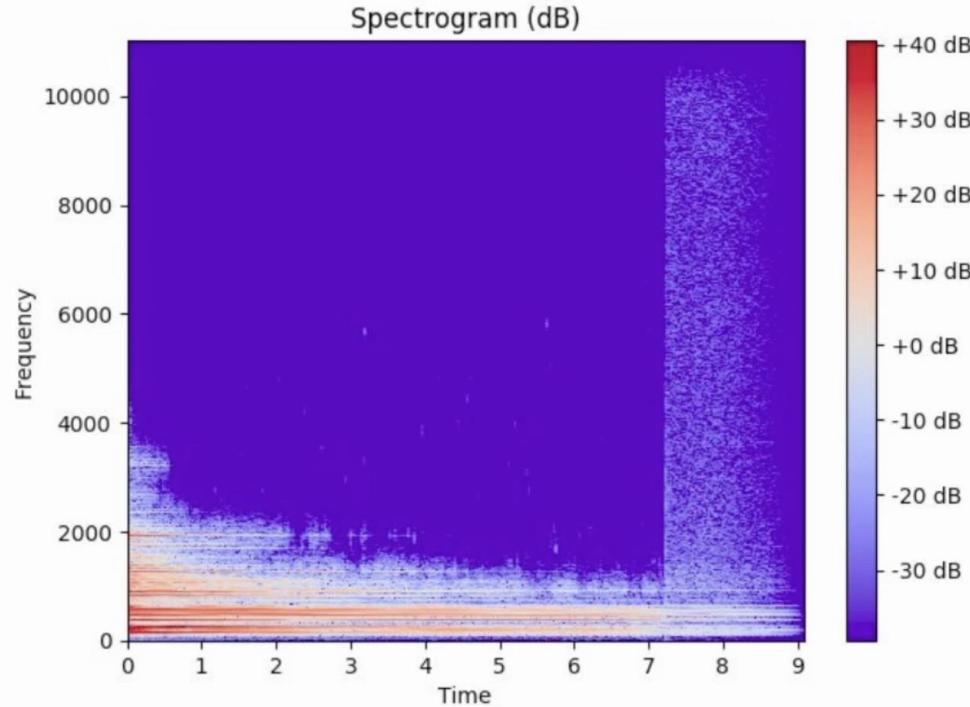


Transformée de Fourier à temps court (STFT)

- Calcul de plusieurs FFT à différents intervalles
- Préserve les informations temporelles
- Taille de la fenêtre de temps fixe
- Donne un spectrogramme (temps + fréquence + amplitude)

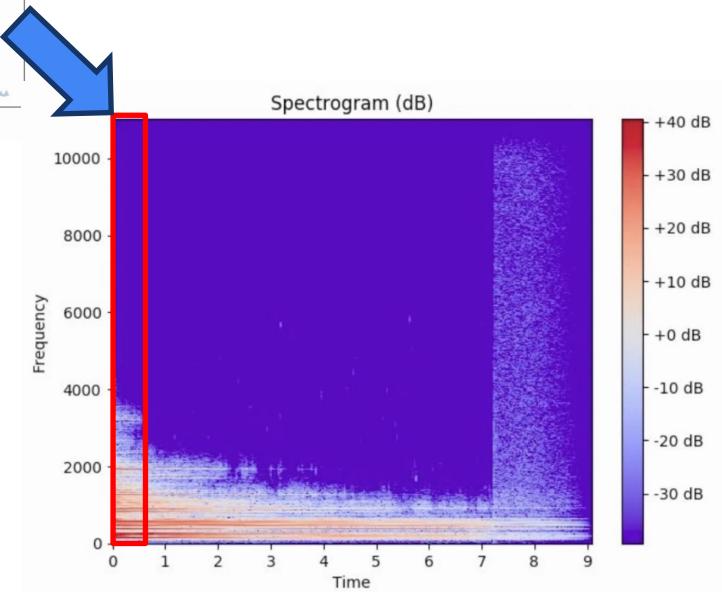
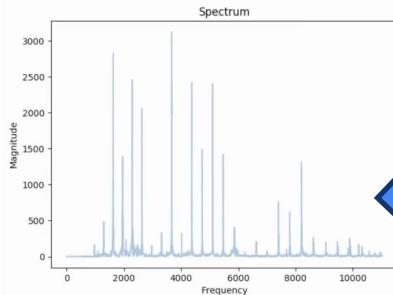
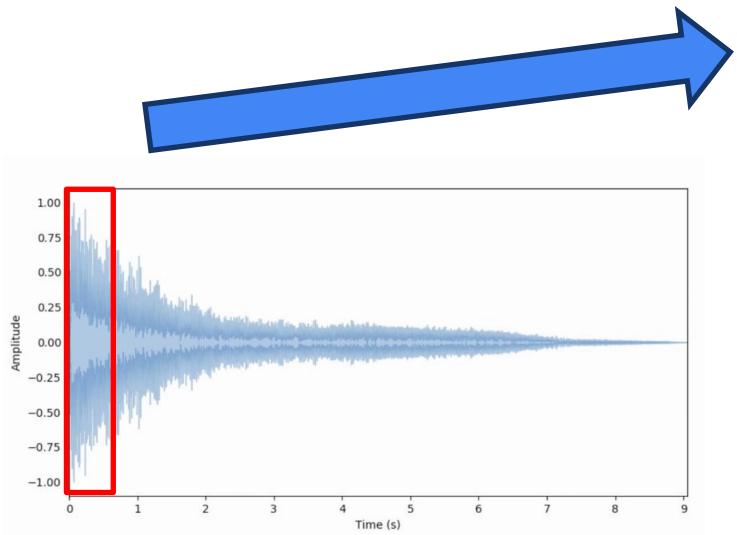


Transformée de Fourier à temps court (STFT)



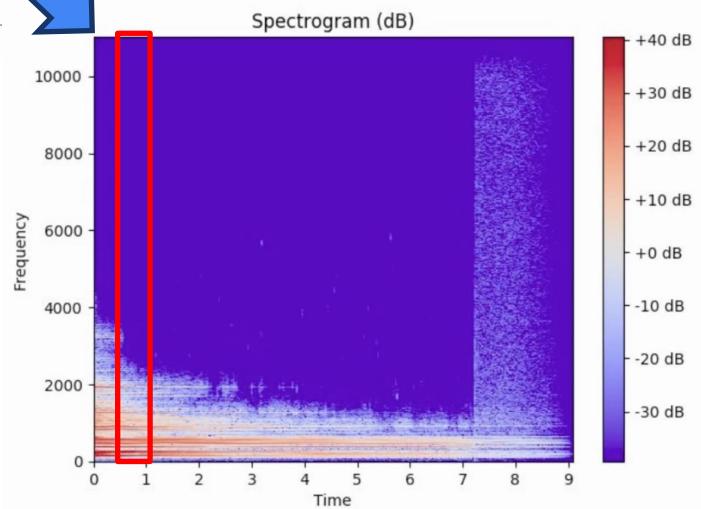
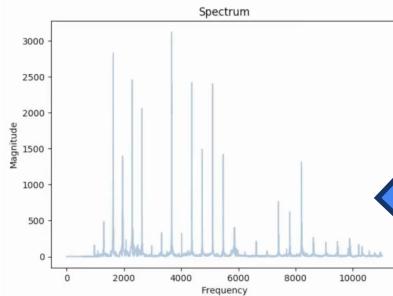
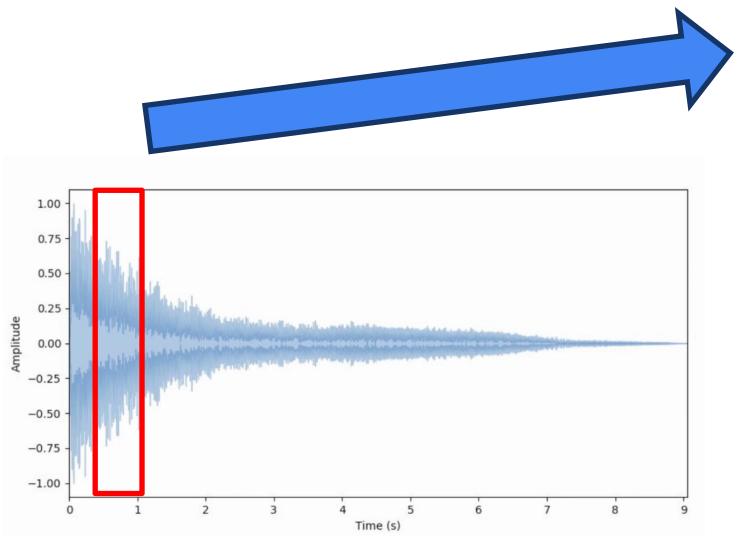


D'une onde au spectrogramme



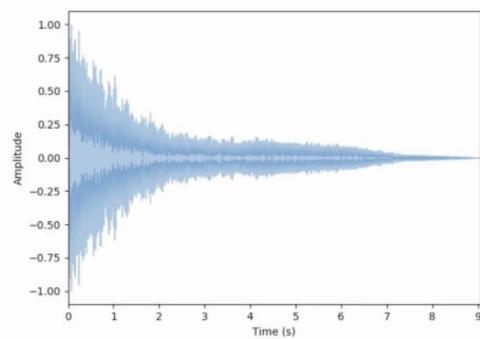


D'une onde au spectrogramme

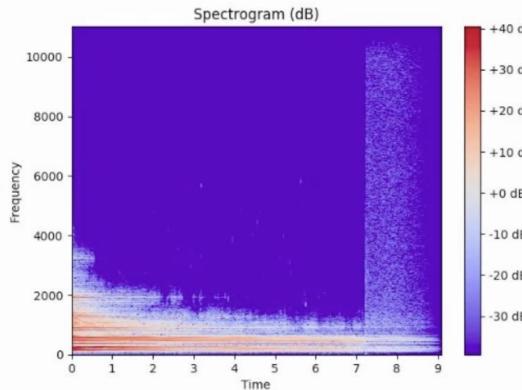




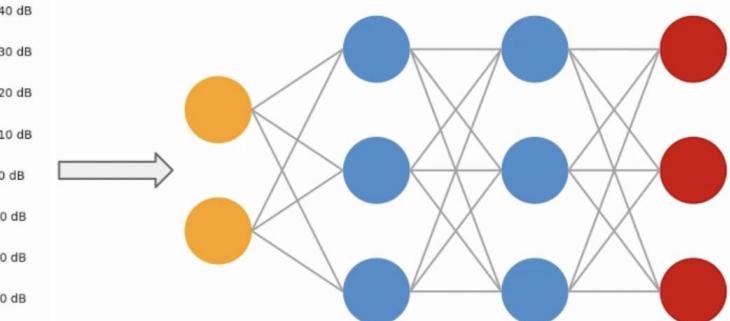
Pourquoi c'est important en deep learning ?



STFT



→



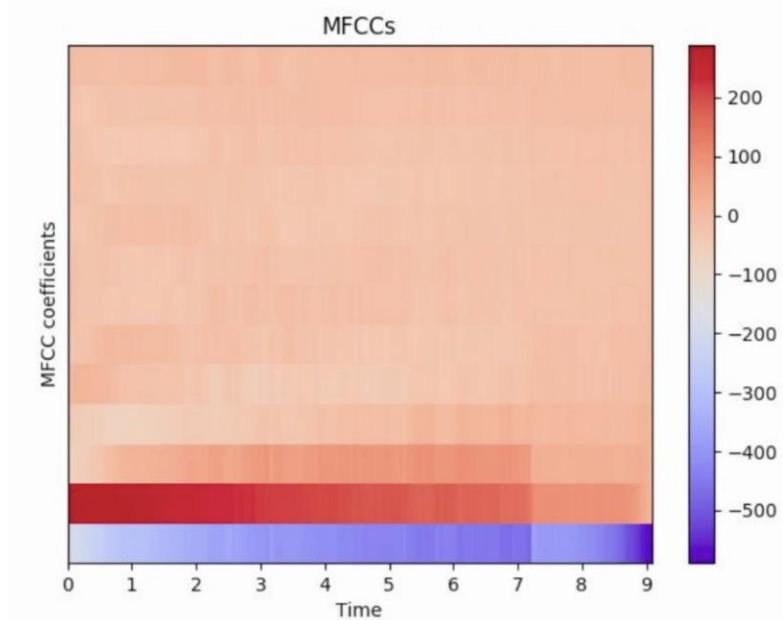
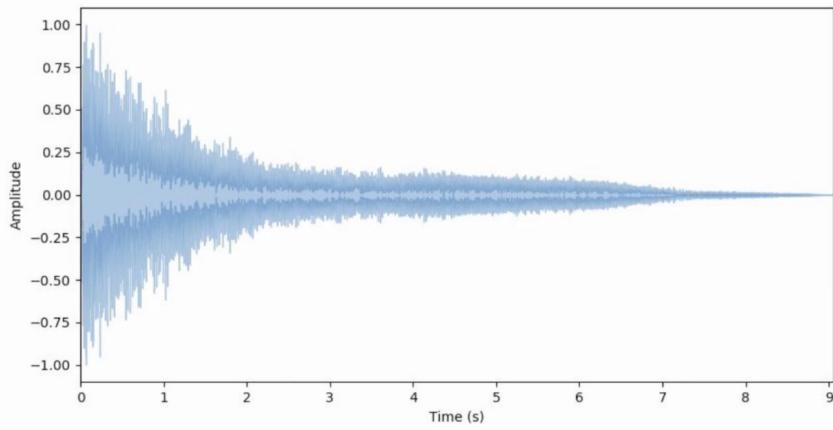


Mel frequency Cepstral Coefficients (MFCCs)

- Capture des aspects de timbres/textures du son
- Caractéristiques du domaine des fréquences
- Approche du système auditif humain
- 12 à 40 coefficients
- Calculés à chaque fenêtre de temps



D'une onde au MFCCs



Deep learning Pytorch

Partie 12 : Le natural language processing



Présenté par **Morgan Gautherot**



Le Big Data

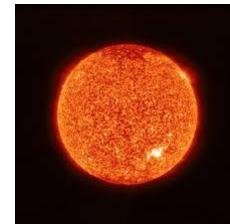
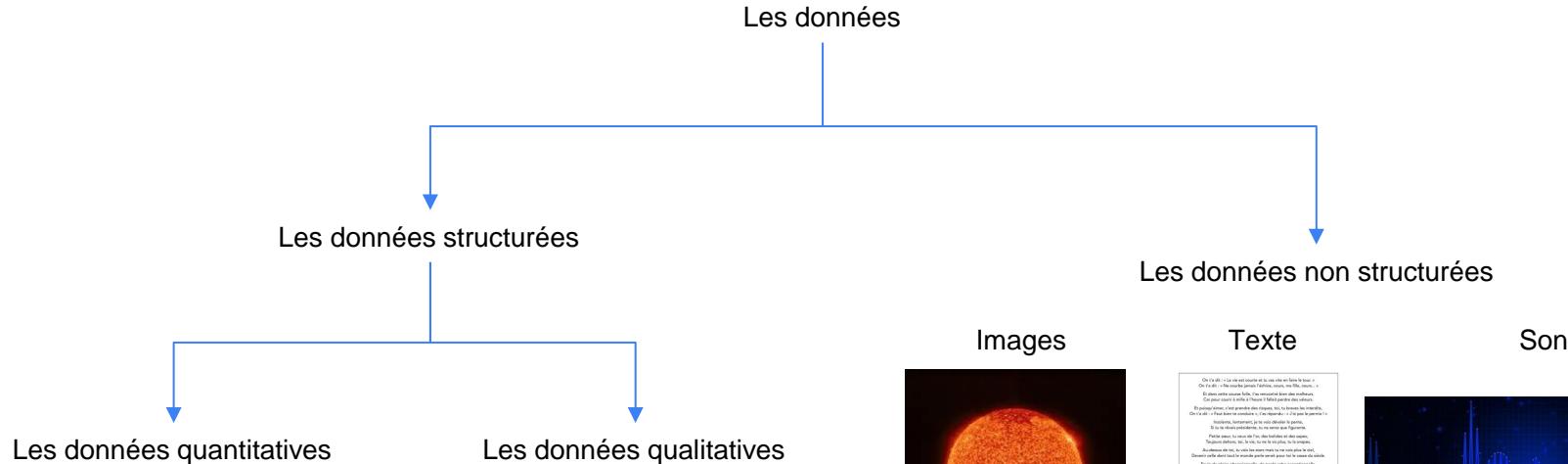
Seulement 21% de ces données sont structurées !

BIG DATA





Les catégories de données



On y va : <je me suis couché et je me suis enfin le bœuf>
On l'a fait : <je me suis couché et je me suis enfin le bœuf>
C'est pas mal : <je me suis couché et je me suis enfin le bœuf>
Ca pour courir à côté à l'heure d'aller prendre des enfants
Et puis je suis allé faire une sieste dans la chambre de ma femme,
On l'a fait : <je me suis couché et je me suis enfin le bœuf>
C'est pas mal : <je me suis couché et je me suis enfin le bœuf>
Si je te disais que j'étais dépendant, tu ne serais que stupide.
Tu n'as pas de problème, tu n'as pas de problème.
Tout ça dépend de ce que tu veux bien dire mais je ne sais pas.
D'accord avec tout ce que tu veux bien dire mais je ne sais pas.
Encore une fois dépendant, je garde cette愚蠢itude.
Tout ça dépend de ce que tu veux bien dire mais je ne sais pas.
Dis ce que tu veux bien dire, pour dire ce que tu veux bien dire.
C'est pas mal : <je me suis couché et je me suis enfin le bœuf>
C'est pas mal : <je me suis couché et je me suis enfin le bœuf>
Ta mère ne m'a rien donné et c'est que lorsque dans l'appartement,
Ta mère ne m'a rien donné et c'est que lorsque dans l'appartement,
A trop court après de venir, j'en ai acheté le Disney
A trop court après de venir, j'en ai acheté le Disney
C'est pas mal : <je me suis couché et je me suis enfin le bœuf>
C'est pas mal : <je me suis couché et je me suis enfin le bœuf>





Le traitement naturel du langage

Texte

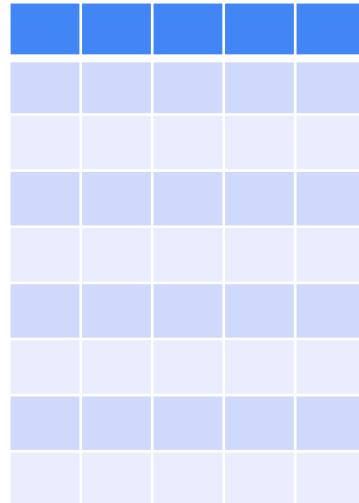
Données non structurées

L'ASSOMMOIR

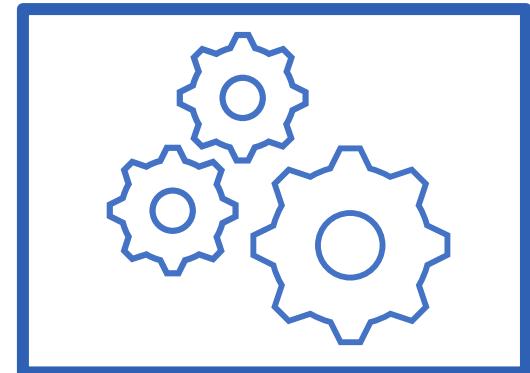
Germaine avait attendu Landier jusqu'à deux heures du matin. Puis, toute frissonnante d'être restée en chemise à l'air vif de la fenêtre, elle s'était assoupi, jetée au travers du lit, heureuse, les joues trempées de larmes. Depuis huit heures, au sortir du Vent d'au-delà, où ils mangiaient, l'envoyait se coucher avec les enfants et ne repartissait que tard dans la nuit, en ressortant qu'il cherchait du travail. Ce soir-là, pendant qu'elle guettait son retour, elle croyait l'avoir vu entrer au bal du Grand-Balcon, dont les dix fenêtres flamboyantes éclairaient d'une nappe d'incendie la coulée soire des boulevards extérieurs; et, derrière lui, elle avait aperçue la petite Adèle, une bruisseuse qui dinait à leur restaurant, marchant à cinq ou six pas, les mains ballantes, comme si elle venait de lui quitter le banc pour ne pas passer ensemble sous la clarté crue des gloires de la porte.

Data frame

Données structurées



Modélisation



Deep learning par la pratique

Leçon 2 : Les prétraitements classiques



Présenté par **Morgan Gautherot**



La tokenization

Transformer un texte en vecteur de mots

« Le machine learning est un apprentissage par minimisation de l'erreur !»



['Le', 'machine', 'learning', 'est', 'un', 'apprentissage', 'par', 'minimisation', 'de', 'l'erreur', '!']



La normalisation

Suppression de la ponctuation, enlever les apostrophes, mettre en minuscule.

```
[ 'Le', 'machine', 'learning', 'est', 'un', 'apprentissage', 'par', 'minimisation', 'de', 'l'erreur', '!' ]
```



```
[ 'le', 'machine', 'learning', 'est', 'un', 'apprentissage', 'par', 'minimisation', 'de', 'le', 'erreur' ]
```



Les stop words

Les mots de liaison qui n'apportent pas de sens mais du bruit

['Le', 'machine', 'learning', 'est', 'un', 'apprentissage', 'par', 'minimisation', 'de', 'l'erreur', '!']



['machine', 'learning', 'apprentissage', 'minimisation', 'erreur']



Doublons

Plusieurs mots avec le même sens mais une orthographe différente.

« Julien est plus grand que toi, mais tu es plus grande que Sophie. Julien et Sophie sont plus grands que Marie. »



[‘grand’, ‘grande’, ‘grands’]



Lemmatisation

Transformer un mot en son sens radical (lemme)

« Julien est plus grand que toi, mais tu es plus grande que Sophie. Julien et Sophie sont plus grands que Marie. »



[‘grand’, ‘grand’, ‘grand’]



Stemming

Supprimer la fin des mots

« Julien est plus grand que toi, mais tu es plus grande que Sophie. Julien et Sophie sont plus grands que Marie. »



[‘gran’, ‘gran’, ‘gran’]



Résumé

- Tokenization
- Normalisation
- Suppression des stop words
- Lemmatisation / Stemming

Deep learning par la pratique

Leçon 3 : Bag of words



Présenté par **Morgan Gautherot**



Le vecteur de vocabulaire

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



D'une phrase à un bag of words

Doc 1

Le thé est bouillant



Traitements du texte

Doc 1

['thé', 'être', 'bouillant']



Transformation en bag of words

$$V = [arachide, \dots, bain, bouillant, \dots, \text{être}, \dots, \text{thé}, \dots, voiture, \dots, zumba]$$

$$D_1 = [\quad 0, \dots, \quad 0, \quad 1, \dots, \quad 1, \dots, \quad 1, \dots, \quad 0, \dots, \quad 0]$$



D'un corpus à un dataframe

toncoeurbat

Tu me donnes envie de faire table rase du passé,
et de t'aider à faire table rase du tien. Tu me
donnes envie d'être pour toi ce que je n'ai jamais
été pour quelqu'un d'autre. Tu me donnes
envie de l'offrir mon cœur, mes ambitions,
mes envies, enfin tout le plaisir du danger. Tu me
donnes envie de t'en faire trembler le moins
possible entier. Tu me donnes envie de parler de
sentiments et de bonheurs tout en empêchant
mes verbes au présent. Tu me donnes envie
de devenir quelqu'un qu'il apportera fierté et
douceur. Tu me donnes envie de décrocher la
Lune, de croire en l'impossible. Tu me donnes
envie de vivre, bordel. Parce que tu es l'éincelle
au fond de mes yeux, mon sourire terriblement
sincère. Parce que je suis plus seule désormais
tu es là. Et c'est si précieux.

Le supermarché

Sophie fait les courses au supermarché avec son épouse, Clément. Après avoir pris un souffle sur le parking, elle s'apprête à chercher dans leur sac pour prendre une pomme. Mais l'épouse de Clément est revenue. Les courses sont remises à bouteilles et denrées. Sophie sort sa liste des courses pour ne pas oublier. Elle regarde les fruits et légumes, puis la viande et les poissons, les fromages, et la pâtes maison qu'elle mettait dans son sac. Clément est turbulente. Il court sur l'île, ne cesse d'arracher des bouteilles de biscuits, de chocolat, de bonbons et de gâteaux. Sophie se sent dépassée par ce ménage. Il est déjà assez heureux. Elle doit se débrouiller pour tenir ce que son épouse a envie. Zénith dévisse chez la voisine pour lui faire une petite visite. Sophie termine sa course, elle se dirige vers la case où sa petite amie cultive son huile. A la case, elle place ses achats sur le rebord de la fenêtre et attend que Sophie ait terminé ses courses. Puis elle débute. Elle débute, que faire pour empêcher Sophie de venir. Elle la repousse au fond de l'île puis se dirige vers son véhicule. Elle enlève la couvercle et met les courses dans le coffre. Sophie se débrouille pour empêcher Sophie d'entrer. Elle part retrouver Zénith et préparer la soirée.

$$= D_1$$

Doc 3

$$= D_3$$

006

L'amitié est un repère,
une ligne d'horizon. Une lumière
et une porte secrète vers les émotions.
Elle nous aide par temps durs, elle fait
rassurer. Elle ne juge pas, elle accepte
elle est là... Tantôt près, tantôt loin,
mais toujours à distance de coeurs, notre
amitié c'est du bonheur. Notre amitié, c'est
une plante que je cultive avec grands soins
le trésor de mon petit jardin. Une pierre
précieuse qui m'éclaire dans le noir, et
quelqu'un qui connaît toute mon histoire.
J'avance sur mon chemin, tu avances devant
le tien. L'horizon devant nous est infini.
tout comme notre amitié, merci.

On l'a dit : « La vie n'est course et tu vas vite en faire le tour. »
On l'a dit : « Ne courre jamais l'adrine, cours, ma fille, cours... »
Et dans cette course fâle, t'es rencontré bien des réalisateurs.
Cours pour courir à mille à l'heure. Il fallait perdre des valeurs.
Et puisque j'aurai, cette prenne des risques, tel, tu bravais les insécurités
On l'a dit : « Faudra bien te convaincre. » C'est pour le plaisir.

Partie tout le temps de la fin, des bûches et des sapins,
Qu'il se déroule pendant que l'on va se promener.
Tous deux écoutent, lui, le son, le bruit, la voix,
Dessine un cercle tout autour de la partie pour venir à ce résultat.
Ensuite de grande étonnement, de grande surprise exceptionnelle,
Pour l'assurance personnelle, l'heureux couple.
Deux personnes qui ont été dans la même situation, mais qui ont fait des choses
Le plus qu'il leur est arrivé si sûre pour que star fût dans leur cœur.
Tu vois alors, tu vois du moins et n'est que de la poussière,
Mais il y a une chose qui est vraiment importante, c'est que tu sois dans la partie.
Les meilleures personnes d'aujourd'hui qui ont été dans l'ambulance.
Dans le service il fallait tenir fermes et dévoués en ambulance.
A trop empêcher alors, c'est, vraiment le Dr.
Ainsi que l'heureux couple, l'heureux couple.
Cela qui, quoi que tu gagnes, tu ne pourras pas,
Qui t'aime alors que tout le monde t'a, tu finiras associé.
Cela que je t'aime, je t'aime, tu domineras sans que gag
L'heureux couple, l'heureux couple, l'heureux couple.
Cet an' l'hôpital qui se repose, tel au seules chances de mourir.
Commerce pour te changer tôt ou tard et tu pourras changer le man

Doc 1

$$= D_2$$

Doc 4

$$= D.$$

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						
mot_m	0	1	0	0	0	0

Deep learning par la pratique

Leçon 4 : TF-IDF



Présenté par **Morgan Gautherot**



Définition du TF-IDF

TF = Term Frequency

IDF = Inverse Document Frequency



Term Frequency

$$tf_{w,d} = \frac{n_{w,d}}{\sum_k n_{k,d}}$$

d est un document de notre jeu de données

w est un mot de notre document

$n_{w,d}$ est le nombre d'occurrences du mot w du document d



Inverse Document Frequency

$$idf_w = \log\left(\frac{|D|}{f(w, D)}\right)$$

D désigne tous les documents de notre jeu de données

w est un mot de notre document

$f(w, D)$ est le nombre de document D contenant le mot w



Calcul du TF-IDF

$$tf \cdot idf_{w,d} = tf_{w,d} * idf_w$$

$$tf \cdot idf_{w,d} = \frac{\text{fréquence du terme dans le document}}{\text{fréquence du terme dans le corpus}}$$



Vecteur de vocabulaire

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



D'une phrase à un vecteur

Doc 1

Le thé est bouillant



Traitements du texte

Doc 1

['thé', 'être', 'bouillant']



Transformation en vecteur

$$V = [arachide, \dots, bain, bouillant, \dots, \text{être}, \dots, \text{thé}, \dots, voiture, \dots, zumba]$$

$$D_1 = [\quad 0, \dots, \quad 0, tf.idf_{bouillant}, \dots, tf.idf_{\text{\^etre}}, \dots, tf.idf_{\text{th\'e}}, \dots, \quad 0, \dots, \quad 0]$$



D'un corpus à une matrice

Doc 1

Tu me donnes envie de faire table rase du passé,
et de t'écouter à faire table rase du tien. Tu me
donnes envie d'être pour toi ce que je n'ai jamais
eu être pour quelqu'un d'autre. Tu me donnes
envie de l'offrir mon cœur, mes ambitions,
mon arrière-garde, ma planète. Tu me
donnes envie de t'ouvrir à t'en faire ta planète le
monde entier. Tu me donnes envie de parler de
sentiments et de bonheur tout en conjuguant
mes verbes au présent. Tu me donnes envie
de devenir quelqu'un qui l'apprécie, qui aime et
désire. Tu me donnes envie de déchirer la
laine de l'assouplir dans mes doigts. Tu me
donnes envie de vivre, bordel. Parce que tu es l'étoile
au fond de mes yeux, mon sourire terriblement
sincère. Parce que je ne suis plus seule désormais,
tu es là. Et c'est si précieux.

$= D_1$

Doc 3

006

L'AMITIÉ

L'amitié est un répère,
une ligne d'horizon,
et une porte secrète vers les émotions...
Elle nous aide par temps durs, elle fait rire,
elle nous juge pas, elle accepte,
elle est là... Tantôt près, tantôt loin,
mais toujours à portée de main, notre
amitié c'est du bonheur. Notre amitié, c'est
une plante que je cultive avec soins
malgré de petits jalons. Une autre
précieuse qui m'éclaire dans le noir, et
quelqu'un qui connaît toute mon histoire...
J'avance sur mon chemin, tu avances sur
le tien. L'horizon devant nous est infini,
tout comme notre amitié, merci.

$= D_3$

Doc 2

Le supermarché

Sophie fait les courses au supermarché avec son nouveau compagnon. Ayant pour peu de temps sur le parking, elle a pris un chariot dans lequel Clément s'est installé. Tout deux pédalent jusqu'au supermarché, où Sophie achète des légumes et des fruits pour préparer un repas pour eux deux et leur rendez-vous à l'heure du dîner. Sophie sort sa liste des courses pour ne rien oublier. Elle regarde les marchés et compare les prix. Elle choisit des légumes frais, des fruits et des aliments pour faire plaisir à Clément, mais dans une chanson Clément est turbulente. Il regarde sur son siège, il ne voit pas Sophie et se demande pourquoi il a choisi de venir avec elle. Sophie le prends gentiment par la main et lui dit : « Clément, tu as été très gentil avec moi, mais je ne veux pas que tu partes. » Clément, il est déjà assez heureux. Elle doit se hâter de rentrer pour préparer le repas. Deux heures plus tard, Sophie et Clément sont dans leur cuisine. Sophie termine les courses, elle se dirige vers la caisse où elle attend calmement les clients. Clément est placé sur le plan de travail où il se troupe malin. Elle se réjouit de voir que ce n'est pas le cas. A présent, elle regagne le parking, cherche quelques instants sa voiture. Elle la regarde et se rend compte qu'il faut faire un tour pour la nettoyer. Elle se dirige vers la caisse et lui demande pour elle ramasser Clément dans son siège pour éviter qu'il soit perdu de retrouver Sophie et de préparer pour la soirée.

$= D_2$

Doc 4

On s'a dit : « La vie me sourit et la vie croit que faire le bonheur, c'est pas mal. » Mais la vie me sourit, mais la vie croit, mais la vie...
Et alors que je me sens belle, je me sens belle, je me sens belle...
Pour me sourire à elle à l'heure où l'habitat perdure des valeurs.
Et alors que je me sens belle, je me sens belle, je me sens belle...
Qu'il a dit : « Faud bien me considérer, j'en dépendrai... » Je ne le pensais pas.
Si je te disais ça, je te mentais. Si je te disais ça, je te mentais.
Partie sous le soleil de l'été, baignée dans les rayons.
Tout comme moi, je me sens belle, je me sens belle, je me sens belle...
Avranches de 3h, tu vois les deux mots tu te vois plus le ciel,
tu vois les deux mots tu te vois plus le ciel, tu vois les deux mots tu te vois plus le ciel.
Ensuite de plusieurs heures, de grande émotion et de douleur,
des moments de tristesse, de tristesse, de tristesse, de tristesse.
Des moments que je ne connaissais, pour être honnête faut dire presque,
que je ne connaissais pas. Des moments que je ne connaissais pas.
Tu vois des larmes, tu vois du mal être et tu vois que je le présente,
que je le présente, que je le présente, que je le présente, que je le présente.
Quand tu me regardes, tu me regardes, tu me regardes, tu me regardes, tu me regardes.
Quand tu me regardes, tu me regardes, tu me regardes, tu me regardes, tu me regardes.

MÉLISSA DITI DÉMÉ

$= D_4$

V	D_1	D_2	D_3	D_4	...	D_n
mot_1	0	$tf.idf_{mot1}$	0	0	...	0
mot_2	0	0	0	$tf.idf_{mot2}$...	0
\vdots	\vdots	\vdots	\vdots	\vdots	\vdots	\vdots
mot_m	0	$tf.idf_{mot1}$	0	0	0	0

Deep learning avec Pytorch

Partie 13 : Capturer le sens des mots



Présenté par **Morgan Gautherot**



Les matrices creuses

TF-IDF

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

Bag of words

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						
mot_m	0	1	0	0	0	0



Proximité des mots

Content

=

Papier

=

Heureux

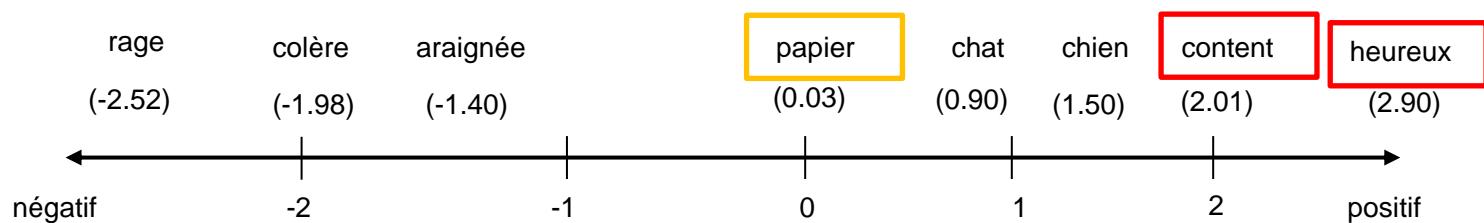
Content = [0, ..., 0, ..., 1, ..., 0, ..., 0, ..., 0]

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

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

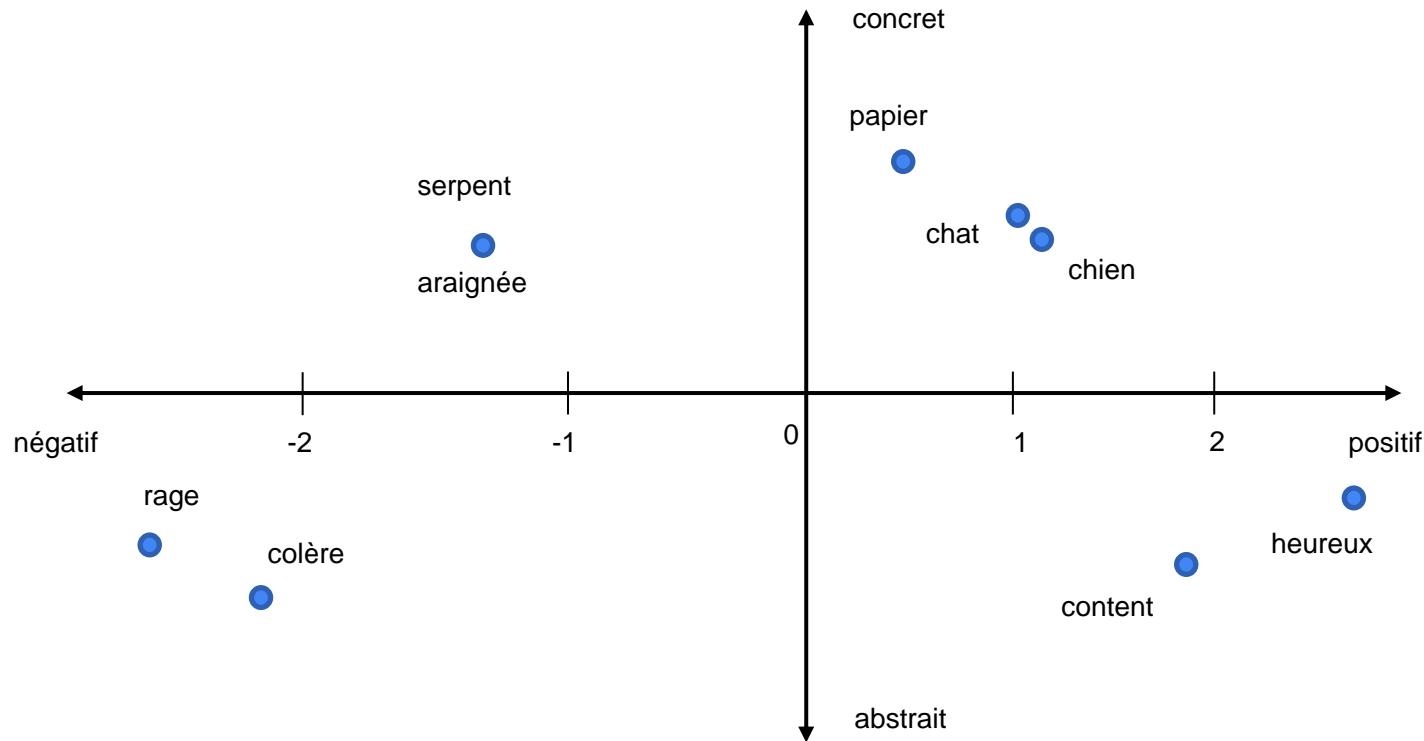


Représentation sous forme vectorielle



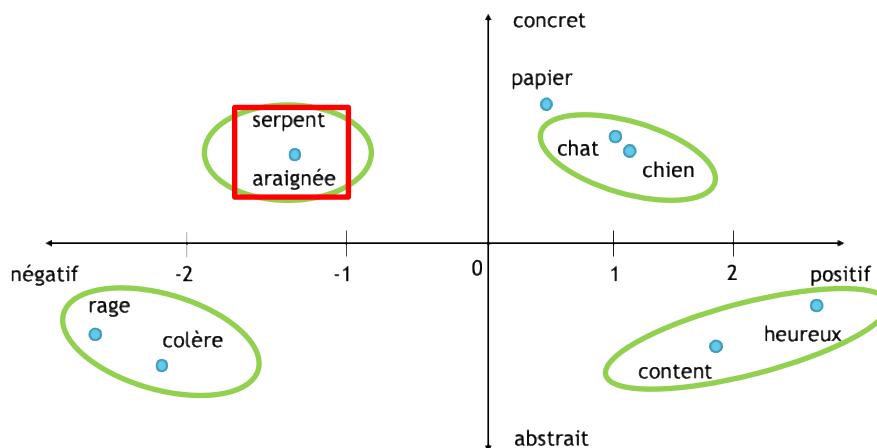


Représentation sous forme vectorielle





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 »



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 le mot pour créer l'embedding

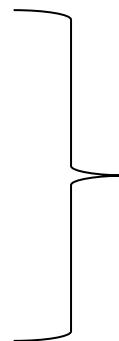
Exemple :

Chaton \approx Chat



Technique plus avancées d'embeddings

- Ces modèles ont des techniques de word embeddings qui changent 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-entraînées de ces modèles sont disponibles sur internet

Deep learning par la pratique

Leçon 6 : Le word2vec



Présenté par **Morgan Gautherot**

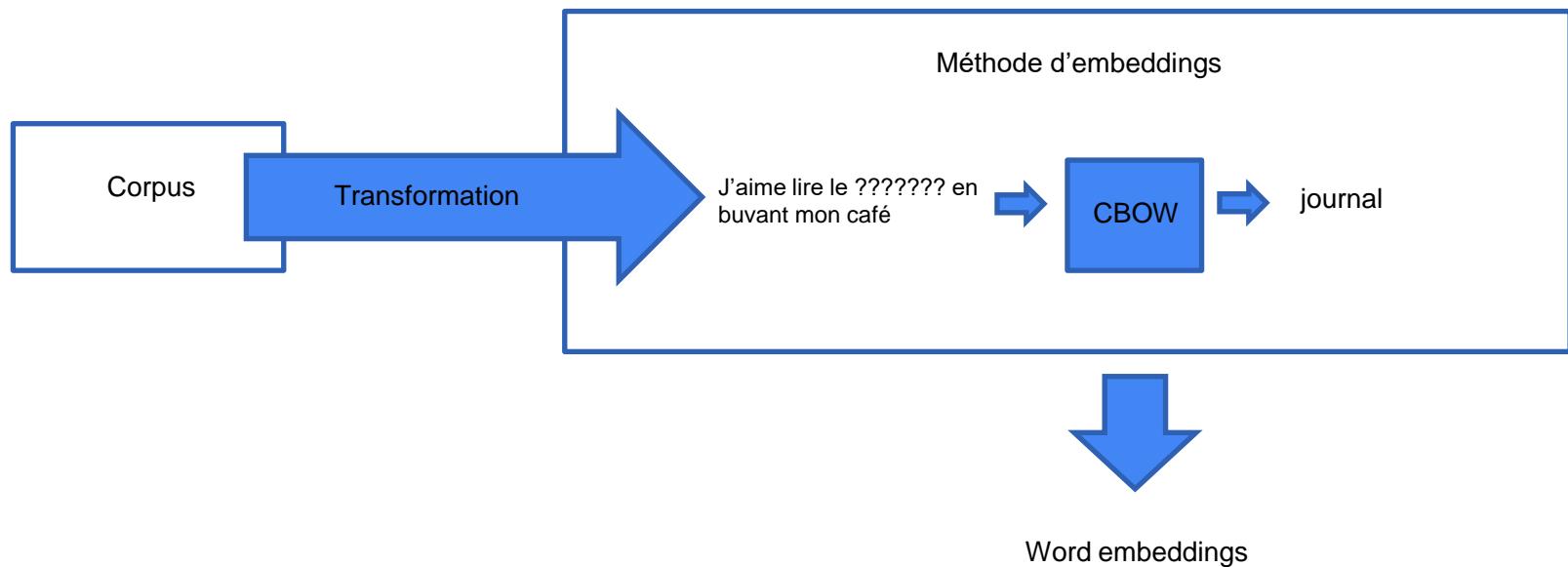


Continuous bag-of-words

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



Continuous bag-of-words word embeddings





Distance entre deux mots

La signification des mots sera déterminée 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

Deep learning par la pratique

Leçon 7 : Entrainer un modèle word2vec



Présenté par **Morgan Gautherot**

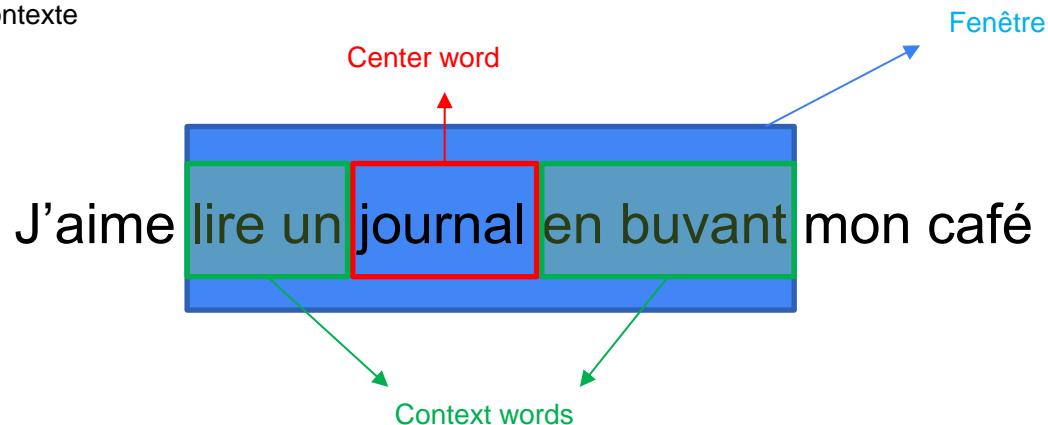


Créer une observation du jeu d'entraînement

$C = 2$

Demi taille du contexte

Taille de la fenêtre = 5





Extraire les mots

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
...



Center word en vecteur

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

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



Context words en vecteur

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

Context vecteur



Données d'entraînement

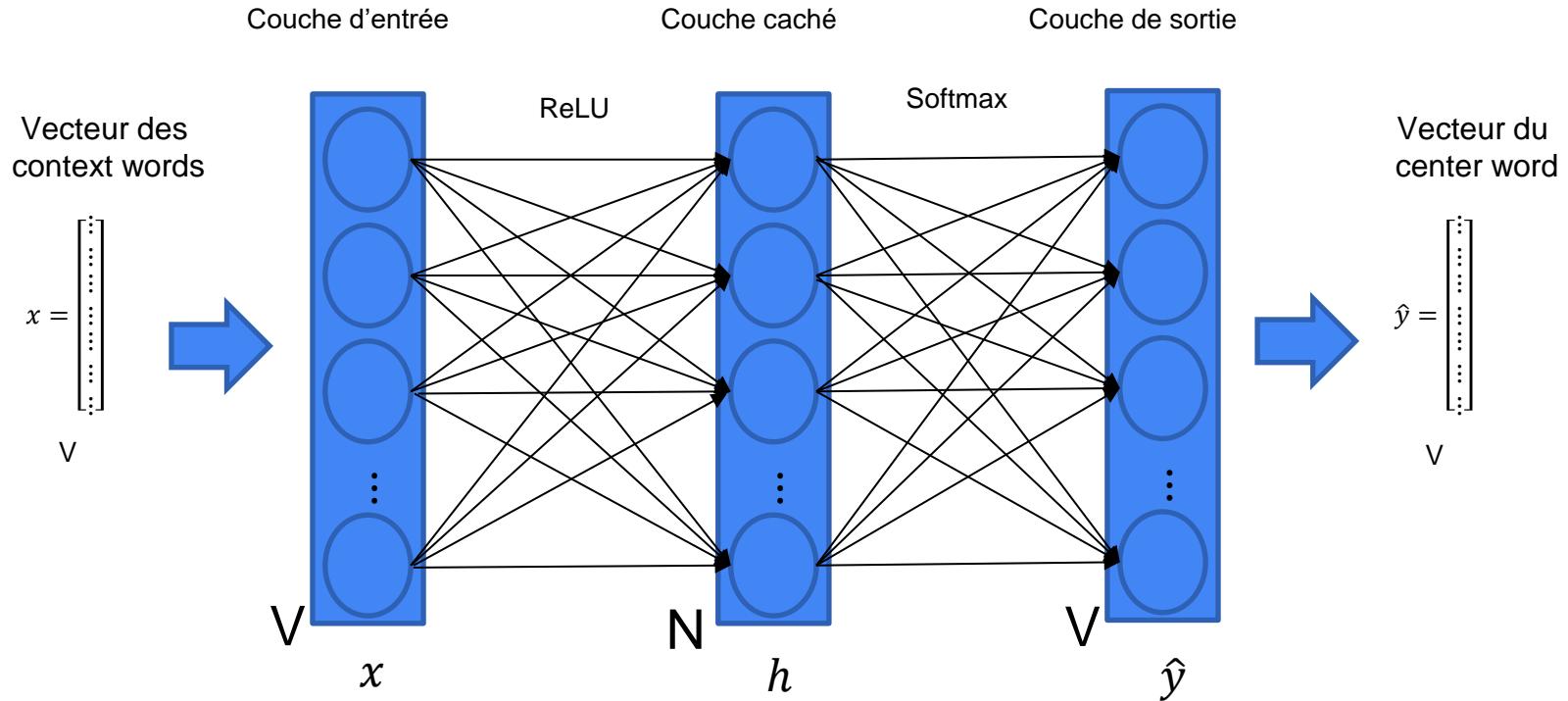
$$V = \begin{bmatrix} arachide \\ \vdots \\ buvant \\ \vdots \\ en \\ \vdots \\ journal \\ \vdots \\ lire \\ \vdots \\ un \\ \vdots \\ zumba \end{bmatrix}$$

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

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



L'architecture du modèle CBOW



Deep learning par la pratique

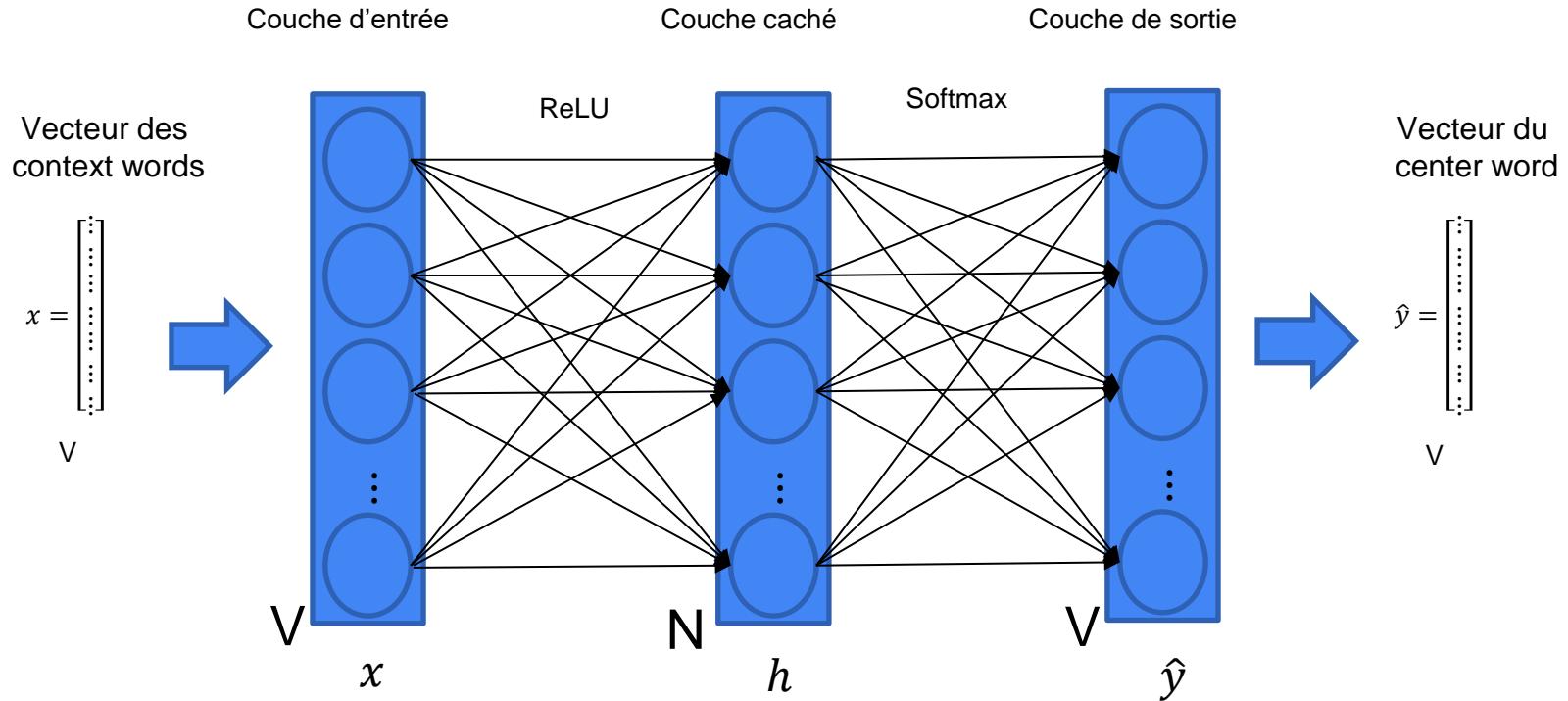
Leçon 8 : Extraire la matrice d'embedding



Présenté par **Morgan Gautherot**



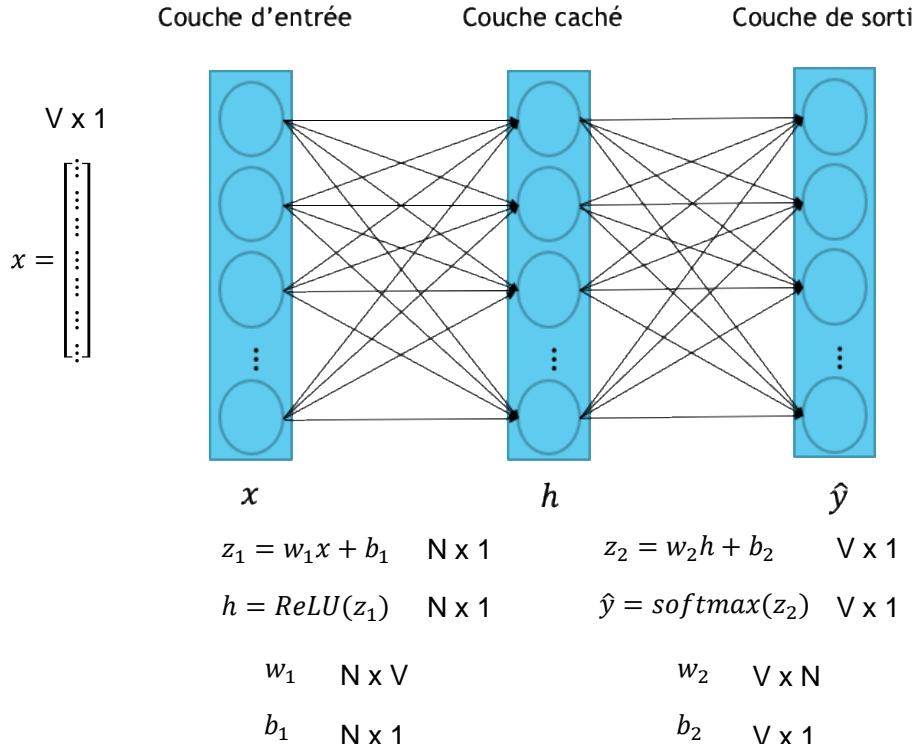
L'architecture du modèle CBOW





Dimension de l'architecture

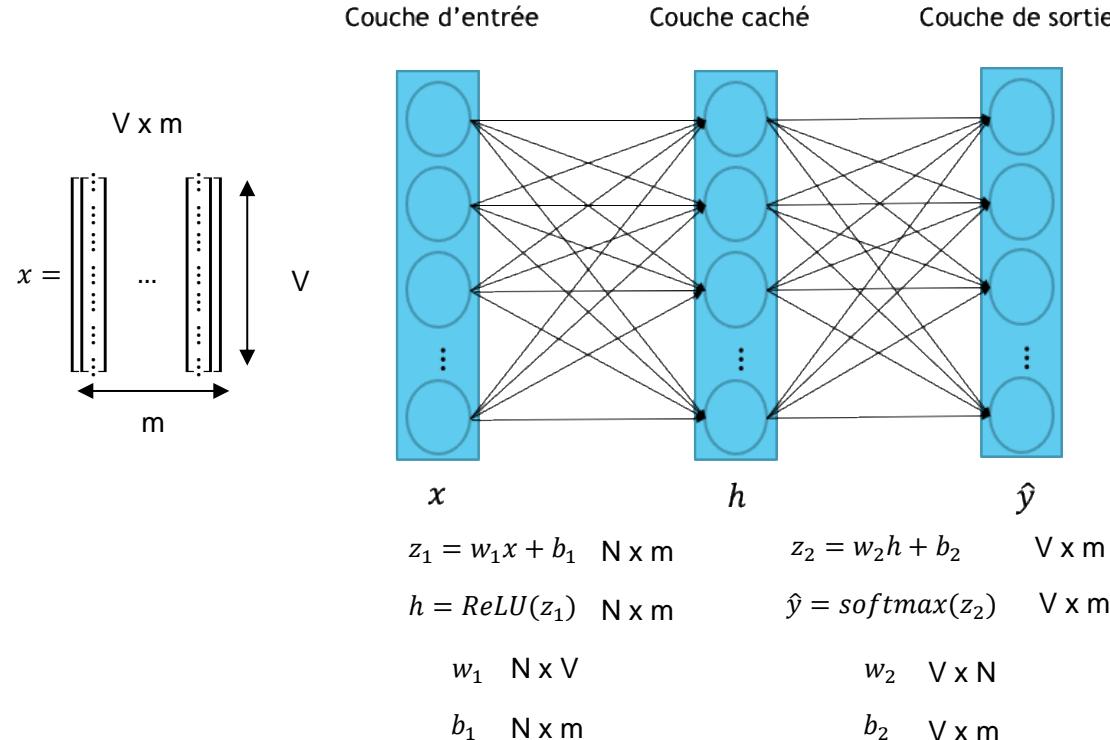
Avec un seul exemple





Dimension de l'architecture

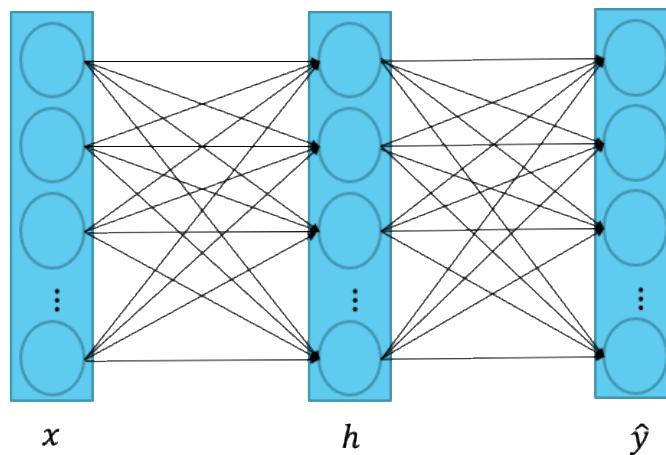
Avec m exemple





Extraire la partie word embeddings

Couche d'entrée Couche caché Couche de sortie



$$w_2 \quad V \times N$$

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

\longleftrightarrow

$$N$$

Deep learning avec Pytorch

Partie 14 : Les Transformers



Presented by **Morgan Gautherot**

Des RNN aux Transformers

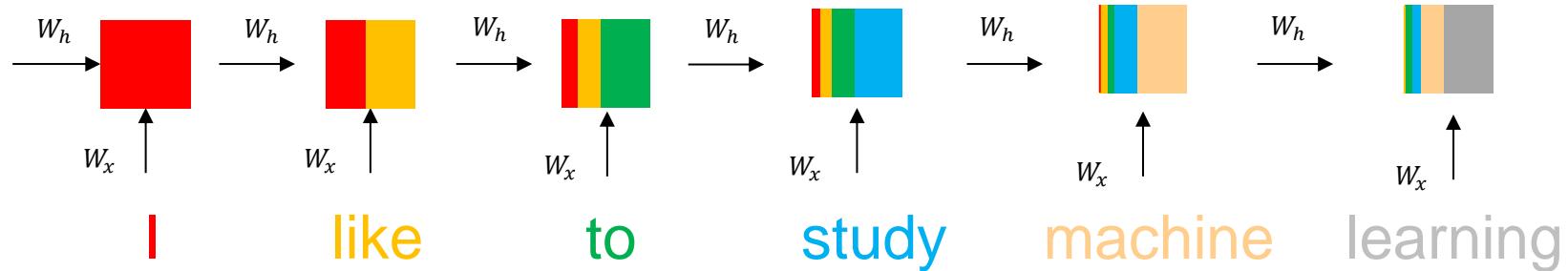


Partie 10 : Les Transformers



L'influence des premiers termes

I like to study machine learning _____





Traduction

How are you today ?

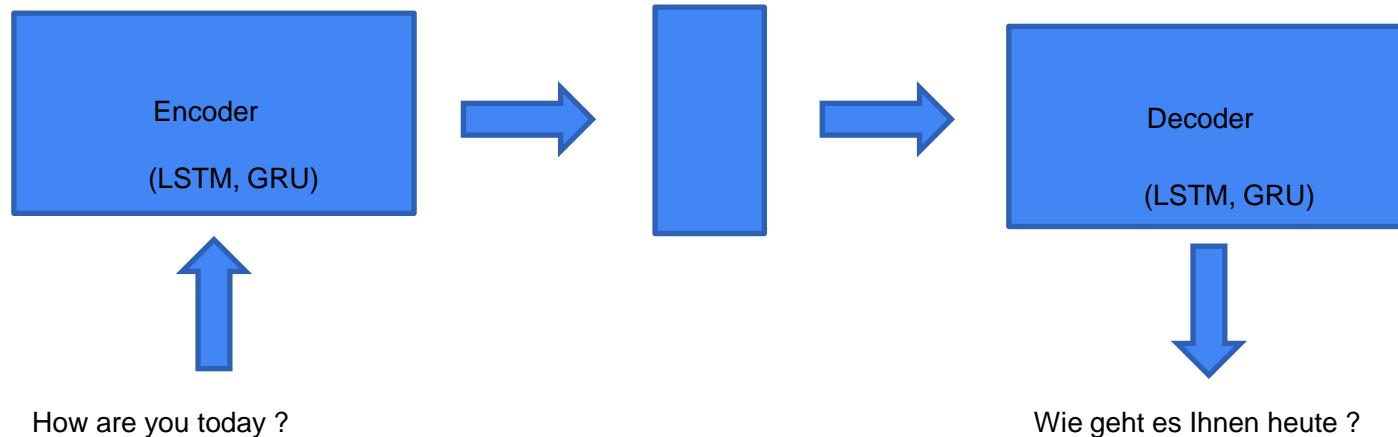


Wie geht es Ihnen heute ?



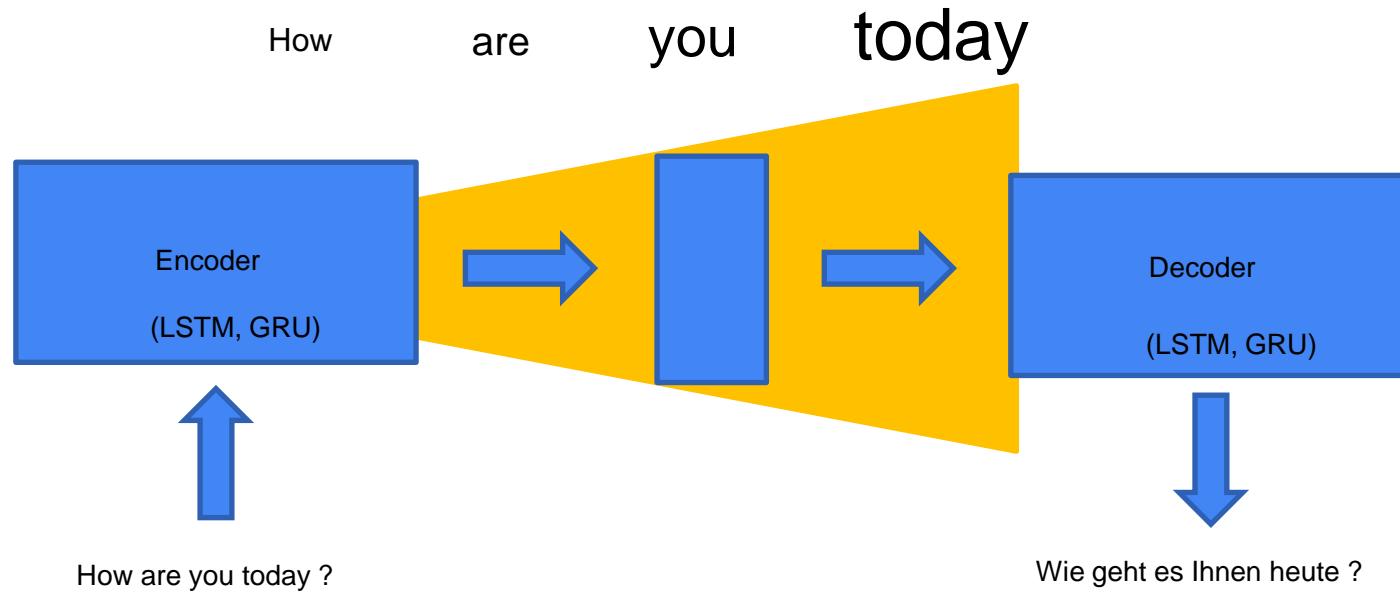


Seq2seq model (Google 2014)





Le bottleneck de l'information



Plus la taille de la séquence augmente, plus les performances du modèle diminuent !



Alignment des mots

I
↓
Ich

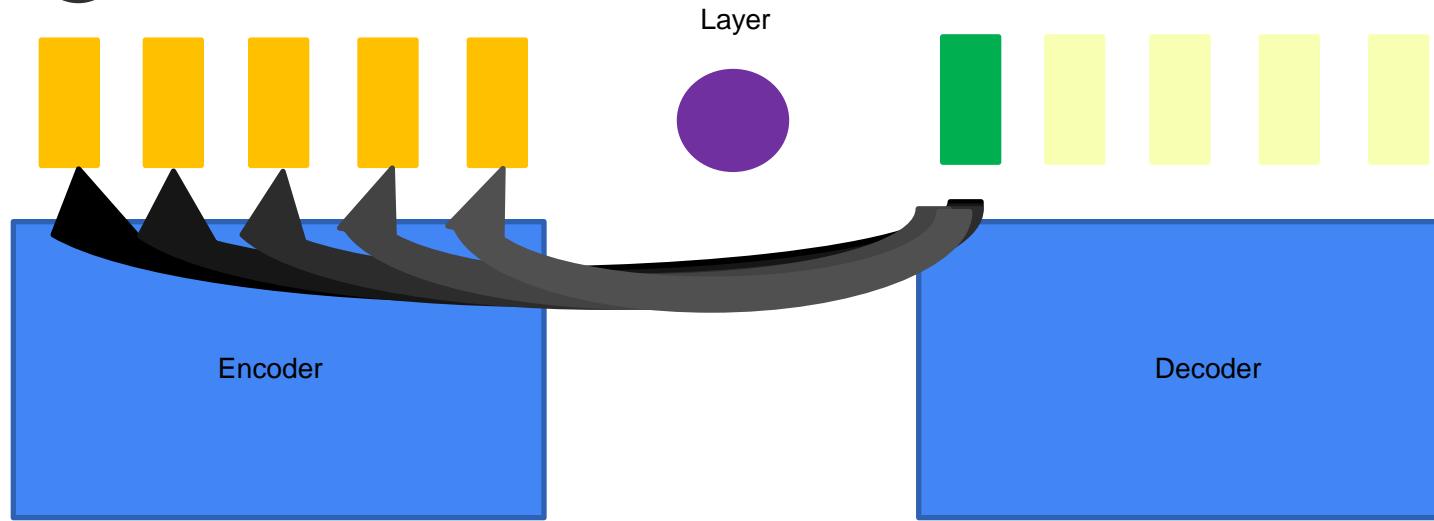
am
↓
bin

a
↓
ein

Jelly
Donut
↓
Berliner



Quel est le mot auquel il faut prêter le plus d'attention ?

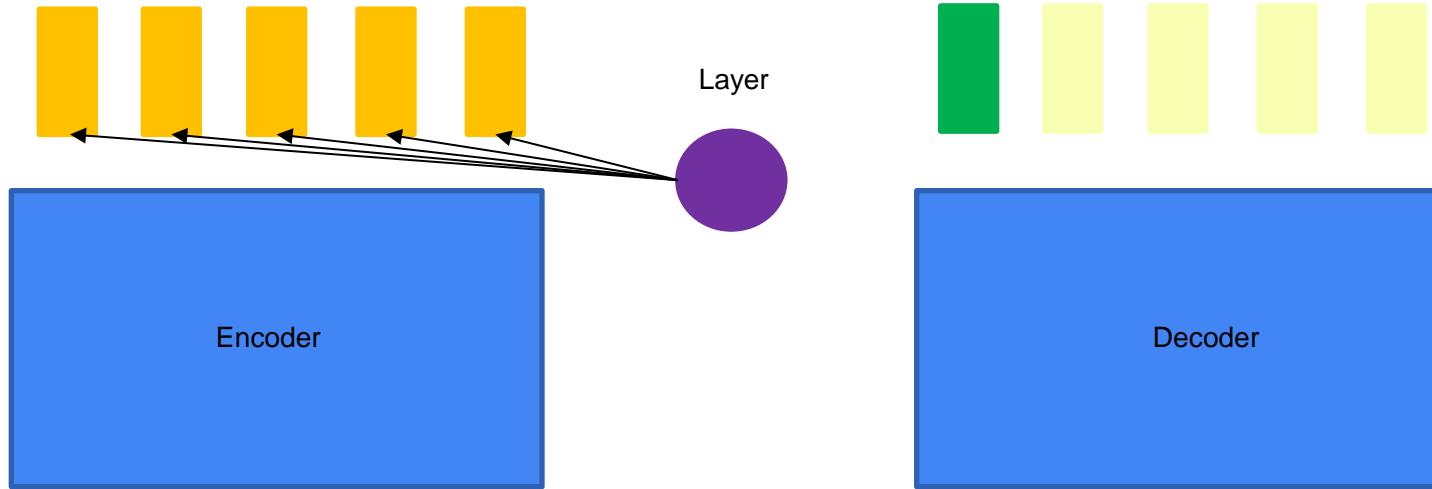


How are you today ?

<start> ? ? ? ?



Quel est le mot auquel il faut prêter le plus d'attention ?

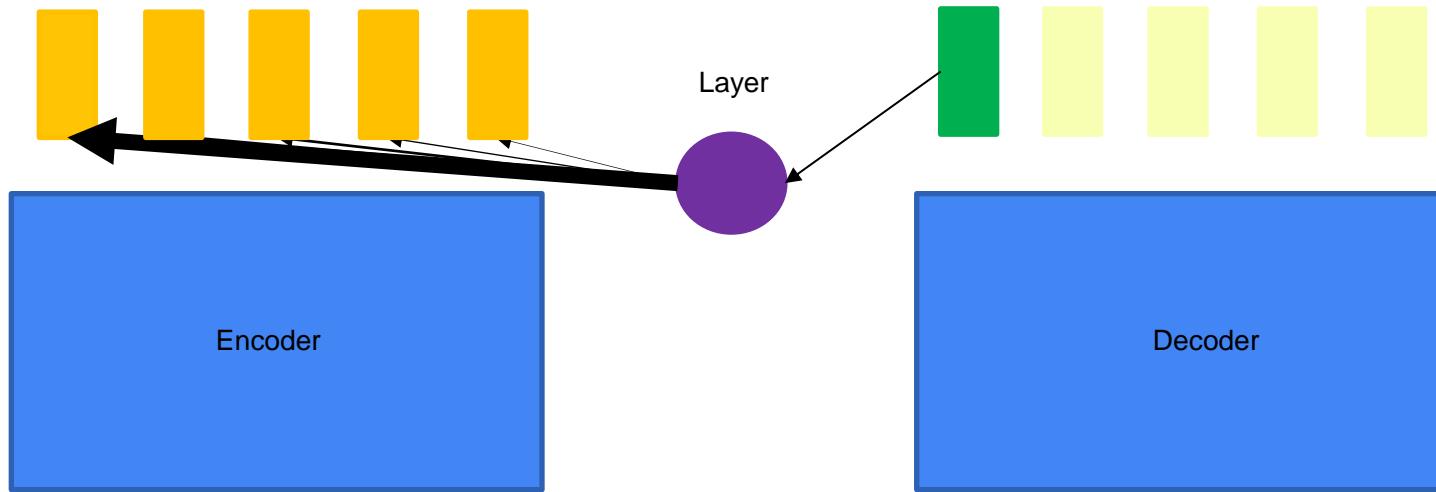


How are you today ?

<start> ? ? ? ?



Plus de poids = plus d'attention

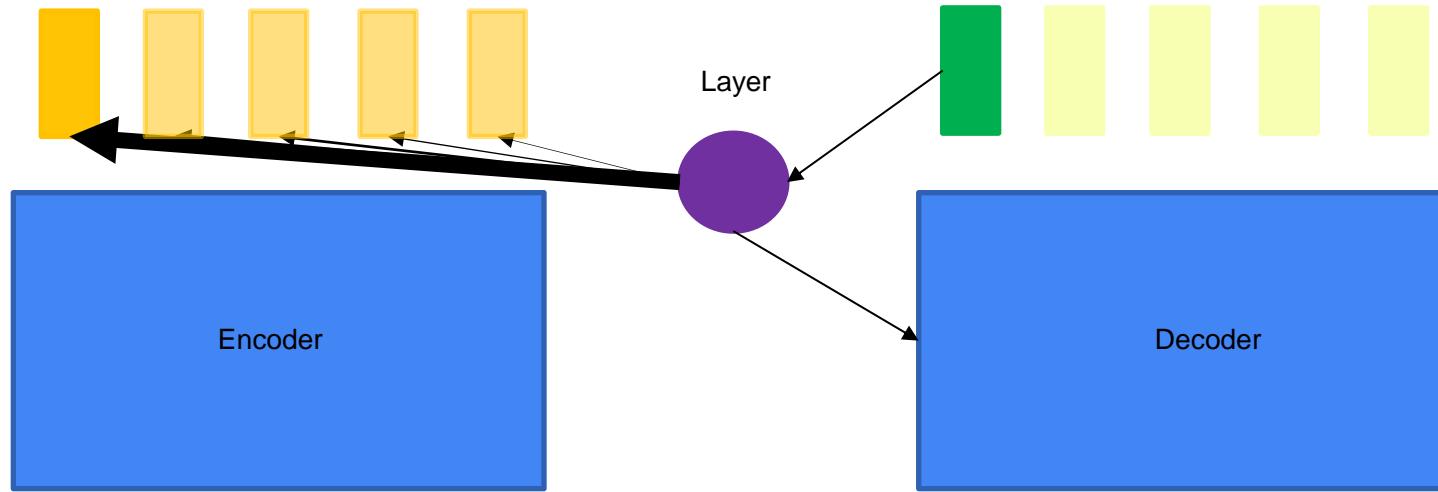


How are you today ?

<start> ? ? ? ?



Plus de poids = plus d'attention



How are you today ?

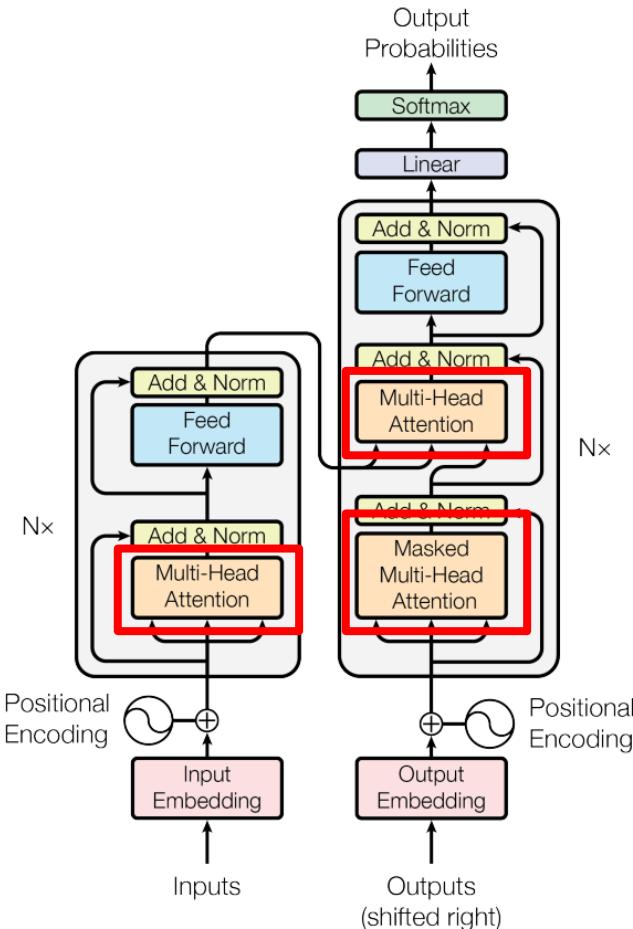
<start> ? ? ? ?



Transformer

Plus rapide que le RNN

Profondément bidirectionnel

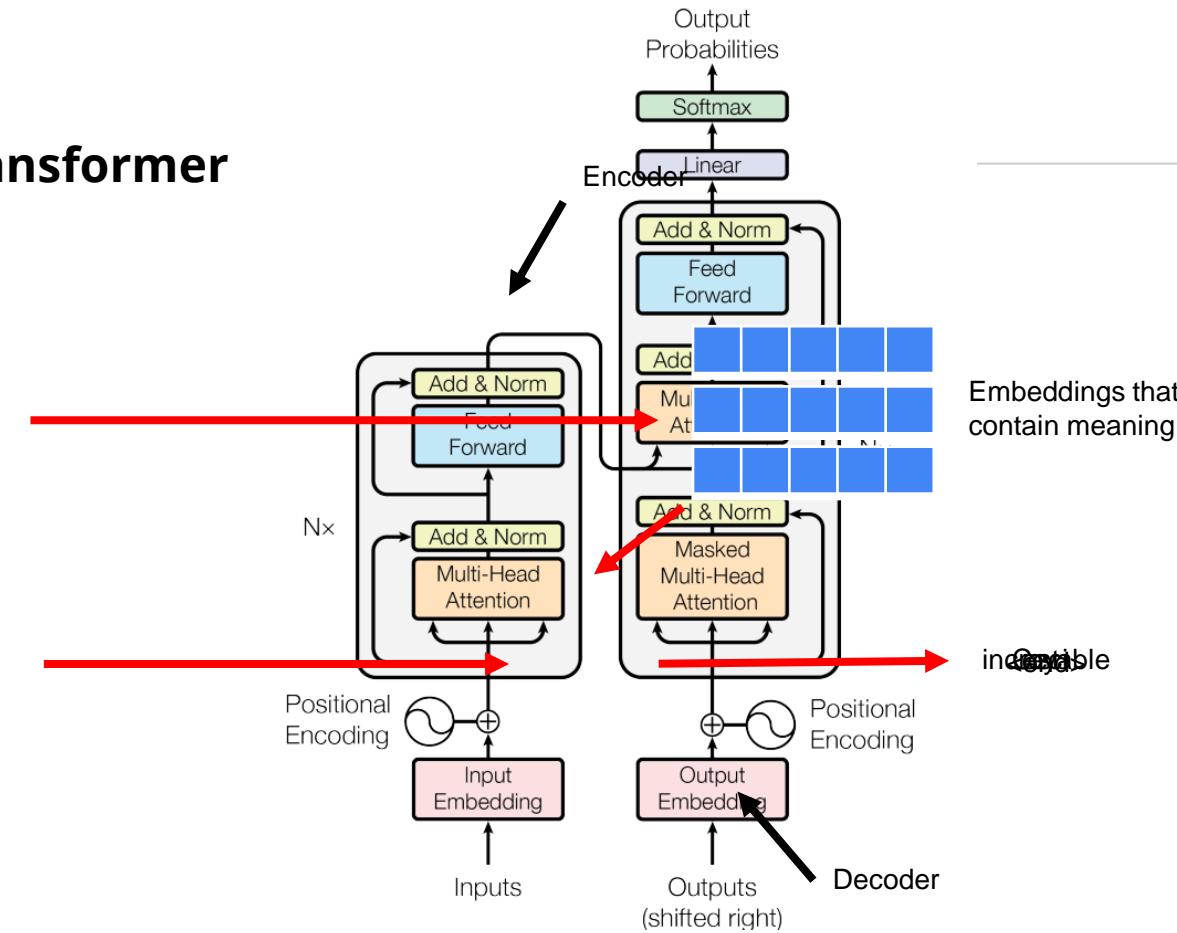




Transformer

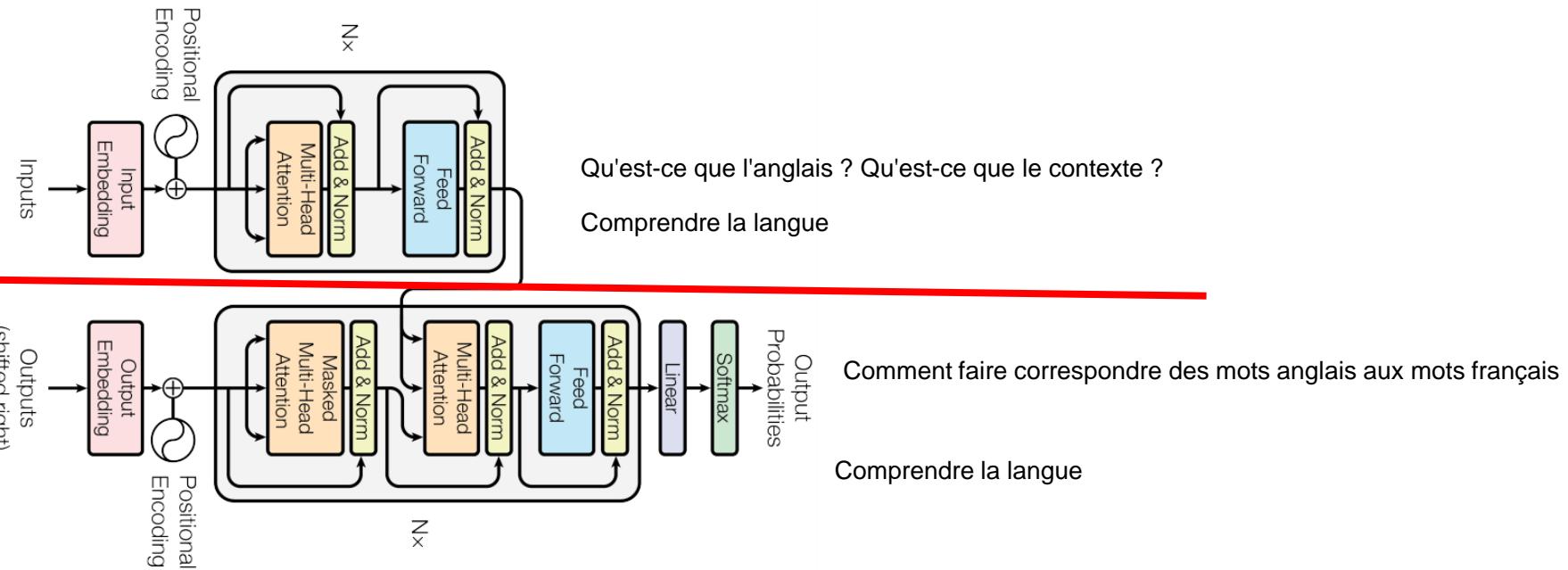
This
is
amazing

↓
<start>





Du transformer à l'embedding



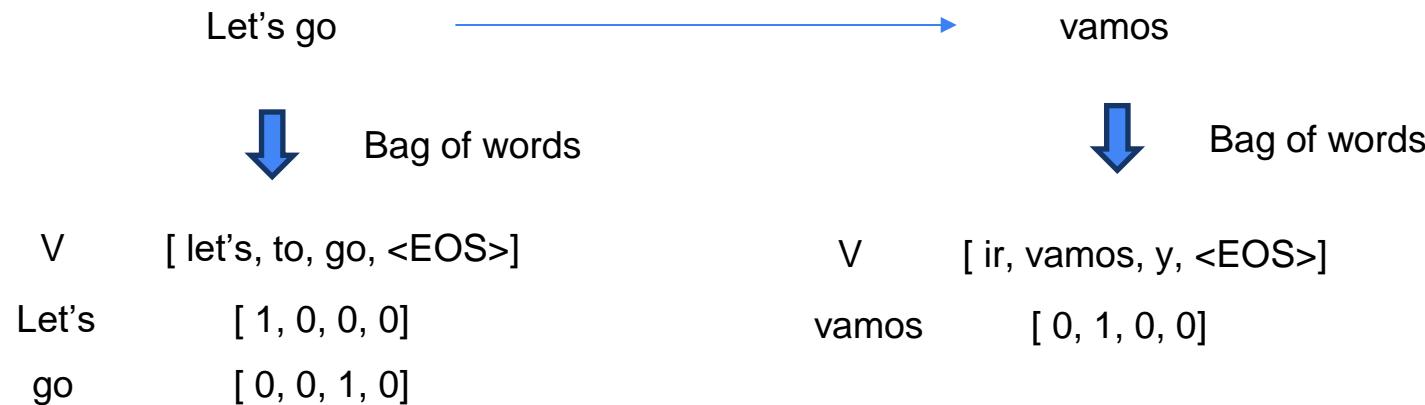
Comprendre les transformateurs



Partie 10 : Les Transformers

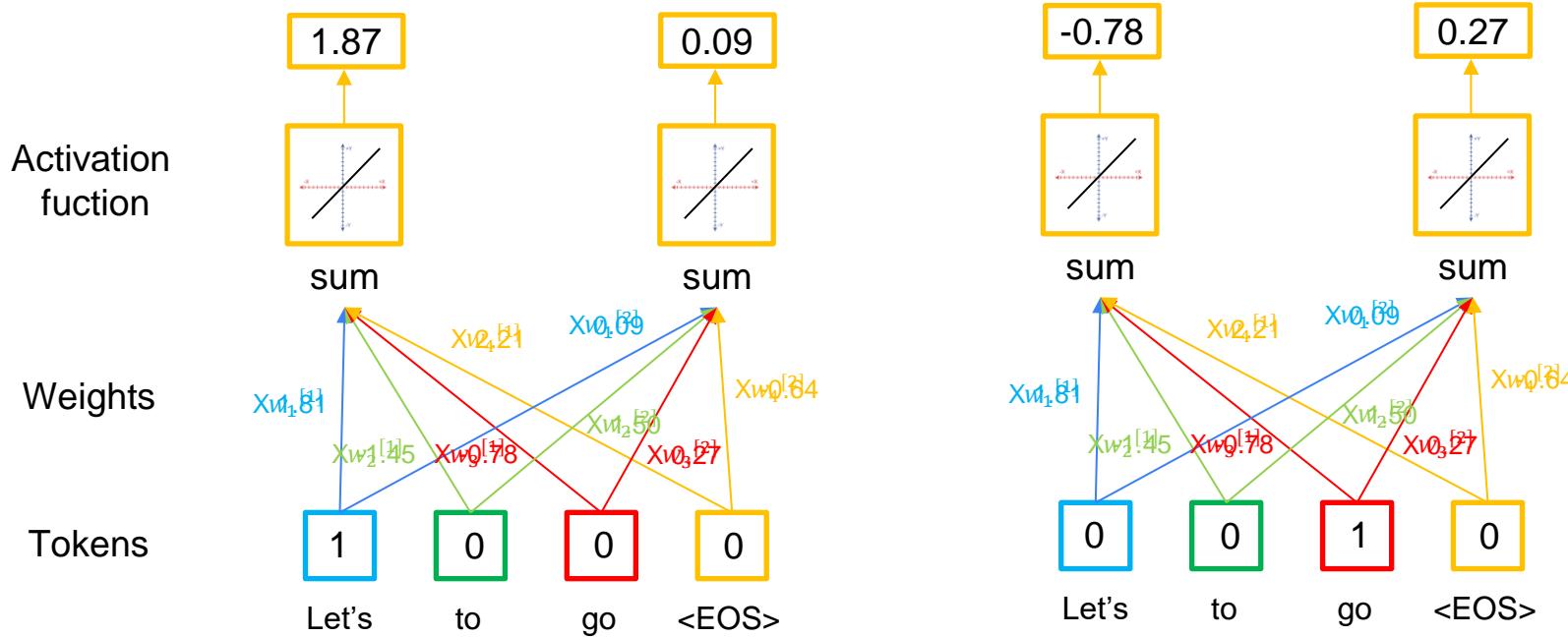


Traduction anglais - espagnol



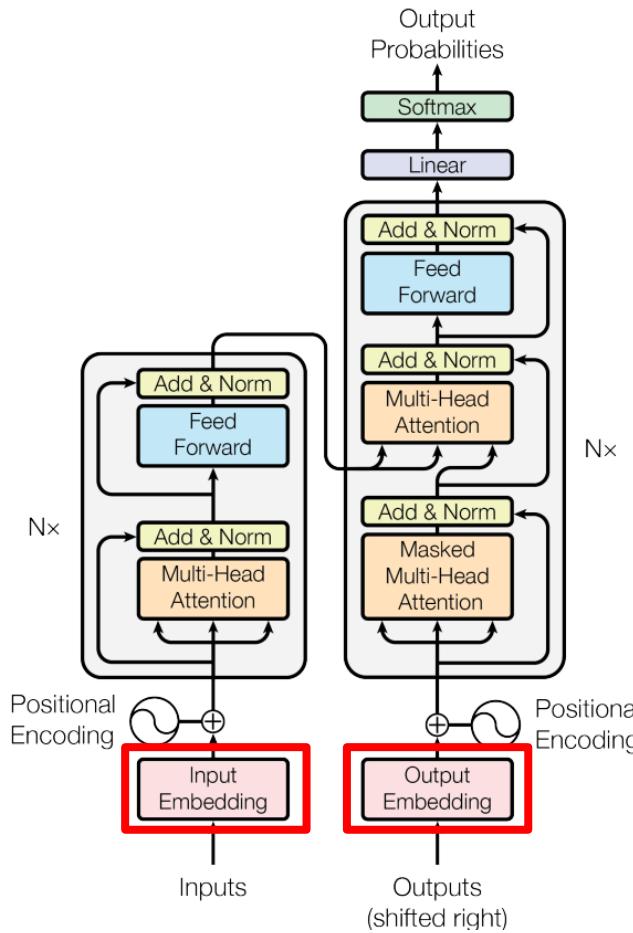


Word embeddings





Transformer





L'ordre des mots

Il est très important de respecter l'ordre des mots

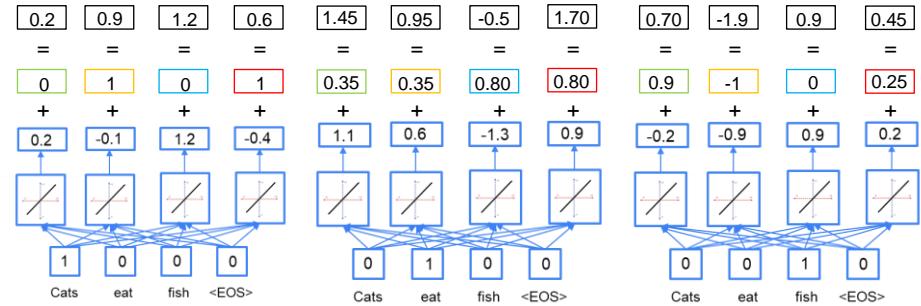
Cats eat fish

!=

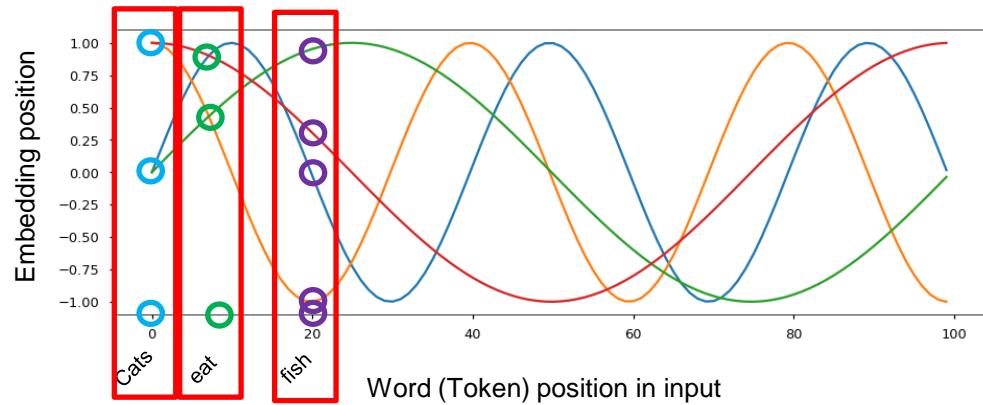
Fish eat cats



Encodage positionnel

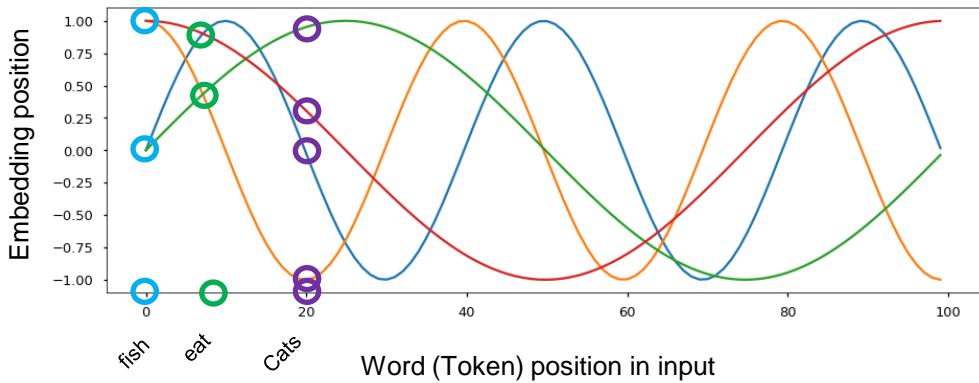
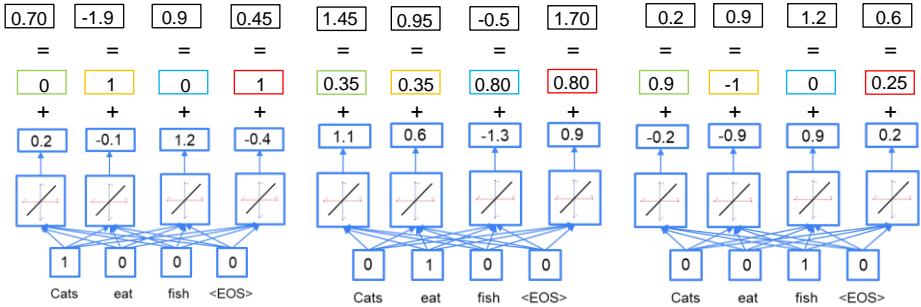


Valeur positionnelle unique





Positional encoding





L'ordre des mots

Cats eat fish

[0.2] [0.9] [1.2] [0.6] [1.45] [0.95] [-0.5] [1.70] [0.70] [-1.9] [0.9] [0.45]

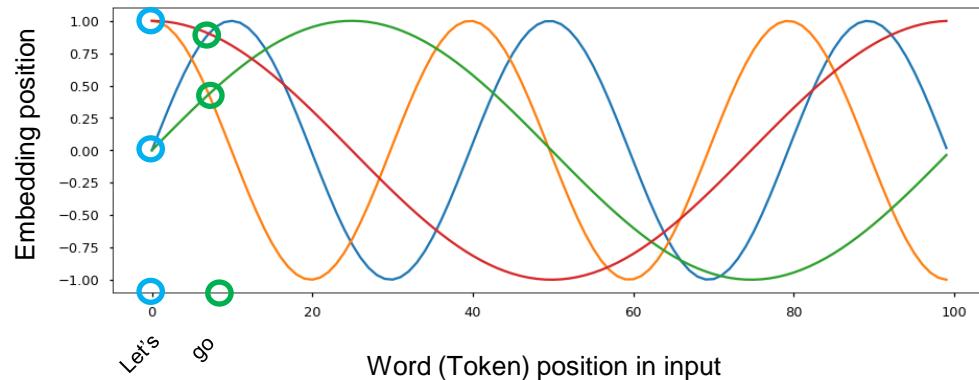
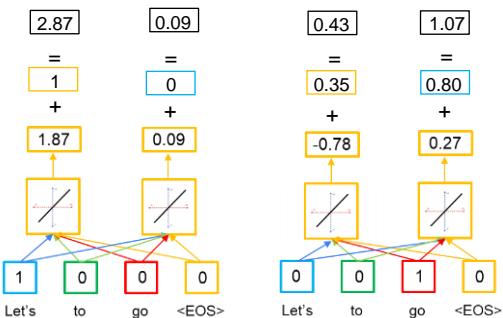
!=

Fish eat cats

[0.70] [-1.9] [0.9] [0.45] [1.45] [0.95] [-0.5] [1.70] [0.2] [0.9] [1.2] [0.6]

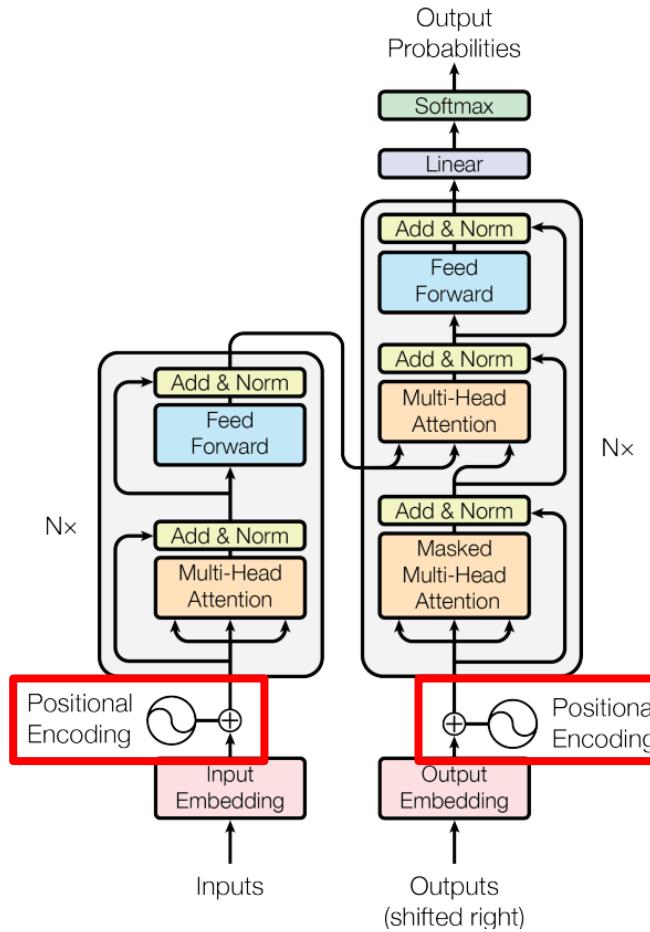


Encodage positionnel





Transformer





Self attention

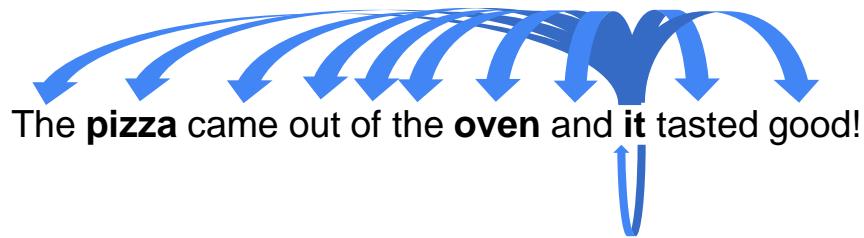
The **pizza** came out of the **oven** and **it** tasted good!





Calcul de la similarité

The **pizza** came out of the **oven** and **it** tasted good!





Calcul de la similarité

Une fois les similitudes calculées, elles sont utilisées pour déterminer comment le transformateur code chaque mot

The **pizza** came out of the **oven** and it tasted good!



Self attention

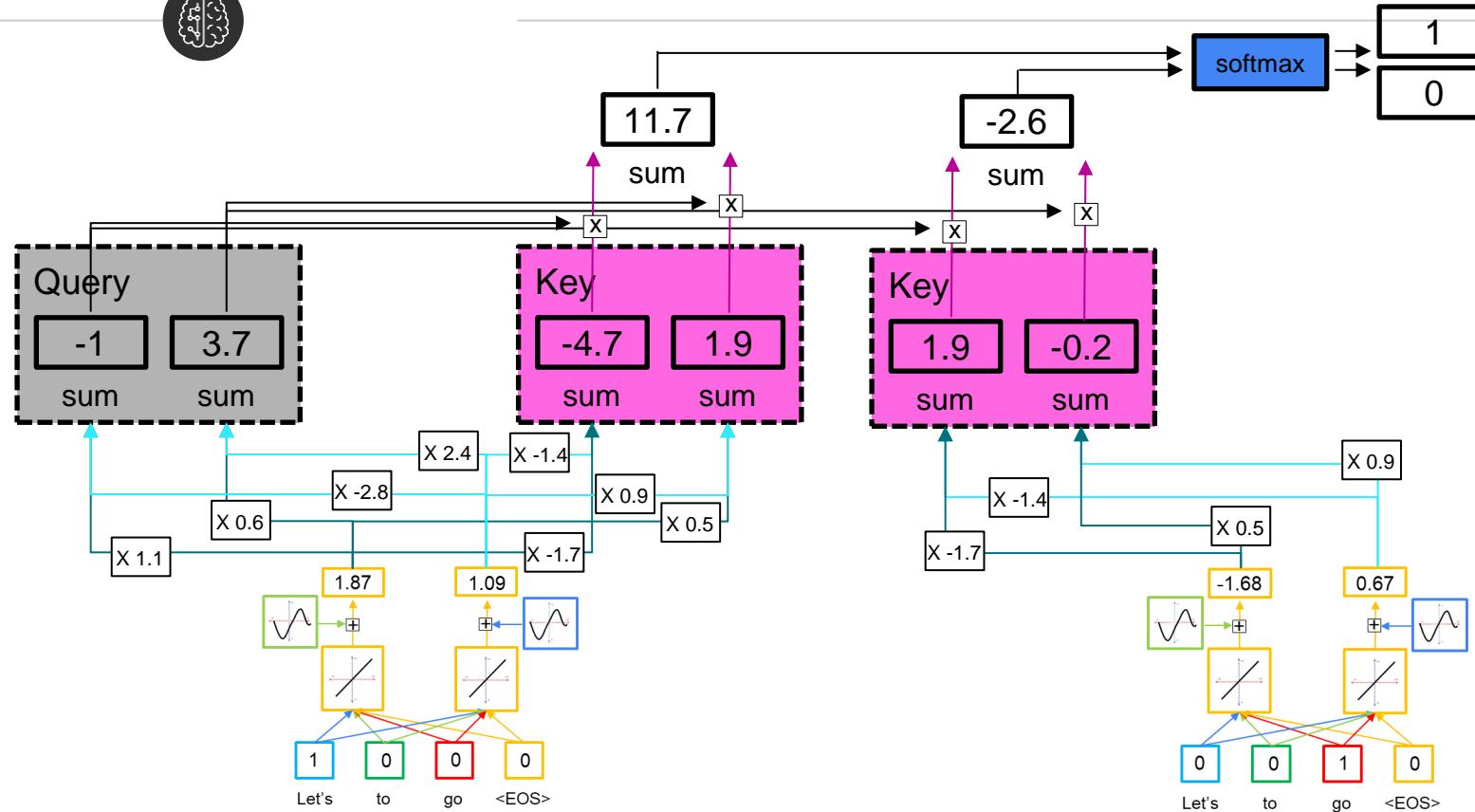
The **pizza** came out of the **oven** and **it** tasted good!



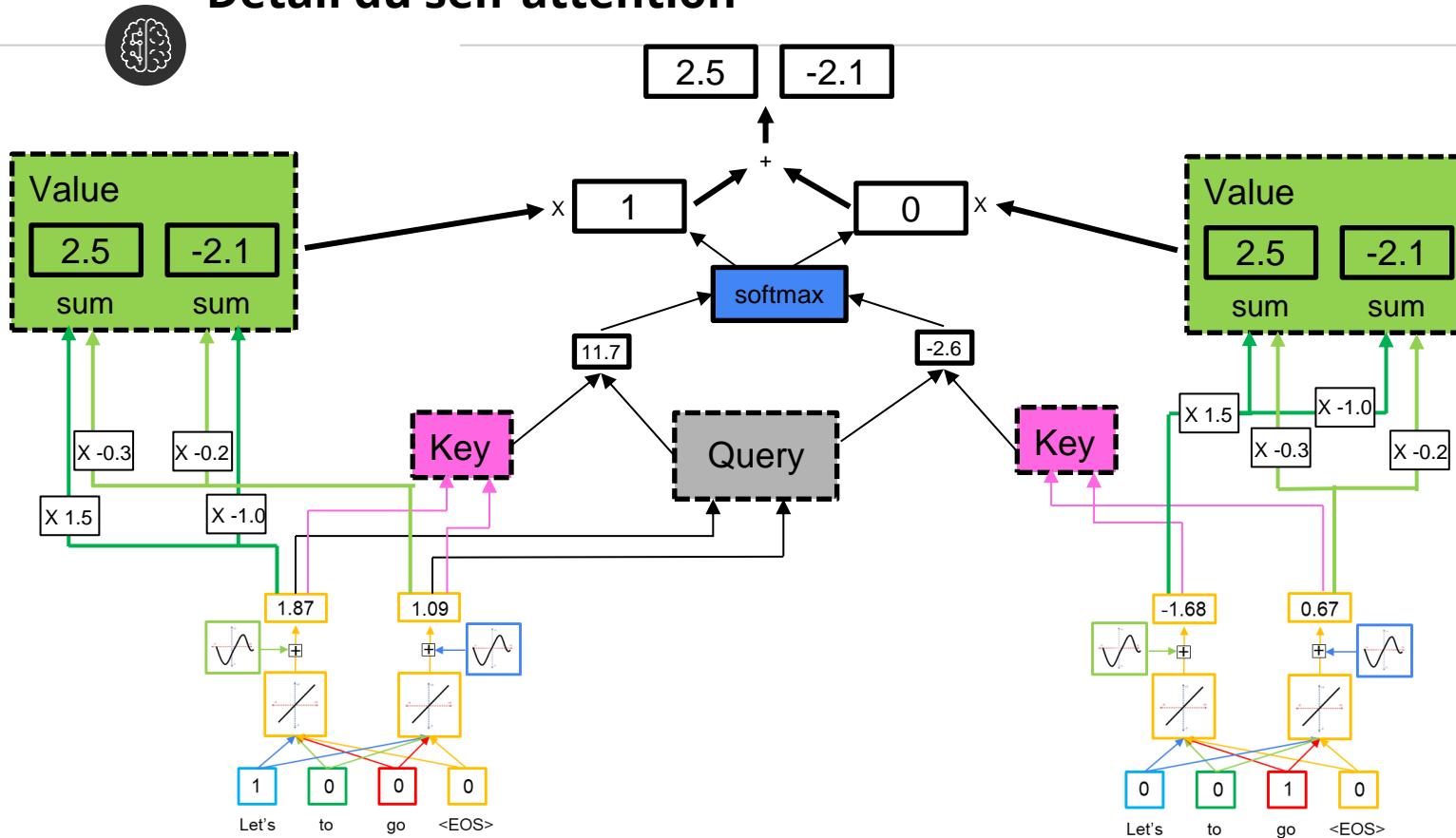
Détail du self-attention



Nous pouvons considérer la sortie de la fonction softmax comme un moyen de déterminer le pourcentage d'utilisation de chaque mot pour encoder le mot « Let's ».



Détail du self-attention



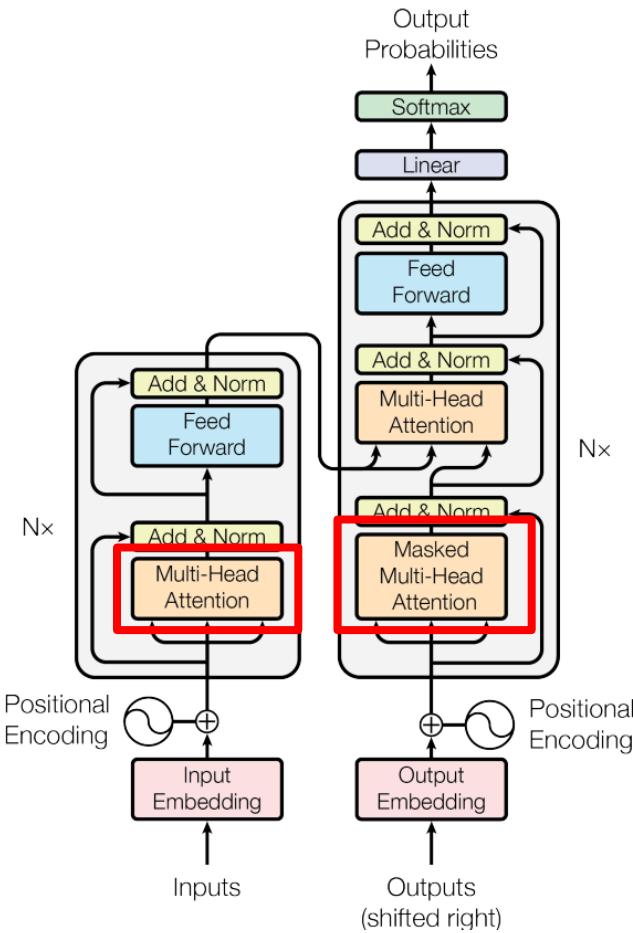


Les avantages de l'auto-attention chez les transformateurs

- Les Transformers utilisent un seul ensemble de poids pour calculer les requêtes d'auto-attention, quel que soit le nombre de mots en entrée
- Les Transformers peuvent calculer les requêtes, les Keys et les valeurs pour chaque mot en même temps.



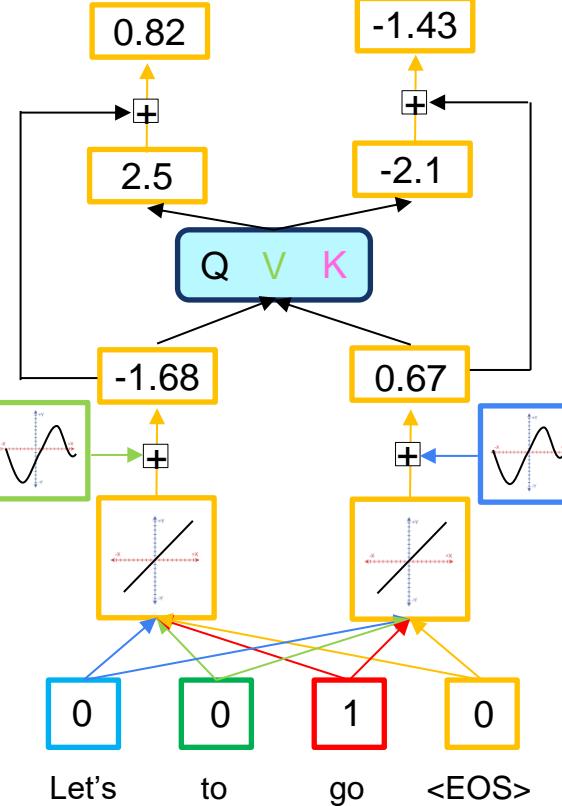
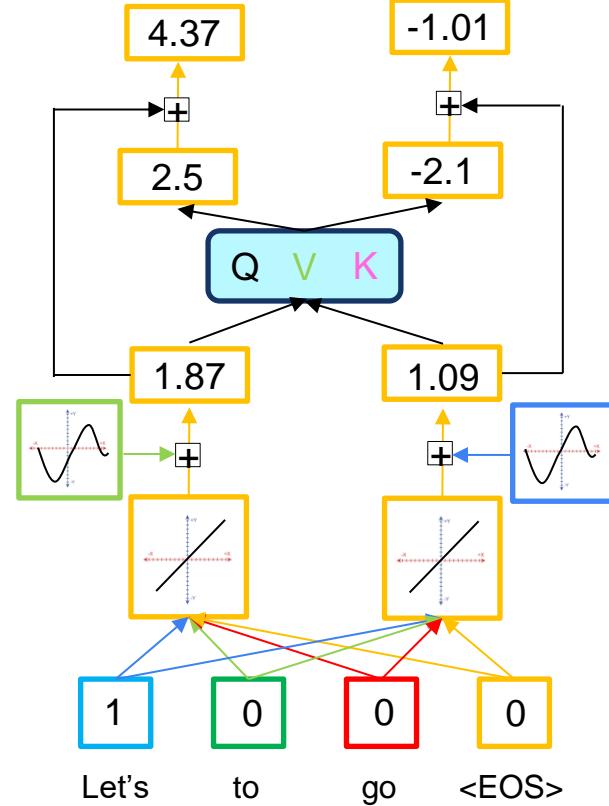
Transformer



Les connexions résiduelles facilitent l'entraînement de réseaux neuronaux complexes en permettant à la couche d'auto-attention d'établir des relations entre les mots en entrée sans avoir à préserver également les informations relatives à l'embedding des mots et à l'encodage de la position.

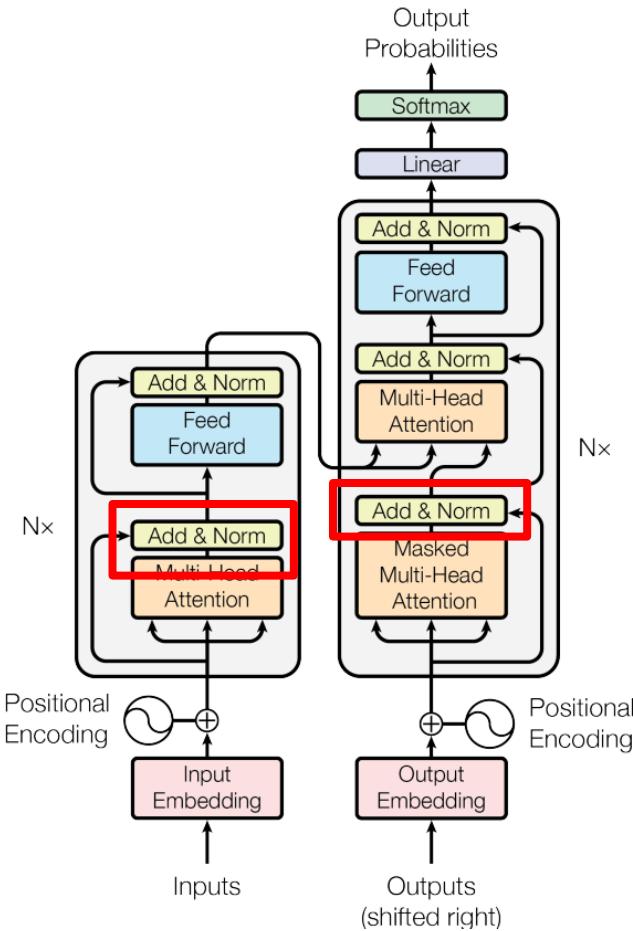


Les connexions résiduelles





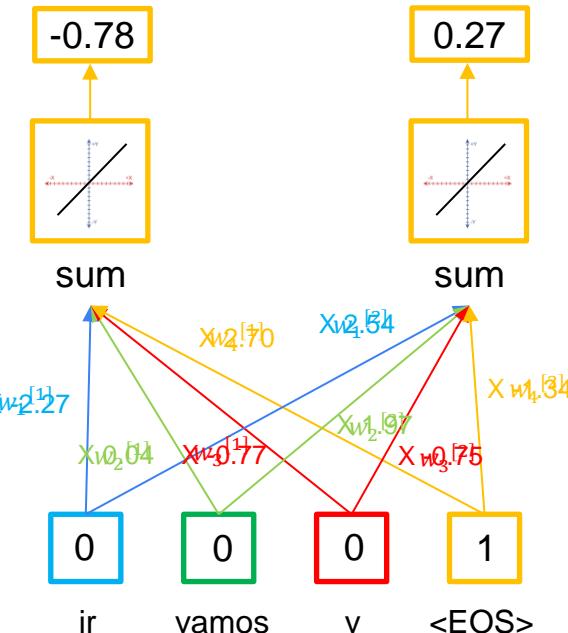
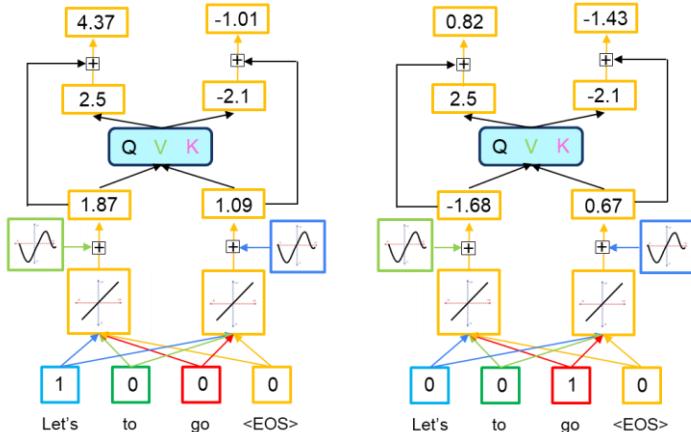
Transformer





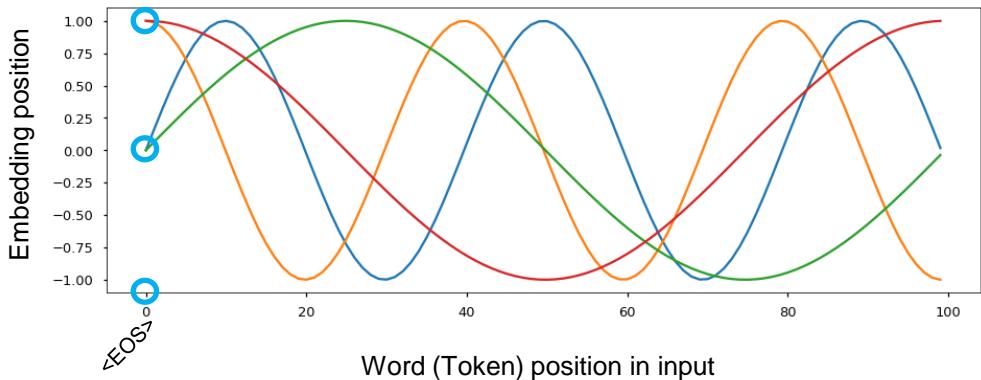
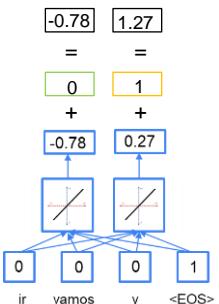
Decoder

Encoder

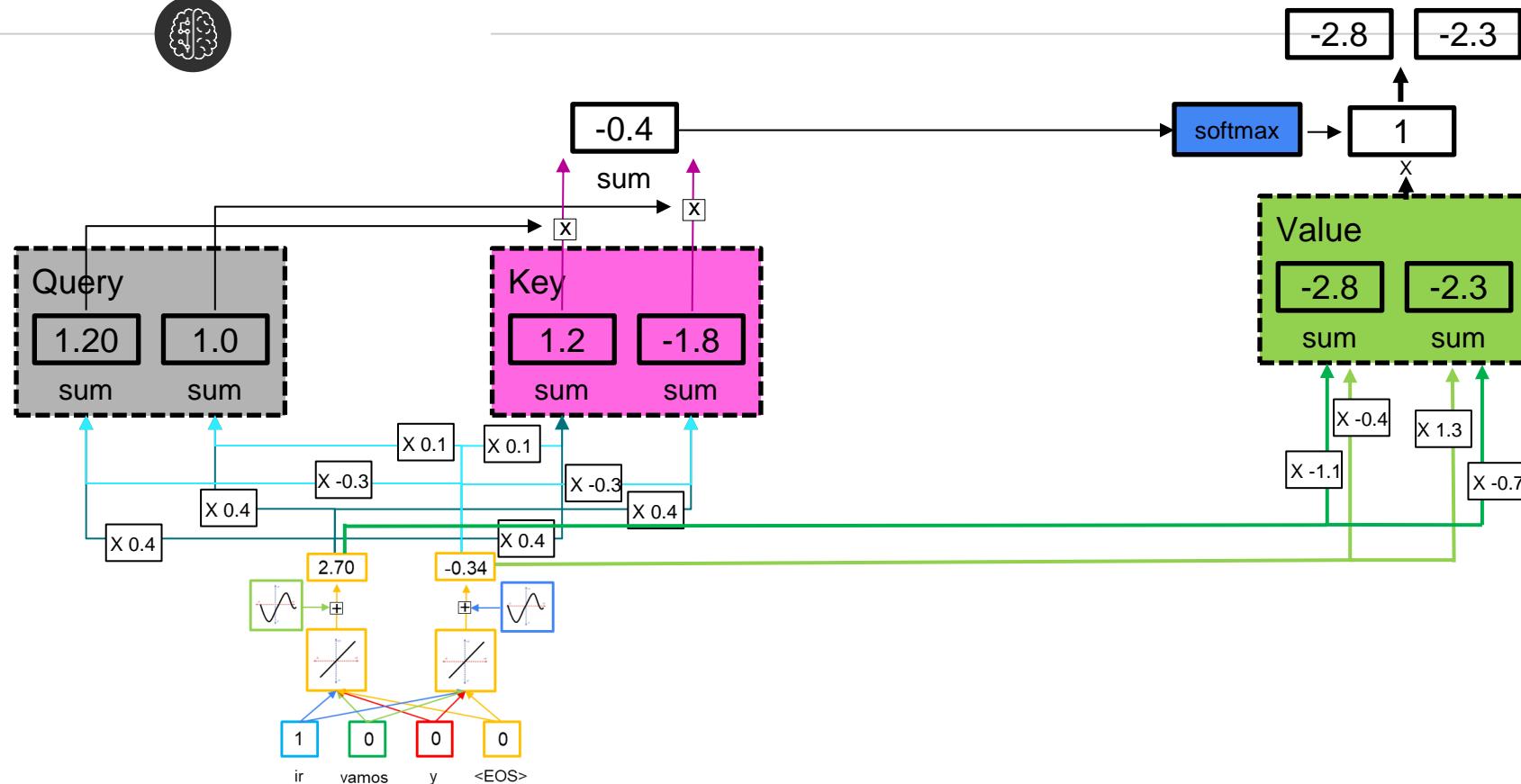




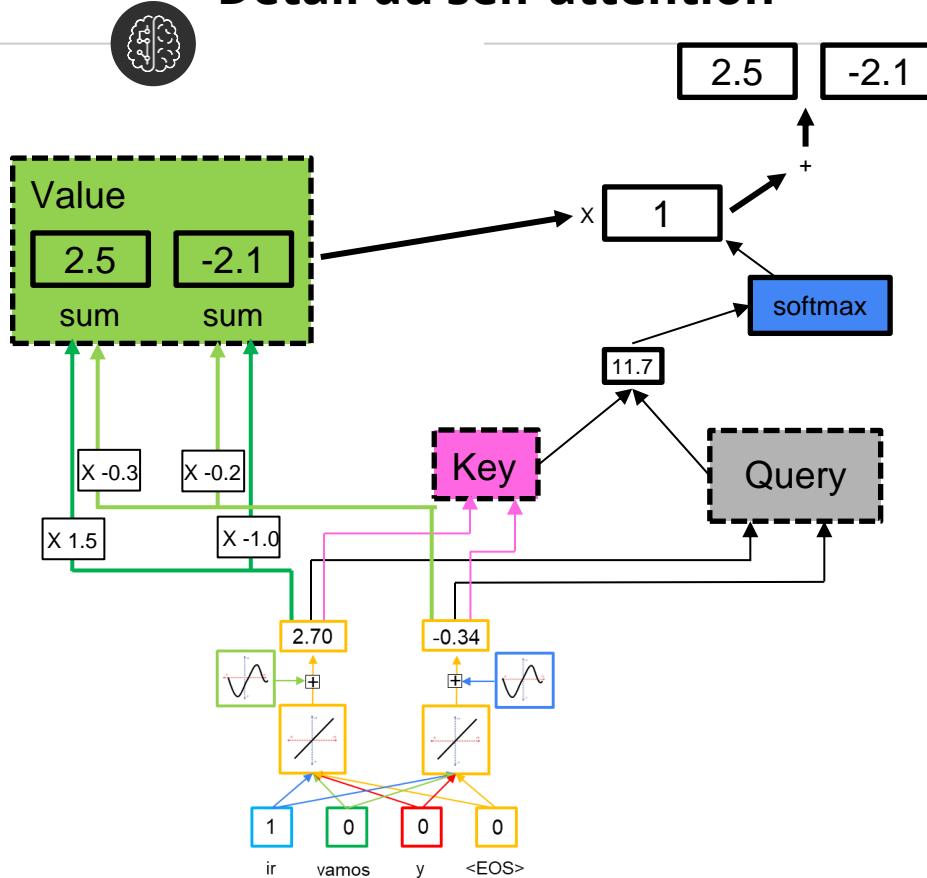
Codage positionnel



Détail du self-attention

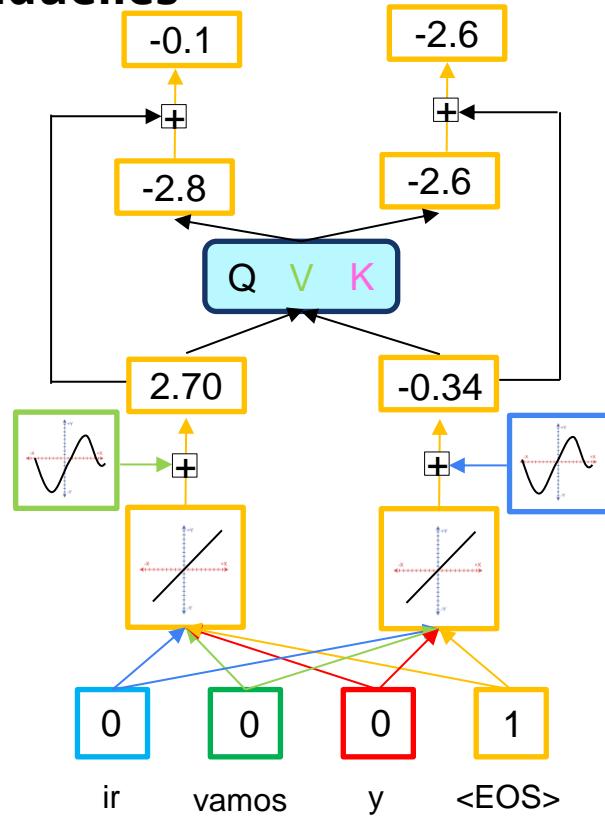


Détail du self-attention





Connections résiduelles

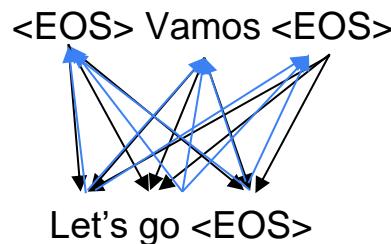




Attention entre l'encoder et le decoder

The **pizza** came out of the **oven** and **it** tasted good!

Nous avons parlé de la façon dont le self-attention aide le transformeur à suivre les relations entre les mots d'une même phrase. Cependant, puisque nous traduisons une phrase, nous devons également tenir compte des relations entre la phrase d'entrée et la phrase de sortie.





Attention entre l'encoder et le decoder

Don't eat the delicious looking and smelling pizza

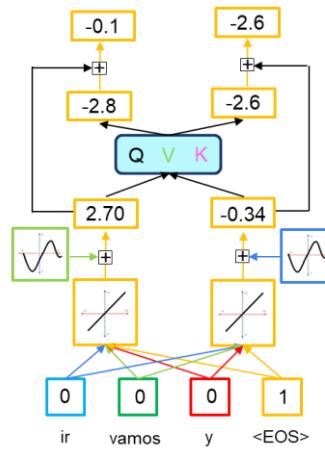
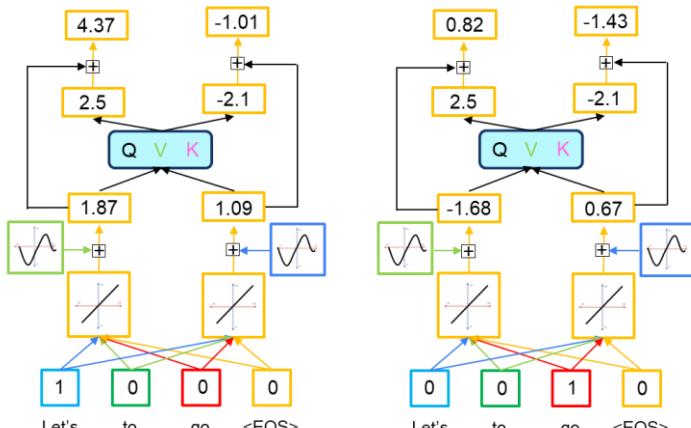


Eat the delicious looking and smelling pizza



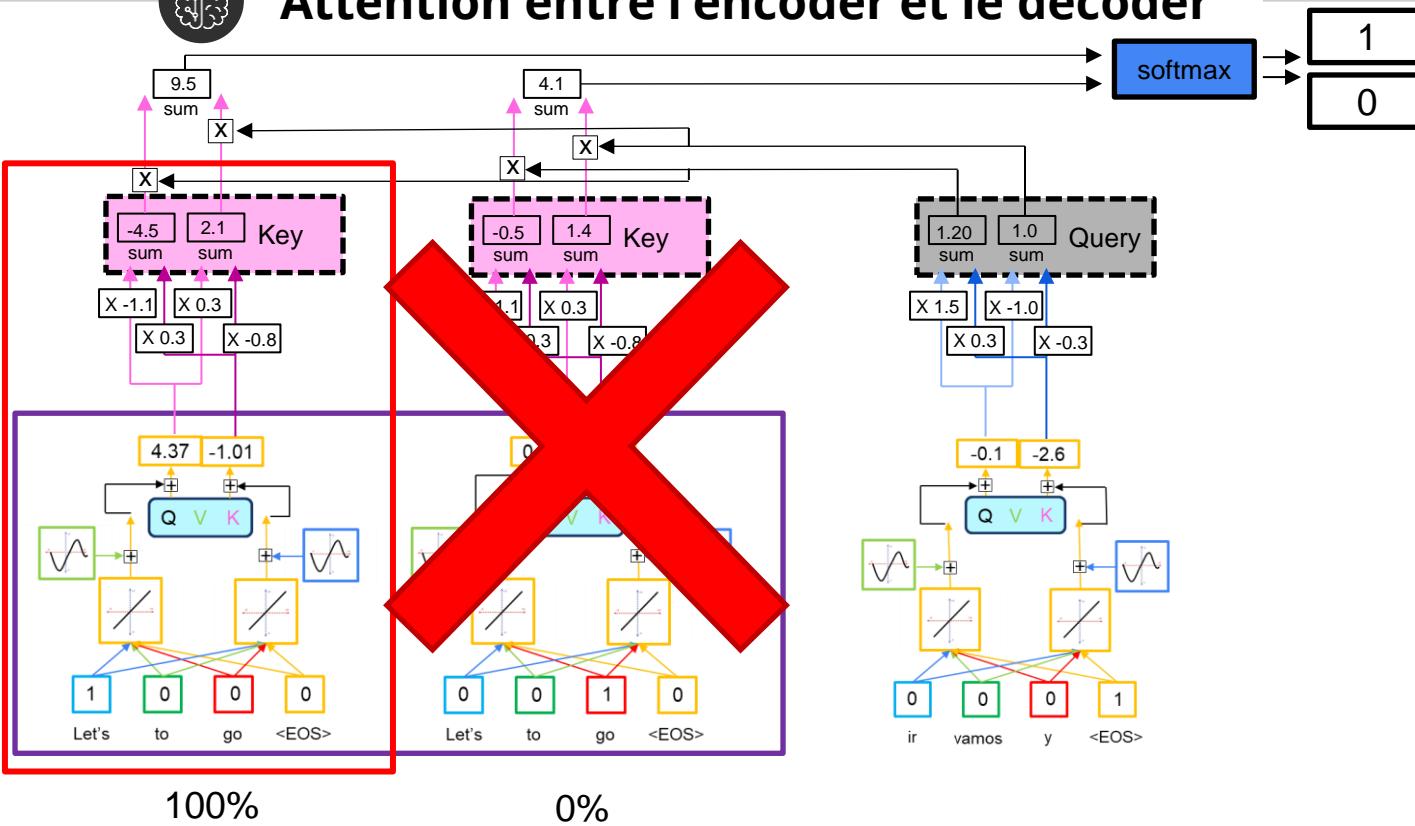
Attention entre l'encoder et le decoder

Encoder



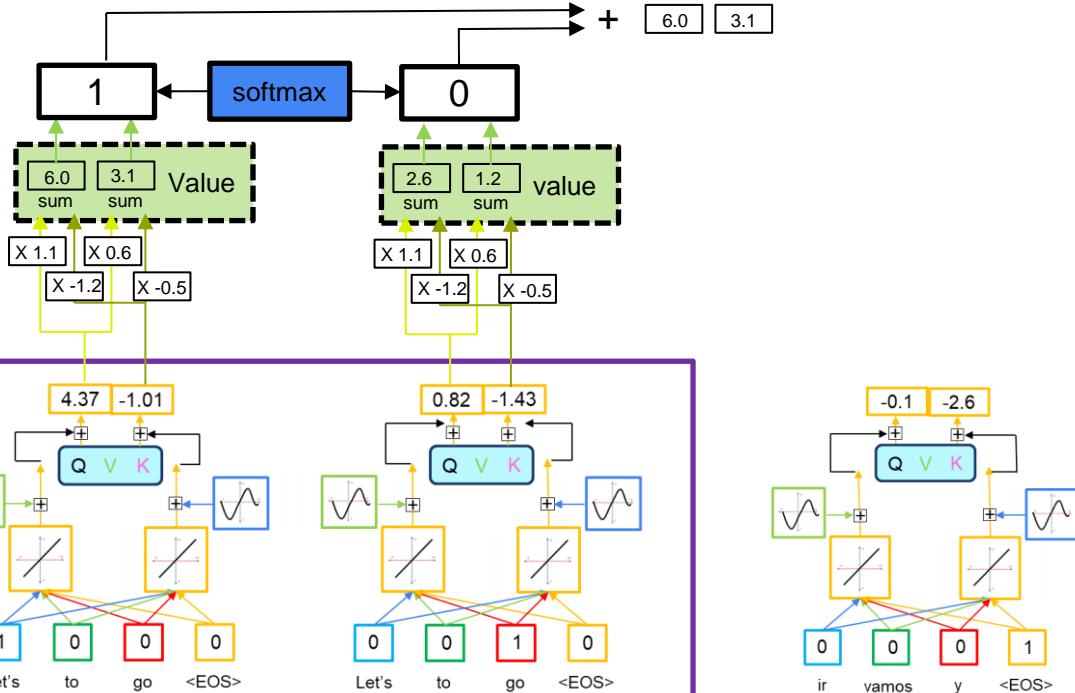


Attention entre l'encoder et le decoder



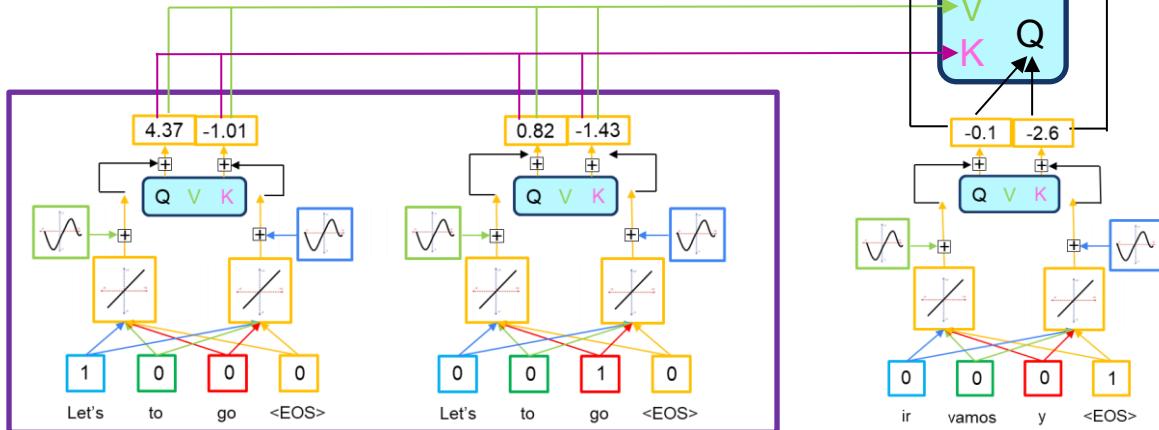


Attention entre l'encoder et le decoder





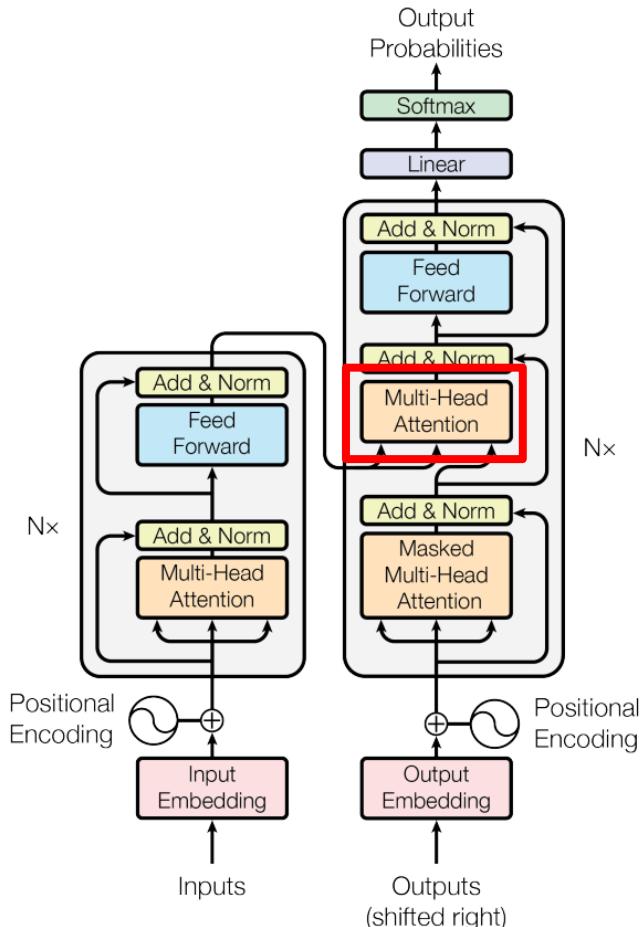
Attention entre l'encoder et le decoder



L'ajout d'une connexion résiduelle permet à l'attention du codeur-décodeur de se concentrer sur les relations entre les mots de sortie et l'entrée sans avoir à préserver le self-attention ou l'embedding du mot et de la position qui a eu lieu précédemment.

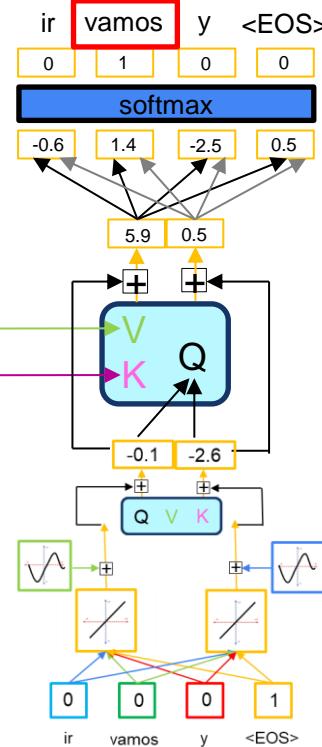
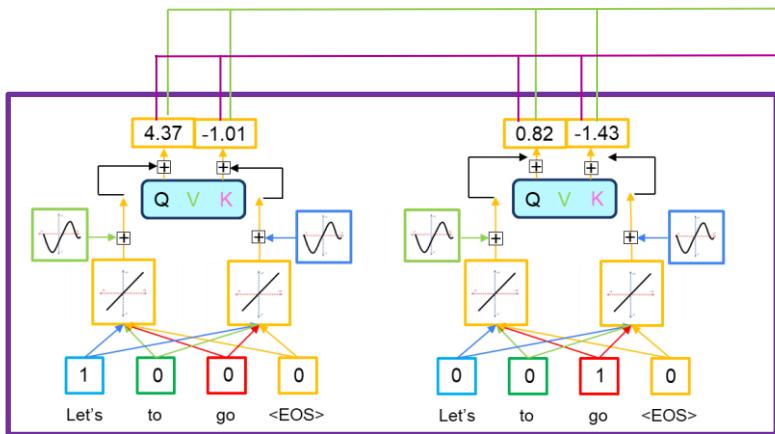


Transformer



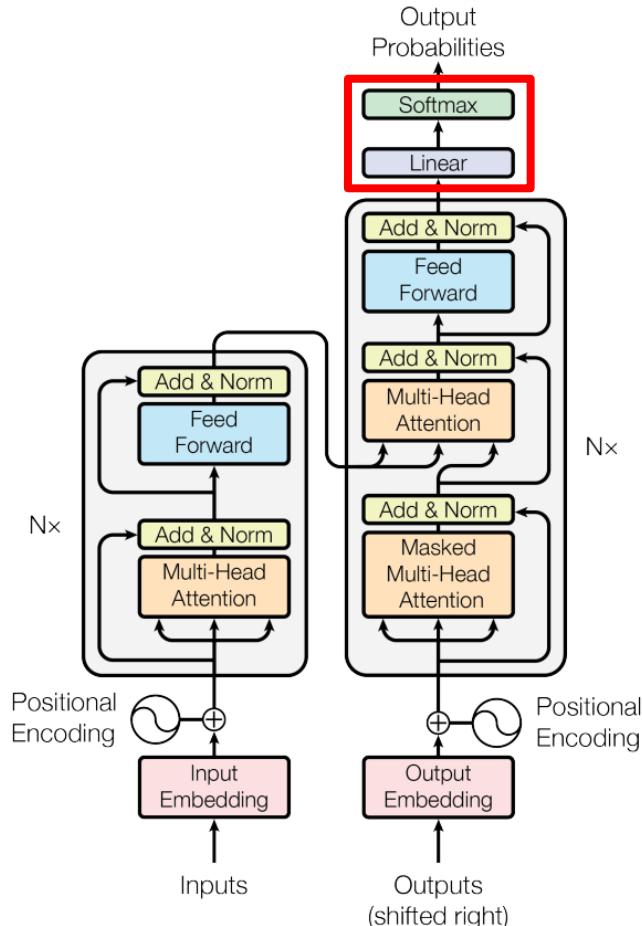


Fully connected layer





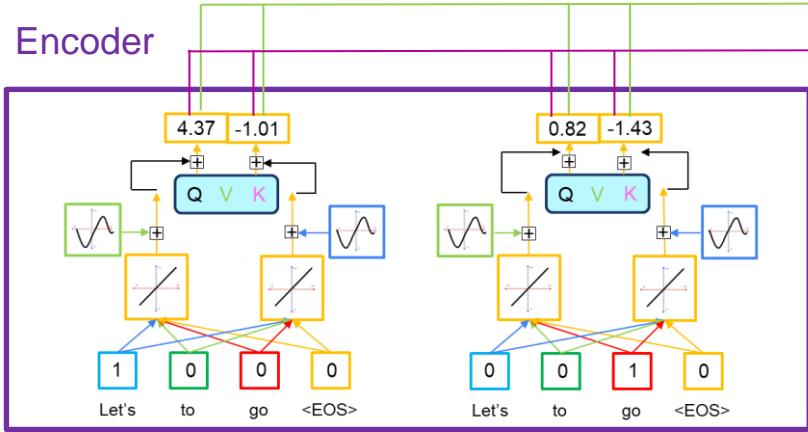
Transformer





Fully connected layer

Encoder



Decoder

