Module d’introduction à l’IA

CONSULTING BSTORM

1.Introduction

CONSULTING BSTORM

1. Introduction

Activité par groupe de deux ou trois:

Essayez de répondre aux questions suivantes:

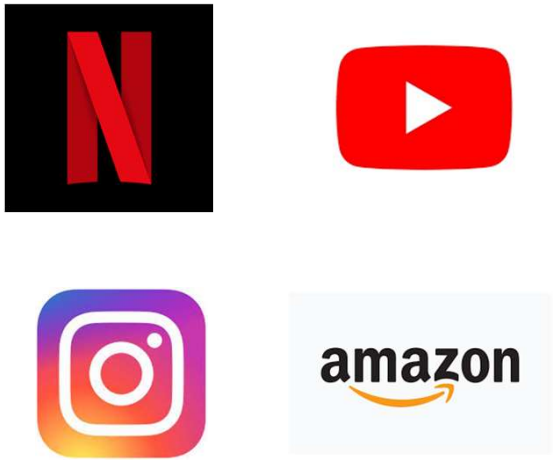
1. Dressez une liste d'applications IA qui vous entourent. Est-ce que vous pouvez regrouper certaines applications entre elles ? Pourquoi ?
2. Essayez d'expliquer avec vos mots comment est-ce que cela fonctionne. De quoi a-t-on besoin ?
3. Qui utilise cette technologie ? Pourquoi ?


ALPHAGO
2017
BFI LONDON FILM FESTIVAL
WINNER
MILL VALLEY FILM FESTIVAL
Official Selection
VIFF

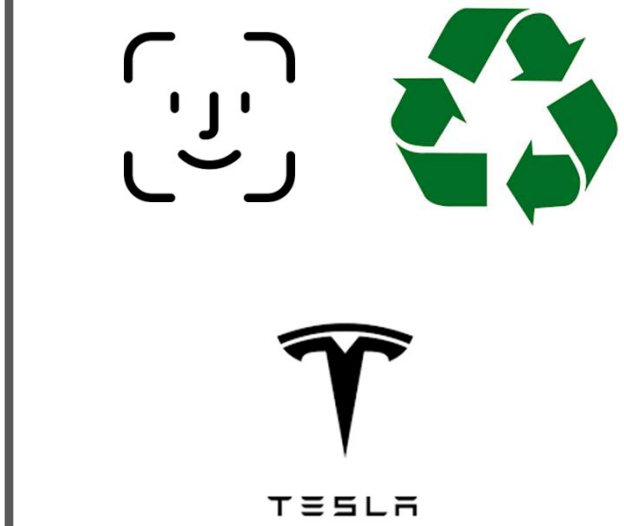

The main agent fails to defend itself.

1. Introduction

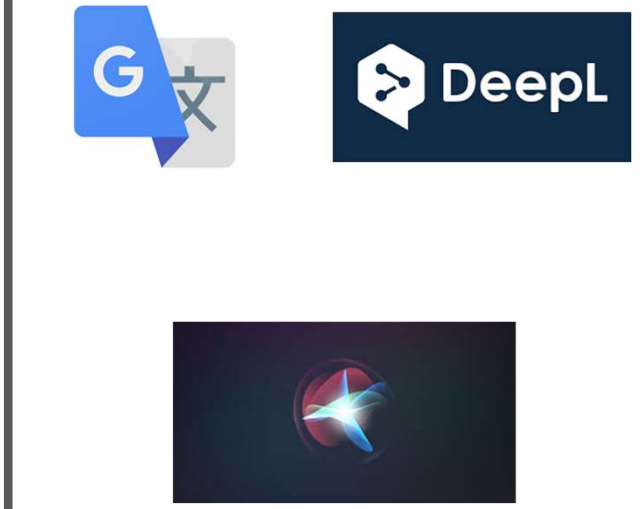
???



???



???



1. Introduction

Système de recommandation




Vision par ordinateur



Traitement automatique du langage (NLP)


1. Introduction

Et bien plus:

- Google Shadow Art
- LSFB
- Waze
- Médecine
- Finance
- Logistique

1. Introduction

Comment un ordinateur peut-il apprendre à reconnaître des images de voitures, d'animaux, jouer aux échecs, etc. ?

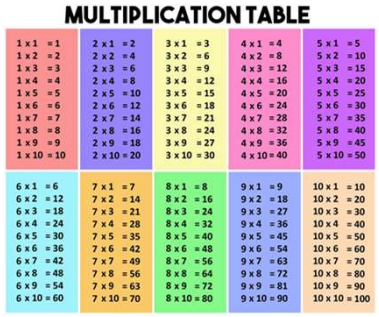
=> avec des données, beaucoup de données

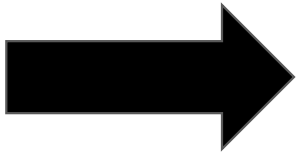
=> des algorithmes

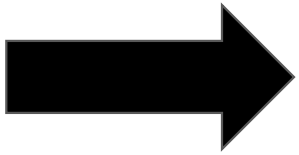
=> beaucoup de puissance de calcul

1. Introduction

On peut faire un rapprochement avec l'apprentissage d'un être humain. Par exemple pour apprendre une langue un être humain a besoin de beaucoup d'exemples, de phrases et de mots utilisés dans des contextes différents.







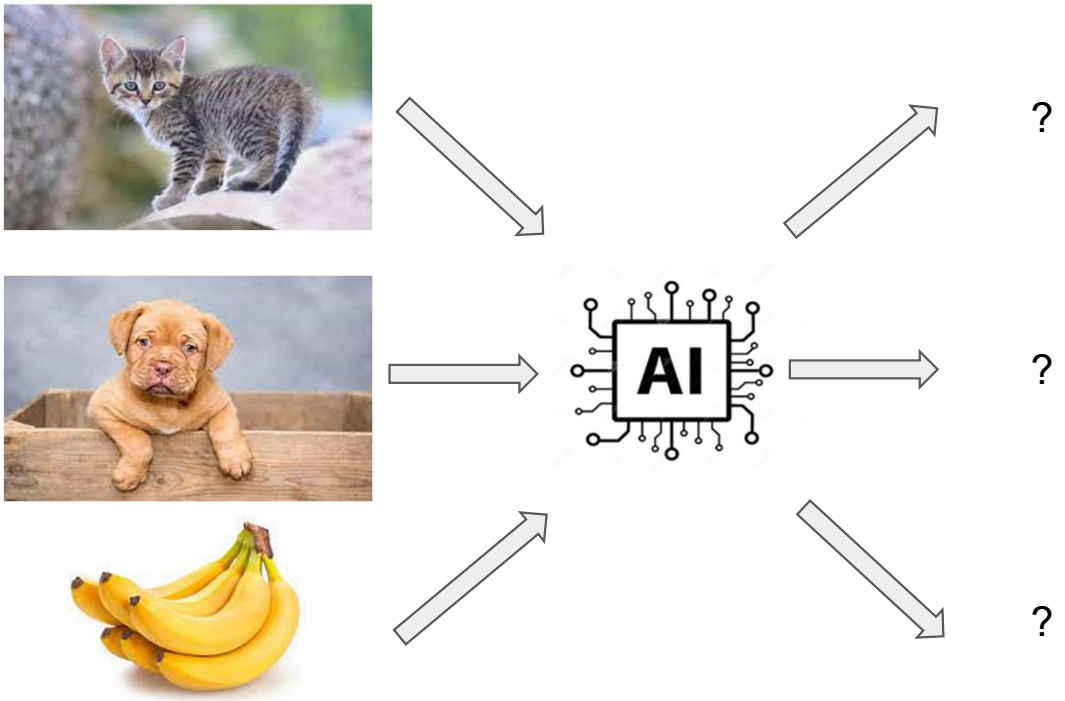
NVIDIA

1. Introduction

Mais est-ce qu'une intelligence artificielle est si intelligente que cela ?

Ci-contre un tweet qui résume assez bien le rapprochement entre intelligence artificielle et humaine.

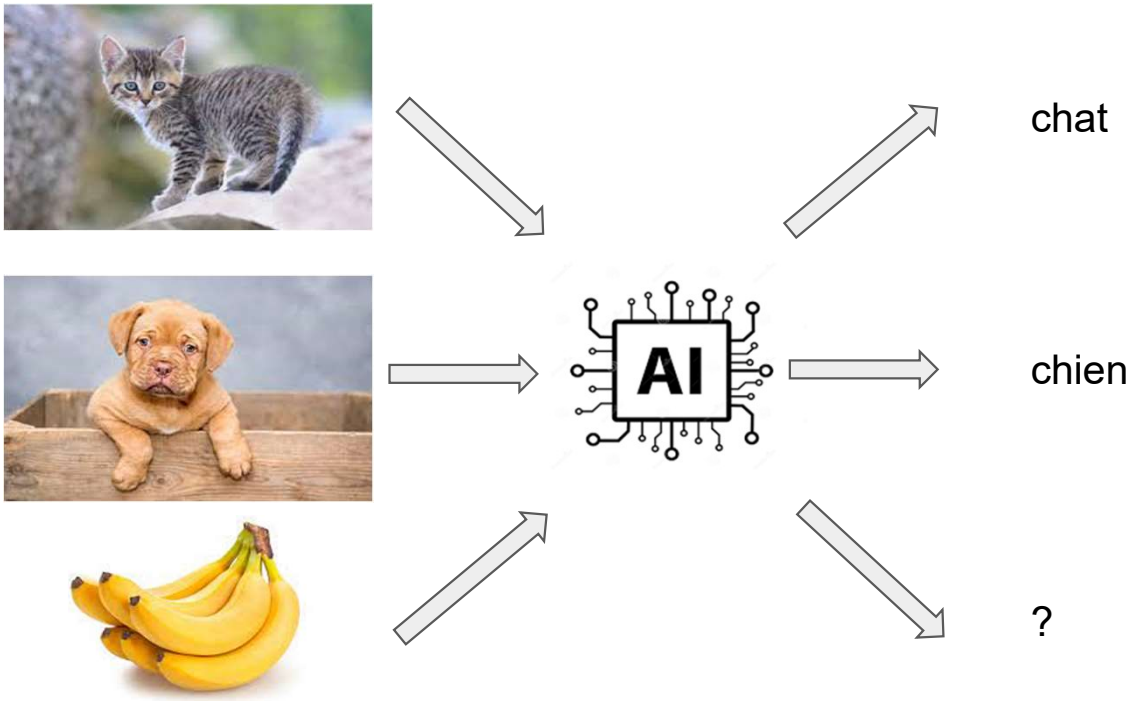
1. Introduction



Imaginons que nous entraînions une IA à différencier des chats et des chiens.

Que se passe-t-il si maintenant on lui présente une image de banane ?

1. Introduction



A priori, si le travail est bien fait, votre IA répondra correctement pour l'image de chat ou de chien.

Si on lui présente une image autre qu'un chien ou un chat, elle répondra l'un des deux.

1. Introduction

- Actuellement l'IA est utilisée afin de résoudre des problèmes très spécifiques. Nous ne sommes pas capables de produire une IA avec une **intelligence générale(AGI)** similaire à celle de l'être humain
- Nous pourrions entraîner une IA à passer des tests de QI mais elle échouerait lamentablement à différencier un lapin d'une carotte
- Si nous entraînons une IA à différencier un chat d'une moto, il faut beaucoup d'images de chats et de motos. Si désormais nous rajoutons des images de vélos, il en faudra tout autant. Or un vélo et une moto sont des objets très similaires

1. Introduction

Un cerveau humain consomme 300 à 500 kcal par jour soit 15 à 25 watts contre 250 watts pour un GPU professionnel (e.g. Tesla V100) mais il ne faut pas oublier certains points:

- Une seule carte graphique ne suffit pas toujours
- Il faut pouvoir les refroidir
- Leur durée de vie est limitée

1. Introduction

We have verified that the results match the paper when training with 1, 2, 4, or 8 GPUs. Note that training FFHQ at 1024×1024 resolution requires GPU(s) with at least 16 GB of memory. The following table lists typical training times using NVIDIA DGX-1 with 8 Tesla V100 GPUs:

Configuration	Resolution	Total kimg	1 GPU	2 GPUs	4 GPUs	8 GPUs	GPU mem
config-f	1024×1024	25000	69d 23h	36d 4h	18d 14h	9d 18h	13.3 GB
config-f	1024×1024	10000	27d 23h	14d 11h	7d 10h	3d 22h	13.3 GB
config-e	1024×1024	25000	35d 11h	18d 15h	9d 15h	5d 6h	8.6 GB
config-e	1024×1024	10000	14d 4h	7d 11h	3d 20h	2d 3h	8.6 GB
config-f	256×256	25000	32d 13h	16d 23h	8d 21h	4d 18h	6.4 GB
config-f	256×256	10000	13d 0h	6d 19h	3d 13h	1d 22h	6.4 GB

https://github.com/NVlabs/stylegan2

Voici les temps d'entraînement nécessaires pour la tâche suivante:

https://thispersondoesnotexist.com/

1. Introduction

Taille d'instance	Processeurs virtuels	Mémoire des instances (Gio)	GPU – A100	Mémoire de GPU	Bande passante du réseau (Gbit/s)	GPUDirect RDMA	Pair à pair GPU	Stockage d'instance (Go)	Bande passante EBS (Gbit/s)	Prix/heure pour les instances à la demande	Tarif horaire effectif des instances réservées sur 1 an*	Tarif horaire effectif des instances réservées sur 3 ans*
p4d.24xlarge	96	1 152	8	320 Go HBM2	400 ENA et EFA	Oui	NVSwitch 600 Go/s	8 disques SSD NVMe de 1 000	19	32,77 USD	19,22 USD	11,57 USD
p4de.24xlarge (version de prévisualisation)	96	1 152	8	640 Go HBM2e	400 ENA et EFA	Oui	NVSwitch 600 Go/s	8 disques SSD NVMe de 1 000	19	40,96 USD	24,01 USD	14,46 USD

1. Introduction

Malgré tout, le fantasme de “reproduire le cerveau humain” reste très prégnant dans nos sociétés pour au moins trois raisons

- Les médias généralistes en parlent comme une technologie presque magique
- Les livres, BD, séries, films nous baignent dans ce fantasme
- Le manque de formations

Le rapprochement entre les deux intelligences semble actuellement hasardeux. Il faudrait commencer par comprendre comment notre cerveau fonctionne mais aussi comment certaines IA prennent des décisions.

2. Les données

CONSULTING BSTORM

2. Les données

Activité par groupe de deux ou trois:

Essayez de répondre aux questions suivantes:

1. Donnez des exemples de données. Sont-elles toutes de même nature ?
2. Comment est-ce que l'on récolte des données ?
3. Comment les stocke-t-on ?
4. Que peut-on faire avec ?
5. Peut-on tout faire avec ?

2. Les données

Données structurées:

- Une livraison uber eats
- Une commande amazon
- Une liste de courses
- Une classe d'élèves et leurs notes

Données non structurées:

- Images, vidéos
- Du texte, livres, tweets, commentaires Youtube, etc.
- Sons
- Données médicales (génome)

2.1 Acquisition de données

Twitter API

2.1 Acquisition des données

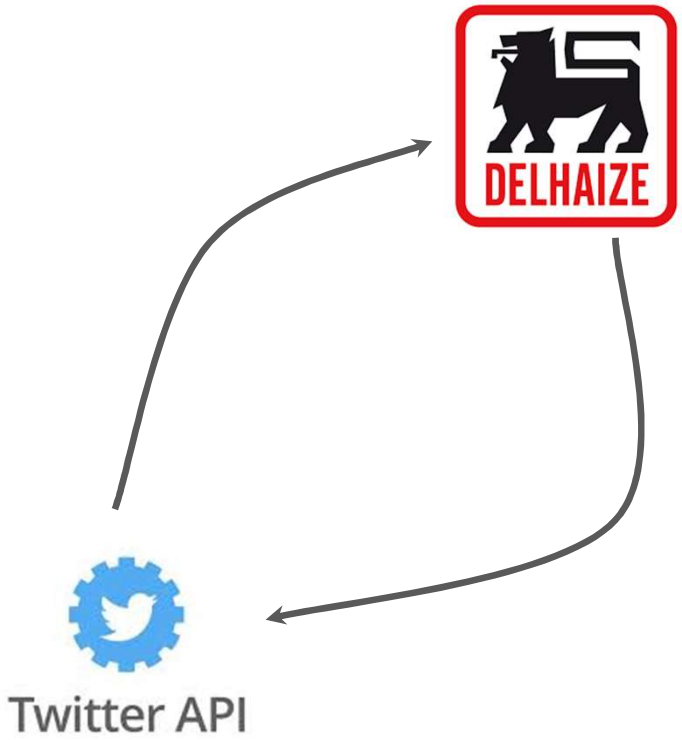
Comment fait-on pour générer, récolter ou obtenir des données ?

- L'encodage manuel
- Sondages, enquêtes de satisfaction, etc.
- Lorsque vous vous créer un compte ou lorsque de la connection
- Lorsque vous commentez, likez, swipez, postez une photo, etc.
- Il est possible d'acheter des données

2.1 Acquisition des données

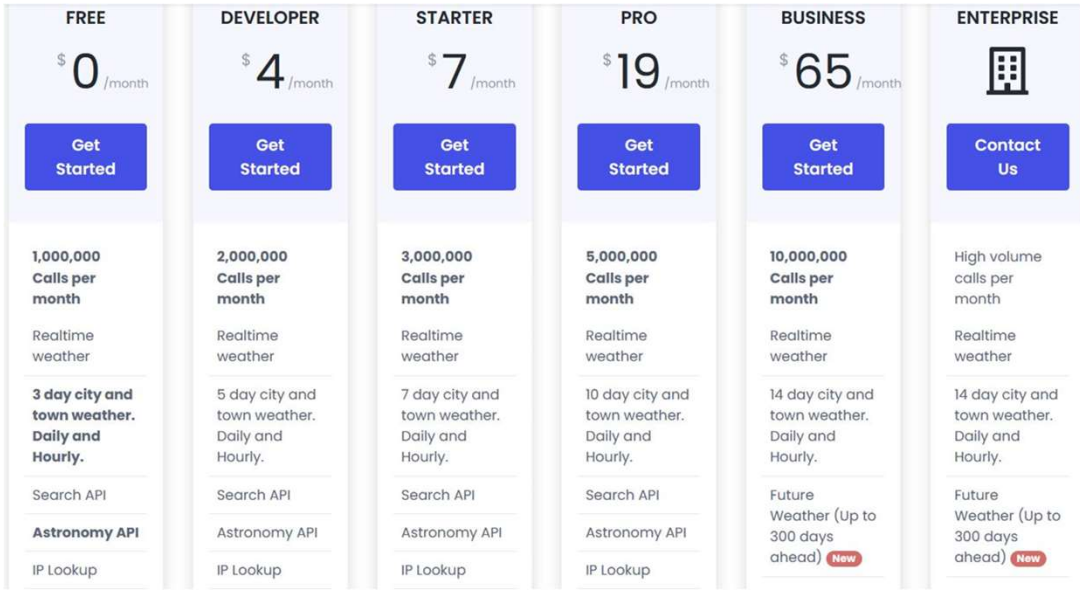
Imaginons que vous travaillez pour Delhaize et vous lancez une nouvelle campagne publicitaire sur les réseaux sociaux.

Comment obtenir les données qui permettraient d'analyser la réception de la campagne ?



2.1 Acquisition des données

Les API sont des services qui permettent à des entreprises, agences gouvernementales, ONG d'exposer vers l'extérieur leurs données.



FREE	DEVELOPER	STARTER	PRO	BUSINESS	ENTERPRISE
$0 /month	$4 /month	$7 /month	$19 /month	$65 /month	
Get Started	Get Started	Get Started	Get Started	Get Started	Contact Us
1,000,000 Calls per month	**2,000,000 Calls per month**	**3,000,000 Calls per month**	**5,000,000 Calls per month**	**10,000,000 Calls per month**	High volume calls per month
Realtime weather	Realtime weather	Realtime weather	Realtime weather	Realtime weather	Realtime weather
3 day city and town weather. Daily and Hourly.	5 day city and town weather. Daily and Hourly.	7 day city and town weather. Daily and Hourly.	10 day city and town weather. Daily and Hourly.	14 day city and town weather. Daily and Hourly.	14 day city and town weather. Daily and Hourly.
Search API	Search API	Search API	Search API	Future Weather (Up to 300 days ahead) New	Future Weather (Up to 300 days ahead) New
Astronomy API	Astronomy API	Astronomy API	Astronomy API		
IP Lookup	IP Lookup	IP Lookup	IP Lookup		

https://www.weatherapi.com/pricing.aspx

2.1 Acquisition des données

Le web scraping est un autre outil permettant d'extraire de manière automatisée à l'aide de robots des informations, photos, vidéos d'une page internet.

Plusieurs inconvénients:

- Les sites s'en protègent de plus en plus
- Pas toujours légal
- Si le site change, le scraping ne fonctionne plus

Quotes to Scrape

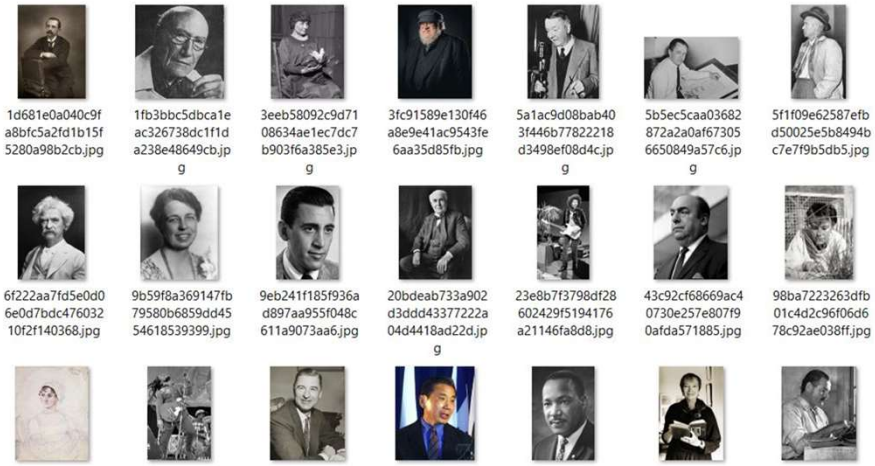
"The world as we have created it is a process of our thinking. It cannot be changed without changing our thinking."
by Albert Einstein (about)
Tags: change deep-thoughts thinking world

"It is our choices, Harry, that show what we truly are, far more than our abilities."
by J.K. Rowling (about)
Tags: abilities choices

"There are only two ways to live your life. One is as though nothing is a miracle. The other is as though everything is a miracle."
by Albert Einstein (about)
Tags: inspirational life live miracle miracles

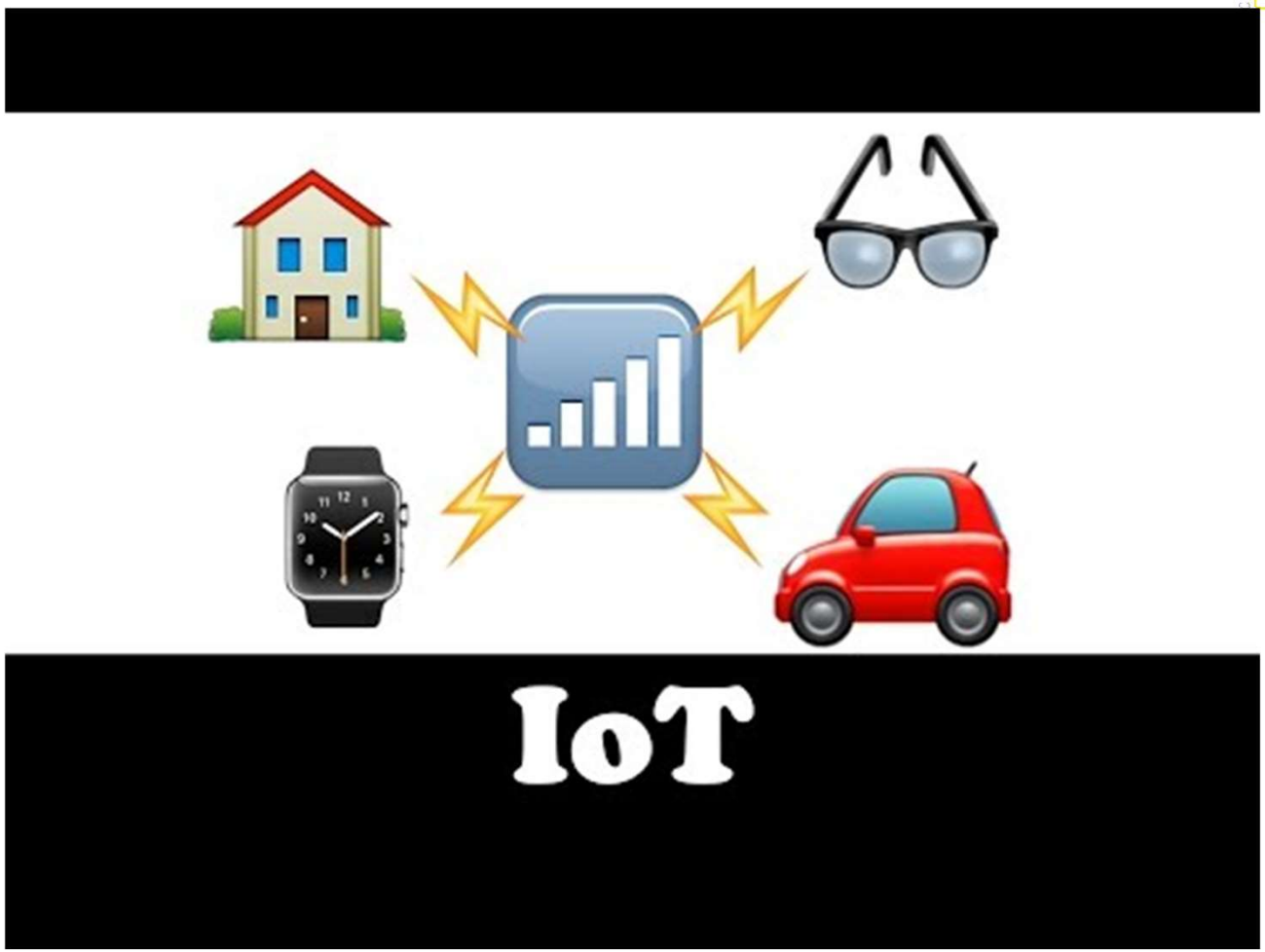
"The person, be it gentleman or lady, who has not pleasure in a good novel, must be intolerably stupid."
by Jane Austen (about)
Tags: aliteracy books classic humor

https://quotes.toscrape.com/



quote	author	tags	born_date	born_place
the person be it gentleman or lady who h...	Jane Austen	aliteracy,books,classic...	1775-12-16 00:00:00.000	in Steventon Rectory, Hampshir...
it is our choices harry that show what we ...	J.K. Rowling	abilities,choices	1965-07-31 00:00:00.000	in Yate, South Gloucestershire, ...
imperfection is beauty madness is geniu...	Marilyn Monroe	be-yourself,inspirational	1926-06-01 00:00:00.000	in The United States
the world as we have created it is a proc...	Albert Einstein	change,deep-thoughts...	1879-03-14 00:00:00.000	in Ulm, Germany
there are only two ways to live your life o...	Albert Einstein	inspirational,life,live,mi...	1879-03-14 00:00:00.000	in Ulm, Germany
try not to become a man of success rath...	Albert Einstein	adulthood,success,val...	1879-03-14 00:00:00.000	in Ulm, Germany
if you want your children to be intelligent ...	Albert Einstein	children,fairy-tales	1879-03-14 00:00:00.000	in Ulm, Germany
a wise girl kisses but doesnt love listens ...	Marilyn Monroe	attributed-no-source	1926-06-01 00:00:00.000	in The United States
love does not begin and end the way we ...	James Baldwin	love	1924-08-02 00:00:00.000	in Harlem, New York, The Unite...
it matters not what someone is born but ...	J.K. Rowling	dumbledore	1965-07-31 00:00:00.000	in Yate, South Gloucestershire, ...
life is like riding a bicycle to keep your ba...	Albert Einstein	life,simile	1879-03-14 00:00:00.000	in Ulm, Germany
the real lover is the man who can thrill yo...	Marilyn Monroe	love	1926-06-01 00:00:00.000	in The United States
anyone who thinks sitting in church can ...	Garrison Keillor	humor,religion	1942-08-07 00:00:00.000	in Anoka, Minnesota, The Unite...
a woman is like a tea bag you never kno...	Eleanor Roosevelt	misattributed-eleanor-...	1884-10-11 00:00:00.000	in The United States
i may not have gone where i intended to ...	Douglas Adams	life,navigation	1952-03-11 00:00:00.000	in Cambridge, England, The Uni...


arte
Le business
des données


IoT

2.2 Stockage des données

2.2 Le stockage des données

Quand on parle de stockage des données, nous avons besoin a minima de trois éléments:

1. D'une base de données
2. Un modèle, une façon de voir et de stocker les données
3. Un système de gestion de base de données (SGBD), un logiciel permettant de gérer une base de données

2.2 Stockage des données

Le modèle le plus répandu pour la conception des bases de données est le modèle relationnel qui s'articule autour d'un système de tables, de relations entre les tables et de règles à respecter.

Il est accompagné de toute une série de SGBD différents prenant en charge le modèle relationnel. Eux-mêmes accompagnés d'un langage (SQL) permettant notamment de créer, interroger et modifier une base de données.

id Client	Nom	Prénom	Adresse
1	De Croo	Alice	Rue des anges, 32
2	Dupont	Eric	Avenue Charles Lestrange, 265
3	Heidegger	Margot	Rue des anges, 15

numéro commande	Date	Client	Produit
1	10/08/2020	1	2
2	02/01/2021	1	2
3	15/05/2021	2	1
4	28/06/2021	3	3

id produit	Nom	Prix
1	Iphone	800€
2	Samsung Galaxy	750€
3	Sony XM4	300€

2.2 Stockage des données

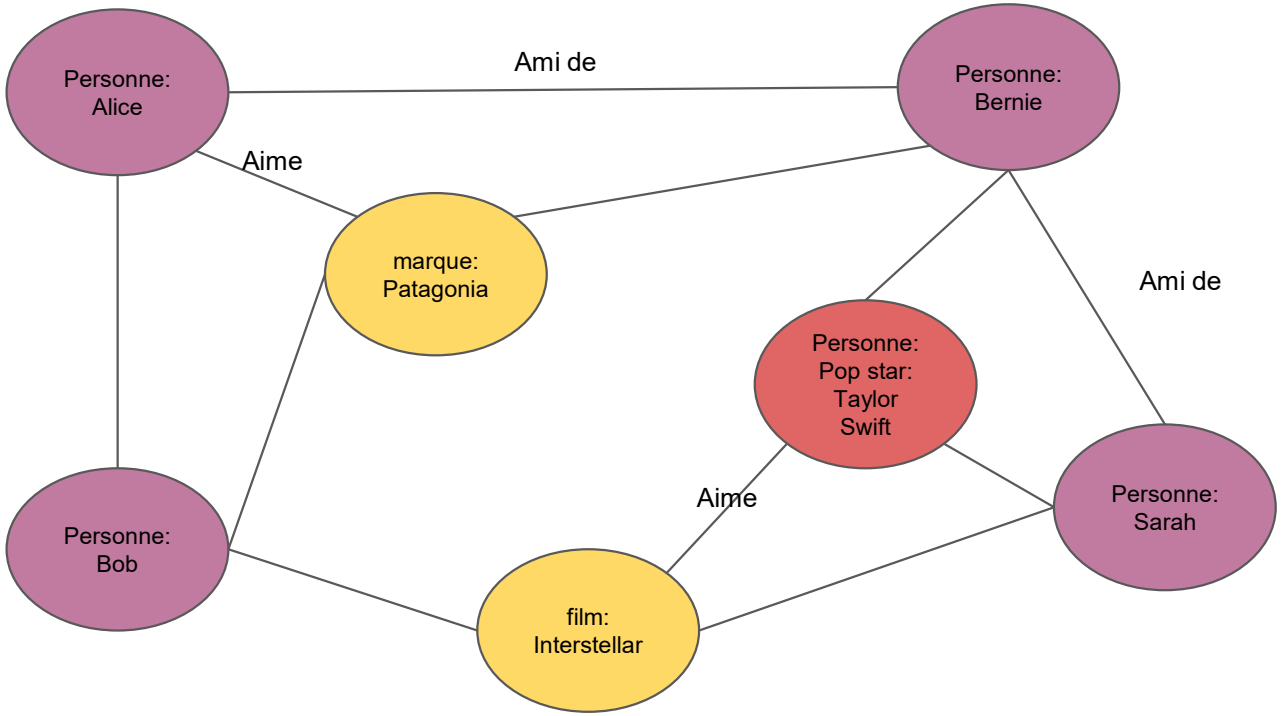
Avantages

- Vision très structurée de ses données
- Prend peu de place sur un disque dur
- Un langage de base commun (SQL)

Inconvénients

- Comment gérer des données non structurées ?
- Problèmes de performance lors de lourdes opérations sur de gros volumes de données

2.2 Stockage des données



A côté des bases de données relationnelles, nous retrouvons des bases de données orientées:

- Documents
- Graphes (*cfr.* ci-contre)
- etc.

2.2 Stockage des données

SGBD SQL

SGBD NoSQL

2.2 Stockage des données

Pourquoi des SGBD ?:

- Sécurité
- Cohérence
- Données de meilleure qualité
- Plus facile pour partager des données
- etc.

2.3 Tout n’est pas IA

2.3 Tout n’est pas IA

L’utilisation des données dépassent largement le cadre de l’intelligence artificielle.

- Elles ont d’abord un rôle opérationnel. Si Amazon demande votre adresse c’est d’abord et avant tout afin de savoir où expédier vos commandes. Cela peut également servir à optimiser les trajets des livreurs par exemple.
- Pour des questions légales, les entreprises doivent conserver leurs données durant un temps donné

2.3 Tout n’est pas IA

Une autre utilisation possible consiste à mettre en évidence certaines informations contenues dans les données à l’aide de graphiques (reporting):

- Permet de rendre intelligible des informations contenues dans vos données
- Permet de traquer ses *Key Performance Indicator* (KPI) en temps réel
- Permet d’aider la prise de décision
- etc.

2.3 Tout n’est pas IA

Tableau ou PowerBi sont des outils qui permettent de produire des dashboards interactifs afin de présenter des données, présenter des analyses, etc.

Exemple:

- Accidents de vélo à Londres
- Most Wanted
- Brugel

3. Les algorithmes

Comment un ordinateur apprend ?

CONSULTING BSTORM

3.1 Exercice introductif

3.1 Exercice introductif

Pour commencer, entraînez une IA à reconnaître cinq objets différents qui vous entourent avec l'outil suivant: https://experiments.withgoogle.com/teachable-machine

- Essayez de l'entraîner à vous reconnaître vous et pas les autres
  - Est-ce exactement le même type de tâche que de reconnaître un objet ? Justifiez
- Essayez de l'entraîner à reconnaître des émotions
  - Est-ce une tâche plus difficile que de reconnaître des objets ? Oui/non, pourquoi ?
- Identifiez les concepts importants pour l'entraînement

3.1 Exercice introductif

Activité par groupe de deux ou trois:

1. Deux problématiques de la vie de tous les jours vous sont proposées
2. Tentez de les résoudre et surtout notez les étapes et les éléments qui vous semblent importants pour résoudre ces problèmes. Plusieurs approches sont possibles, n'hésitez pas à les répertorier

3.1 Exercice introductif

Choisir sa voiture

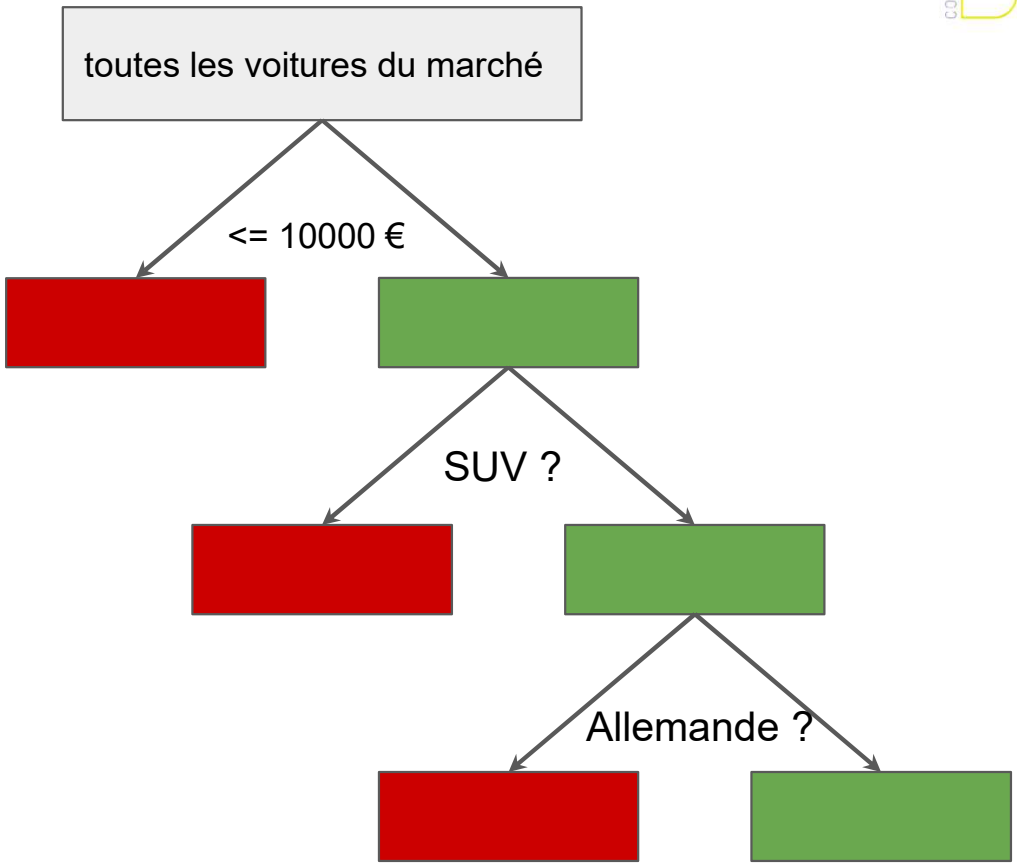
Problématique:

- Sur le marché automobile les acheteurs peuvent vite se perdre lors de leur prochain achat de voiture
- Comment feriez-vous pour choisir votre prochaine voiture dans ce catalogue presque inépuisable de voitures ?

Optimiser sa récolte

Problématique:

- Vous êtes agriculteurs ou novices en jardinerie et vous ne savez pas quand planter vos graines afin de maximiser votre récolte
- Comment feriez-vous pour déterminer la meilleure période de plantation ?

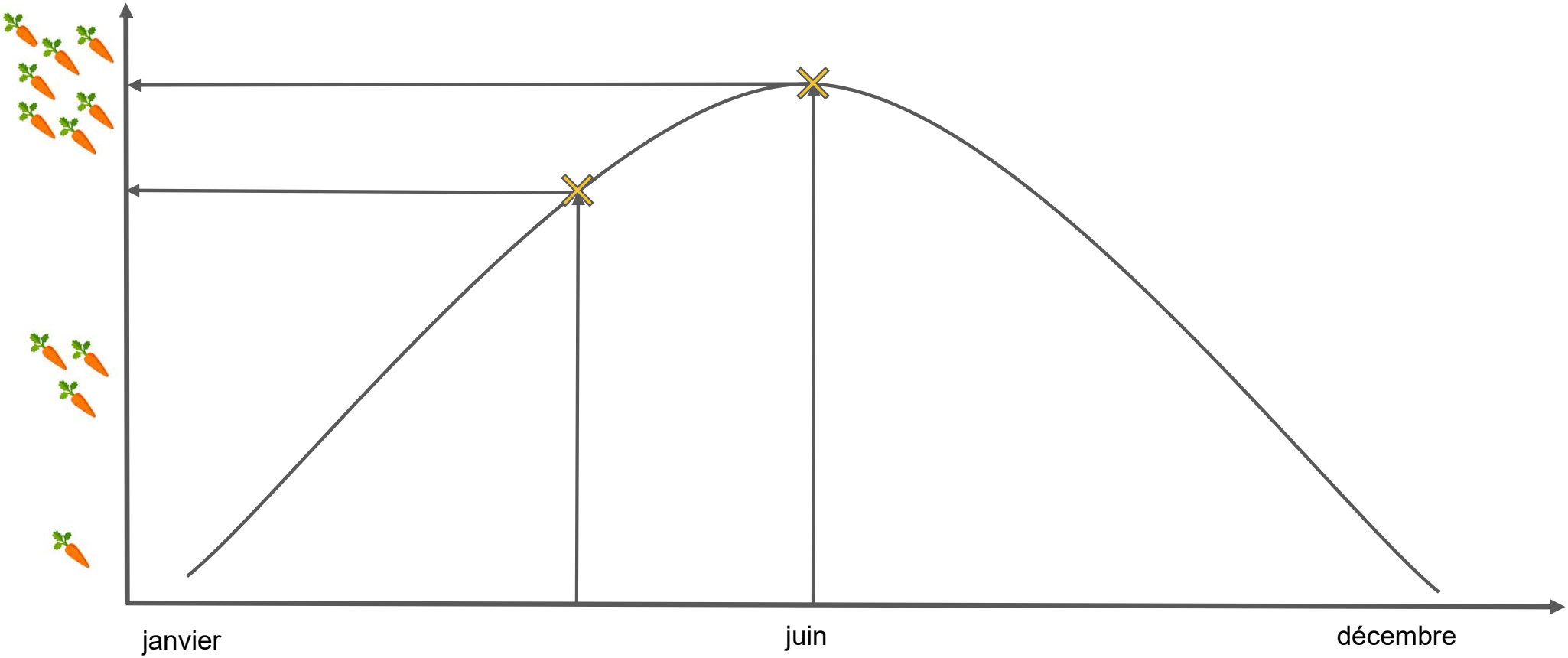

toutes les voitures du marché
<= 10000 €
SUV ?
Allemande ?

3.1 Exercice introductif

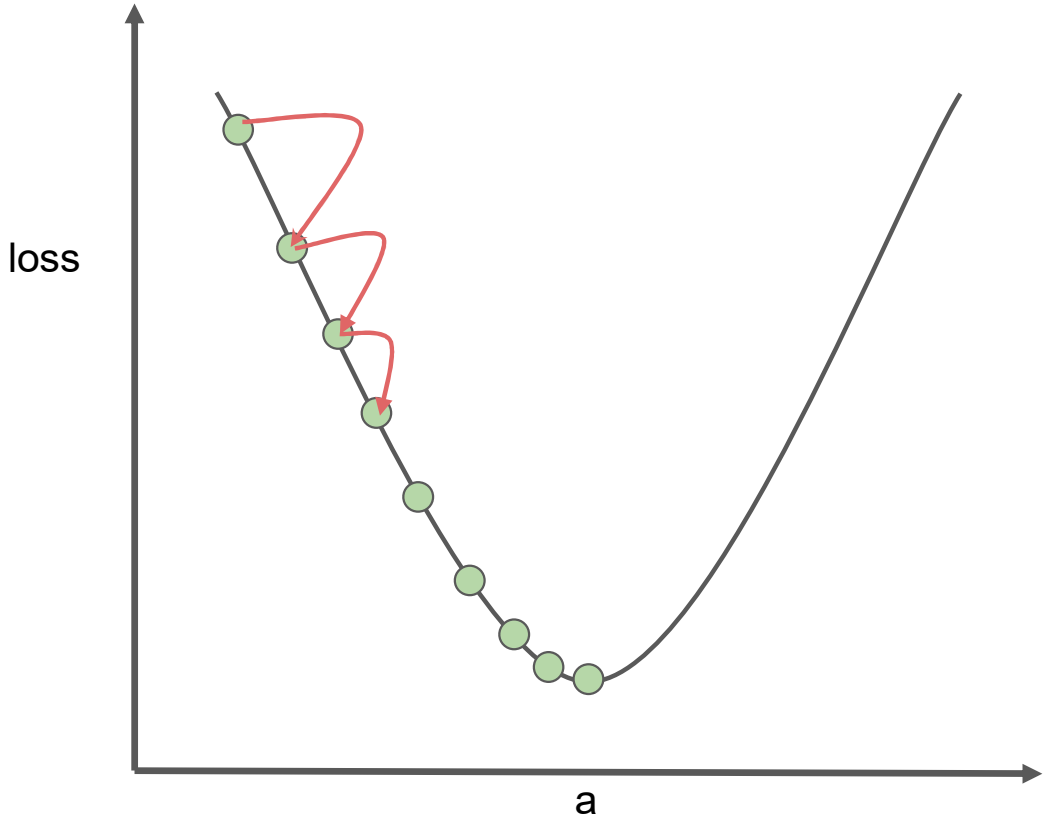
Les arbres de décisions sont des algorithmes très répandus pour diverses raisons:

- Facile à comprendre
- Facile à construire (suite de décisions binaires)
- Très efficace malgré leur apparente naïveté
- On peut les utiliser pour divers problèmes. Estimer le prix d'une maison, d'une voiture, pour choisir sa prochaine voiture. On les utilise également dans certains jeux vidéos

3.1 Exercice introductif



3.1 Exercice introductif



L'algorithme de la descente de gradient est un algorithme d'optimisation permettant de minimiser ce que l'on appelle une fonction coût (loss).

Dans notre exemple notre fonction coût pourrait être le nombre de carotte perdues ou qui n'ont pas poussé

3.2 Concepts de base

3.2 Concepts de base

Pour rappel, un ordinateur a besoin de beaucoup d'exemples pour son apprentissage. C'est que l'on appelle un dataset

Il ne suffit pas de lui donner des exemples. Il faut également lui donner une méthode pour apprendre (un algorithme)

Cependant, il existe beaucoup d'algorithmes différents qui répondent à des situations différentes

3.2 Concepts de base

Un dataset est constitué:

- Samples: les **observations**, les lignes
- Features: les **variables explicatives**, les colonnes, les **X**
- Label(target): facultatif, une **valeur cible** que l'on souhaite déterminer, **y**

Deux grandes catégories d'apprentissage:

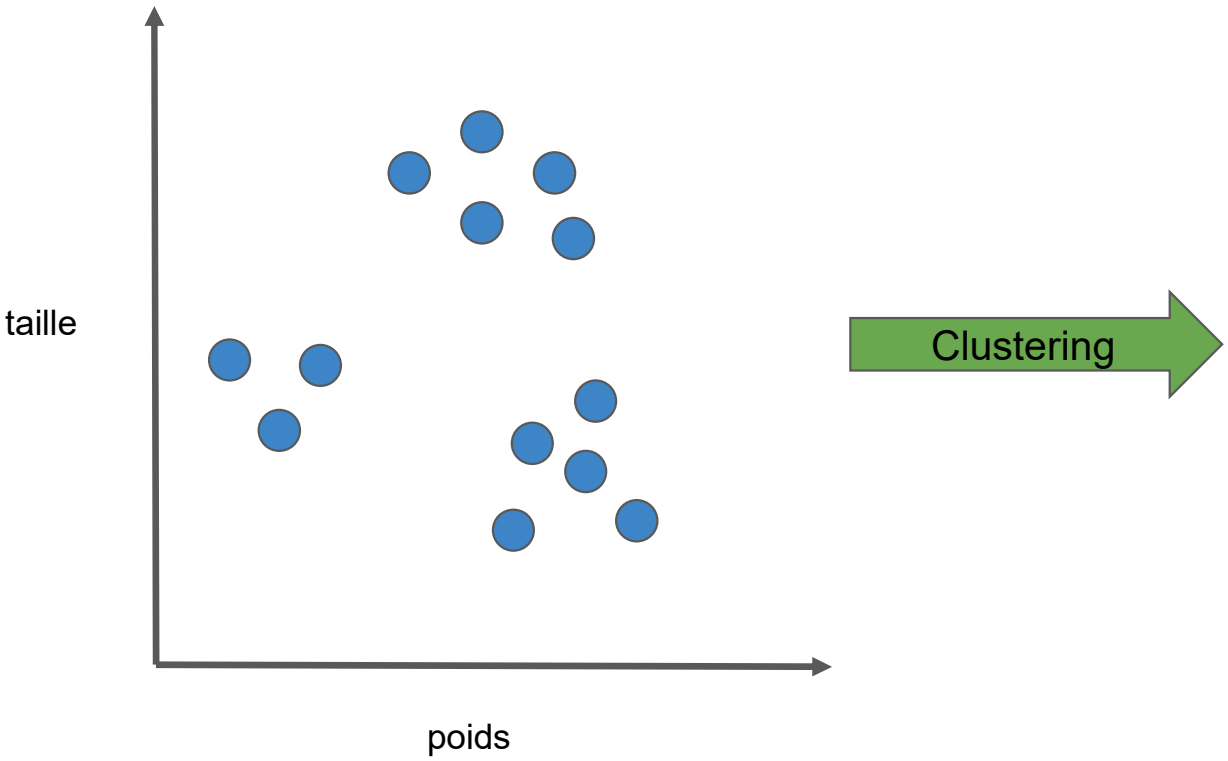
- **Non supervisé**: pas de target
- **Supervisé**: une ou plusieurs targets

3.2 Concepts de base

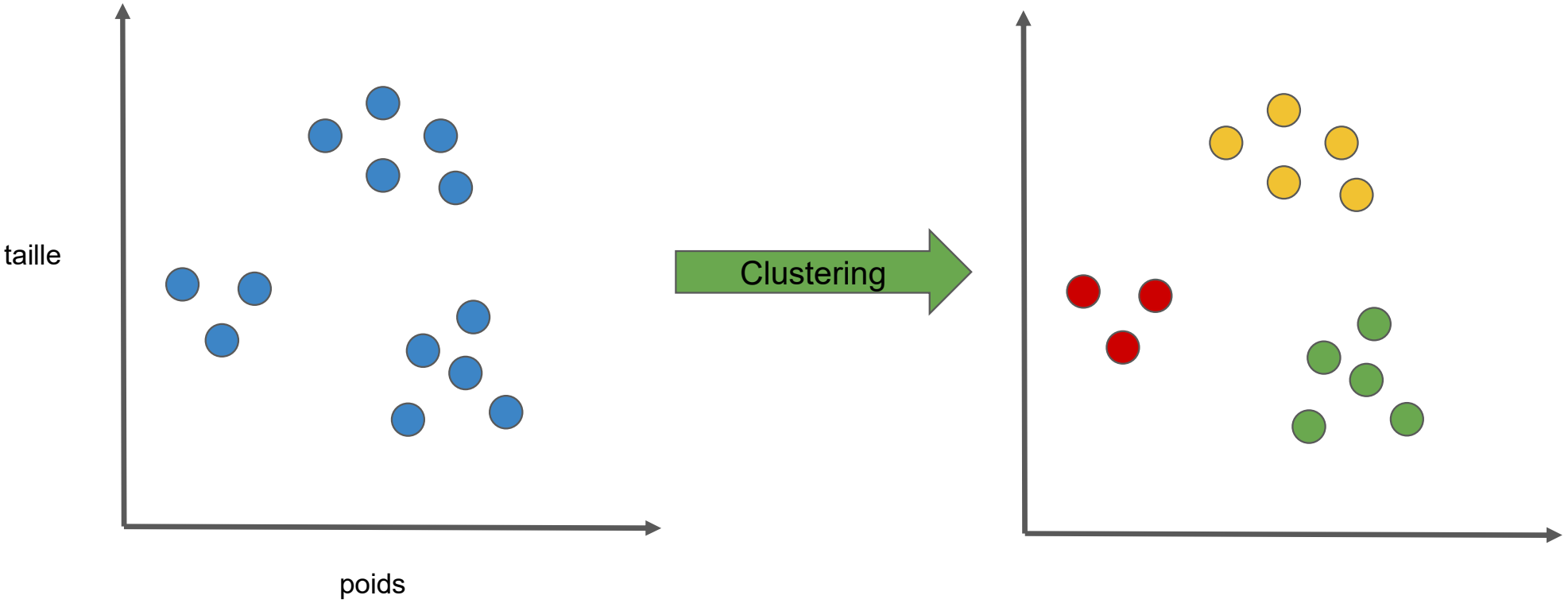
3 grandes catégories de problèmes:

- **Clustering**: non supervisé, on regroupe les individus semblables entre eux
- **Régression**: supervisé, on cherche à déterminer la valeur d'une variable quantitative continue(e.g. le prix d'une maison)
- **Classification**: supervisé, on cherche à déterminer la valeur d'une variable discrète(e.g. chien ou chat)

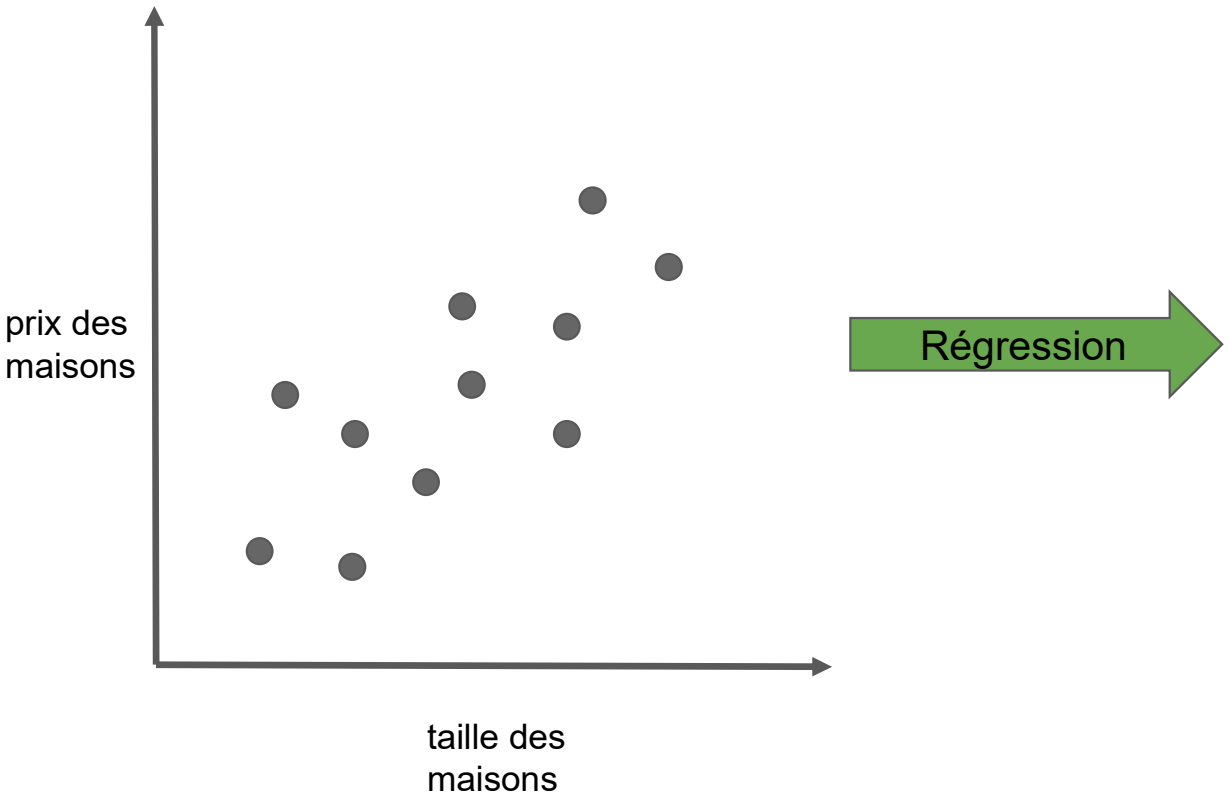
3.2 Concepts de base



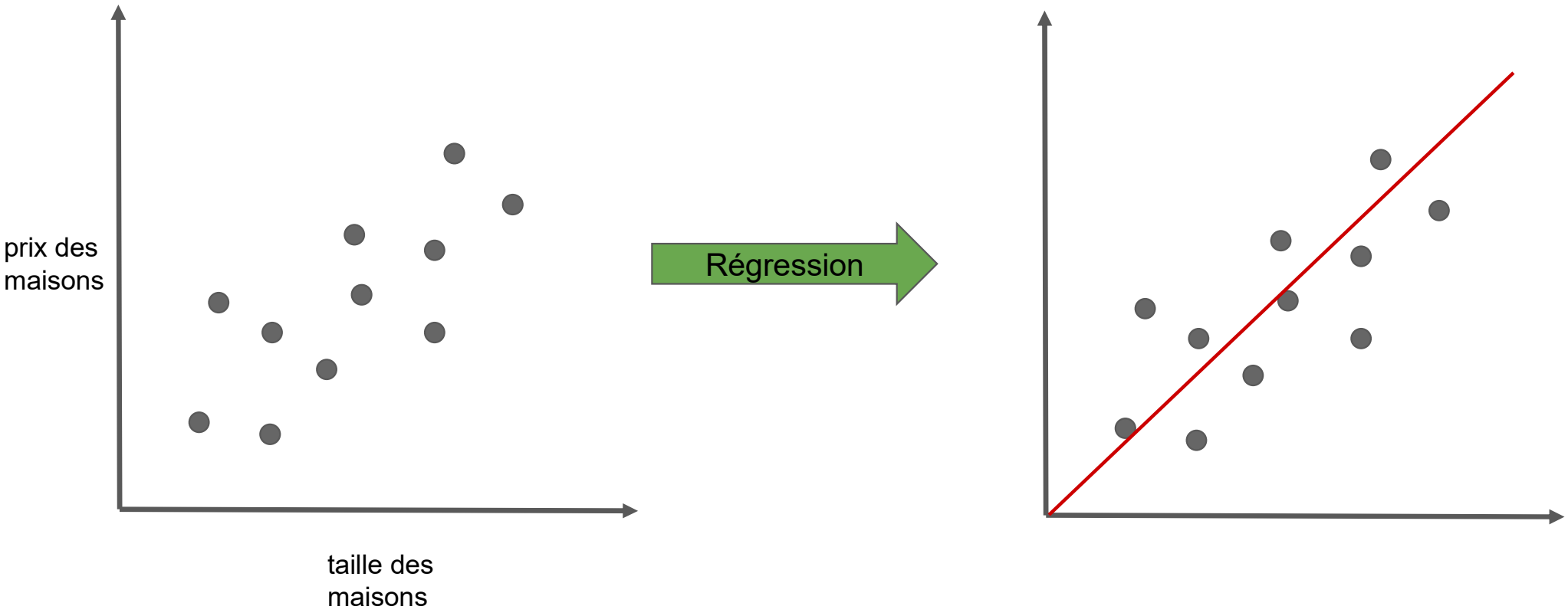
3.2 Concepts de base



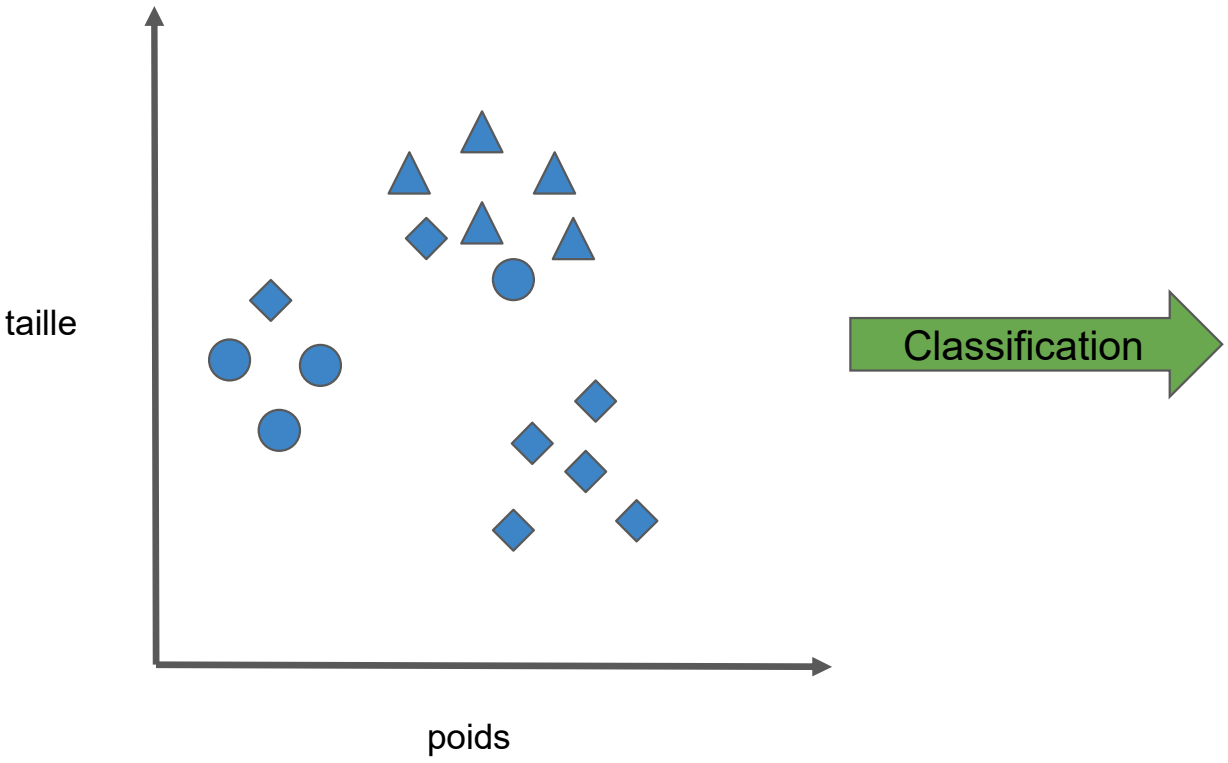
3.2 Concepts de base



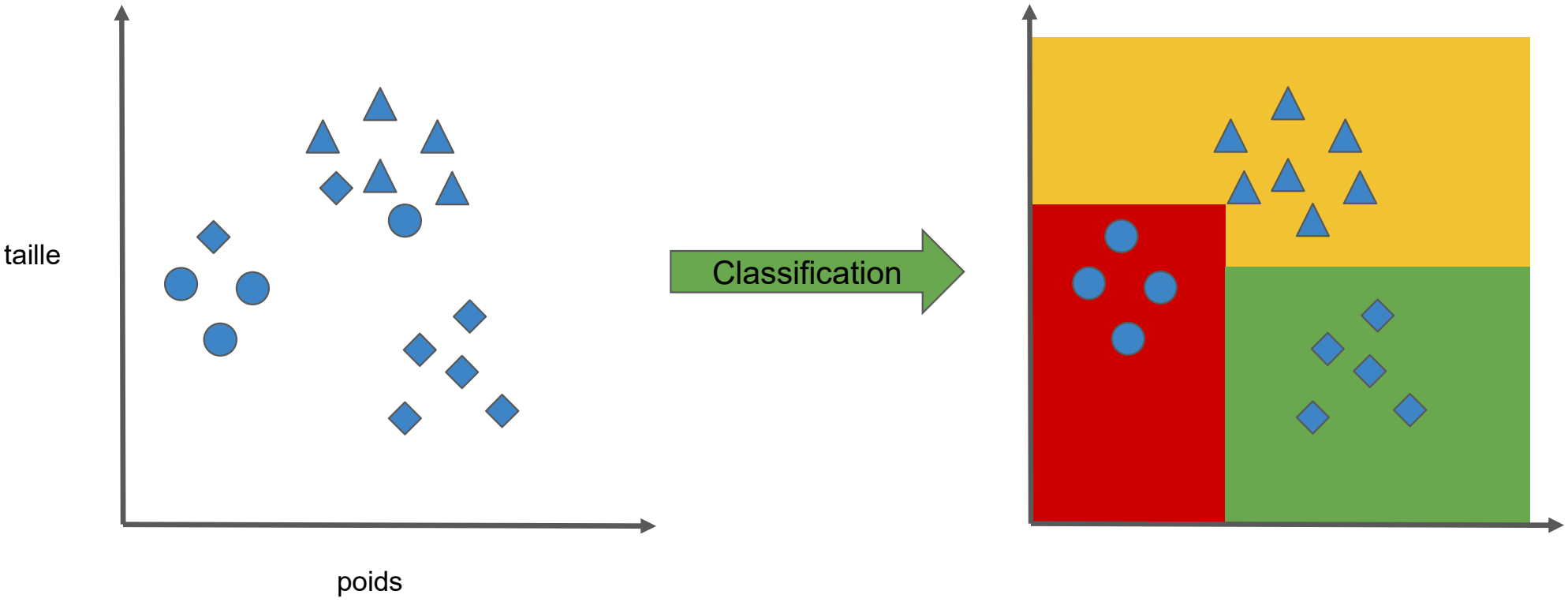
3.2 Concepts de base



3.2 Concepts de base



3.2 Concepts de base



3.2 Concepts de base

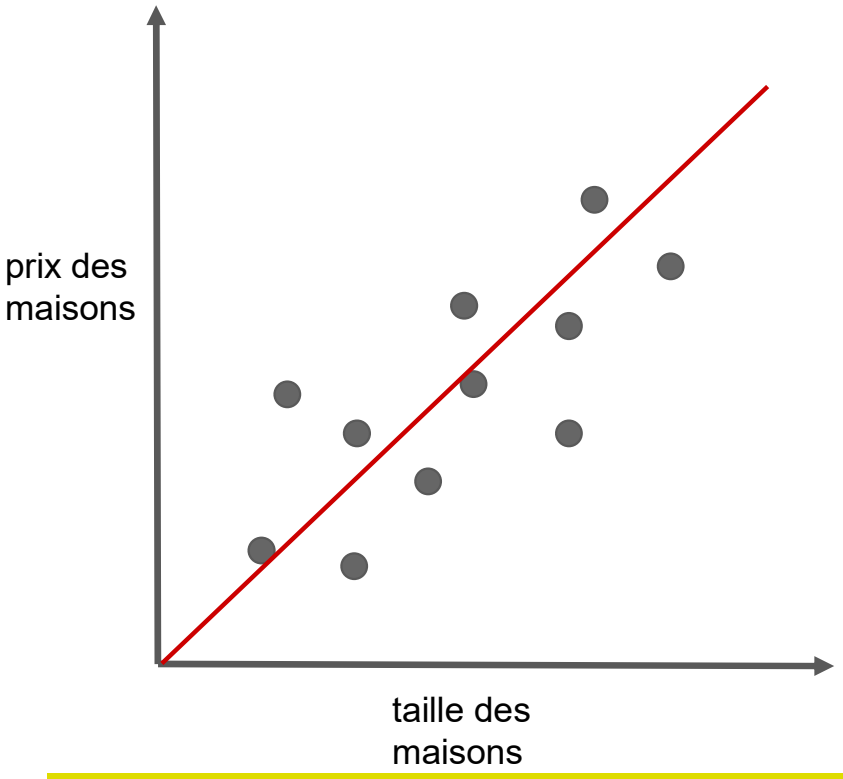
Concrètement que signifie entraîner un algorithme?

Deux exemples pour mieux comprendre:

- La régression linéaire et de manière plus générale les modèles linéaires
- Les arbres de décision

3.3 La régression linéaire

3.3 La régression linéaire



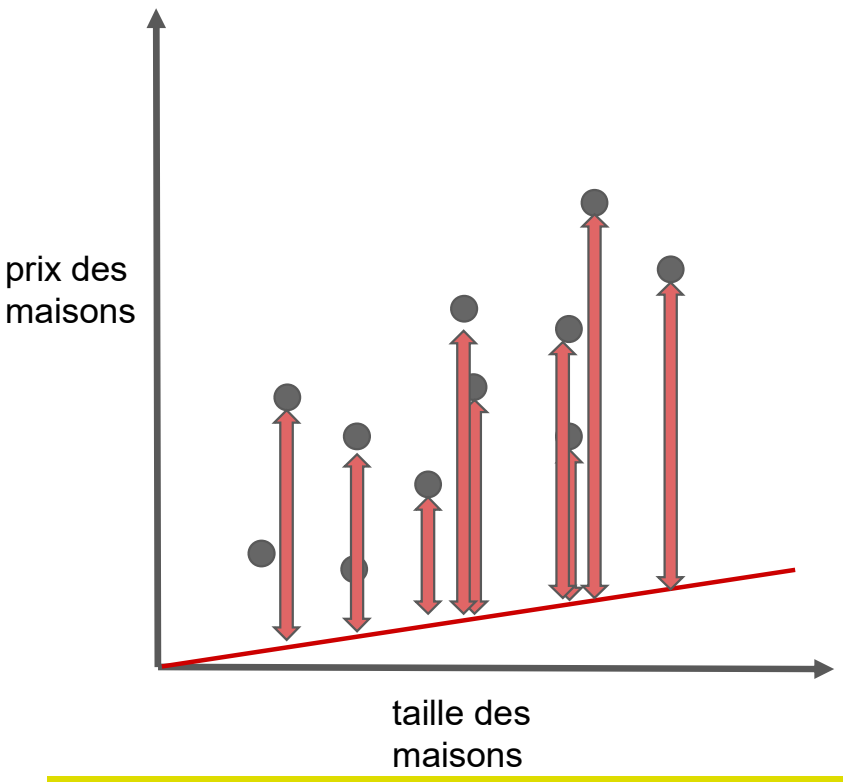
Nous avons tracé une **ligne** passant entre nos points

Rappel, une droite:

f(x) = ax+b

- a = coefficient
- b = constante
- x = variable (taille de notre maison)
- f(x) = pour faire simple, ici notre target

3.3 La régression linéaire

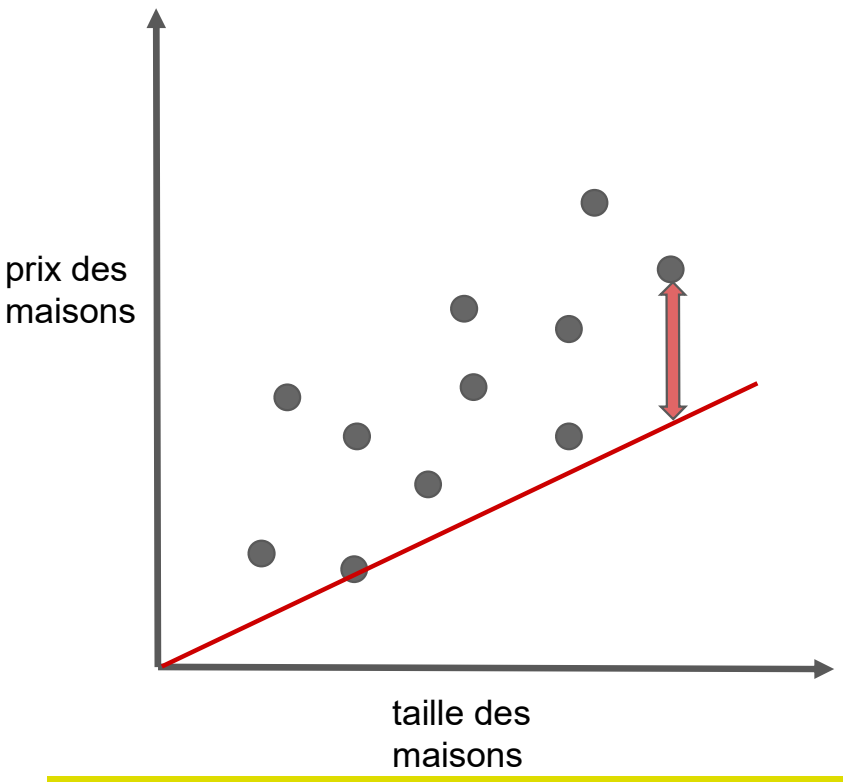


Comment trouver le bon coefficient ?

Par essai-erreur:

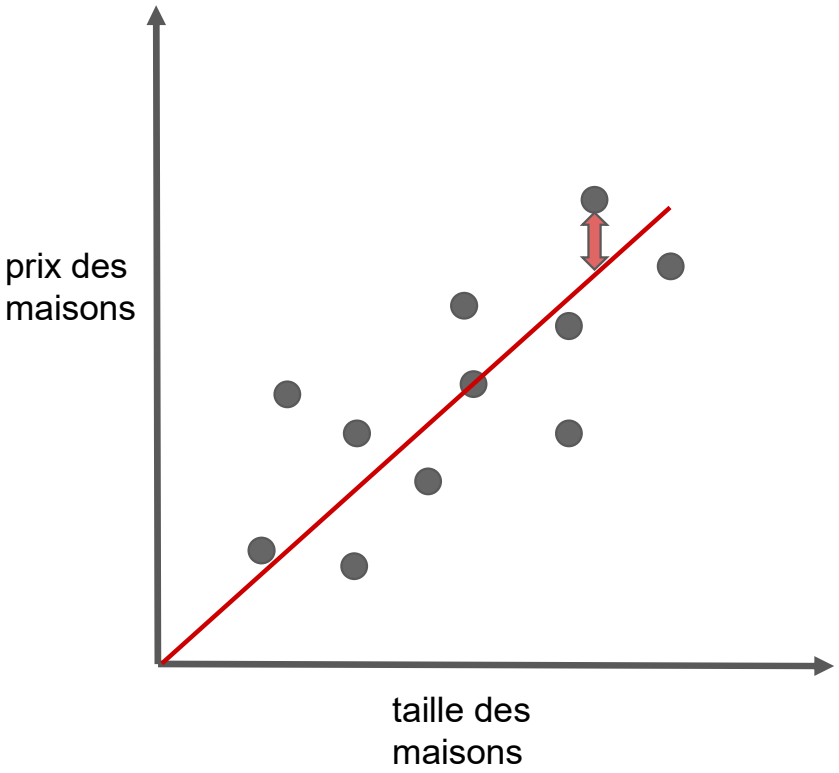
- on choisit la première fois un coefficient aléatoirement
- on regarde la quantité d'erreur
- on modifie notre coefficient

3.3 La régression linéaire

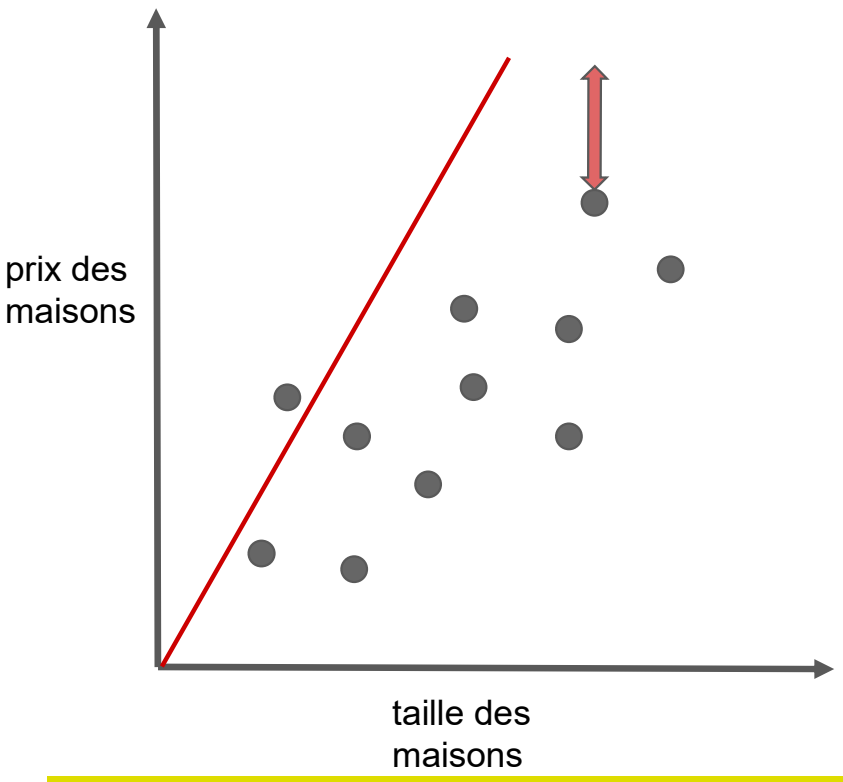


A chaque itération, on calcule nos erreurs et on modifie notre coefficient

3.3 La régression linéaire



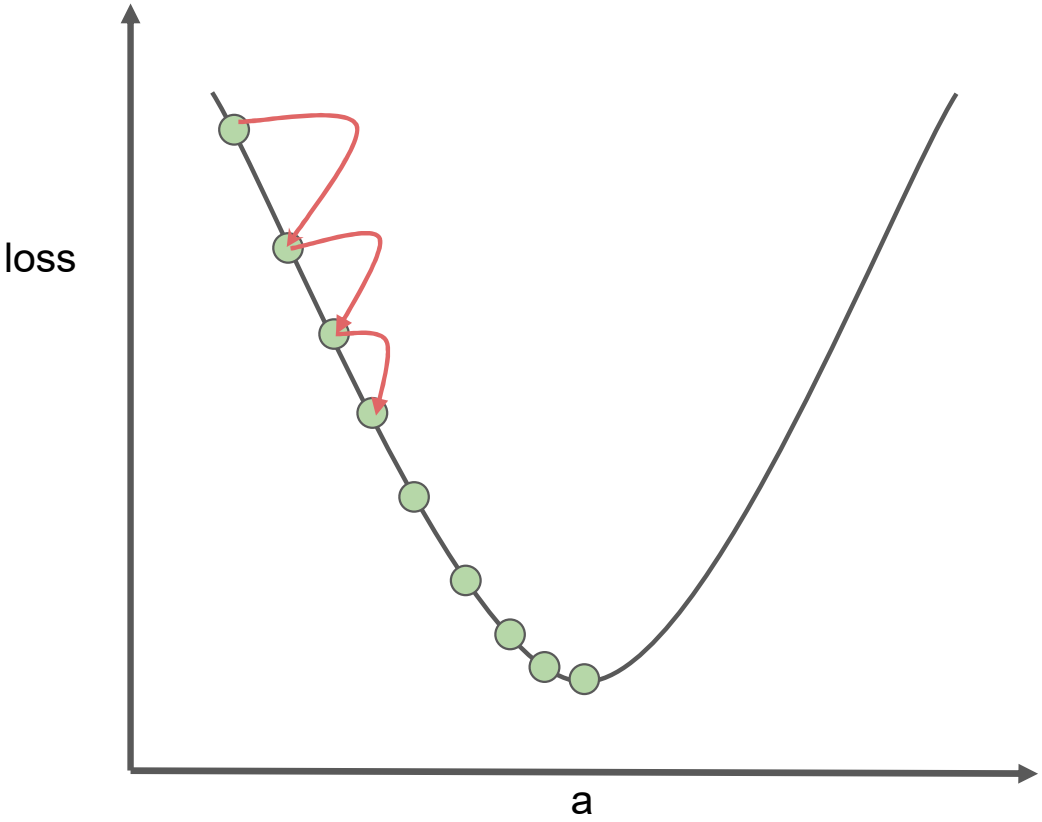
3.3 La régression linéaire

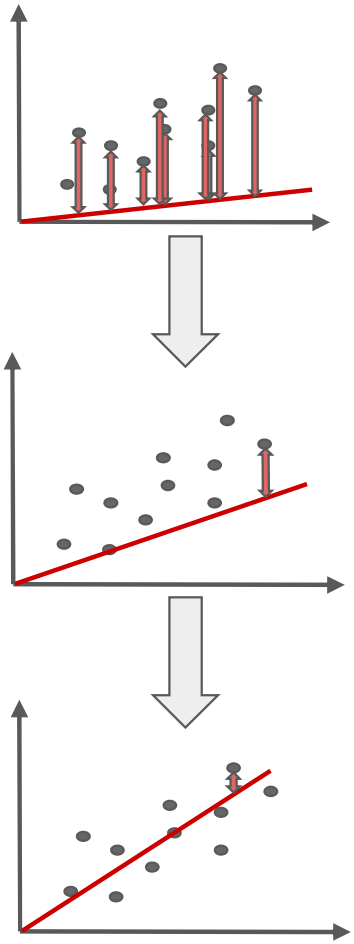


Nos erreurs remontent, nous sommes allés trop loin

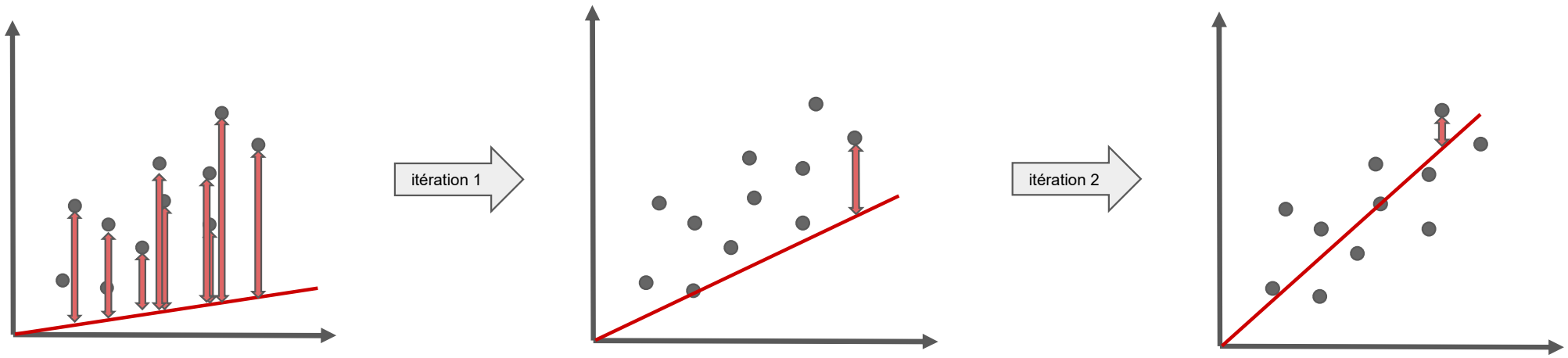
L'entraînement est terminé

3.3 La régression linéaire

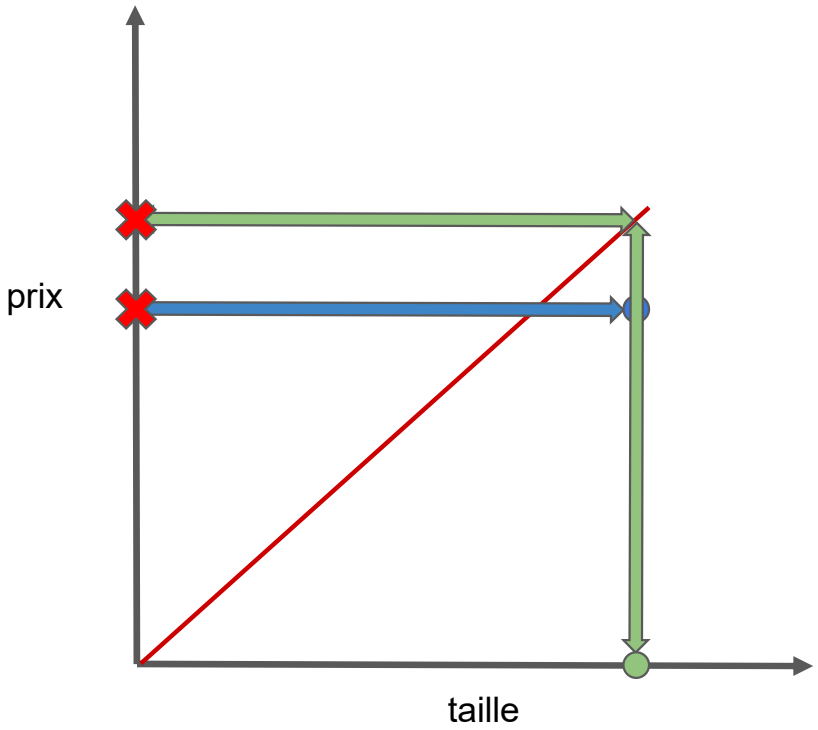




3.3 La régression linéaire



3.3 La régression linéaire



Désormais nous pouvons prédire le prix d'une maison

Avec en bleu la valeur réelle et en vert la prédiction de notre modèle

3.3 La régression linéaire

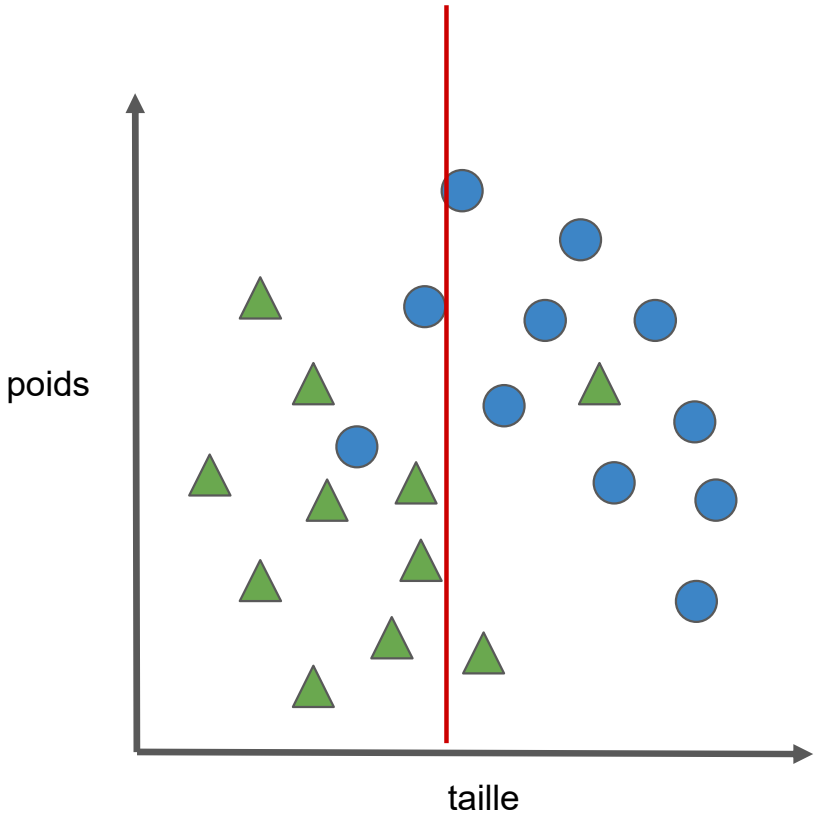
Si on a plus qu'une variable?

$f(x) = a_1x_1+a_2x_2+b$

prix = a_1*taille maison+a_2* taille jardin+b

3.4 La régression logistique

3.4 La régression logistique



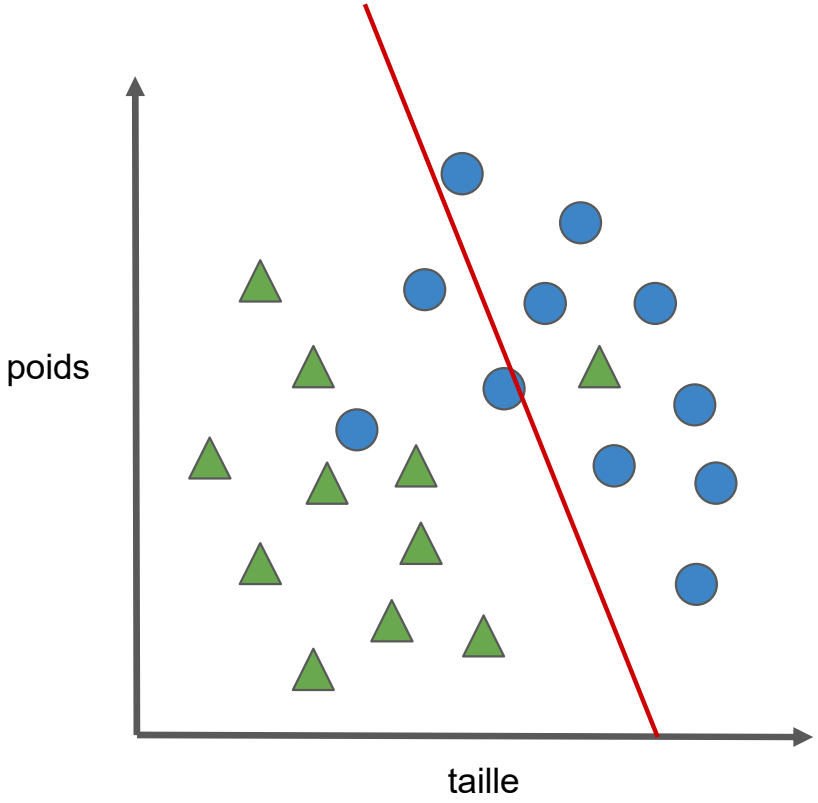
=> Même principe que la régression linéaire

=> **Classification**

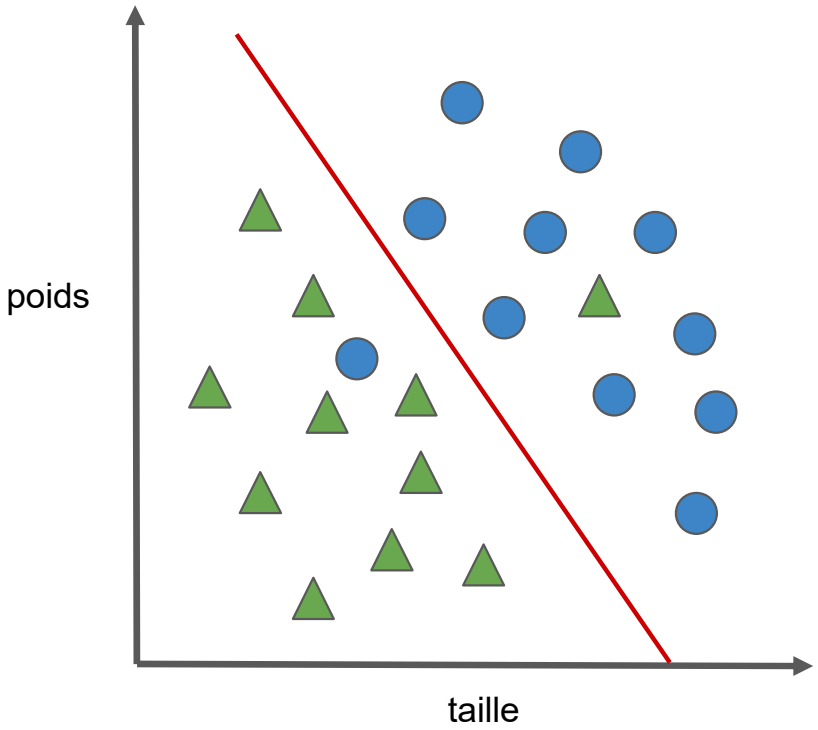
Avec la régression linéaire on **minimise les erreurs**

Avec la régression logistique on cherche à **maximiser la distance** séparant nos classes

3.4 La régression logistique



3.4 La régression logistique



3.5 Les arbres de décision

3.5 Les arbres de décision

=> Pour la régression ou la classification

=> Non linéaire

=> Très simple à interpréter

3.5 Les arbres de décision

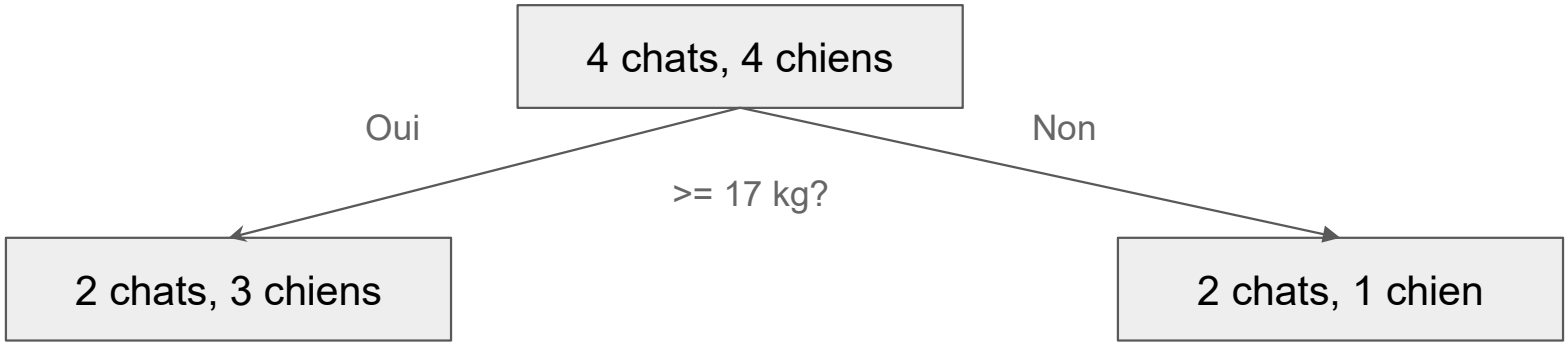
poids(kg)	target(classe)
13	chat
17	chat
26	chien
32	chien
18	chien
8	chat
15	chien
20	chat

Nous avons deux classes possibles et une seule feature pour les discriminer

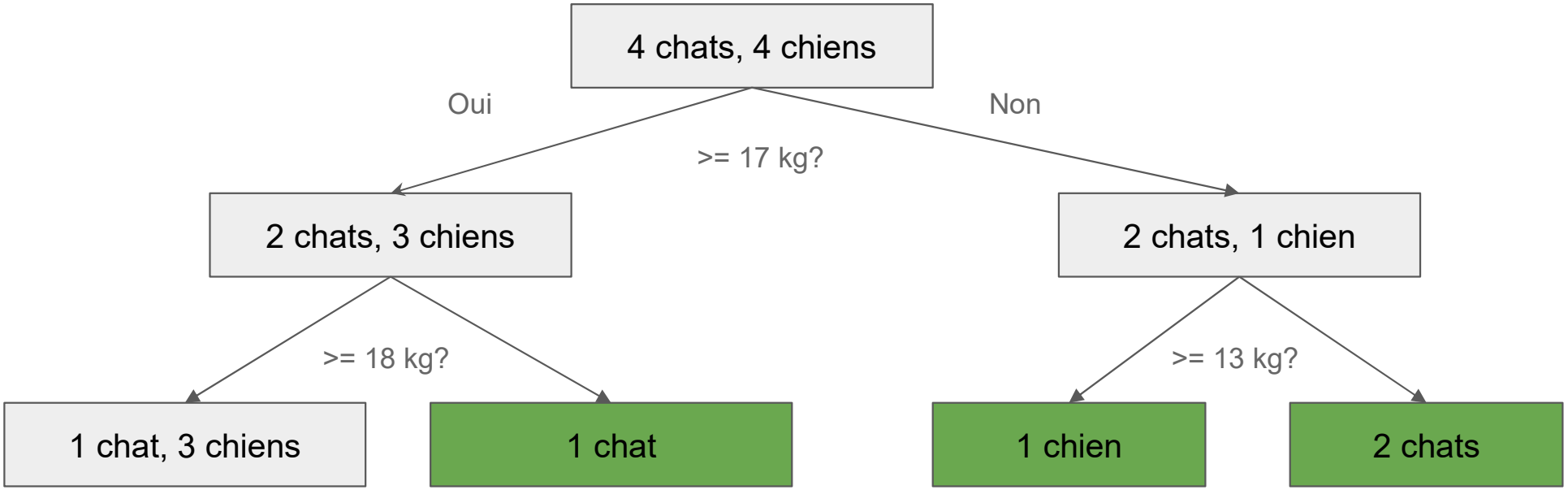
=> Les arbres de décision se construisent en se posant **une succession de questions**

Nous prenons aléatoirement une valeur pour séparer nos données en deux

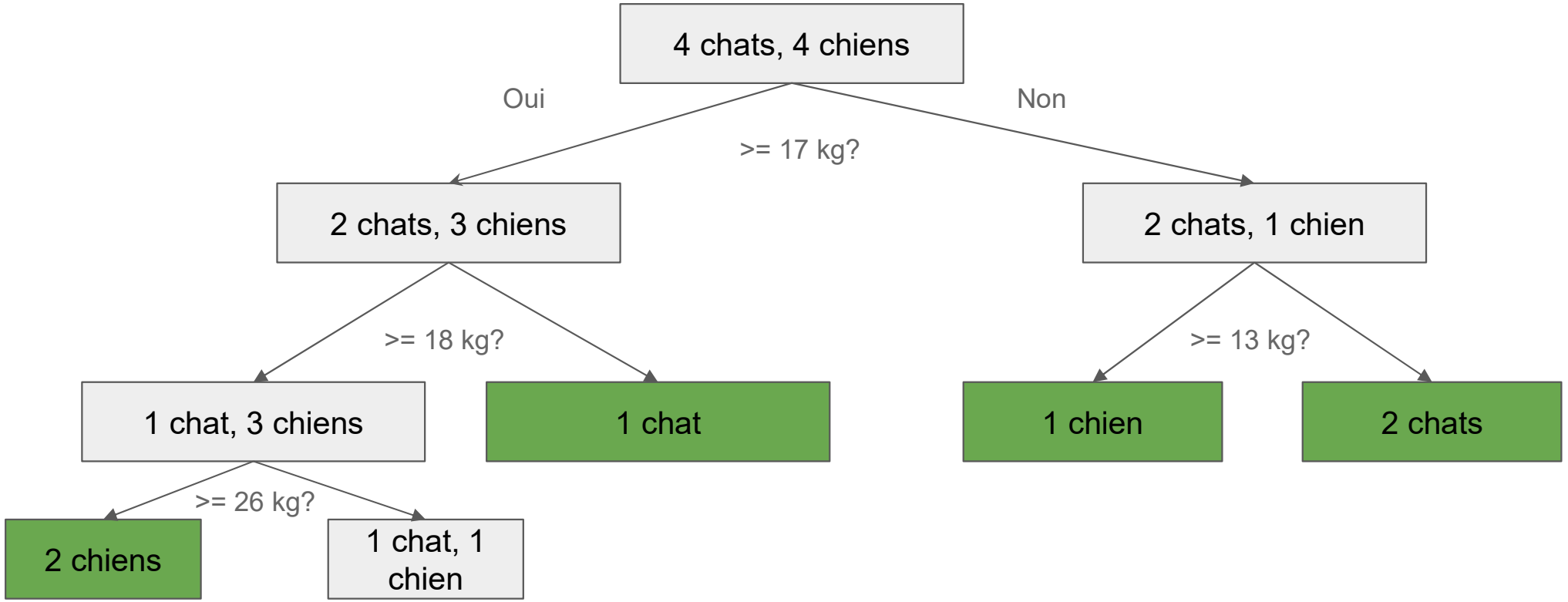
3.5 Les arbres de décision



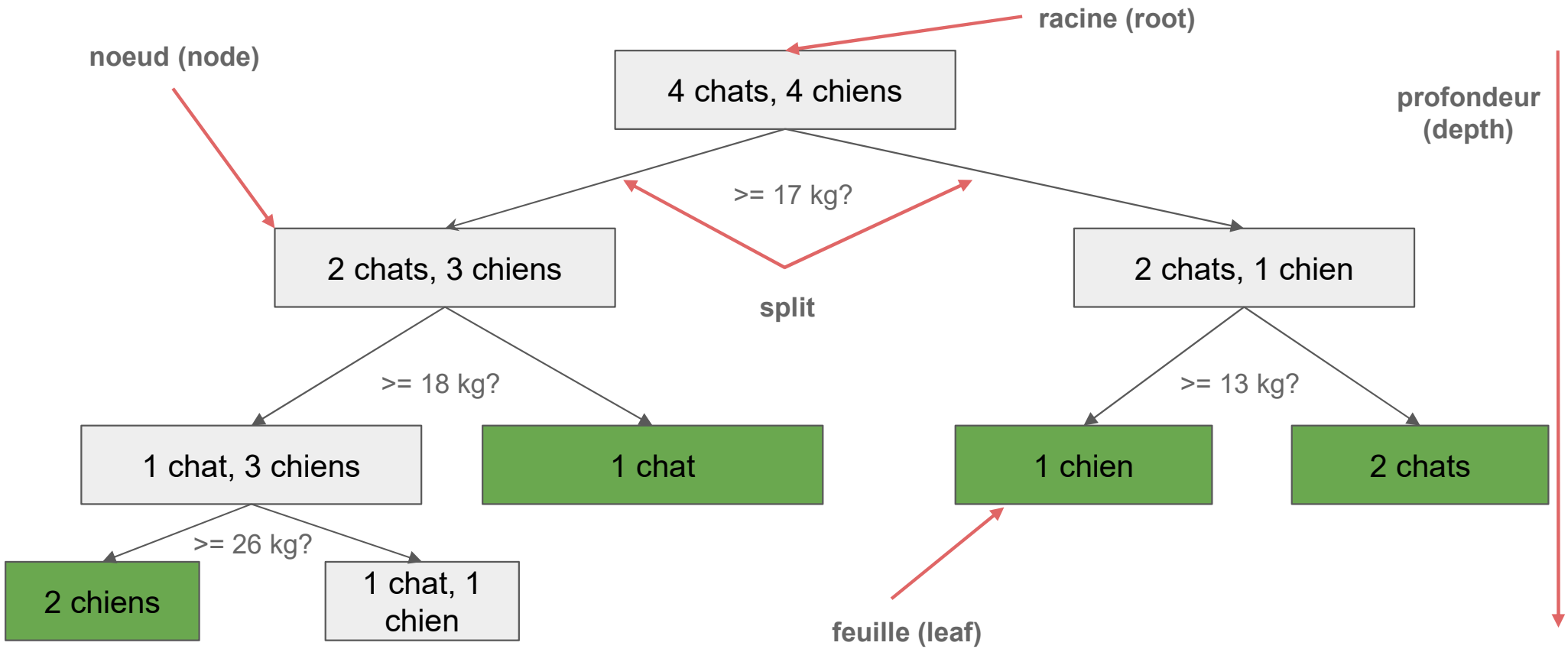
3.5 Les arbres de décision



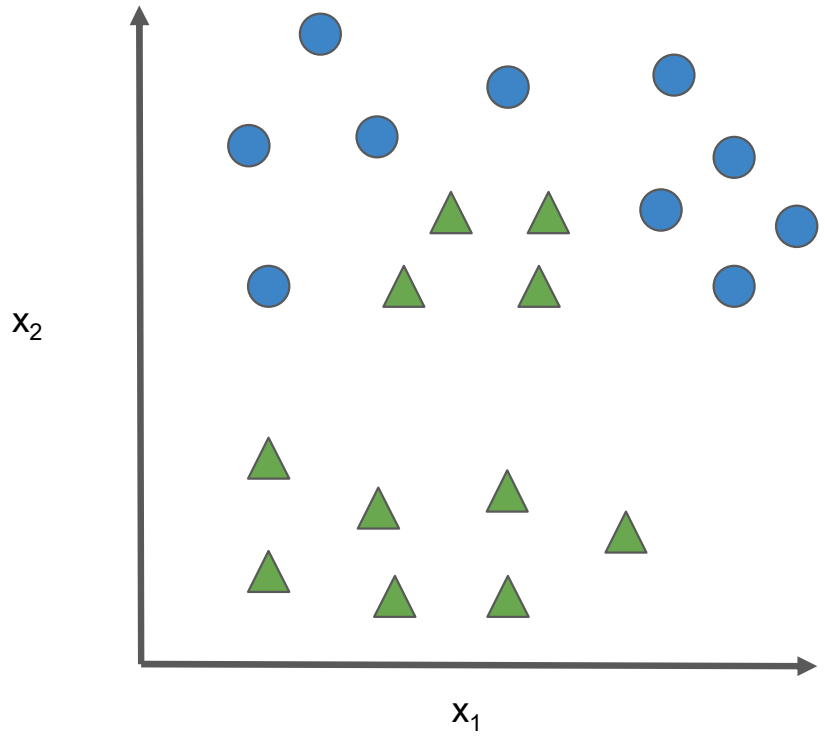
3.5 Les arbres de décision



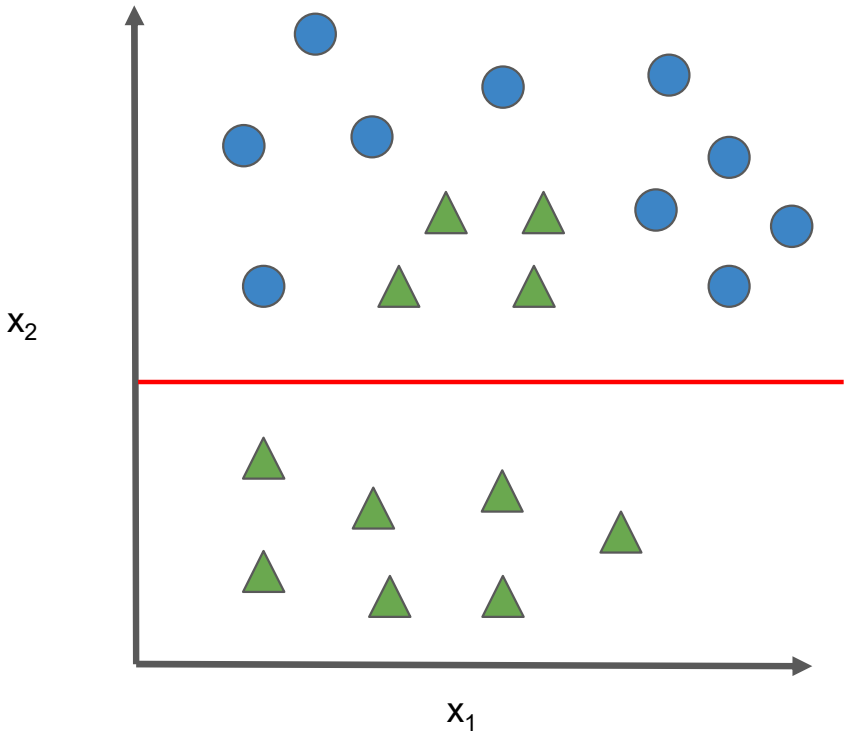
3.5 Les arbres de décision



3.5 Les arbres de décision

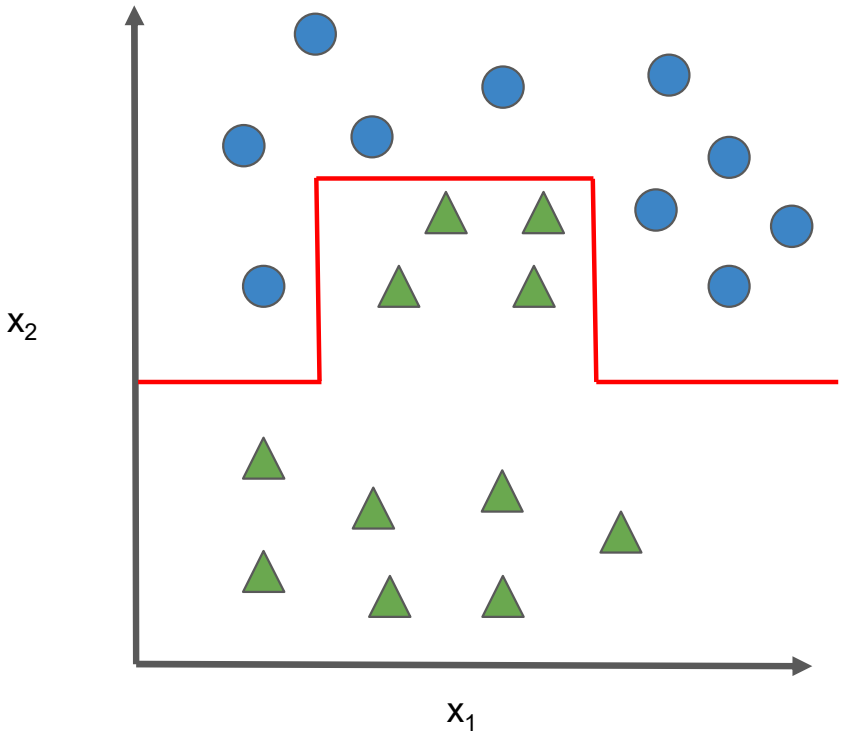


3.5 Les arbres de décision



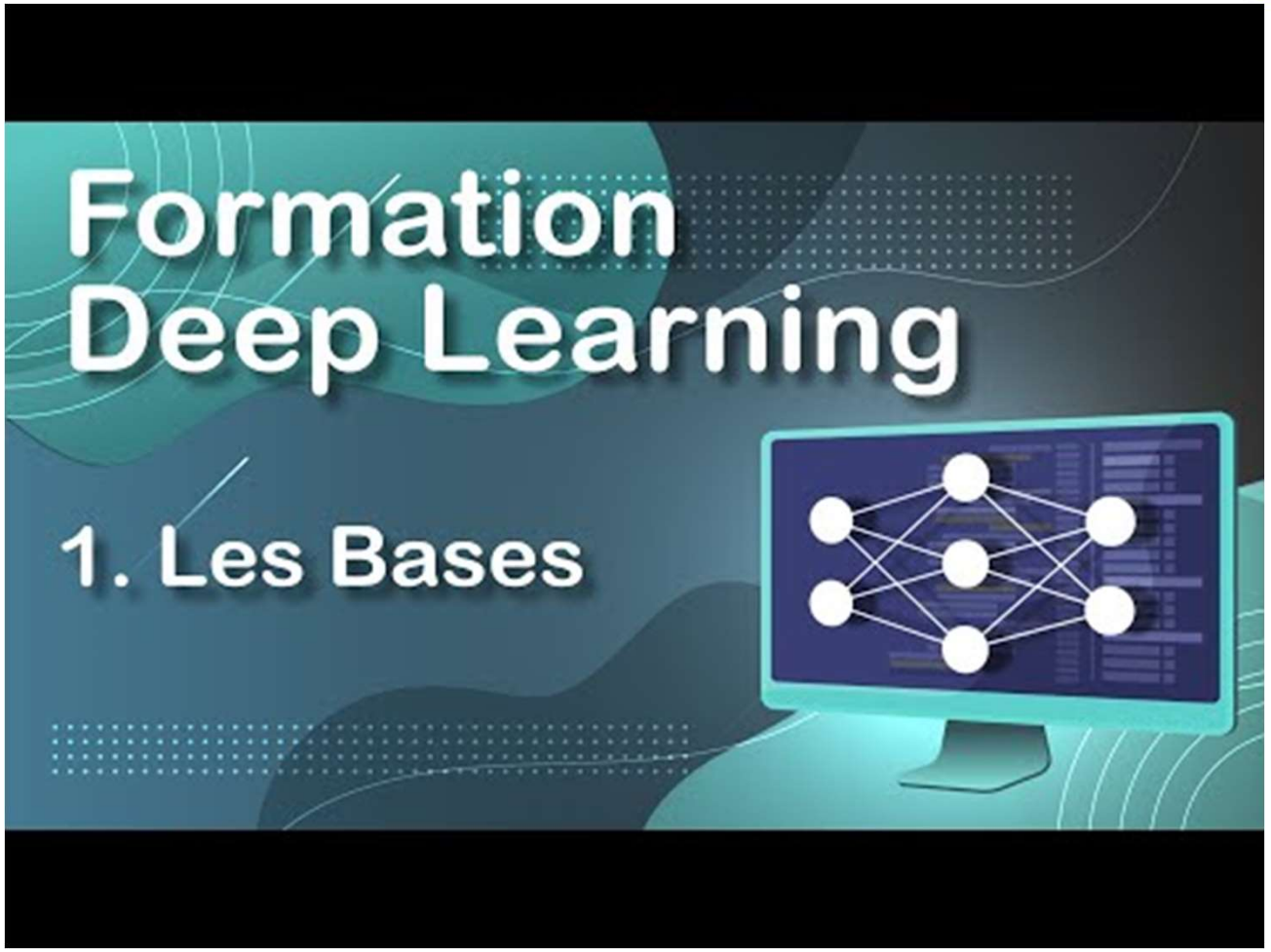
Régression logistique

3.5 Les arbres de décision



Arbre de décision

3.6 Machine learning et deep learning ?


Formation
Deep Learning
1. Les Bases

3.6 ML et DL

Tentons de comprendre l'intuition qui se cache derrière certaines architectures de deep learning:

- convolution Neural Network (**CNN**)
- recurent Neural Network (**RNN**)
- long Short Term Memory (**LSTM**)
- generative Adversarial Network (**GAN**)
- etc.

3.6 ML et DL

Les réseaux de neurones à convolution sont principalement utilisés pour traiter des images.

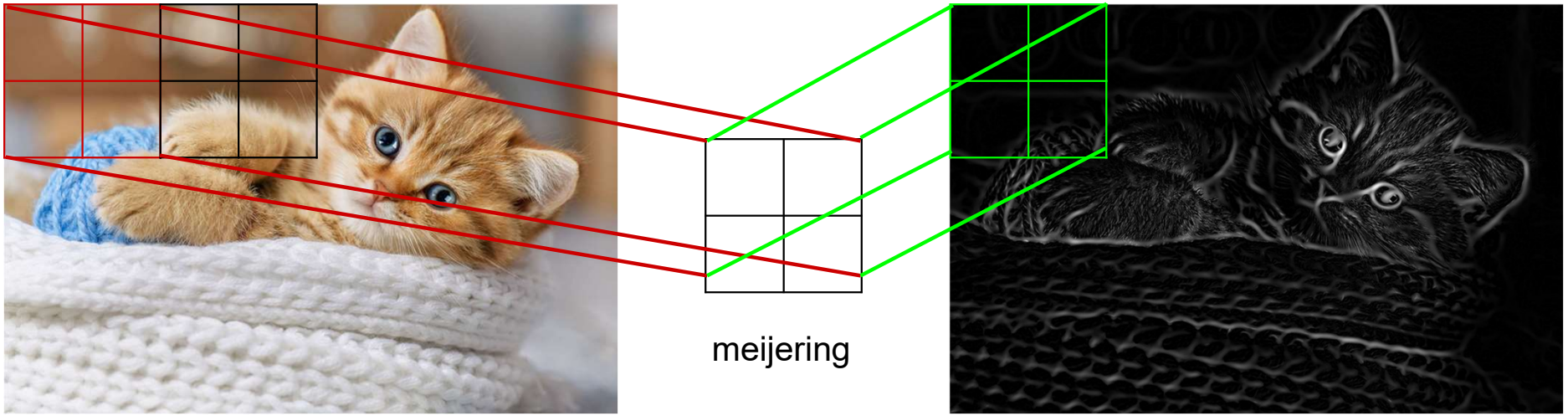
Il peut s’agir de:

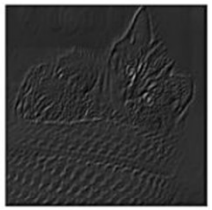
- classification
- segmentation
- débruiteur
- etc.

3.6 ML et DL

Une convolution désigne le fait de balayer une image avec un filtre

Ce filtre est petit réseau de neurones qui cherche à “résumer” l’image

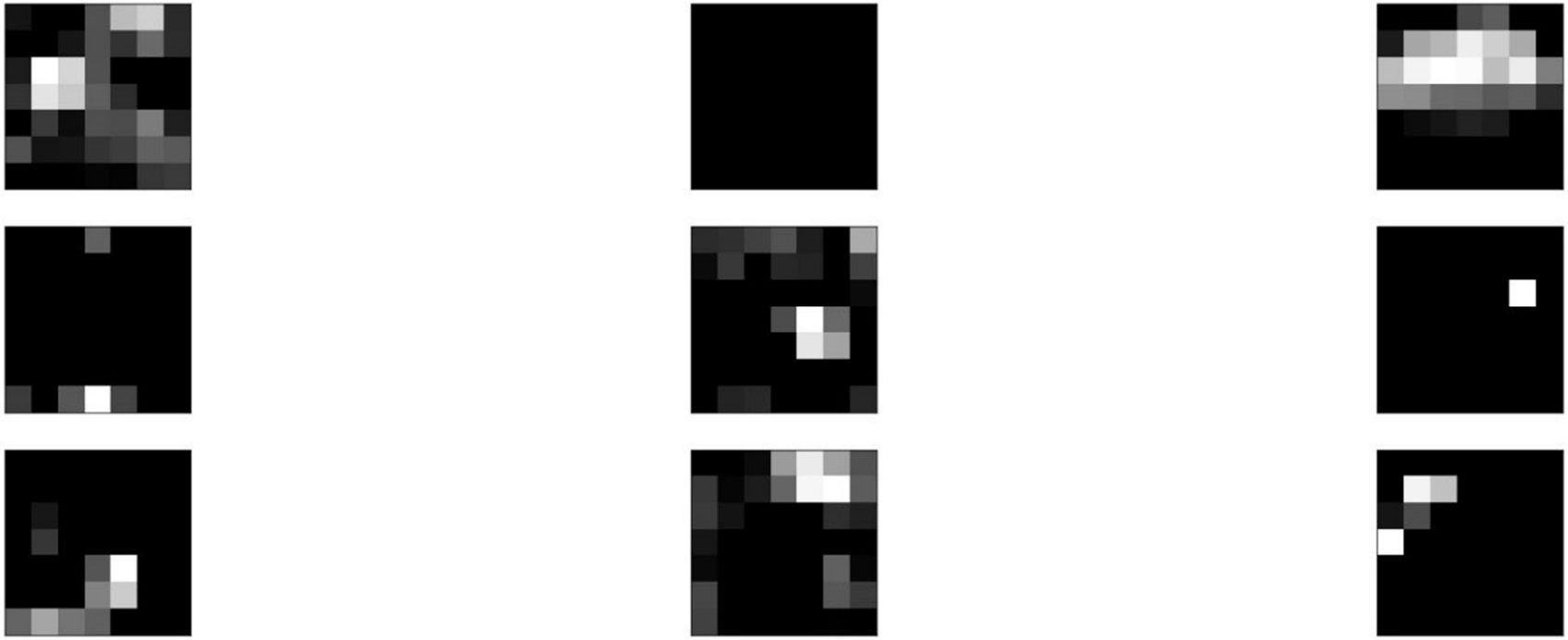












3.6 ML et DL

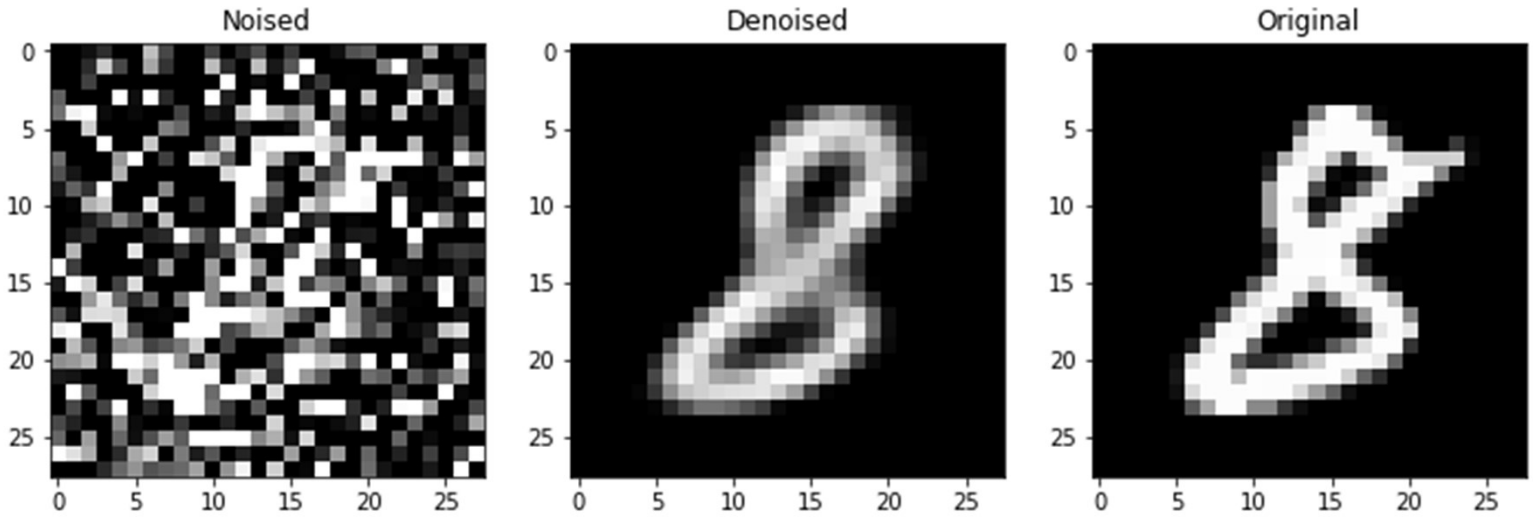
Les auto encodeurs (**AE**) correspondent à une architecture particulière de réseau de neurones qui possède deux entités:

- L'encodeur qui a pour vocation de projeter l'information dans un nouvel espace réduit (espace latent). Il s'agit donc d'une réduction de dimensions
- Le décodeur reconstruit l'information sur base de l'espace latent

=> Il s'agit d'un apprentissage non supervisé

3.6 ML et DL

8



3.6 ML et DL

Les **RNN** sont des réseaux capables de gérer des séquences:

- Des lettres, des mots
- Des images
- Du son, des paroles
- …

3.6 ML et DL

Les **RNN** sont des réseaux capables de gérer des séquences:

- Des lettres, des mots
- Des images
- Du son, des paroles
- …

3.6 ML et DL

Les **RNN** sont des réseaux capables de gérer des séquences:

- Des lettres, des mots
- Des images
- Du son, des paroles
- …

3.6 ML et DL

Les **RNN** sont des réseaux capables de gérer des séquences:

- Des lettres, des mots
- Des images
- Du son, des paroles
- …

3.6 ML et DL

=> Il faut avoir une mémoire pour prédire le déplacement d'un objet

Où les utilise-t-on ?

- Reconnaissance vocale
- voiture autonome
- génération de texte
- génération de musique

LSTM, BI-LSTM sont des **RNN** améliorés

3.6 ML et DL

Les **GAN** sont des architectures récentes (2014)

L'idée est simple, on met deux réseaux en compétition:

- Un générateur
- Un discriminateur

L'objectif du générateur est de produire de l'information que le discriminateur ne pourra pas identifier comme fausse

3.6 ML et DL

Où les utilise-t-on ?

- Deepfake:
  - Vidéo 1
  - Vidéo 2
- Faux visages
- Génération de collections de mode
- Modélisation 3D (Architecture, chimie, pharmacie)
- ...

3.6 ML et DL

L'**apprentissage par renforcement (RL)** est un cas particulier puisque le réseau de neurone doit se débrouiller:

- sans donnée
- sans règle

Les données sont inhérentes à l'environnement

On ne fournit qu'une seule chose: une **récompense**

3.6 ML et DL

Difficile à mettre en oeuvre dans la réalité:

- simulation
- réalité

3.7 Les outils

3.7 Les outils

Quels outils utilise-t-on pour entraîner une IA ?

1. nous l'avons déjà vu: des bases de données
2. un langage de programmation
3. des librairies

3.7 Les outils

Un langage de programmation permet de donner une série d'instructions que l'ordinateur est capable d'interpréter, d'exécuter et éventuellement de renvoyer un résultat

Il en existe beaucoup et chacun possède ses spécificités

```python
def get_cnn(initializer='glorot_uniform') -> tf.keras.Model:
    """
    Build Simple CNN architecture for siamese Network.
    We Tried a lot of different architectures with Optuna for HP optimization but it seems really
    hard to find other viable architectures.

    2 Convolution layers and one Dense

    Args:
        initializer (str): kernel initializer Used for layers. Defaults to glorot_uniform

    Returns:
        tf.keras.Model: CNN model
    """
    inputs = Input(IMG_SHAPE)

    x = Conv2D(64, 4, activation='relu', kernel_initializer=initializer)(inputs)
    x = MaxPool2D()(x)

    x = Conv2D(64, 4, activation='relu', kernel_initializer=initializer)(x)
    x = MaxPool2D()(x)

    x = Flatten()(x)
    # x = GlobalAvgPool2D()(x)
    outputs = Dense(128, activation='relu', kernel_initializer=initializer)(x)

    cnn = keras.Model(inputs, outputs)
    return cnn
```

3.7 Les outils

Dans le domaine de l'intelligence artificielle, un langage se démarque des autres, Python. Pourquoi ?

- Relativement simple à apprendre
- Beaucoup de librairies bien fournies et gratuites
- Grosse communauté active
- Multitâche

3.7 Les outils

Cependant, il existe d'autres langages qui peuvent également être utilisés (selon les cas, selon les entreprises, selon la formation suivie, etc.)

- R, langage avec des librairies de statistiques très fournies
- C ++, langage plus bas niveau mais plus performant que Python. Très utile pour de l'embarqué
- Julia, langage assez jeune qui a pour ambition de réunir la facilité d'utilisation que l'on retrouve en Python mais avec les performances du C++

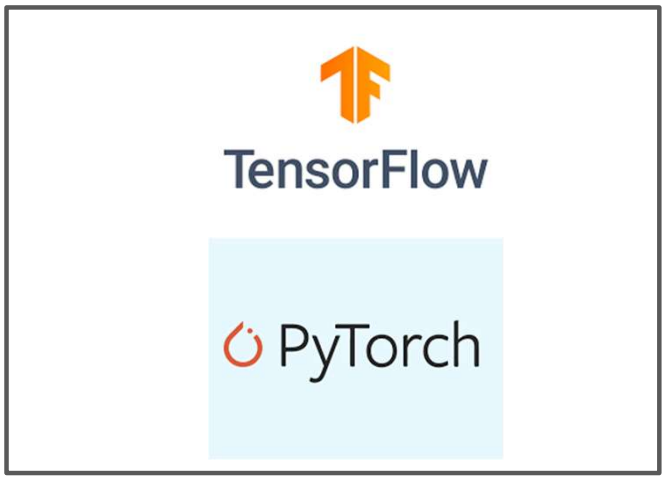


3.7 Les outils

Il faut voir les librairies comme des boîtes à outils. Il s'agit de code déjà implémenté que vous pouvez réutiliser. Par exemple il n'est pas nécessaire de recoder un arbre de décision ou une régression logistique

machine learning

deep learning

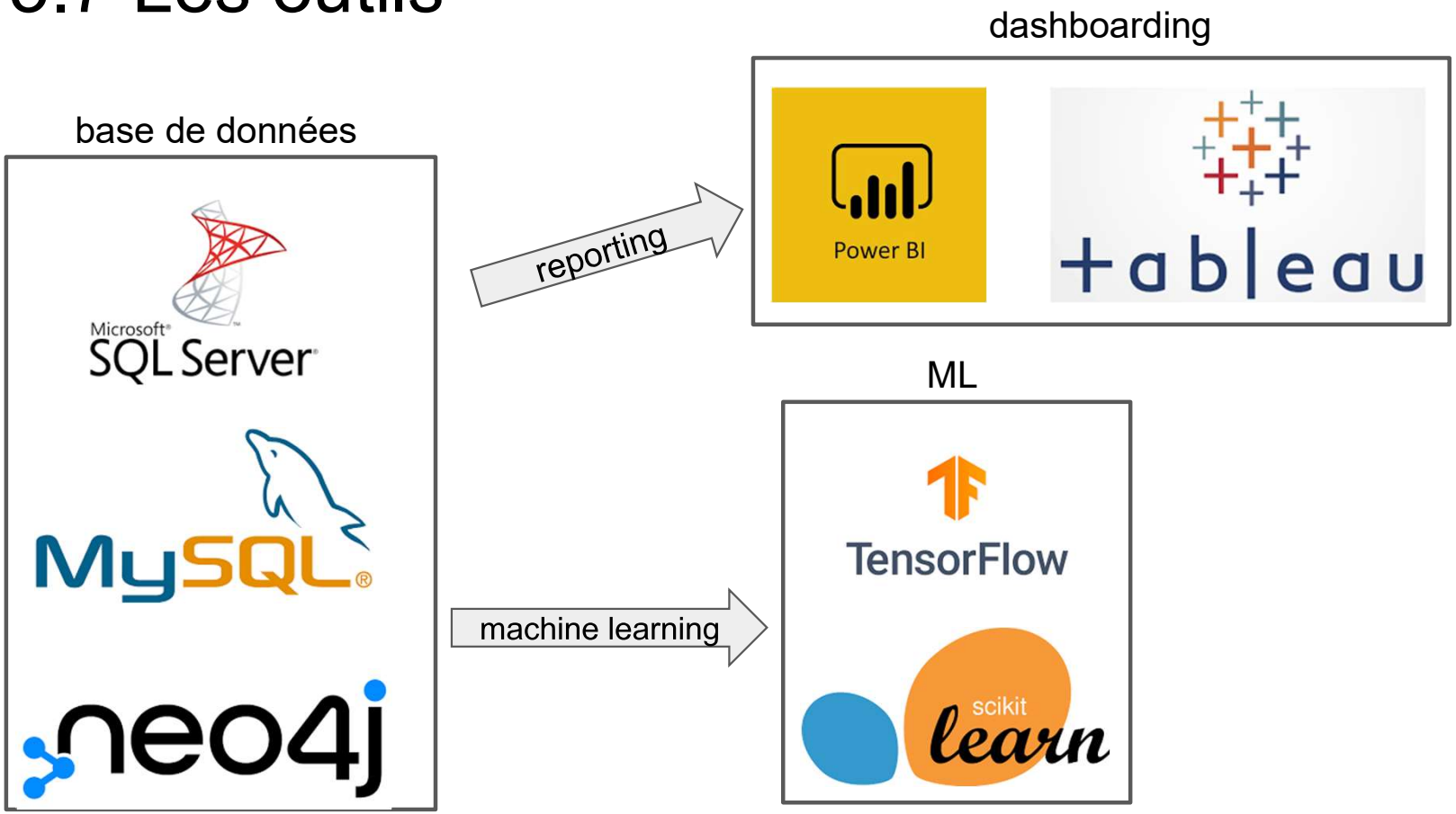


3.7 Les outils

Comment choisir correctement le bon outil avec cette multitude de langages, de librairies, d'algorithmes, etc. Plusieurs éléments à prendre en compte:

1. La quantité de données
2. La nature du problème (classification, régression, etc.)
3. La complexité du problème (2 classes vs 1000 classes)
4. La puissance de calcul disponible (serveurs clouds ou bien smartphone)
5. etc.

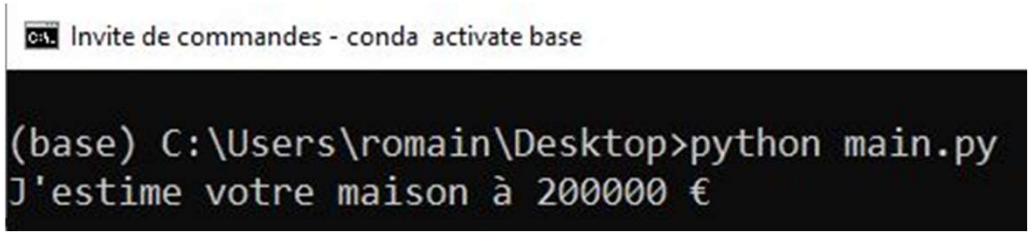
3.7 Les outils



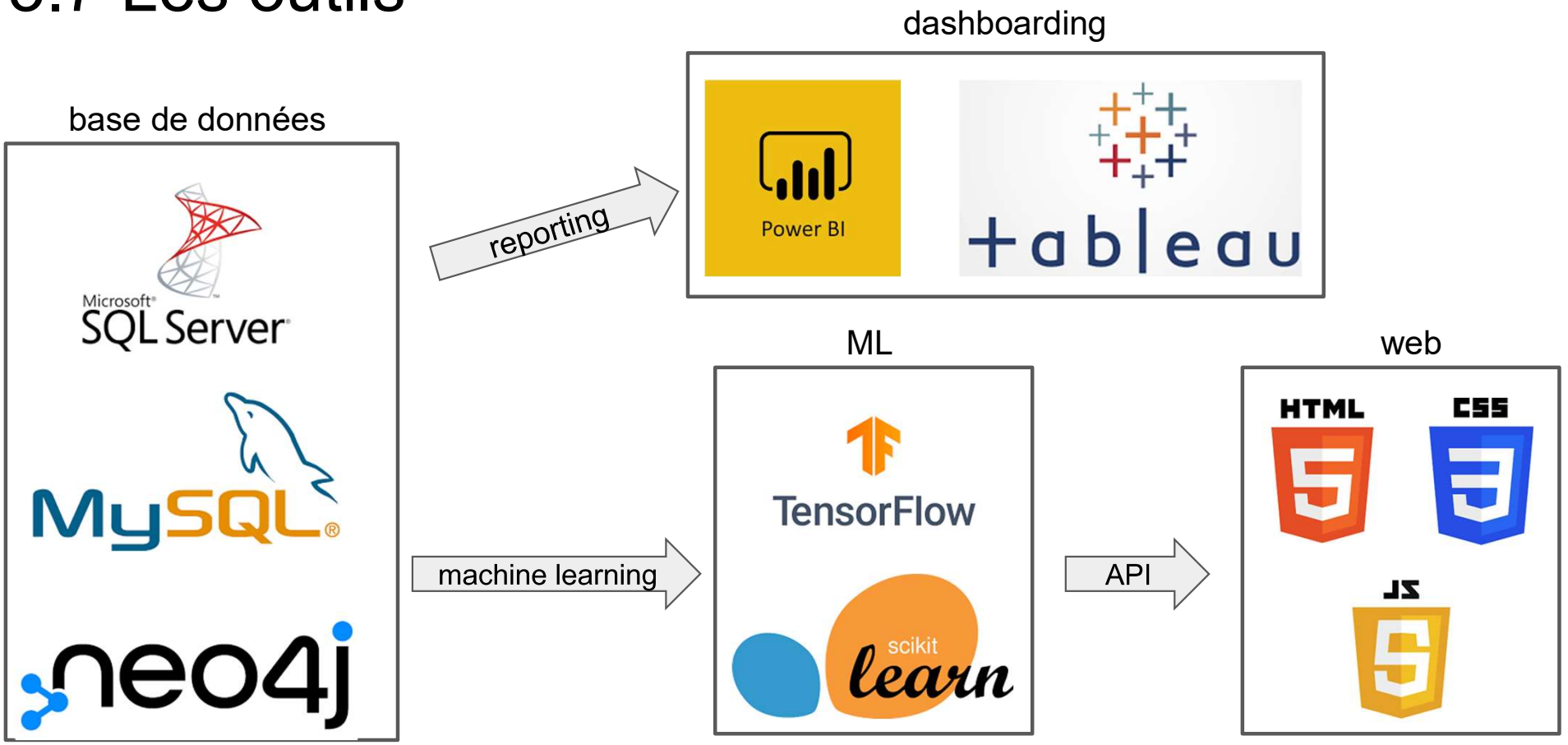
3.7 Les outils

Il manque une pierre à l'édifice. Comment intégrer notre IA dans un service, un site internet, une application ? On ne peut pas demander à quelqu'un de prendre un bout de code Python, d'entrer ses données et d'exécuter lui-même le code.

Pour faire ceci nous avons besoin des technologies du web comme HTML, CSS, Javascript, Angular, etc.

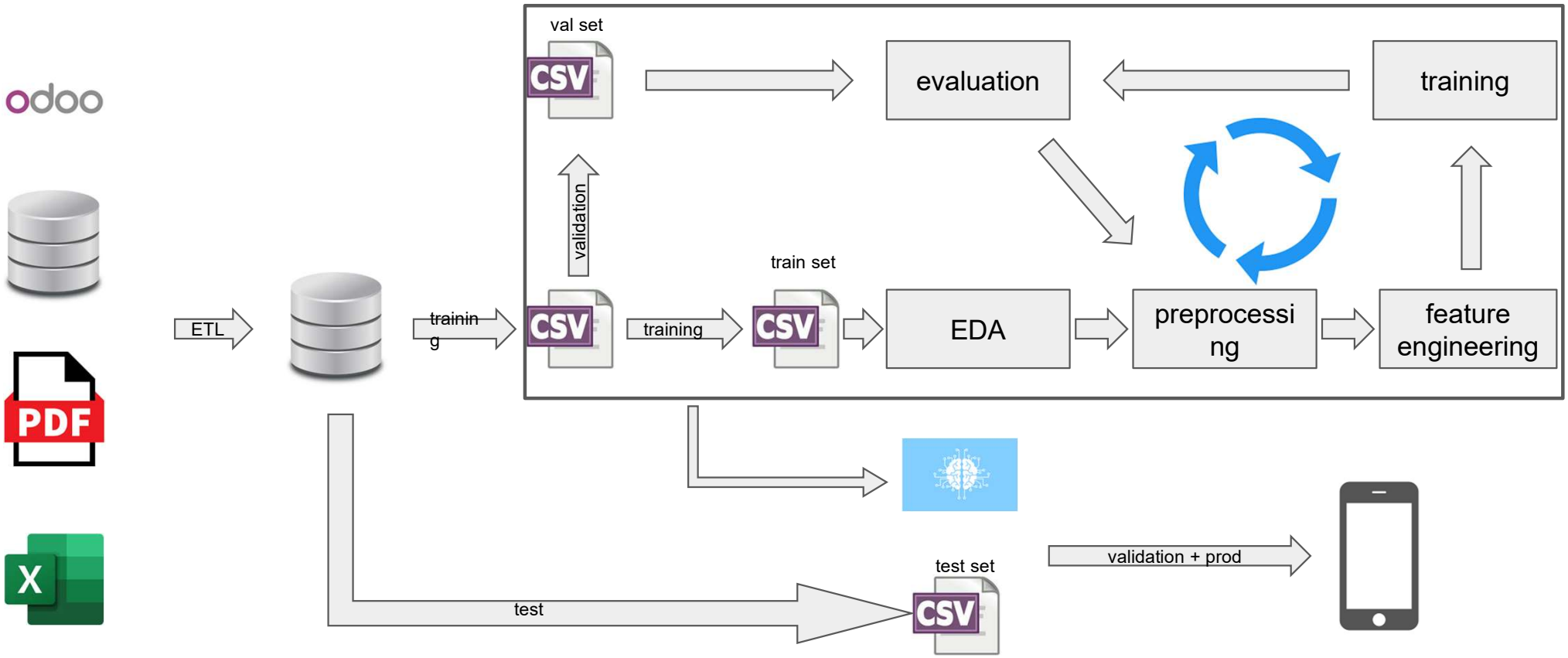


3.7 Les outils



3.8 Déroulement d’un projet

3.8 Déroulement d'un projet



3.8 Déroulement d'un projet

1. Identification de la problématique

- Quel est le problème ?
- Peut-on le régler avec une solution machine learning ?
- Est-ce la solution la plus adéquate ?
- A-t-on les moyens de le faire ?

3.8 Déroulement d'un projet

2. Audit sur les données

- A-t-on les données nécessaires ? En suffisance ? Labellisées ?
  - Si non, peut-on les récolter, comment, combien ?
- Quels sont leur nature ?
- Où sont-elles ?
- Peut-on tout utiliser ?

3.8 Déroulement d'un projet

3. ETL

- **E**xtract: extraire les données nécessaires dans les différentes sources disponibles
- **T**ransform: cette partie consiste à nettoyer vos données et à les homogénéiser
- **L**oad: charger toutes les données dans un datawarehouse par exemple

=> Il existe différents outils dédiés à cette tâche: Python, Talend, Pentaho, SSIS, etc.

3.8 Déroulement d'un projet

4. Constitution des datasets

Après la constitution de votre dataset il est **IMPERATIF** de le subdiviser en deux sous datasets:

- trainset
- testset

Il faut à tout prix garder un jeu de données de côté que votre modèle n'aura jamais vu

3.8 Déroulement d'un projet

5. **E**xploratory **D**ata **A**nalysis (**EDA**)

Il s'agit d'analyser vos données pour diverses raisons:

- comprendre de quoi elles parlent précisément
- identifier des variables potentiellement intéressantes
- identifier des modifications à effectuer pour l'étape suivante
- identifier quelques algorithmes potentiels
- etc.

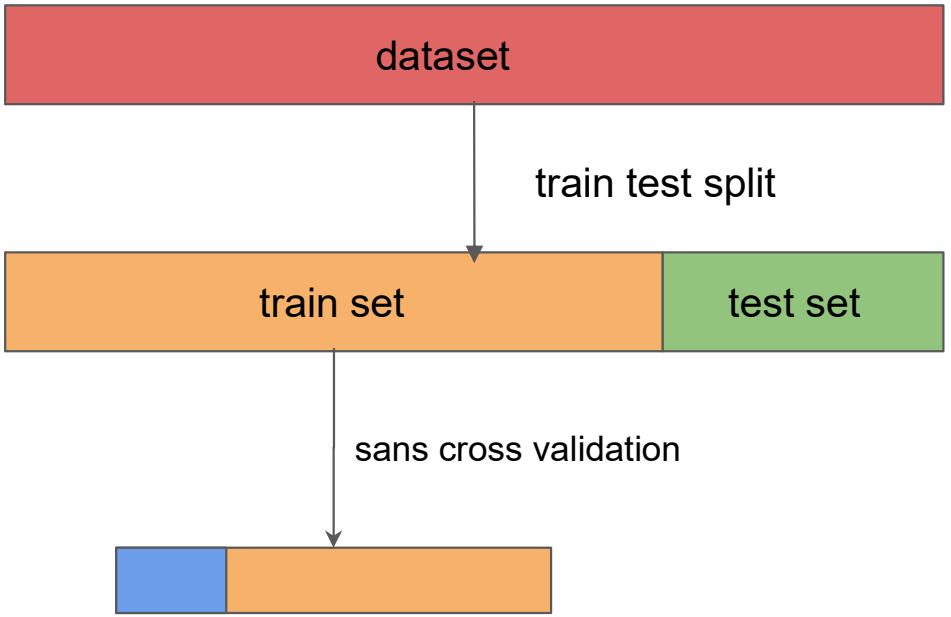
3.8 Déroulement d'un projet

6. Constitution d'un validation set

A ce stade, vous avez le choix entre plus algorithmes et il est impossible de savoir à l'avance lequel se débrouillera le mieux.
Une seule solution, mettre à l'épreuve une poignée de candidats.

C'est à cela que sert votre validation set

3.8 Déroulement d'un projet



3.8 Déroulement d'un projet

7. Preprocessing

Avec l'étape suivante, il s'agit du cœur du travail. Des données de mauvaise qualité ne peuvent que donner de mauvais résultats. (Garbage in, Garbage out)

- Que faire des variables textuelles, des catégories ?
- Que faire si je rencontre des valeurs manquantes ?
- Il y a-t-il des erreurs d'encodage ?
- etc.

3.8 Déroulement d'un projet

8. Features engineering

Nous aurions pu fusionner cette étape avec la précédente mais cette étape consiste entre autres à:

- sélectionner les variables
- discrétiser certaines variables
- extraire des variables
- etc.

3.8 Déroulement d'un projet

9. Sélection d'un modèle et mise en production

Quelques points d'attention lors du déploiement:

- Les calculs sont-ils effectués sur des serveurs cloud ?
  - Que se passe-t-il sans connexion internet ?
- Les calculs sont-ils effectués sur les appareils des utilisateurs ?
  - faut-il refactorer le code ?
- Une solution hybride ?

3.8 Déroulement d'un projet

10. Monitoring

Une fois en production, il faut garder un oeil sur votre modèle et souvent le réentraîner pour deux raisons principalement:

- Les données varient (data drift)
- De nouvelles classes apparaissent

4. L’éthique

Peut-on tout faire ?

CONSULTING BSTORM

4. L’éthique

L’IA pose beaucoup de questions, l’idée dans ce chapitre n’est pas d’y répondre mais d’initier une prise de conscience au travers de quelques exemples. Commençons par un jeu:

- moral machine

4. L’éthique

Essayez de répondre à ces questions suivantes:

- Votre position tient-elle en toutes circonstances ?
  - Identifier des cas limites
  - Proposez ces cas limites à vos collègues
- En dehors des voitures autonomes, pensez à des évènements, des applications, des cas qui pourraient également soulever ce genre de questions ?


UTILITARISTE JUSQU'AU BOUT ?

4. L’éthique

En septembre 2021 sur Facebook, après avoir regardé des vidéos avec des personnes noires, l’algorithme de Facebook demande s’ils veulent voir d’autres vidéos de singes et cet évènement s’inscrit dans une continuité. Twitter, Google ont également déjà été confrontés à cette problématique.

La question qui se pose est comment réussir à ne pas reproduire les discriminations présentes dans la société ?

4. L’éthique

Les algorithmes de recommandation vous proposent le contenu qui est censé le plus vous plaire. Quand il s’agit de Netflix, Spotify, cela peut sembler anecdotique.

Par contre lorsqu’il s’agit de Politique, notamment lors d’élections, Twitter a par exemple tendance à favoriser certaines idées, à enfermer les utilisateurs dans une bulle.

Quand Facebook bannit des comptes, supprime des postes, etc. Est-ce toujours souhaitable ?

4. L’éthique

Ces algorithmes se basent sur des données et y sont donc sensibles.

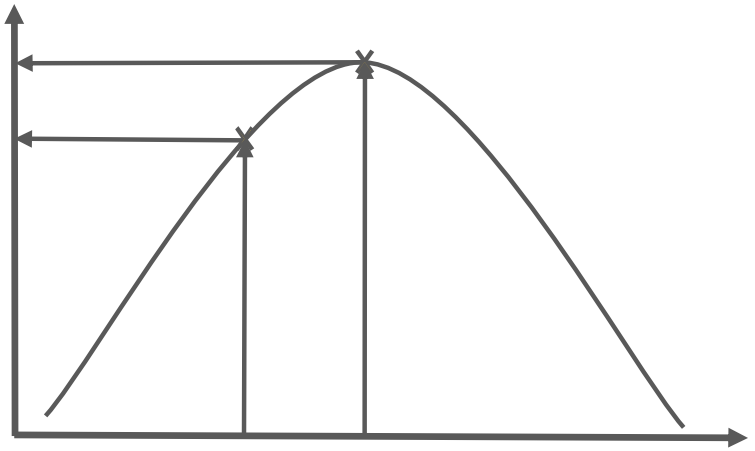
Ce qui veut dire qu’il est totalement possible d’altérer leur comportement en leur fournissant de mauvaises données (data poisoning).

C’est exactement ce que pourrait faire certains partis politiques en ayant recours à des robots.

4. L’éthique

Ces algorithmes de recommandation remettent en cause la question de l’autonomie individuelle.

Puisque Netflix vous propose le meilleur choix possible, en réalité vous n’avez plus de choix. Ces algorithmes tendent à vous faire choisir le choix optimal ce qui implicitement revient à vous retirer l’action de choisir.




EXPLICATION
L'IA va-t-elle trop loin ?

4. L’éthique

Où en est-on actuellement ?

- Les réglements se mettent en place
- La recherche se développe doucement dans les universités
- Des librairies naissantes fournissent des outils pour aider à résoudre ces problèmes

Mais malheureusement, on constate régulièrement des pas en arrière

4. L’éthique

A côté de ceci, certains autres règlements comme le Règlement général sur la protection des données (RGPD/GDPR) encadre la collecte et l’utilisation des données:

- Vous pouvez demander à supprimer vos données
- Les recensements ethniques sont interdits
- Demander ou chercher à connaître l’état de santé (VIH, grossesse, etc.) d’une personne est interdit. Sauf pour des raisons de recherches.
- Les entreprises ne peuvent récolter que les données dont elles ont besoin

4. L’éthique

Le privacy shield est un dispositif permettant aux données des citoyens européens d’être envoyées aux Etats-Unis.

Le 16 juillet 2020 cet accord est invalidé par une décision de la cour de justice de l’Union européenne

5. Profils métiers

Qui fait quoi ?

CONSULTING BSTORM

5. Profils métiers

Pour chaque métier présentez:

1. Son rôle, ses responsabilités, sa place dans un projet, *etc.*
2. Les compétences nécessaires:
    a. outils, donnez des exemples et catégorisez les
    b. langages
    c. soft skills
3. L'expérience nécessaire
4. D'éventuels liens avec d'autres métiers. Comment bien les distinguer, etc.
5. Evolution
6. *etc.*

5. Profils métiers

1. Développeur front
2. Développeur back
3. Développeur fullstack
4. Business Analyst
5. Administrateur de bases de données
6. Data Engineer
7. Data Scientist
8. Machine Learning Engineer
9. Architecte informatique
10. Data Architect
11. Data Protection Officer
12. Chief Technical Officer

5. Profils métiers

Ressources utiles et diverses:

- Les différents sites d'offres d'emplois, agences régionales (e.g. glassdoor, ICTjob, forem, *etc.*)
- clémentine
- ADP, CNIL
- Blog de Dataiku, (e.g. infographie sur les technologies)
- Enquête Github
- Enquête sur les salaires
- etc.

6. Business case

Exercice récapitulatif

CONSULTING BSTORM