Machine Learning

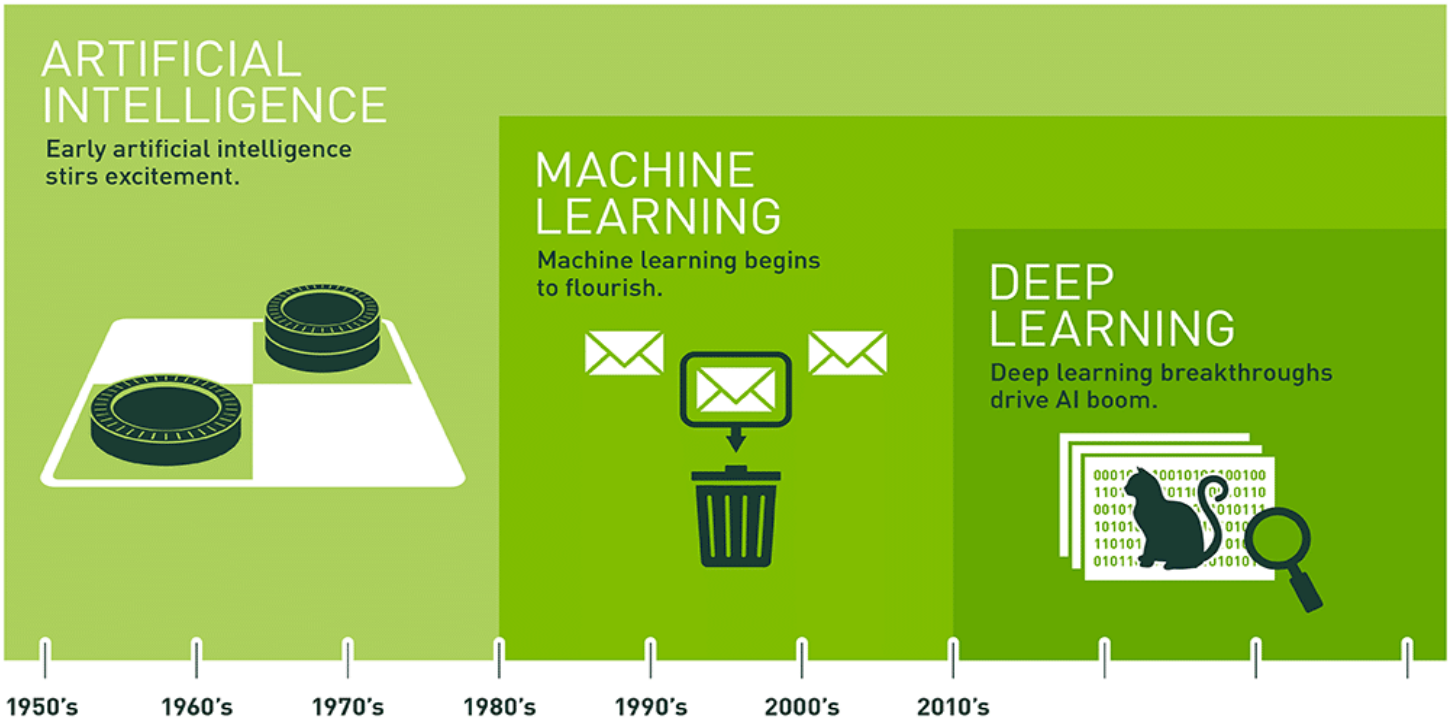
Présenté par : Rodrique Kafando
Doctorant en Science de Données & IA
Email : kafando.rodrique@gmail.com

Décembre 2020

SOMMAIRE

L'apprentissage machine est une branche de l'intelligence artificielle qui permet aux systèmes informatiques d'apprendre et de s'améliorer automatiquement à partir de l'expérience sans être explicitement programmés.

Un peu d'histoire...[1]



Since an early flush of optimism in the 1950s, smaller subsets of artificial intelligence – first machine learning, then deep learning, a subset of machine learning – have created ever larger disruptions.

Figure 1: L'évolution de l'IA dans la recherche

Quelques définitions

- Intelligence artificielle - L'intelligence humaine exposée par des machines.
  - Doter à la machine, la capacité à résoudre les problèmes complexes, en imitant le principe de raisonnement de l'être humain
- Apprentissage automatique (Machine Learning) - Une approche pour atteindre les objectifs de l'intelligence artificielle.
  - Pratique consistant à utiliser des algorithmes pour analyser des données, en tirer des informations, puis faire une détermination ou une prédiction sur un individu donné.
- Apprentissage profond (Deep Learning) - Une technique pour mettre en œuvre l'apprentissage automatique, dans le but d'atteindre des performances très élevées.
  - il s'agit d'une autre approche algorithmique, celle des réseaux de neurones artificiels. Les réseaux de neurones sont inspirés par notre compréhension de la biologie de notre cerveau - toutes ces interconnexions entre les neurones, mais avec des couches discrètes dans son architecture.

Les types de modèles en Machine Learning

Types de modèles les plus utilisés

Il existe principalement trois types de modèles en machine learning. Nouus avons les modèles :

- supervisés, non-supervisés et par renforcement (non abordé dans ce cours).
- Et c'est quoi la différence ?

La principale différence entre ces types est le niveau de disponibilité des données de vérité de base, c'est-à-dire la connaissance préalable de ce que devrait être la sortie du modèle pour une entrée donnée.

Les types de modèles en Machine Learning

Types de modèles les plus utilisés

Il existe principalement trois types de modèles en machine learning. Nouus abordons les modèles supervisés et non-supervisés dans cette partie du cours :

- L'apprentissage supervisé vise à apprendre une fonction qui, à partir d'un échantillon de données et des résultats souhaités, se rapproche d'une fonction qui met en correspondance les entrées et les sorties.
- L'apprentissage non supervisé n'a pas (ou n'a pas besoin) de sorties étiquetées, son but est donc de déduire la structure naturelle présente dans un ensemble de points de données.

Les types de modèles en Machine Learning

Apprentissage supervisé

- L'apprentissage supervisé implique l'entraînement des systèmes informatiques à l'aide de données **explicitement étiquetées**.
- Les données étiquetées ici signifient que l'entrée a été étiquetée avec les étiquettes de sortie correspondantes souhaitées. L'algorithme d'apprentissage (modèle) apprend de ces données étiquetées par un **processus itératif** qui lui permet ensuite d'effectuer des prédictions futures.

Les types de modèles en Machine Learning

Apprentissage supervisé

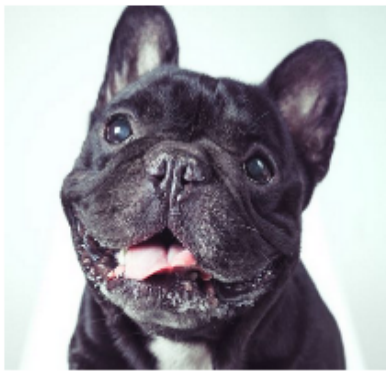
Les problèmes d'apprentissage supervisé peuvent être divisés en 2 sous-classes - la Classification et Régression. La seule différence entre ces 2 sous-classes est le type de sortie ou de cible que l'algorithme vise à prédire.

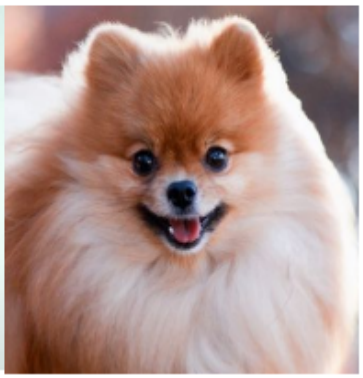
- **Classification** : lorsque la cible à prédire est une variable de type catégoriel
- **Régression** : lorsque la cible à prédire est une variable de type numérique

Les types de modèles en Machine Learning

Apprentissage supervisé

Les problèmes d'apprentissage supervisé peuvent être divisés en 2 sous-classes - la Classification et Régression. La seule différence entre ces 2 sous-classes est le type de sortie ou de cible que l'algorithme vise à prédire.

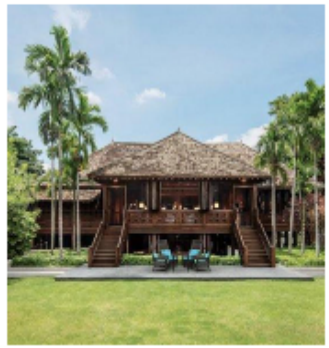




Type de chien

↓

Classification



Prix d'une location

↓

Regréssion

Figure 2:

Les types de modèles en Machine Learning

Apprentissage supervisé

- Problème de classification : dans la classification, l'objectif est d'identifier à quelle catégorie appartient un objet (input).
  - ex : l'image contient-elle un chat ou un chien. Le mail est-il un spam ou non. Un patient est-il séropositif ou non.
  - On distingue deux types de problèmes classification:
    - **Binaire** : lorsque la cible catégorielle ne peut prendre que 2 valeurs (2 classes)
    - **Multiclasse** : lorsque la cible peut peux prendre plus de 2 valeurs. Par exemple un problème de classification de types de médicaments, de types de fleurs, etc.

Les types de modèles en Machine Learning

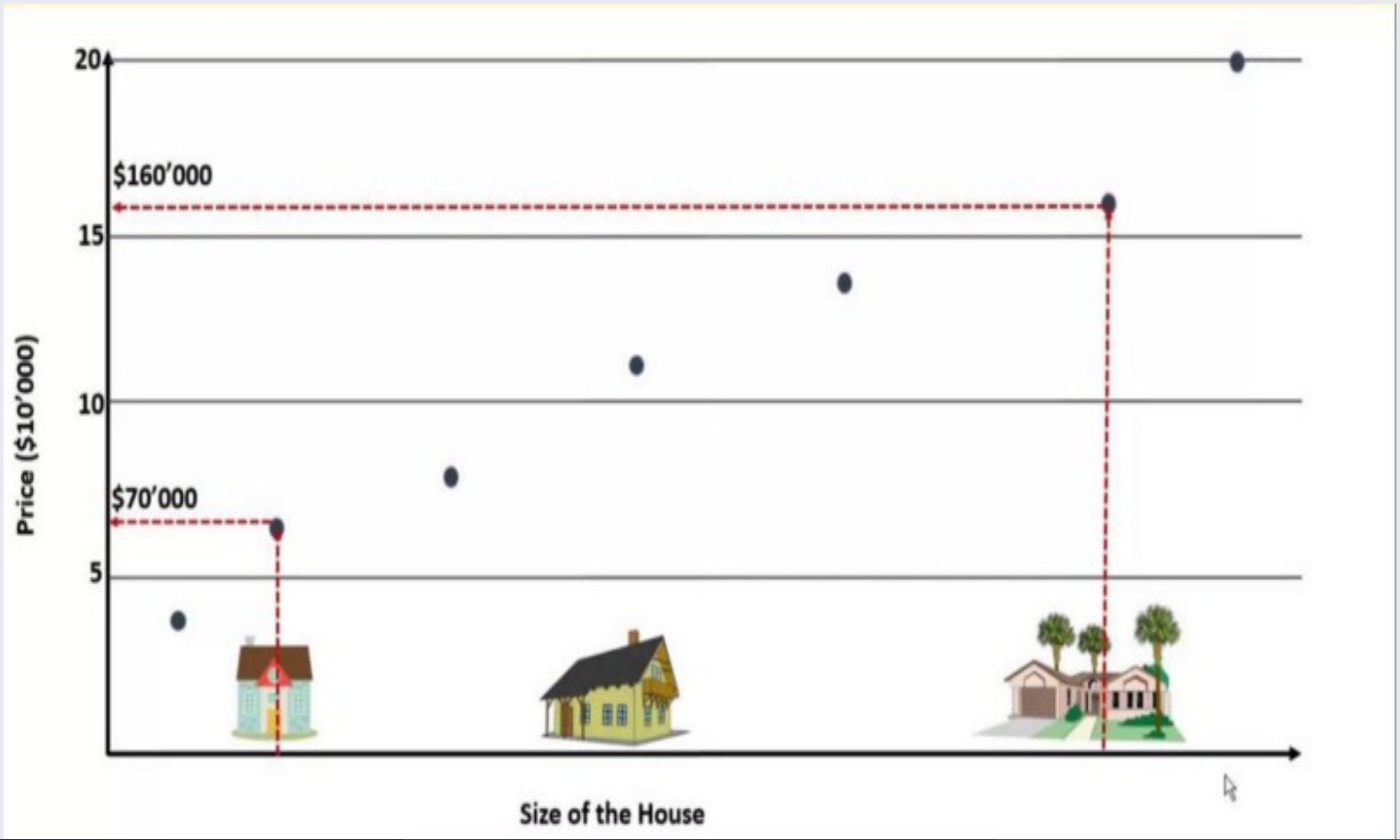
Apprentissage supervisé

- Problème de Régression : un problème de régression se pose lorsque la variable de sortie à prévoir est une valeur numérique.
  - ex : prix des produits dans une pharmacie, le poids des enfants dans un hôpital, les salaires des infirmiers dans un centre de santé, etc.
  - le prix, les poids et les salaires sont des valeurs numériques, comme l'illustre l'image ci-dessous qui montre les prix des maisons sur l'axe vertical.

Les types de modèles en Machine Learning

Apprentissage supervisé

- Problème de Régression : un problème de régression se pose lorsque la variable de sortie à prévoir est une valeur numérique.



Les types de modèles en Machine Learning

Apprentissage supervisé

- Quiz: Quel(s) est/sont le(s) problème(s) de classification dans les propositions suivantes ?
  - Prédire le sexe d'une personne par son style d'écriture
  - Prédire le prix d'une maison en fonction de la zone
  - prédire la maladie d'une personne à partir d'une liste de symptômes
  - Prévoir le nombre de boîtes d'une catégorie de médicaments qui sera vendu le mois prochain.

Les types de modèles en Machine Learning

Apprentissage supervisé

- Quelques algorithmes les plus utilisés pour l'apprentissage supervisé
  - Support Vector Machines (SVM), Linear regression, Logistic regression, Naive Bayes, Linear discriminant analysis, Decision trees, Random Forest, K-nearest neighbor algorithm, Neural Networks (Multilayer perceptron), Similarity learning.
  - Lien utile : Scikit-learn

Les types de modèles en Machine Learning

Apprentissage non-supervisé

- Contrairement à l'apprentissage supervisé, le modèle est alimenté par des données qui **n'ont pas d'étiquettes** humaines prédéfinies. C'est à l'algorithme de trouver des structures, des modèles ou des relations cachées dans les données.
- Tout comme l'apprentissage supervisé, l'apprentissage non supervisé peut être divisé en deux :
  - pour des questions de regroupement - **clustering**
  - pour des questions d'association - **rule based association**

Les types de modèles en Machine Learning

Apprentissage non-supervisé

- questions de regroupement - **clustering**: Dans ce cas, l'algorithme est alimenté par des données non étiquetées et vise à diviser les données en groupes appelés clusters en fonction de certains critères de **similarité** ou de **dissimilarité**. Les points de données qui sont similaires sont regroupés sous un même groupe.

Figure 4:

Les types de modèles en Machine Learning

Apprentissage non-supervisé

- questions de regroupement - **clustering**
  - Les regroupements peuvent être classés en:
    - regroupements exclusifs,
    - chevauchements,
    - hiérarchiques,
    - probabilistes
  - Les algorithmes les plus utilisés en clustering sont entre autres:
    - Hierarchical clustering, K-means clustering, K-NN (k nearest neighbors), Principal Component Analysis, Singular Value Decomposition, Independent Component Analysis

Les types de modèles en Machine Learning

Apprentissage non-supervisé

- pour des questions d'association - **rule based association**:
  Dans les problèmes d'association, le modèle **apprend la relation** entre les données et **propose ensuite certaines règles**.
  - Cette technique d'apprentissage non supervisée consiste à découvrir des relations intéressantes entre les variables dans de grandes bases de données.
  - ex : les personnes qui achètent une nouvelle maison sont plus susceptibles d'acheter de nouveaux meubles, quelqu'un qui achète une brosse à dents est susceptible d'acheter du dentifrice, etc.

Les types de modèles en Machine Learning

Apprentissage non-supervisé

- pour des questions d'association - **rule based association**:
  - De nombreux algorithmes permettant de générer des règles d'association ont été proposés. De nombreux algorithmes ont été proposés pour générer des règles d'association. Certains algorithmes bien connus le sont :
    - Apriori algorithm, Eclat algorithm and Frequent Pattern-Growth
  - C/C L'apprentissage non supervisé est très utile dans l'analyse exploratoire car il permet d'identifier automatiquement la structure des données. Par exemple, des sous-groupes de patients atteints de cancer sont regroupés en fonction de la mesure de l'expression génétique, des groupes d'acheteurs en fonction de leur historique de navigation et d'achat, des films sont regroupés en fonction de la note donnée par les spectateurs.

Les types de modèles en Machine Learning

Apprentissage non-supervisé

- pour des questions d'association - **rule based association**:

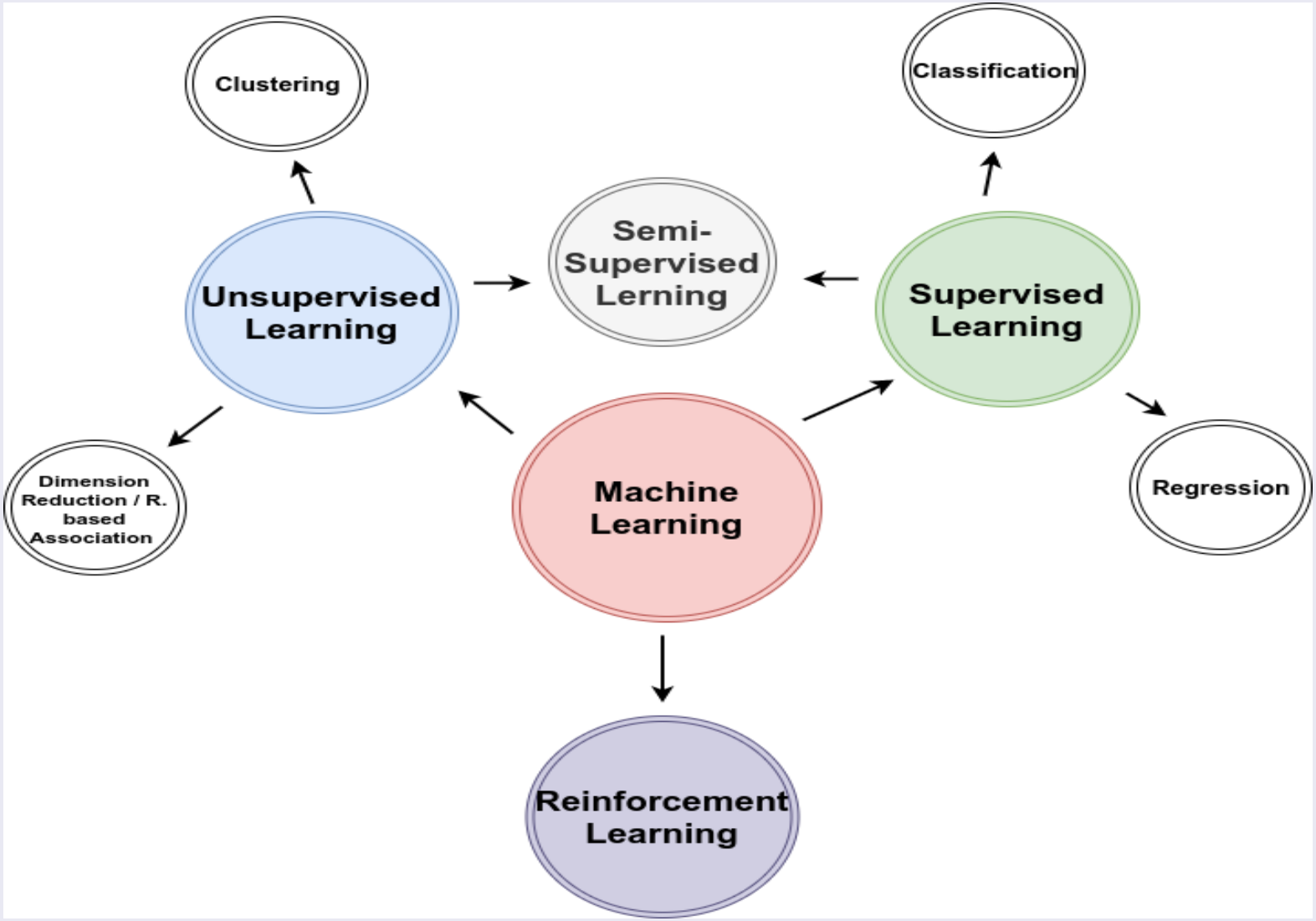
Les types de modèles en Machine Learning

À retenir

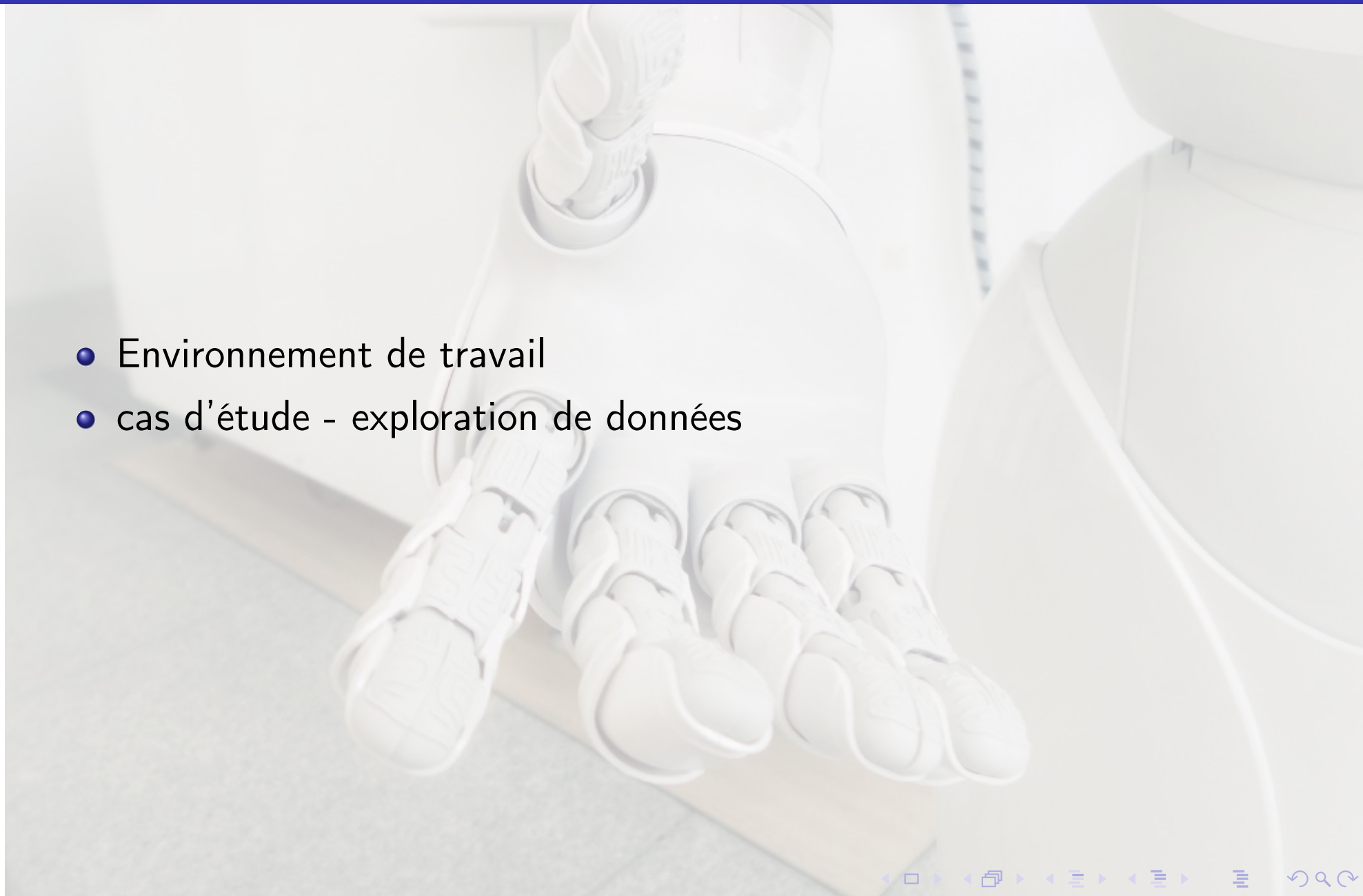
- Comment trouver le bon modèle pour une étude ?
  - 1) jouer avec plusieurs modèles et garder celui qui nous donne de meilleur résultats
  - 2) avoir des cartes de références pour guider le choix du modèle.
    - elle consiste à savoir au préalable, les forces et les faiblesses de chaque algorithme. Et en fonction des données et de l'objectif de l'étude, certains algorithmes seront visiblement adaptés, et d'autres non (confère *ml_cards.pdf*).
  - Faire la différence entre clustering et classification

Les types de modèles en Machine Learning

À retenir



Séance de TD



- Environnement de travail
- cas d'étude - exploration de données

MERCI POUR VOTRE ATTENTION!

RÉFÉRENCES


M. Copeland – "What's the difference between artificial intelligence", *Machine Learning, and Deep Learning* (2016).