



université PARIS-SACLAY

IUT de Vélizy-Rambouillet

CAMPUS DE VÉLIZY-VILLACOUBLAY

APPRENTISSAGE AUTOMATIQUE (ML)

Nahid.Emad@uvsq.fr



Laboratoire
d'Informatique
Parallélisme
Réseaux
Algorithmique
Distribuée



MAISON DE LA SIMULATION

PLAN

- 1. ENVIRONNEMENT : LANGAGES, LOGICIELS, RÉFÉRENCES, ETC.**
- 2. INTRODUCTION : GÉNÉRALITÉ SUR L'APPRENTISSAGE AUTOMATIQUE**
- 3. DIFFÉRENTS TYPES DE MÉTHODES D'APPRENTISSAGE AUTOMATIQUE**
- 4. APPRENTISSAGE SUPERVISÉ**
 - a) RAPPEL DE CALCUL MATRICIEL**
 - b) RÉGRESSION (NON) LINÉAIRE**
- 5. APPRENTISSAGE NON-SUPERVISÉ**
- 6. APPRENTISSAGE PAR RENFORCEMENT**
- 7. APPRENTISSAGE (NON)SUPERVISÉ AVEC EXEMPLES (TD/TP)**

1. ENVIRONNEMENT : LANGAGES, LIBRAIRIES, RÉFÉRENCES

Langages, librairies et environnement utilisés:

- Langage **Python** avec des librairies **NumPy**, **Matplotlib**, etc. (<https://courspython.com/introduction-python.html>),
- L'environnement **Anaconda** (<https://www.anaconda.com/download/>) avec l'application **Jupyter notebook**, etc.

RÉFÉRENCES

1. Azencott, Chloé-Agathe. Introduction au Machine Learning (InfoSup) (French Edition) (p. v). Dunod.
2. Vincent Charvillat, Apprentissage Artificiel (cours, exercices et travaux pratiques), INP-ENSEEIH
3. Vidéo <https://www.youtube.com/watch?v=DrjkjPVf7Bw>
4. Guillaume Saint-Cirgue. Apprendre le Machine Learning en une semaine.

2. INTRODUCTION

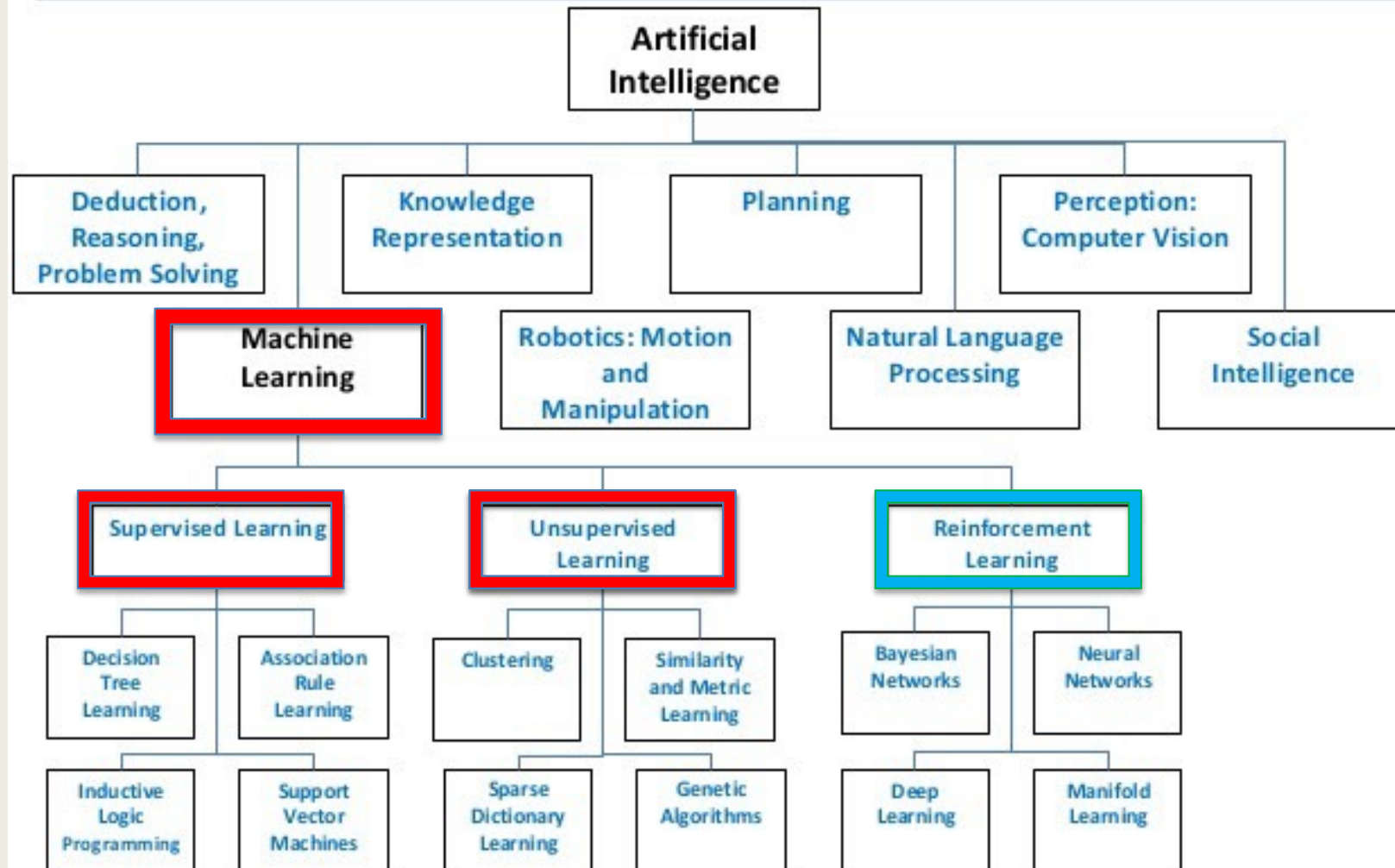
- Généralités
- Analyse de données
- Data, dataset
- Qualité des méthodes ML

2. INTRODUCTION : GÉNÉRALITÉ

- Les systèmes de recommandations,
- Le e-commerce (Amazon, ...),
- La détection des maladies (cancer, etc.),
- L'épidémiologie, la propagation des maladie infectieuses, ...
- La détection d'anomalie (cyberattacks ...),
- La prise de notes dans le domaine médical. ...
- Les assistants numériques (Siri, Google Assistant, Alexa, ...),
- ...

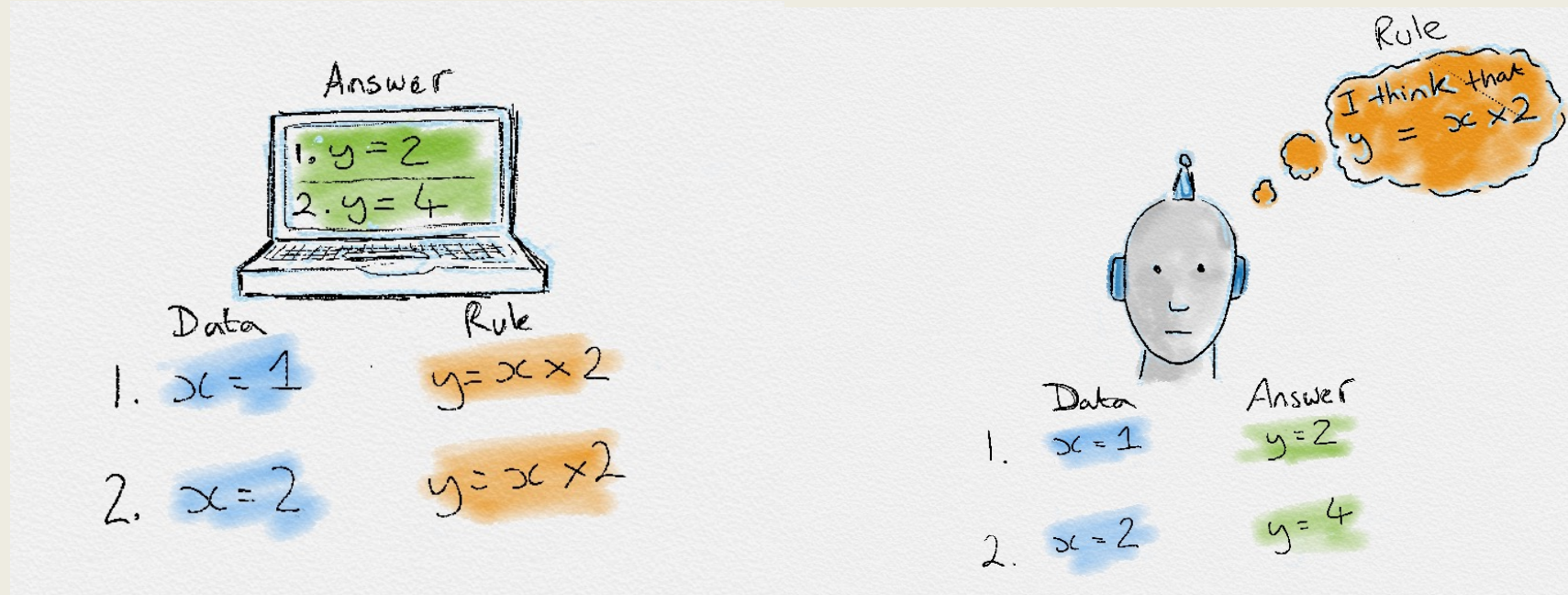
Calcul des probabilités

Artificial Intelligence / Machine Learning Classification



<http://image.slidesharecdn.com/deepdiveinaimlventurelandscape-150831132221-lva1-app6891/95/deepdive-in-aiml-venture-landscape-by-ajit-nazre-rahul-garg-3-638.jpg?cb=1441027412>

Qu'est-ce l'apprentissage automatique (ML)?



Pour apprendre les règles qui caractérisent un problème, les machines doivent passer par un processus d'apprentissage, essayer différentes règles et apprendre de leurs performances.

Machine learning:

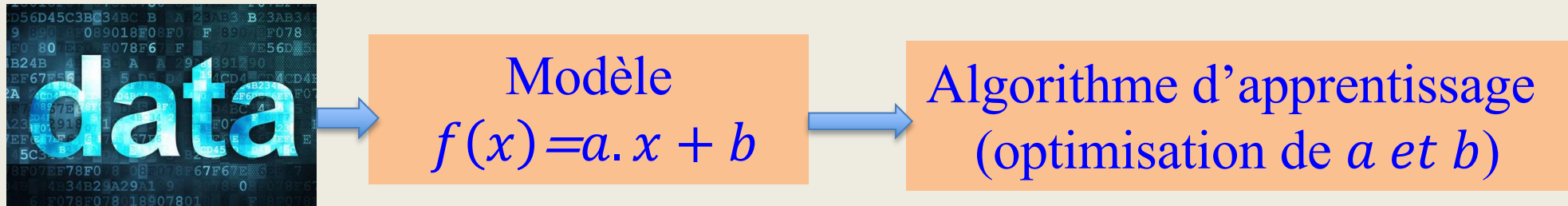
- Permet de transformer les données en connaissances. Cela par l'analyse de ces données afin de trouver les modèles (règles) cachés à l'intérieur.
- L'objectif est de prédire les événements futurs et aider à la prise de décisions complexes. Plus précisément, il s'agit de l'apprentissage automatique d'une fonction qui permet de prédire à partir d'un jeu d'observations de données (labélisées ou non).
- Quelques exemples d'application :
 - ✓ Reconnaissance de caractères/mots/formes, analyse des réseaux sociaux, la publicité ciblée, traduction automatique, finance, etc.

Quand faut-il faire de l'apprentissage automatique ?

- Quand la programmation est trop complexe pour qu'elle soit faite seulement par un programme (une séquence d'instructions).
- Quand on a trop peu d'informations sur le problème ciblé.
- Quand la quantité de données est trop importante.

Quelles sont les grandes étapes d'apprentissage automatique ?

1. **Applications ciblées** : industries, réseaux sociaux, santé / médecine, etc.
2. **Collecte de données** : capteurs, cliques, documents, etc.
3. **Préparation de données** : nettoyage, gestion d'erreurs, etc.
4. **Analyse de données** : exploration, fouille, visualisation, etc.



Big Data : augmentation continue production de données

- Twitter : 50M tweets /jour (=7 téraoctets)
- Facebook : 10 téraoctets /jour
- Youtube : 50h vidéos chargées /minute
- 2.9 millions de e-mails /seconde

Multiples and divisors

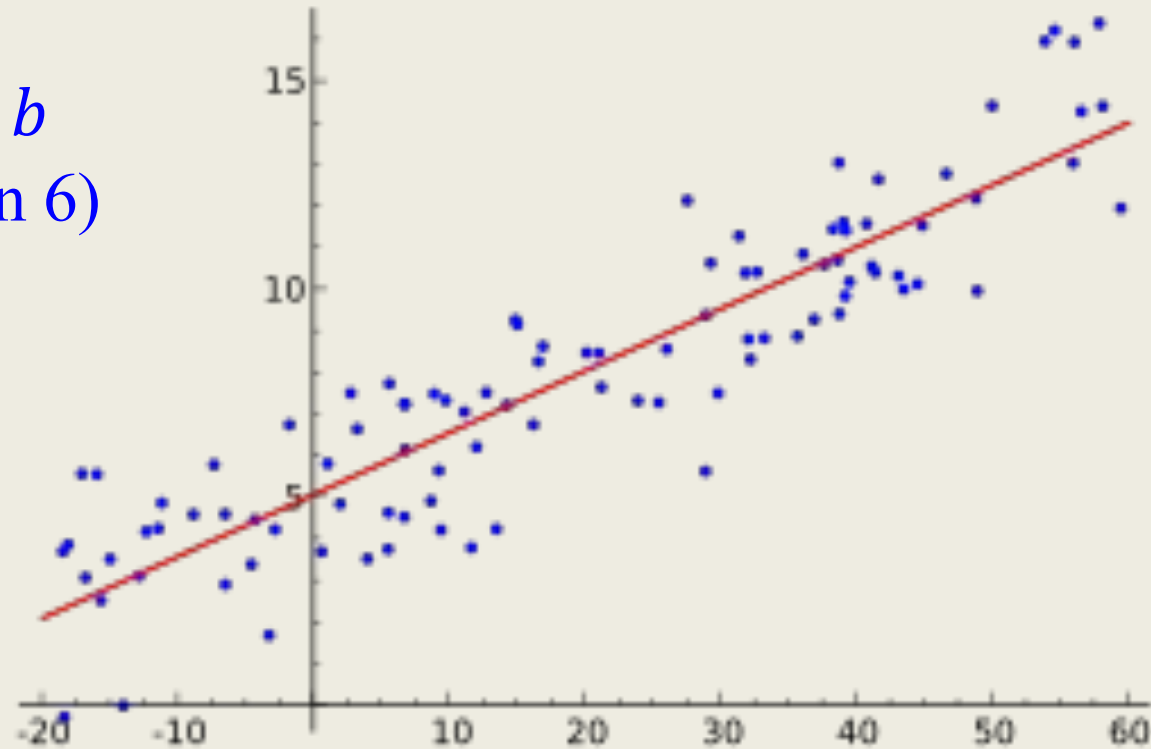
| Prefix | Symbol | Multiples |
|--------|--------|-----------|
| deca | da | 10 |
| hecto | h | 10^2 |
| kilo | k | 10^3 |
| mega | M | 10^6 |
| giga | G | 10^9 |
| tera | T | 10^{12} |
| peta | P | 10^{15} |
| exa | E | 10^{18} |
| zetta | Z | 10^{21} |
| yotta | Y | 10^{24} |

| Prefix | Symbol | Multiples |
|--------|--------|------------|
| deci | de | 10^{-1} |
| centi | c | 10^{-2} |
| milli | m | 10^{-3} |
| micro | μ | 10^{-6} |
| nano | n | 10^{-9} |
| peco | p | 10^{-12} |
| femto | f | 10^{-15} |
| atto | a | 10^{-18} |
| zepto | z | 10^{-21} |
| yocto | y | 10^{-24} |

EXEMPLE DE MODÈLE (RÉGRESSION LINÉAIRE)

$$f(x) = a \cdot x + b$$

(a=5, b environ 6)



[https://fr.wikipedia.org/wiki/R%C3%A9gression_\(statistiques\)](https://fr.wikipedia.org/wiki/R%C3%A9gression_(statistiques))

Les données peuvent représenter le texte, le son, l'image, la vidéo, ...

Dataset/ensemble de données représente des informations obtenues suite à des expériences ou des observations. Chacune des expériences ou des observations est un ***échantillon***. Nous supposons que notre *dataset* est composé de ***m*** échantillons.

Feature/Attribut/variable représente une caractéristique d'un phénomène observé.

Sample/échantillon est une entité caractérisant un objet. Il est constitué d'un ensemble d'attributs ou *features* (on utilise également: instance, point ou vecteur dans \mathbb{R}^n).

Label/étiquette est le résultat constaté ou de la prédiction de celui-ci.

Dataset: ensemble de couples (échantillon, étiquette) = $(x^{(i)}, y^{(i)})$
pour $i=1, m$

EXEMPLES DE DATASET

$m=10$ échantillons d'une seule valeur ou *feature* et une étiquette

X : note sur 20 d'un étudiant,

Y : diplôme (non)validé

| x | y |
|------------|------------|
| $x^{(1)}$ | $y^{(1)}$ |
| $x^{(2)}$ | $y^{(2)}$ |
| $x^{(3)}$ | $y^{(3)}$ |
| $x^{(4)}$ | $y^{(4)}$ |
| $x^{(5)}$ | $y^{(5)}$ |
| $x^{(6)}$ | $y^{(6)}$ |
| $x^{(7)}$ | $y^{(7)}$ |
| $x^{(8)}$ | $y^{(8)}$ |
| $x^{(9)}$ | $y^{(9)}$ |
| $x^{(10)}$ | $y^{(10)}$ |

EXEMPLES DE DATASET

| | | x | y | | |
|---|--|------------|------------|--|--|
| <i>Feature ou Caractéristique ou Une valeur input</i> | | $x^{(1)}$ | $y^{(1)}$ | | |
| | | $x^{(2)}$ | $y^{(2)}$ | | |
| | | $x^{(3)}$ | $y^{(3)}$ | | |
| | | $x^{(4)}$ | $y^{(4)}$ | | |
| | | $x^{(5)}$ | $y^{(5)}$ | | |
| | | $x^{(6)}$ | $y^{(6)}$ | | |
| | | $x^{(7)}$ | $y^{(7)}$ | | |
| | | $x^{(8)}$ | $y^{(8)}$ | | |
| | | $x^{(9)}$ | $y^{(9)}$ | | |
| | | $x^{(10)}$ | $y^{(10)}$ | | |
| | | | | <i>Target ou Label ou résultat</i> | |

EXEMPLES DE DATASET

$m=10$ échantillons de
5 caractéristiques ou
Features et une
étiquette

| x_1 | x_2 | x_3 | x_4 | x_5 | y |
|--------------|--------------|--------------|--------------|--------------|------------|
| $x_1^{(1)}$ | $x_2^{(1)}$ | $x_3^{(1)}$ | $x_4^{(1)}$ | $x_5^{(1)}$ | $y^{(1)}$ |
| $x_1^{(2)}$ | $x_2^{(2)}$ | $x_3^{(2)}$ | $x_4^{(2)}$ | $x_5^{(2)}$ | $y^{(2)}$ |
| $x_1^{(3)}$ | $x_2^{(3)}$ | $x_3^{(3)}$ | $x_4^{(3)}$ | $x_5^{(3)}$ | $y^{(3)}$ |
| $x_1^{(4)}$ | $x_2^{(4)}$ | $x_3^{(4)}$ | $x_4^{(4)}$ | $x_5^{(4)}$ | $y^{(4)}$ |
| $x_1^{(5)}$ | $x_2^{(5)}$ | $x_3^{(5)}$ | $x_4^{(5)}$ | $x_5^{(5)}$ | $y^{(5)}$ |
| $x_1^{(6)}$ | $x_2^{(6)}$ | $x_3^{(6)}$ | $x_4^{(6)}$ | $x_5^{(6)}$ | $y^{(6)}$ |
| $x_1^{(7)}$ | $x_2^{(7)}$ | $x_3^{(7)}$ | $x_4^{(7)}$ | $x_5^{(7)}$ | $y^{(7)}$ |
| $x_1^{(8)}$ | $x_2^{(8)}$ | $x_3^{(8)}$ | $x_4^{(8)}$ | $x_5^{(8)}$ | $y^{(8)}$ |
| $x_1^{(9)}$ | $x_2^{(9)}$ | $x_3^{(9)}$ | $x_4^{(9)}$ | $x_5^{(9)}$ | $y^{(9)}$ |
| $x_1^{(10)}$ | $x_2^{(10)}$ | $x_3^{(10)}$ | $x_4^{(10)}$ | $x_5^{(10)}$ | $y^{(10)}$ |

EXEMPLES DE DATASET

| | x_1 | x_2 | x_3 | x_4 | x_5 | y | |
|---|--------------|--------------|--------------|--------------|--------------|------------|--------------------------------|
| <i>Features / caractéristiques de 10 observations</i> | $x_1^{(1)}$ | $x_2^{(1)}$ | $x_3^{(1)}$ | $x_4^{(1)}$ | $x_5^{(1)}$ | $y^{(1)}$ | <i>Target ou Label</i> |
| | $x_1^{(2)}$ | $x_2^{(2)}$ | $x_3^{(2)}$ | $x_4^{(2)}$ | $x_5^{(2)}$ | $y^{(2)}$ | |
| | $x_1^{(3)}$ | $x_2^{(3)}$ | $x_3^{(3)}$ | $x_4^{(3)}$ | $x_5^{(3)}$ | $y^{(3)}$ | |
| | $x_1^{(4)}$ | $x_2^{(4)}$ | $x_3^{(4)}$ | $x_4^{(4)}$ | $x_5^{(4)}$ | $y^{(4)}$ | |
| | $x_1^{(5)}$ | $x_2^{(5)}$ | $x_3^{(5)}$ | $x_4^{(5)}$ | $x_5^{(5)}$ | $y^{(5)}$ | |
| | $x_1^{(6)}$ | $x_2^{(6)}$ | $x_3^{(6)}$ | $x_4^{(6)}$ | $x_5^{(6)}$ | $y^{(6)}$ | |
| | $x_1^{(7)}$ | $x_2^{(7)}$ | $x_3^{(7)}$ | $x_4^{(7)}$ | $x_5^{(7)}$ | $y^{(7)}$ | |
| | $x_1^{(8)}$ | $x_2^{(8)}$ | $x_3^{(8)}$ | $x_4^{(8)}$ | $x_5^{(8)}$ | $y^{(8)}$ | |
| | $x_1^{(9)}$ | $x_2^{(9)}$ | $x_3^{(9)}$ | $x_4^{(9)}$ | $x_5^{(9)}$ | $y^{(9)}$ | |
| | $x_1^{(10)}$ | $x_2^{(10)}$ | $x_3^{(10)}$ | $x_4^{(10)}$ | $x_5^{(10)}$ | $y^{(10)}$ | |

EXEMPLES DE DATASET

$m=10$
échantillons
d'une seule valeur
ou feature et
Une étiquette

| x | y |
|------------|------------|
| $x^{(1)}$ | $y^{(1)}$ |
| $x^{(2)}$ | $y^{(2)}$ |
| $x^{(3)}$ | $y^{(3)}$ |
| $x^{(4)}$ | $y^{(4)}$ |
| $x^{(5)}$ | $y^{(5)}$ |
| $x^{(6)}$ | $y^{(6)}$ |
| $x^{(7)}$ | $y^{(7)}$ |
| $x^{(8)}$ | $y^{(8)}$ |
| $x^{(9)}$ | $y^{(9)}$ |
| $x^{(10)}$ | $y^{(10)}$ |

$m=10$ échantillons de 5 caractéristiques /
features et une étiquette

| x_1 | x_2 | x_3 | x_4 | x_5 | y |
|--------------|--------------|--------------|--------------|--------------|------------|
| $x_1^{(1)}$ | $x_2^{(1)}$ | $x_3^{(1)}$ | $x_4^{(1)}$ | $x_5^{(1)}$ | $y^{(1)}$ |
| $x_1^{(2)}$ | $x_2^{(2)}$ | $x_3^{(2)}$ | $x_4^{(2)}$ | $x_5^{(2)}$ | $y^{(2)}$ |
| $x_1^{(3)}$ | $x_2^{(3)}$ | $x_3^{(3)}$ | $x_4^{(3)}$ | $x_5^{(3)}$ | $y^{(3)}$ |
| $x_1^{(4)}$ | $x_2^{(4)}$ | $x_3^{(4)}$ | $x_4^{(4)}$ | $x_5^{(4)}$ | $y^{(4)}$ |
| $x_1^{(5)}$ | $x_2^{(5)}$ | $x_3^{(5)}$ | $x_4^{(5)}$ | $x_5^{(5)}$ | $y^{(5)}$ |
| $x_1^{(6)}$ | $x_2^{(6)}$ | $x_3^{(6)}$ | $x_4^{(6)}$ | $x_5^{(6)}$ | $y^{(6)}$ |
| $x_1^{(7)}$ | $x_2^{(7)}$ | $x_3^{(7)}$ | $x_4^{(7)}$ | $x_5^{(7)}$ | $y^{(7)}$ |
| $x_1^{(8)}$ | $x_2^{(8)}$ | $x_3^{(8)}$ | $x_4^{(8)}$ | $x_5^{(8)}$ | $y^{(8)}$ |
| $x_1^{(9)}$ | $x_2^{(9)}$ | $x_3^{(9)}$ | $x_4^{(9)}$ | $x_5^{(9)}$ | $y^{(9)}$ |
| $x_1^{(10)}$ | $x_2^{(10)}$ | $x_3^{(10)}$ | $x_4^{(10)}$ | $x_5^{(10)}$ | $y^{(10)}$ |

EXEMPLES DE DATASET

$m=10$
échantillons
d'une seule valeur
ou feature et
une étiquette

$$(x, y) =$$

| x | y |
|------------|------------|
| $x^{(1)}$ | $y^{(1)}$ |
| $x^{(2)}$ | $y^{(2)}$ |
| $x^{(3)}$ | $y^{(3)}$ |
| $x^{(4)}$ | $y^{(4)}$ |
| $x^{(5)}$ | $y^{(5)}$ |
| $x^{(6)}$ | $y^{(6)}$ |
| $x^{(7)}$ | $y^{(7)}$ |
| $x^{(8)}$ | $y^{(8)}$ |
| $x^{(9)}$ | $y^{(9)}$ |
| $x^{(10)}$ | $y^{(10)}$ |

$m=10$ échantillons de 5 caractéristiques ou
Features et une étiquette

$$(X, y) =$$

| x_1 | x_2 | x_3 | x_4 | x_5 | y |
|--------------|--------------|--------------|--------------|--------------|------------|
| $x_1^{(1)}$ | $x_2^{(1)}$ | $x_3^{(1)}$ | $x_4^{(1)}$ | $x_5^{(1)}$ | $y^{(1)}$ |
| $x_1^{(2)}$ | $x_2^{(2)}$ | $x_3^{(2)}$ | $x_4^{(2)}$ | $x_5^{(2)}$ | $y^{(2)}$ |
| $x_1^{(3)}$ | $x_2^{(3)}$ | $x_3^{(3)}$ | $x_4^{(3)}$ | $x_5^{(3)}$ | $y^{(3)}$ |
| $x_1^{(4)}$ | $x_2^{(4)}$ | $x_3^{(4)}$ | $x_4^{(4)}$ | $x_5^{(4)}$ | $y^{(4)}$ |
| $x_1^{(5)}$ | $x_2^{(5)}$ | $x_3^{(5)}$ | $x_4^{(5)}$ | $x_5^{(5)}$ | $y^{(5)}$ |
| $x_1^{(6)}$ | $x_2^{(6)}$ | $x_3^{(6)}$ | $x_4^{(6)}$ | $x_5^{(6)}$ | $y^{(6)}$ |
| $x_1^{(7)}$ | $x_2^{(7)}$ | $x_3^{(7)}$ | $x_4^{(7)}$ | $x_5^{(7)}$ | $y^{(7)}$ |
| $x_1^{(8)}$ | $x_2^{(8)}$ | $x_3^{(8)}$ | $x_4^{(8)}$ | $x_5^{(8)}$ | $y^{(8)}$ |
| $x_1^{(9)}$ | $x_2^{(9)}$ | $x_3^{(9)}$ | $x_4^{(9)}$ | $x_5^{(9)}$ | $y^{(9)}$ |
| $x_1^{(10)}$ | $x_2^{(10)}$ | $x_3^{(10)}$ | $x_4^{(10)}$ | $x_5^{(10)}$ | $y^{(10)}$ |

EXEMPLES DE DATASET

<https://learningactors.com/data-preprocessing-in-python/>

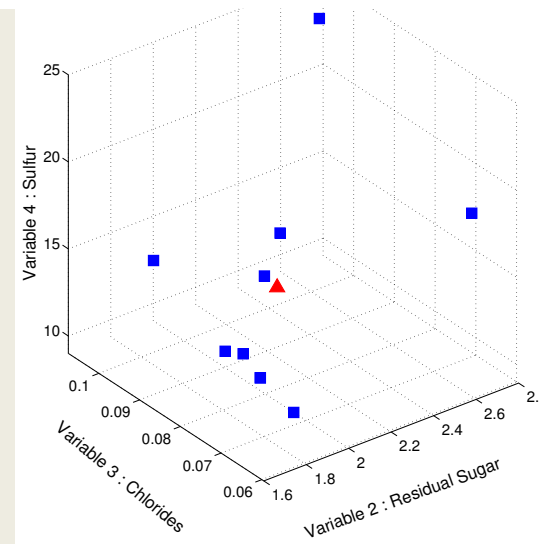
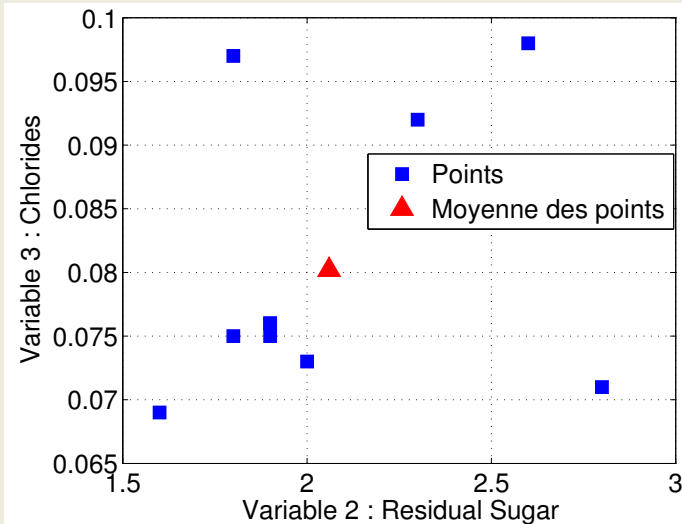
| 1 | Region | Age | Income | Online Shopper |
|----|--------|-----|--------|----------------|
| 2 | India | 49 | 86400 | No |
| 3 | Brazil | 32 | 57600 | Yes |
| 4 | USA | 35 | 64800 | No |
| 5 | Brazil | 43 | 73200 | No |
| 6 | USA | 45 | | Yes |
| 7 | India | 40 | 69600 | Yes |
| 8 | Brazil | | 62400 | No |
| 9 | India | 53 | 94800 | Yes |
| 10 | USA | 55 | 99600 | No |
| 11 | India | 42 | 80400 | Yes |

Quelle est la corrélation entre les features (région géographique, âge et salaire) des personnes et d'être un acheteur en ligne?

EXEMPLES DE DATASET (VISUALISATION)

Élément d'information pour la prédiction de la qualité du vin:

| Points x \ Features | citric acid | residual sugar | chlorides | sulfur dioxide |
|----------------------------|-------------|----------------|-----------|----------------|
| 1 | 0 | 1.9 | 0.076 | 11 |
| 2 | 0 | 2.6 | 0.098 | 25 |
| 3 | 0.04 | 2.3 | 0.092 | 15 |
| Point $x \in \mathbb{R}^4$ | 0.56 | 1.9 | 0.075 | 17 |
| 5 | 0 | 1.9 | 0.076 | 11 |
| 6 | 0 | 1.8 | 0.075 | 13 |
| 7 | 0.06 | 1.6 | 0.069 | 15 |
| 8 | 0.02 | 2 | 0.073 | 9 |
| 9 | 0.36 | 2.8 | 0.071 | 17 |
| 10 | 0.08 | 1.8 | 0.097 | 15 |



Reconnaissance automatique de l'écriture manuscrite

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

Un prédicteur est une fonction f qui, pour une image $x^{(i)} \in X$ prédit une classe $y_i = f(x^{(i)}) \in \{0, 1, \dots, 9\}$.

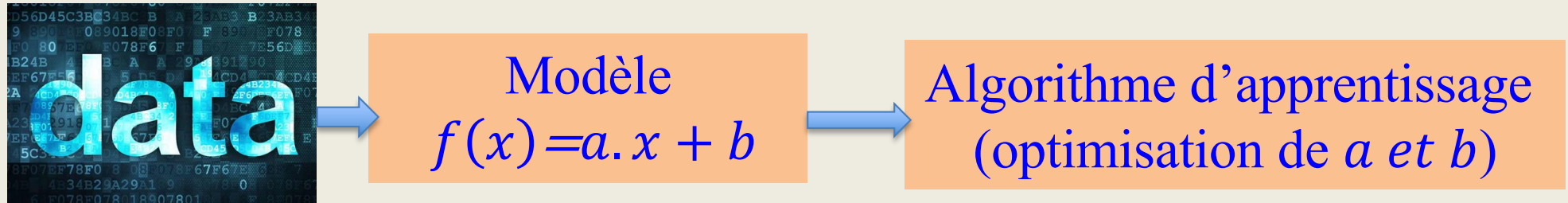
$x^{(i)} \in X = \mathbb{R}^{64 \times 64} \rightarrow$ algo. ML \rightarrow prédit la classe de $x^{(i)} \in \{0, \dots, 9\}$

Comment savoir si f fournie une bonne prédiction?

Soit $\{(x^{(i)}, y^{(i)}) \in X \times y\}_{i=1, \dots, m}$ l'ensemble de m données connues. On peut les utiliser pour tester f :

si $E_i = |f(x^{(i)}) - y^{(i)}|$ est **petit**, alors f est un bon prédicteur.

E indique l'erreur de prédiction.



- Le modèle peut être une régression (linéaire ou non), un parcours d'arbre, ...
- L'objectif d'un algorithme d'apprentissage est de minimiser l'erreur de prédiction sur l'ensemble des données. Autrement dit, pour cet exemple, il faut qu'il trouve le a et le b tels que l'erreur:

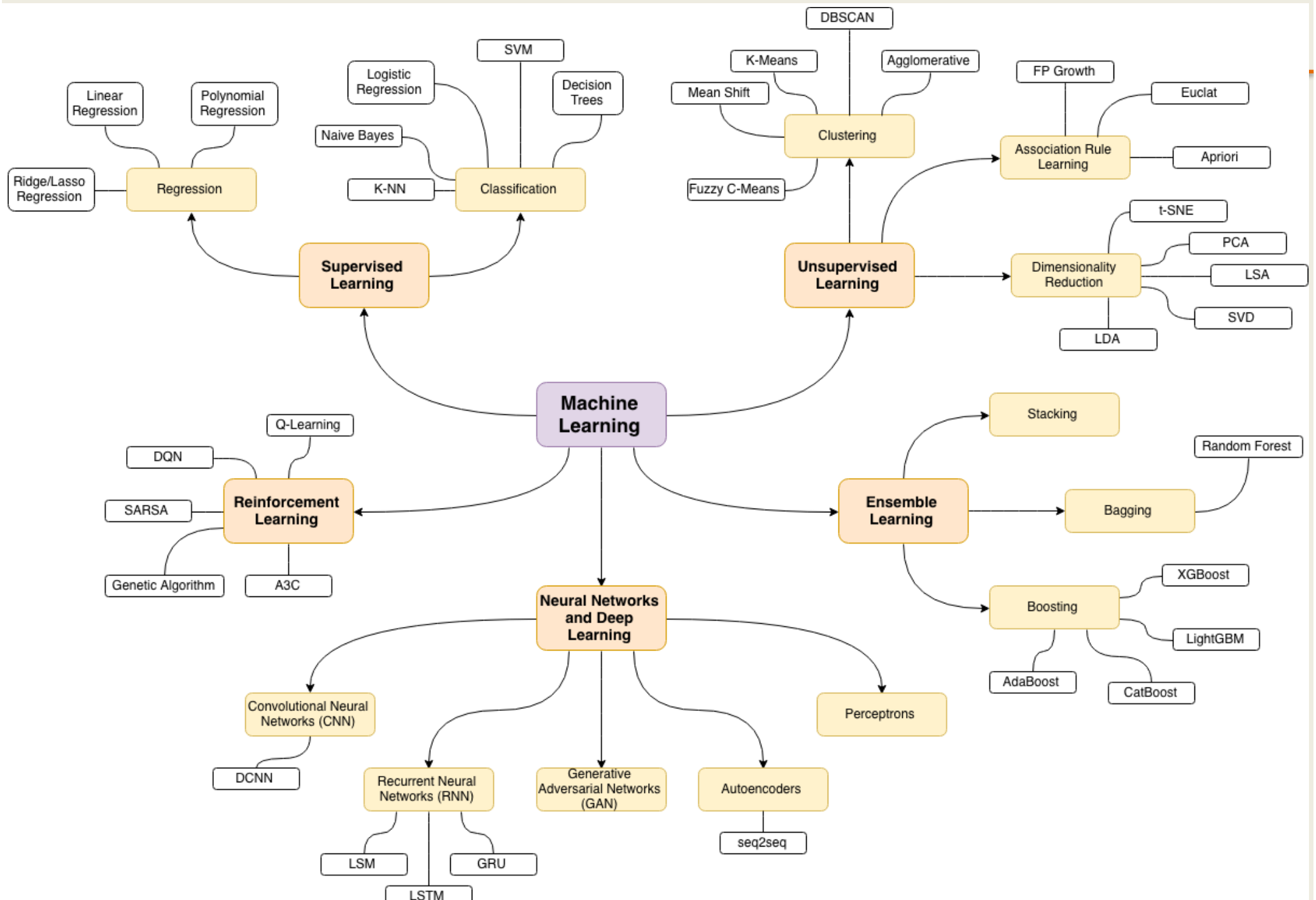
$$E_i = |f(x^{(i)}) - y^{(i)}|$$

soit aussi petite que possible pour toutes les données

3. DIFFÉRENTS TYPES DE MÉTHODES ML

Trois types principaux d'apprentissage :

1. **Supervisé** : Le but est d'apprendre à faire des prédictions à l'aide d'exemples *labélisés*. Les étiquettes supervisent l'apprentissage (classification binaire ou multiple, régression, etc.)
2. **Non-supervisé** : Il s'agit de rechercher des formes cachées dans les données. Par exemples des sous-groupes dans des graphes d'individus (partitionnement, réduction de dimension, etc.).
3. **Par renforcement** : Apprentissage par l'utilisation de l'expérience (issue de l'environnement). Une bonne action est récompensée par une valeur positive sinon négative. Ainsi le système utilise ces récompenses pour s'améliorer (échecs, go, etc.).



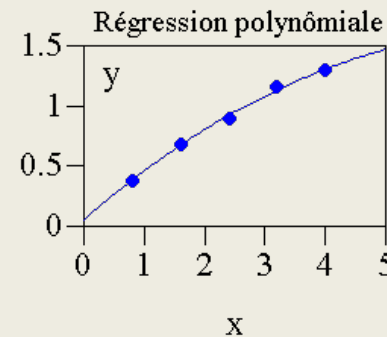
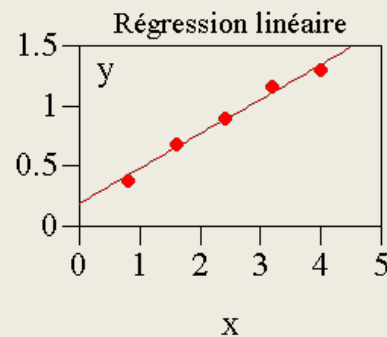
1. ML supervisé: Quand on connaît les *labels* d'un ensemble d'échantillons (un *dataset*), on les utilise pour définir (prédire) de labels pour d'autres échantillons.

Soit $\{(x^{(i)}, y^{(i)})\}_{i=1,\dots,m}$ un *dataset* d'entraînement. Une méthode ML supervisée consiste à définir une fonction f de prédiction tel que: $f(x)$ pour tout x de ce *dataset* soit aussi proche que possible de y . Autrement dit, la machine va utiliser les entrées/sorties $x^{(i)}, y^{(i)}$ afin de trouver la formule optimale de la fonction f .

✓ **Exemples :** classification d'images, détection d'objets, détection d'anomalies, etc.

An algorithm ML est un algorithme qui optimise un modèle (une fonction f) à partir d'exemples.

- Régression :** Quand y est un vecteur à valeurs réelles (entre $-\infty$ et $+\infty$). Une régression peut être *linéaire* (une droite) ou *polynomiale* (une courbe).



- Classification :** Quand y est un vecteur à valeurs discrètes (entre $0, 1, \dots, c$).
 - Binaire :** $y = \{a, b\}$ (par exemple $a = -1, b = +1$)
 - Multi-classe :** $y = \{1, 2, \dots, c\}$ (avec $c > 2$).

2. ML non-supervisé: Quand on connaît les entrées mais on ne connaît ni les résultats ni la fonction. C'est à la machine de trouver l'organisation des données et extraire des sous-ensembles, etc.

Dans ce cas, seul un *dataset* $\{(x^{(i)})\}$, pour $i=1, m\}$ est connu. Ex: une méthode ML non-supervisée essaie de trouver des structures cachées dans ce *dataset* et d'extraire des sous-ensembles.

✓ **Exemples:** partitionnement d'un réseau d'individus, classification, etc.

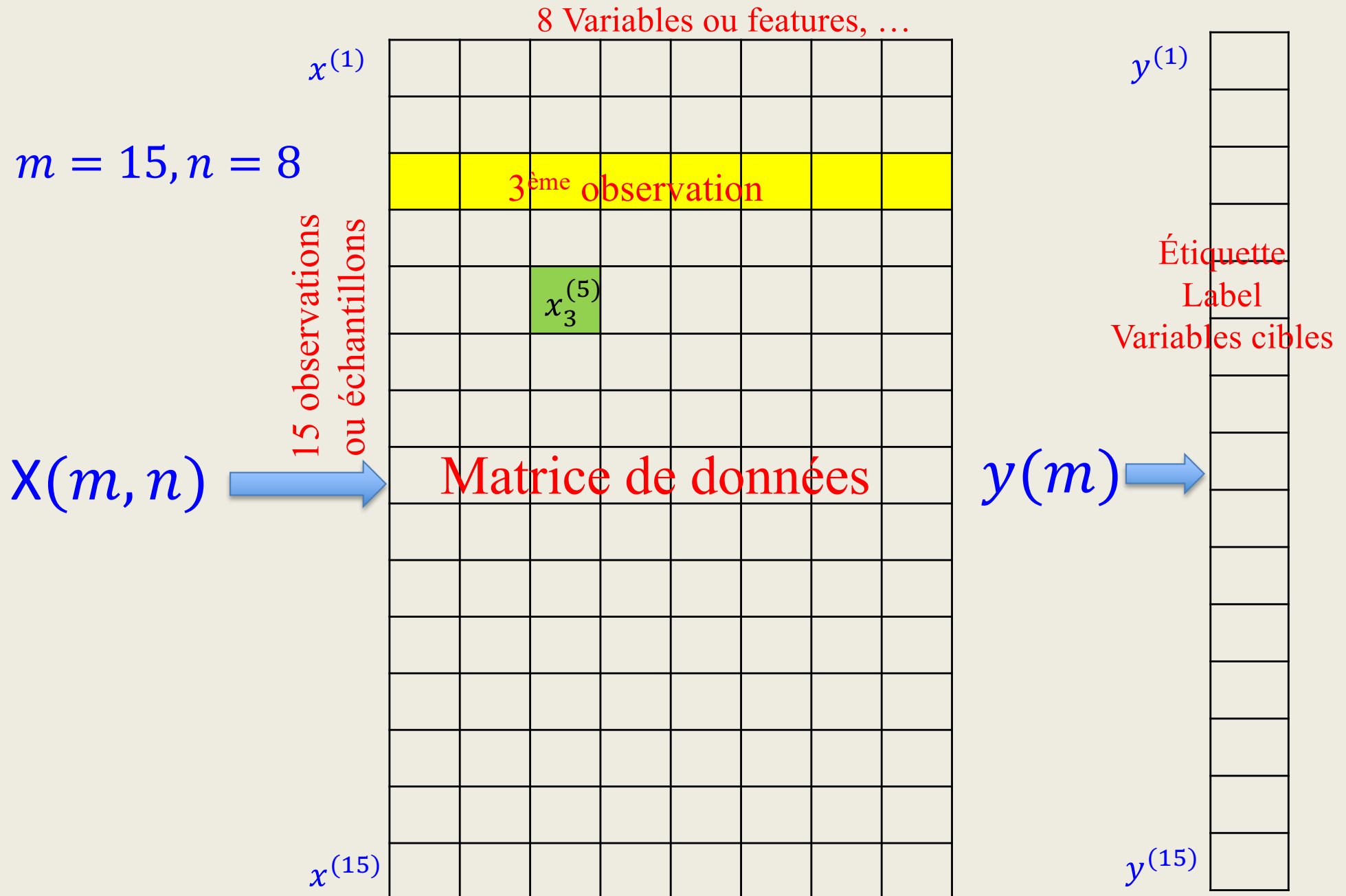
3. ML semi-supervisé: une combinaison de SML et UML.

4. APPRENTISSAGE SUPERVISÉ

Régression linéaire, polynomiale :

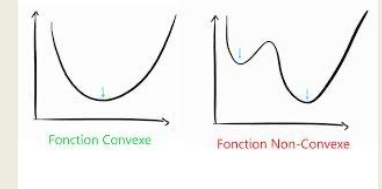
1. **Dataset** : Soit $\{(x^{(1)}, y^{(1)}), (x^{(2)}, y^{(2)}), \dots, (x^{(m)}, y^{(m)})\}$ notre dataset (cf: page suivante)
2. **Modèle + paramètres** : Pour les modèle RL et RP, la fonction f est définie par :
 - $f(x) = a.x + b$ avec a, b initiaux choisis aléatoirement
 - $f(x) = a.x^2 + b.x + c$ avec a, b et c initiaux choisis aléatoirement

MLS: 1. DATASET



4. APPRENTISSAGE SUPERVISÉ

Régression linéaire, polynomiale (suite) :



3. **Fonction de coût** : Soit $E_i = (f(x^{(i)}) - y^{(i)})^2$ (pour tout $i=1$ à m).

Soit $J(a, b) = \frac{1}{2 \cdot m} (E_1 + E_2 + \dots + E_m)$ l'erreur quadratique moyenne. L'objectif de l'AAA est de trouver les paramètres a et b minimisant la fonction $J(a, b)$.

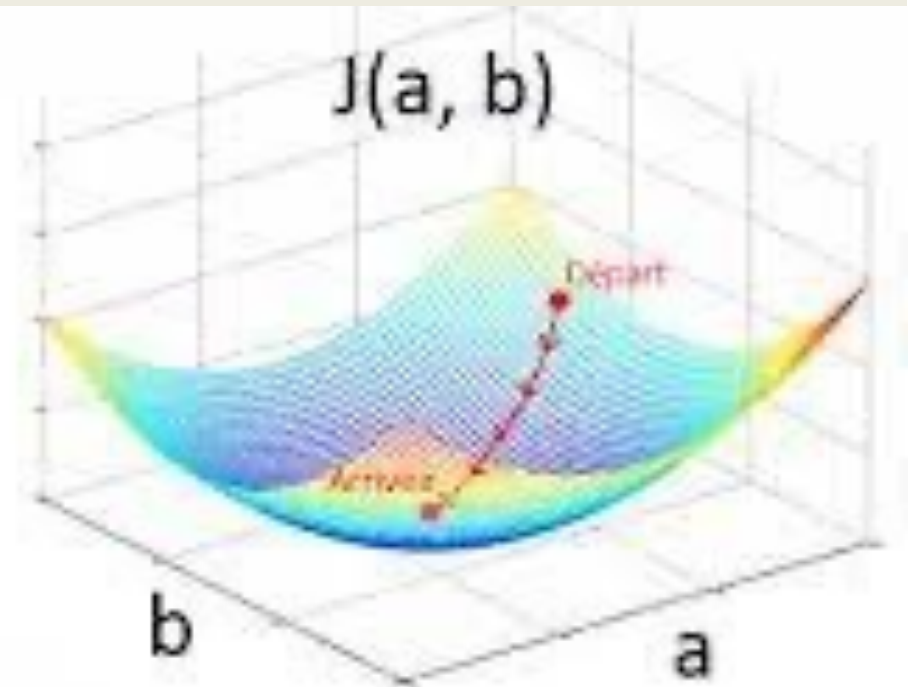
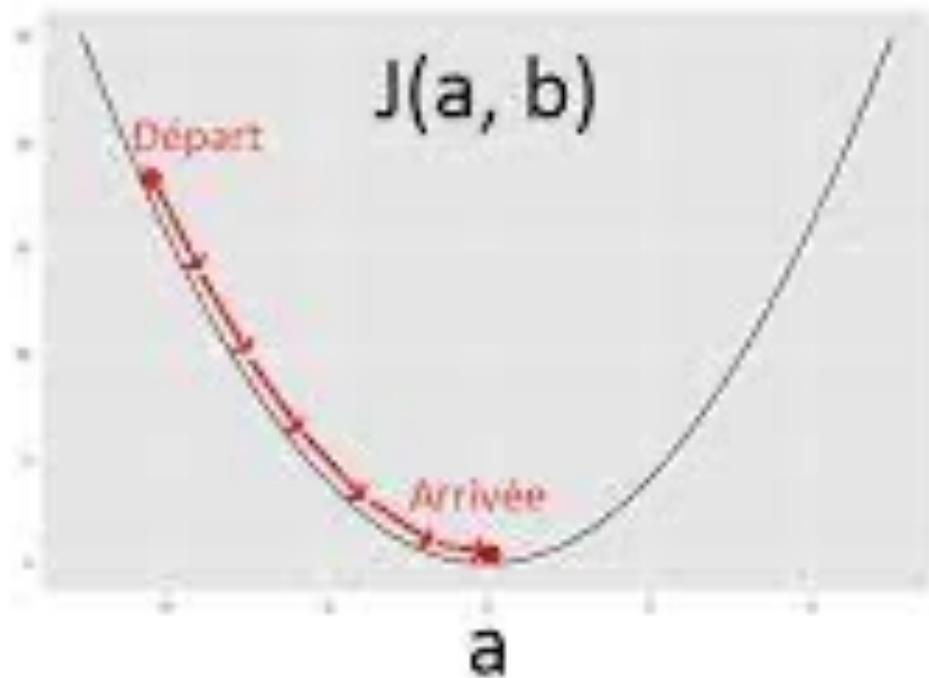
a) Méthode des moindres carrés permet de faire cette minimisation. Cela par minimiser la dérivée de J relativement à a et b ($\frac{\partial J}{\partial a} = 0$ et $\frac{\partial J}{\partial b} = 0$). Cependant, elle nécessite l'inversion de la matrice des données. Ce qui peut coûter très cher lorsque le dataset est de grande taille.

b) Méthode de descente de gradient (DG) permet de faire cette minimisation pas à pas et est moins coûteuse (J devra être convexe).

4. APPRENTISSAGE SUPERVISÉ

Régression linéaire, polynomiale (suite) :

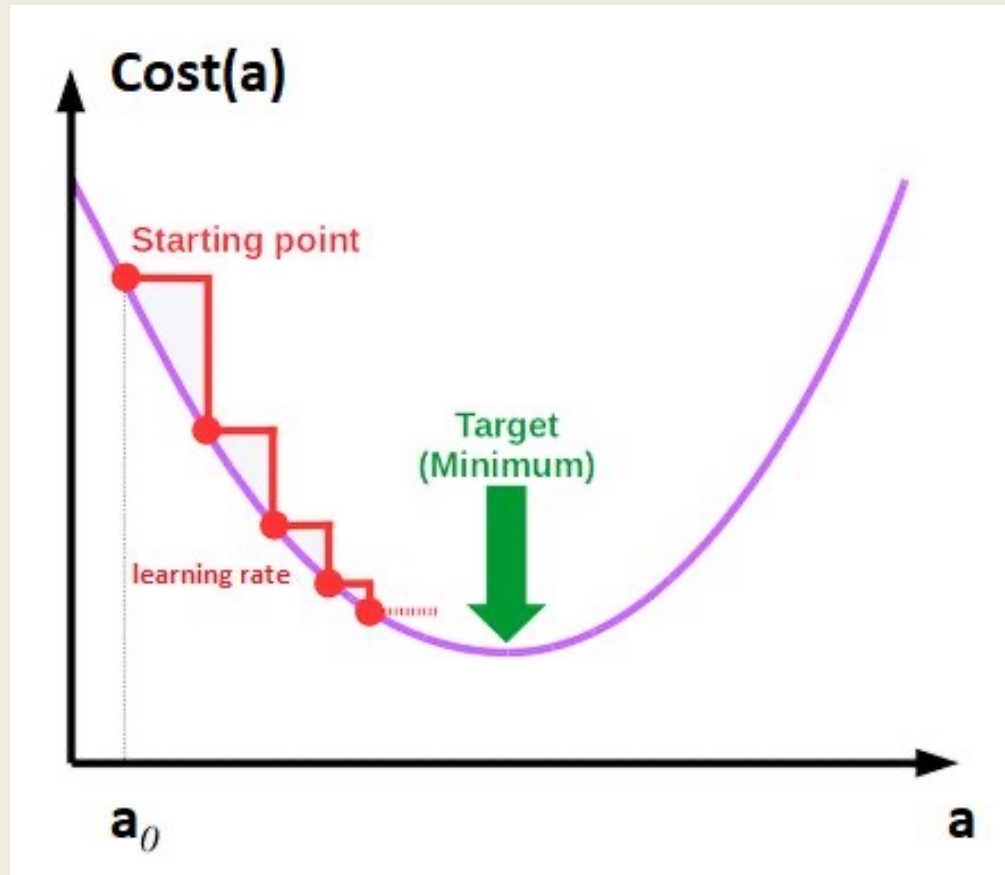
4. **Algorithme d'apprentissage** : La méthode de DG permet de minimiser n'importe quelle fonction convexe en un temps raisonnable.



4. APPRENTISSAGE SUPERVISÉ

Régression linéaire, polynomiale (suite) :

4. **Algorithme d'apprentissage** : La méthode de DG permet de minimiser n'importe quelle fonction convexe en un temps raisonnable.



$$a = a + \alpha \frac{\partial J(a, b)}{\partial a}$$
$$b = b + \alpha \frac{\partial J(a, b)}{\partial b}$$

4. APPRENTISSAGE SUPERVISÉ

Régression linéaire (suite) :

- 4. Algorithme d'apprentissage :** La méthode de DG permet de minimiser n'importe quelle fonction convexe en un temps raisonnable.
1. Initialisation: choix aléatoire de a et b (a_0 et b_0)
 2. Calculer la pente de la Fonction Coût : $(\frac{\partial J}{\partial a}$ et $\frac{\partial J}{\partial b})$.
 3. Mise à jour de a_i et b_i : Avancer d'un pas dans la direction de la pente la plus forte (modification des paramètres a et b).
 4. Répétition des étapes 1 et 2 jusqu'à l'obtention du minimum de $J(a, b)$.

4. APPRENTISSAGE SUPERVISÉ

4. Algorithme d'apprentissage ; calcul des dérivés/gradients :

$$J(a, b) = \frac{1}{2 \cdot m} \sum_{i=1}^{i=m} (a \cdot x^{(i)} + b - y^{(i)})^2$$

$$\frac{\partial J(a, b)}{\partial a} = \frac{1}{m} \sum_{i=1}^{i=m} (a \cdot x^{(i)} + b - y^{(i)}) \cdot x^{(i)}$$

$$\frac{\partial J(a, b)}{\partial b} = \frac{1}{m} \sum_{i=1}^{i=m} (a \cdot x^{(i)} + b - y^{(i)})$$

Supposons qu'un échantillon $x^{(i)}$ est constitué d'une seule valeur et que le label correspondant est : $y^{(i)}$. Alors

4. APPRENTISSAGE SUPERVISÉ

4. Algorithme d'apprentissage ; calcul des dérivés/gradients :

| X = | | $\theta =$ | $y =$ |
|-----------|---|------------|-----------|
| x | 1 | | |
| $x^{(1)}$ | 1 | a | $y^{(1)}$ |
| $x^{(2)}$ | 1 | b | $y^{(2)}$ |
| $x^{(3)}$ | 1 | | $y^{(3)}$ |
| ... | 1 | | $y^{(4)}$ |
| $x^{(m)}$ | 1 | | $y^{(m)}$ |

$$F(X, y) = X \cdot \theta \quad \frac{\partial J(\theta)}{\partial \theta} = \frac{1}{m} X^T \cdot (X \cdot \theta - y)$$

$$\text{La mise à jour de : } \theta = \theta - \alpha \cdot \frac{\partial J(\theta)}{\partial \theta}$$

EXERCICES

Visionnez la vidéo <https://www.youtube.com/watch?v=DrjkjPVf7Bw>
Sélectionnez un des exemples d'IA qui vous paraît le plus intéressant et expliquez la raison de votre intérêt..