

Introduction to Machine Learning

E. A. ABDELLAOUI

ENS-UMI, Meknès
Morocco

03-03-2022

Table of Contents

1. Machine Learning

2. Fondamentaux de ML

3. Evaluation d'un modèle de ML







C'qoui le Machine Learning

Machine Learning

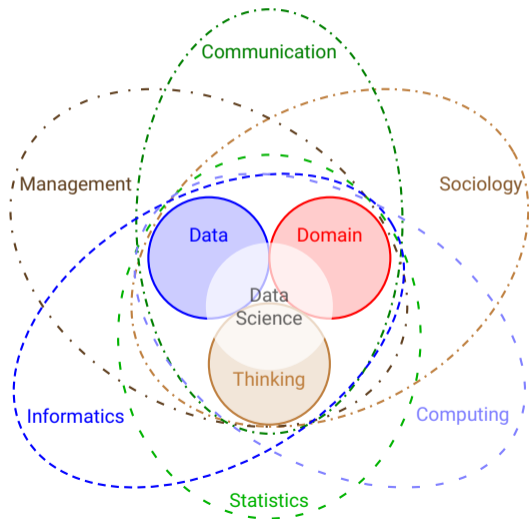


Figure 1: Areas of Machine Learning

Machine Learning



Définition

L'apprentissage automatique peut-être vu comme l'ensemble des techniques permettant à une machine d'apprendre à réaliser une tâche sans avoir à la programmer explicitement pour cela.



Figure 2: Principe de Machine Learning

Machine Learning



Exemple 1

- Voici 1000 photos des animaux, retrouve celles ou se trouve mon chien (voir la figure 3).



Figure 3: Photo des animaux



Fondamentaux de ML

Fondamentaux de ML

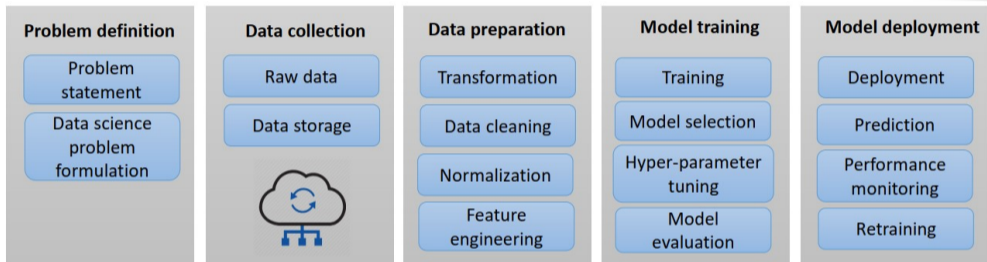


Figure 4: Étapes de ML

Dataset



Table 1: Exemple d'une base données (Dataset)

Prix	Surface	Qualité	postale
30000	50	4	9500
70550	80	2	10120
50950	65	3	5000
80480	95	1	45000
...

Dataset



Exemple de Dataset sur des appartements

Target y	Features		
	x_1	x_2	x_3
Prix	Surface	Qualité	Adresse postale
313,000	90	3	95000
720,000	110	5	93000
250,000	40	4	44500
290,000	60	3	67000
190,000	50	3	59300
...

A blue double-headed vertical arrow on the left of the table is labeled m , indicating the number of rows (examples). A yellow double-headed horizontal arrow below the table is labeled n , indicating the number of columns (features).

Par convention:

m : nombre d'exemples

n : nombre de features

Figure 5: Étapes de ML

Dataset



Dataset (x, y)

y	x_1	x_2	x_3	...	x_n
$y^{(1)}$	$x_1^{(1)}$	$x_2^{(1)}$	$x_3^{(1)}$...	$x_n^{(1)}$
$y^{(2)}$	$x_1^{(2)}$	$x_2^{(2)}$	$x_3^{(2)}$...	$x_n^{(2)}$
$y^{(3)}$	$x_1^{(3)}$	$x_2^{(3)}$	$x_3^{(3)}$...	$x_n^{(3)}$
...
$y^{(m)}$	$x_1^{(m)}$	$x_2^{(m)}$	$x_3^{(m)}$...	$x_n^{(m)}$

Figure 6: Étapes de ML

Prétraitement de dataset



Table 2: Exemple d'une base données (Dataset)

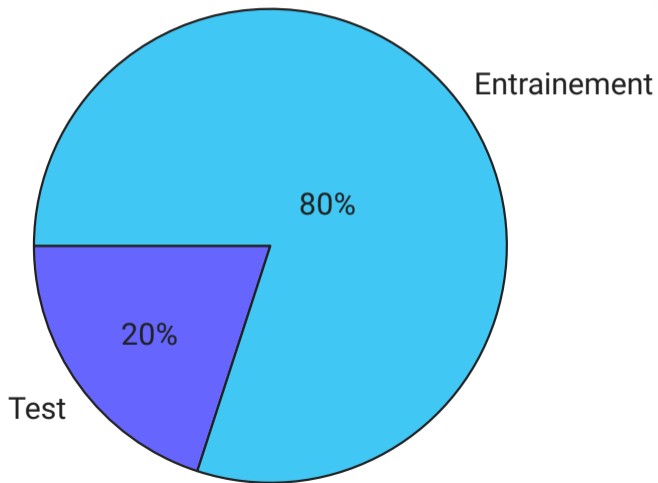
Prix	Surface	Qualité	postale
30000	50	4	D
70550	??	2	D
50950	65	3	C
??	95	1	D
50950	65	3	C
...

Prétraitement de dataset



Figure 7: Prétraitement de dataset

Séparation de dataset



Séparation de dataset

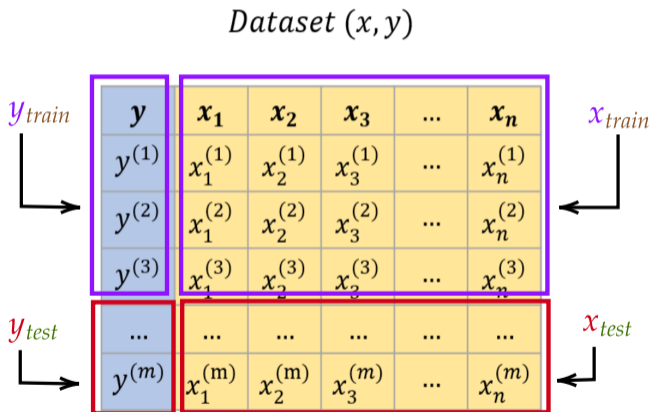


Figure 9: Séparation de dataset

Modèle

Le modèle est en quelque sorte le cœur de votre programme, c'est lui qui va effectuer la tâche que vous cherchez à accomplir, par exemple reconnaître un animal sur une photo ou prédire le prix d'un appartement.

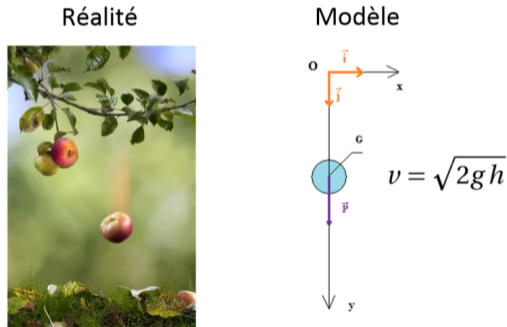


Figure 10: Modélisation mécanique

Modèle

Mais pour prédire le prix d'un appartement en fonction de toutes ses caractéristiques, quelle est l'équation mathématique à entrer dans la machine ? Et quelle est l'équation pour reconnaître un chat sur une photo ? Nul ne le sait !

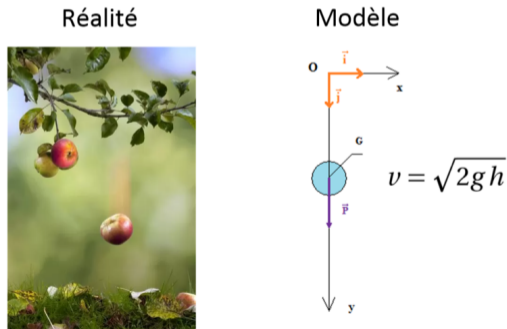


Figure 11: Modélisation mécanique

Modèle

Laisser la machine trouver le modèle qui correspond le mieux à votre objectif en utilisant la Dataset (x, y)

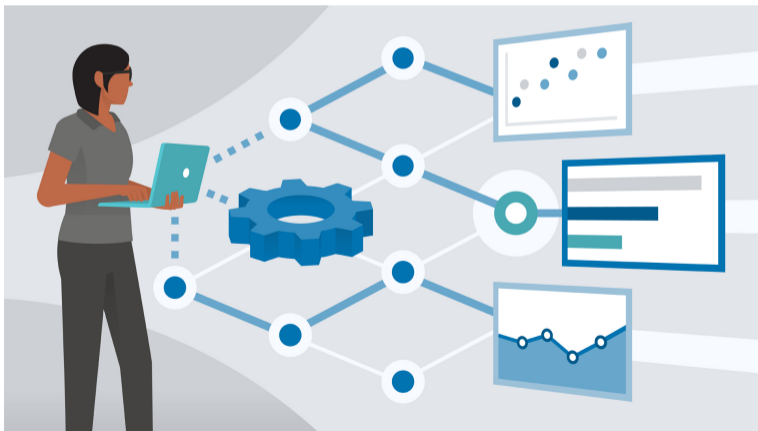


Figure 12: Modélisation par machine

Modèle

Laisser la machine trouver le modèle qui correspond le mieux à votre Dataset (x, y)

$$y = f(x) + \epsilon \quad (1)$$



Algorithme de ML

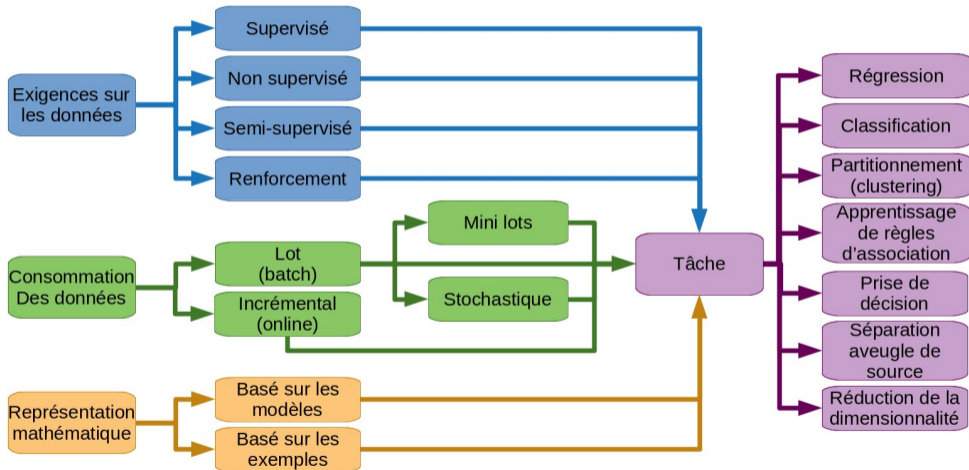


Figure 13: Algorithmes de ML

Apprentissage supervisé



Définition

Un algorithme d'apprentissage supervisé nécessite la présence, pendant la phase d'apprentissage, des résultats corrects attendus pour chaque élément de l'ensemble initial de données. Nous dirons dans ce cas que les éléments sont **annotés** ou **étiquetés**.

Apprentissage supervisé

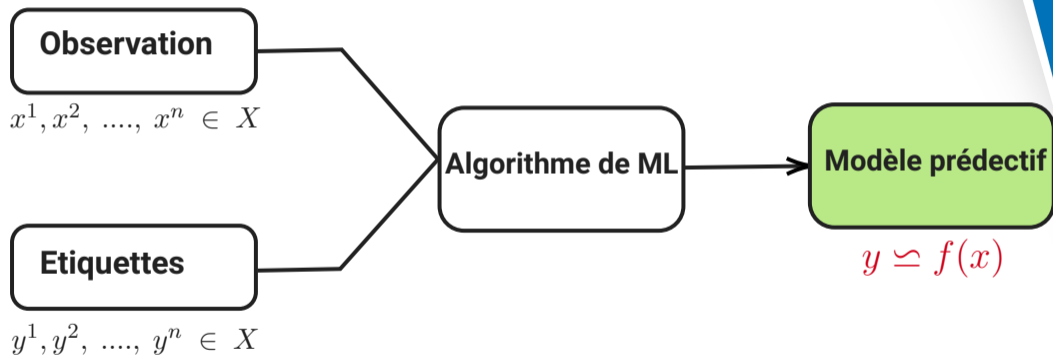


Figure 14: Apprentissage supervisé

Apprentissage NON-supervisé



Définition

Contrairement à l'apprentissage supervisé, une technique d'apprentissage non supervisé ne nécessite pas la présence d'exemples de résultats attendus (ou d'étiquettes).

Apprentissage NON-supervisé



Figure 15: Apprentissage NON supervisé



Apprentissage NON-supervisé



Figure 16: Apprentissage NON supervisé

Régression vs classification



L'identifier de type de sortie attendu de notre programme nous aideons à choisir l'algorithme d'apprentissage automatique. Cela répond également aux questions suivantes: est-ce une **valeur continue** (un nombre)? Est-ce une **valeur discrète** (une catégorie)? Le premier cas s'appelle la régression et le second est la classification

Régression vs classification



L'identifier de type de sortie attendu de notre programme nous aideons à choisir l'algorithme d'apprentissage automatique. Cela répond également aux questions suivantes: est-ce une **valeur continue** (un nombre)? Est-ce une **valeur discrète** (une catégorie)? Le premier cas s'appelle la régression et le second est la classification

Mathématiquement

$$f(x) = \begin{cases} \textit{Rgression}, & \text{si } y \in R \\ \textit{Classification}, & \text{si } y \in \{0, 1\} \end{cases}$$

Régression vs classification

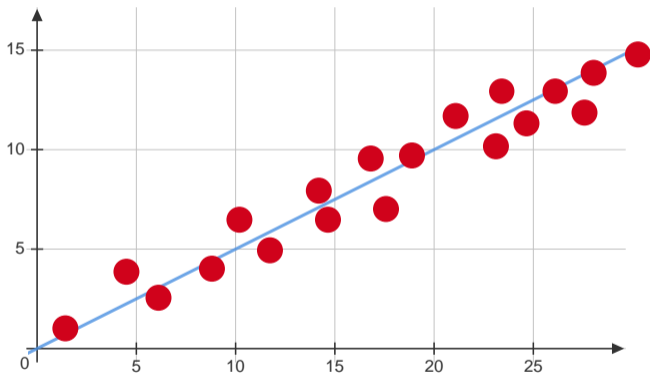


Figure 17: Régression

Régression vs classification

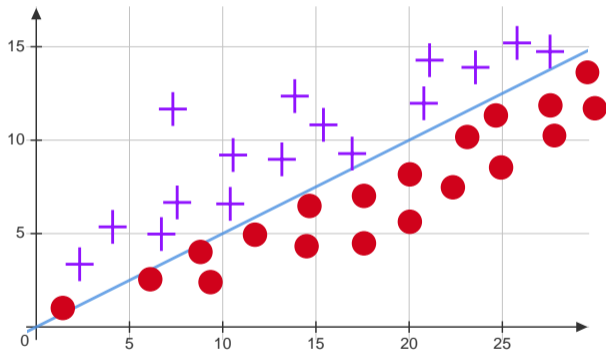


Figure 18: Classification

Algorithmes de prédiction



Algorithmes

- Bayesian;
- Gradient Boost;
- Random Forest;
- Extra Trees ;
- SVM;
- etc.



Evaluation

Métriques de performance (Régression)



1

$$RMSE = \sqrt{\frac{\sum_{k=1}^N ((k) - t_m(k))^2}{N - 1}}$$

2

$$MAE = \frac{\sum_{k=1}^N | (k) - t_m(k) |}{N}$$

3

$$R^2 = 1 - \frac{\sum_{k=1}^N ((k) - t_m(k))}{\sum_{k=1}^N ((k) - t_m(k))}$$

4

etc.

Métriques de performance (Classification)



1

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN}$$

2

$$Precision = \frac{TP}{TP + FP}$$

3

$$Recall = \frac{TP}{TP + FN}$$

4

$$f - score = \frac{2 * (Recall * Precision)}{Recall + Precision}$$

Métriques de performance (Classification)



		Label prédit	
		Faux	Réel
Vrai label	Faux	Vrai Positif (VP)	Faux Négatif (FN)
	Réel	Faux Positif (FP)	Vrai Négatif (VN)

Table 3: Matrice de confusion

- Vrai** la prédiction est vraie (label prédit = vrai label)
- Faux** la prédiction est fausse (label prédit \neq vrai label)
- Positif** la prédiction est fausse
- Négatif** la prédiction est vraie



Bref

Résumé

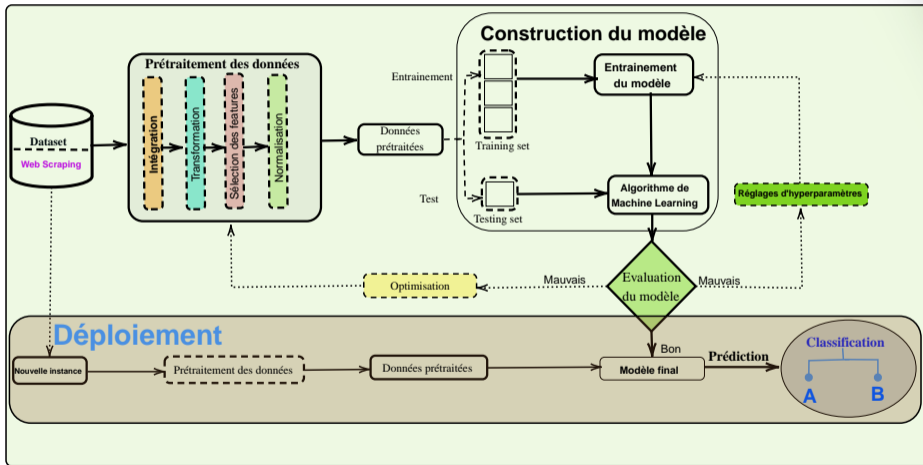


Figure 19: Processus d'un projet de ML



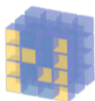
Pratique

Outils pour pratiquer ML



Figure 20: Pratiquer ML

Quelques Bibliothèques de Python



NumPy



plotly

matplotlib

theano



TensorFlow



Keras



python™

Natural Language Analyses
with NLTK



Thanks

Doubts and Suggestions

abdellaoui.e@gmail.com

www.abdellaoui.net

Introduction to Machine Learning

E. A. ABDELLAOUI

ENS-UMI, Meknès
Morocco

03-03-2022


جامعة مولاي إسماعيل
UNIVERSITÉ MOULAY ISMAÏL


المدرسة العليا للإساتجة
ÉCOLE NORMALE SUPÉRIEURE