

# Intelligence Artificielle 2

Cours 11 à 13 - Algorithmes de ML

Steve Lévesque, Tous droits réservés © où applicables

# Table des matières

- 1 Algorithmes de Machine Learning
  - Définition
  - Supervisé - Classification et la Régression
  - Non-supervisé
    - Non-supervisé - Clustering
    - Non-supervisé - Dimensionality Reduction (PCA)
  - Semi-supervisé
  - L'apprentissage par renforcement

# Définition

Un algorithme de Machine Learning est un type de modèle avec des propriétés statistiques, logiques ou mathématiques qui permet d'accomplir une tâche spécifique ou générale.

Chacun est doté d'un atout et défaut, donc il faut bien les choisir en fonction du problème.

Certains peuvent tout simplement ne pas fonctionner avec un domaine ou une tâche précise.

# Types et domaines d'application de Machine Learning - Topologie brève

Question : Où est-ce que la rédaction de chiffres écrits (MNIST) se retrouverait-elle ?

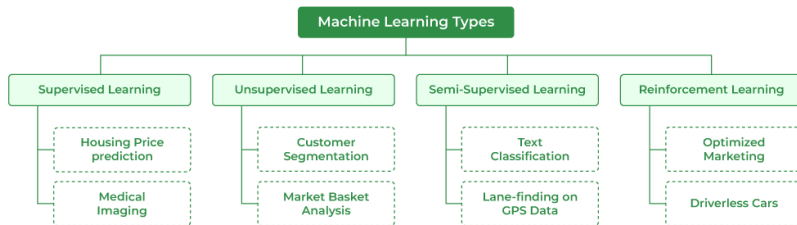


Figure: <https://www.geeksforgeeks.org/machine-learning-algorithms/>

# Supervisé - Classification et la Régression

Pour le type le plus standard, supervisé, il existe deux (2) types de méthodes de prédictions.

La Classification et la **Régression**

La grande différence est que la Classification est pour prédire des valeurs discrètes (précises) tandis que la **Régression** **est pour prédire des valeurs continues (un nombre non catégorisé).**

# Supervisé - Classification et la Régression - Différences (Classification)

## Classification (partie 1) :

- Dans cet énoncé de problème, les variables cibles sont discrètes.
- Des problèmes tels que la classification des spams, la prédiction des maladies sont résolus à l'aide d'algorithmes de classification.
- Dans cet algorithme, nous essayons de trouver la meilleure limite de décision possible qui peut séparer les deux classes avec la séparation maximale possible.
- Les métriques d'évaluation telles que Précision ("Accuracy"), Recall et F1-Score sont utilisées ici pour évaluer les performances des algorithmes de classification.
- Ici, nous sommes confrontés à des problèmes tels que la classification binaire ou les problèmes de classification multi-classes.

# Supervisé - Classification et la Régression - Différences (Classification)

## Classification (partie 2) :

- Les données d'entrée sont des variables indépendantes et des variables dépendantes catégorielles.
- La tâche de l'algorithme de classification mappe la valeur d'entrée de  $x$  avec la variable de sortie discrète de  $y$ .
- La sortie est des étiquettes catégorielles.
- L'objectif est de prédire les étiquettes catégorielles/classes.
- Exemples de cas d'utilisation : détection de spam, reconnaissance d'images, analyse des sentiments
- Exemple d'algorithmes : Logistic Regression, Decision Trees, Random Forest, Support Vector Machines (SVM), K-Nearest Neighbors (K-NN), Naive Bayes, Neural Networks, K-Means Clustering, Multi-layer Perceptron (MLP), etc.

# Supervisé - Classification et la Régression - Différences (Régression)

## Régression (partie 1) :

- Dans cet énoncé de problème, les variables cibles sont continues.
- Des problèmes tels que la prévision des prix des maisons, les problèmes de prévision des précipitations sont résolus à l'aide d'algorithmes de régression.
- Dans cet algorithme, nous essayons de trouver la ligne la mieux adaptée qui peut représenter la tendance globale dans les données.
- Des métriques d'évaluation telles que l'erreur quadratique moyenne, le score  $R^2$  et le MAPE sont utilisées ici pour évaluer les performances des algorithmes de régression.
- Ici, nous sommes confrontés à des problèmes tels que les modèles de régression linéaire ainsi que les modèles non linéaires.



# Supervisé - Classification et la Régression - Différences (Régression)

Régression (partie 2) :

- Les données d'entrée sont des variables indépendantes et des variables dépendantes continues.
- La tâche de l'algorithme de régression consiste à mapper la valeur d'entrée ( $x$ ) avec la variable de sortie continue ( $y$ ).
- La sortie est des valeurs numériques continues.
- L'objectif est de prédire des valeurs numériques continues.
- Des exemples de cas d'utilisation sont la prévision du prix des actions, la prévision du prix de l'immobilier, la prévision de la demande.
- Exemple d'algorithmes : Linear Regression, Polynomial Regression, Ridge Regression, Lasso Regression, Support Vector Regression (SVR), Decision Trees for Regression, Random Forest Regression, K-Nearest Neighbors (K-NN) Regression, Neural Networks for Regression, etc.

# Non supervisé - Clustering

L'apprentissage non supervisé est un type d'algorithme d'apprentissage automatique où les algorithmes sont utilisés pour trouver les modèles, la structure ou la relation au sein d'un ensemble de données à l'aide d'un ensemble de données non étiquetées.

Il explore la structure inhérente des données sans catégories ni étiquettes prédéfinies.

# Non supervisé - Clustering

La tâche consistant à regrouper les points de données en fonction de leur similitude les uns avec les autres est appelée clustering ou analyse de cluster.

Cette méthode est définie dans la branche de l'apprentissage non supervisé, qui vise à obtenir des informations à partir de points de données non étiquetés, c'est-à-dire que, contrairement à l'apprentissage supervisé, nous n'avons pas de variable cible.

# Non supervisé - Clustering

Le clustering vise à former des groupes de points de données homogènes à partir d'un ensemble de données hétérogènes.

Il évalue la similarité en fonction d'une métrique telle que la distance euclidienne, la similarité cosinusoidienne, la distance de Manhattan, etc., puis regroupe les points avec le score de similarité le plus élevé.

# Non supervisé - Clustering

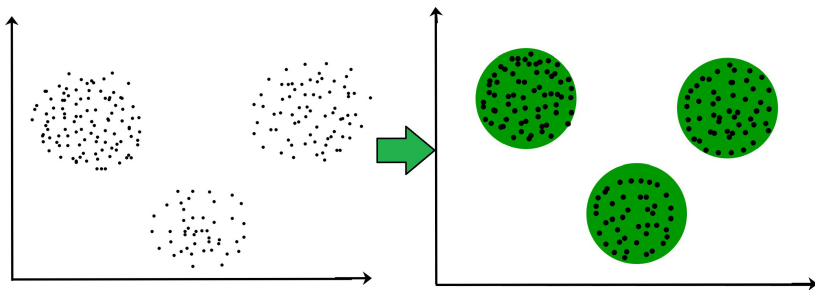


Figure: <https://www.geeksforgeeks.org/clustering-in-machine-learning/>

# Non supervisé - Dimensionality Reduction (PCA)

La réduction de la dimensionnalité est une technique utilisée pour réduire le nombre de caractéristiques dans un ensemble de données tout en conservant autant d'informations importantes que possible.

En d'autres termes, il s'agit d'un processus de transformation de données de grande dimension en un espace de dimension inférieure qui préserve toujours l'essence des données d'origine.

# Non supervisé - Dimensionality Reduction (PCA)

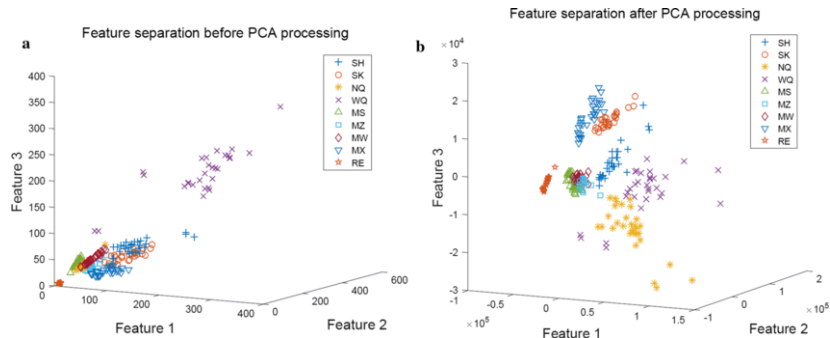


Figure: [https://www.researchgate.net/figure/Comparison-of-feature-separation-before-and-after-PCA-dimension-reduction\\_fig8-331794885](https://www.researchgate.net/figure/Comparison-of-feature-separation-before-and-after-PCA-dimension-reduction_fig8-331794885)

# Non supervisé - Dimensionality Reduction (PCA)

PCA sur MNIST pour réduire de  $8 \times 8 = 64$  dimensions (chaque pixel compte comme une dimension) à 2 dimensions

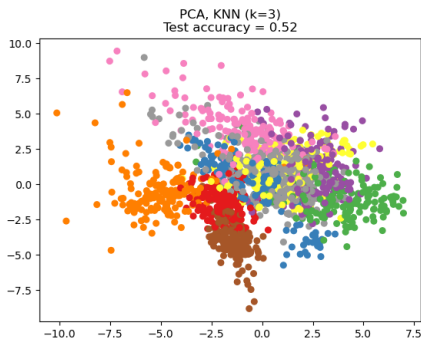


Figure: [https://scikit-learn.org/stable/auto\\_examples/neighbors/plot\\_nca\\_dim\\_reduction.html](https://scikit-learn.org/stable/auto_examples/neighbors/plot_nca_dim_reduction.html)



# Semi-supervisé

L'apprentissage semi-supervisé est un type d'apprentissage automatique qui se situe entre l'apprentissage supervisé et l'apprentissage non supervisé.

Il s'agit d'une méthode qui utilise une petite quantité de données étiquetées et une grande quantité de données non étiquetées pour entraîner un modèle.

# Semi-supervisé

L'objectif de l'apprentissage semi-supervisé est d'apprendre une fonction capable de prédire avec précision la variable de sortie en fonction des variables d'entrée, comme l'apprentissage supervisé.

Cependant, contrairement à l'apprentissage supervisé, l'algorithme est entraîné sur un ensemble de données qui contient à la fois des données étiquetées et non étiquetées.

## Semi-supervisé

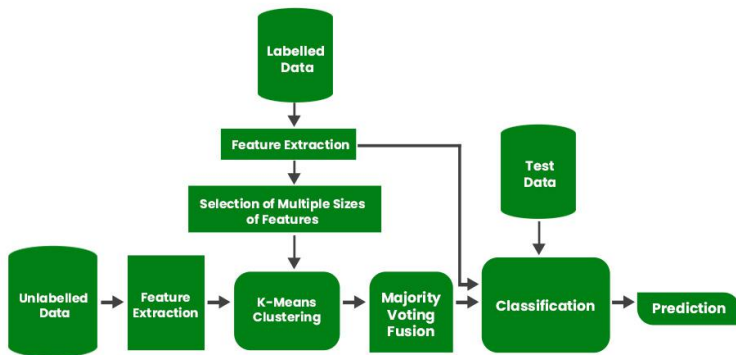


Figure: <https://www.geeksforgeeks.org/ml-semi-supervised-learning/>

# L'apprentissage par renforcement

L'apprentissage par renforcement est un type d'algorithmes d'apprentissage automatique où un agent apprend à prendre des décisions successives en interagissant avec son environnement.

L'agent reçoit le feedback sous forme d'incitations ou de sanctions en fonction de ses actions. Le but de l'agent est de découvrir des tactiques optimales qui maximisent les récompenses cumulatives au fil du temps par essais et erreurs.

# L'apprentissage par renforcement

L'apprentissage par renforcement est fréquemment utilisé dans des scénarios dans lesquels l'agent doit apprendre à naviguer dans un environnement, à jouer à des jeux, à gérer des robots ou à porter des jugements dans des situations incertaines.

# L'apprentissage par renforcement

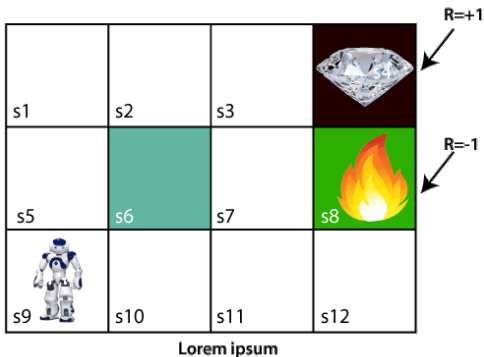


Figure: <https://www.javatpoint.com/reinforcement-learning>

# Bibliographie

- <https://www.geeksforgeeks.org/ml-classification-vs-regression/>