

<p><b>Les CONCEPTS, Les DEFINITIONS :</b></p> <p><b>MACHINE LEARNING</b></p>
--

**Les Concepts, les définitions :**

- **apprentissage automatique**

Technique consistant à entraîner des modèles informatiques à partir de données afin qu'ils puissent faire des prédictions ou prendre des décisions sans être explicitement programmés.

- **apprentissage supervisé**

Méthode d'apprentissage où les données sont étiquetées (avec des résultats connus) et le modèle apprend à prédire les sorties à partir des entrées.

- **apprentissage non supervisé**

Méthode où les données ne sont pas étiquetées. L'objectif est de découvrir des structures ou des regroupements cachés dans les données.

- **apprentissage semi supervisé**

Méthode combinant apprentissage supervisé et non supervisé, utilisant un petit ensemble de données étiquetées et un grand ensemble non étiqueté.

- **apprentissage par renforcement**

Méthode d'apprentissage basée sur un système de récompense: un agent apprend à interagir avec un environnement pour maximiser une récompense cumulative.

- **regression**

Technique utilisée pour prédire une valeur continue à partir de données d'entrée.

- **principaux algo de regression**

Régression linéaire, régression polynomiale, regression Ridge, Lasso, SVR (Support Vector Regression), arbres de décision, forêts aléatoires.

- **classification**

Tâche d'apprentissage supervisé consistant à attribuer une étiquette ou une classe à une entrée.

- **principaux algo de classification**

K-plus proches voisins, SVM, régression logistique, arbres de décision, forêts aléatoires, Naïve Bayes, réseaux de neurones.

- **algo qui font regression et classification**

SVM, arbres de décision, forêts aléatoires, réseaux de neurones, KNN.

- **partitionnement**

Technique d'apprentissage non supervisé visant à diviser un jeu de données en groupes (ou clusters) selon leur similarité.

- **apprentissage hors ligne**

Entraînement d'un modèle sur un ensemble de données fixe en une seule fois, avant toute utilisation.

- **apprentissage incremental**

Méthode où le modèle est mis à jour progressivement à mesure que de nouvelles données arrivent.

- **réseau de neurone**

Modèle inspiré du fonctionnement du cerveau humain, composé de couches de neurones artificiels connectés entre eux pour apprendre des relations complexes.

- **reseau de neurone artificiel**

Architecture de machine learning constituée de neurones organisés en couches (entrée, cachée, sortie) et utilisée pour modéliser des fonctions complexes.

- **algo parametrique**

Algorithme qui suppose une forme prédéfinie pour le modèle et apprend un nombre fini de paramètres (ex : régression linéaire).

- **algo non parametrique**

Algorithme ne faisant pas d'hypothèse forte sur la forme du modèle et pouvant s'adapter à plus de complexité (ex : KNN, arbres de décision).

- **difference modele linéaire et non linéaire**

Un modèle linéaire suppose une relation linéaire entre variables, tandis qu'un modèle non linéaire peut capturer des relations plus complexes (non proportionnelles).

- **bias / variance**

Le biais est l'erreur due à des suppositions trop simples du modèle. La variance est l'erreur causée par une trop grande sensibilité aux variations des données d'entraînement.

- **pourquoi réduire la dimension d'un jeu de données**

Pour améliorer la performance du modèle, réduire le surapprentissage, diminuer le temps de calcul et visualiser plus facilement les données.

- **reduction de dimension**

Processus consistant à projeter les données dans un espace de dimension plus faible tout en conservant l'information la plus significative (ex : PCA).

- **Régularisation**

Technique visant à pénaliser les modèles trop complexes afin de réduire le surapprentissage (ex : L1, L2).

- **les différents type de partitionnements**

K-means, clustering hiérarchique, DBSCAN, MeanShift, Affinity Propagation.