

Abonnez-vous à DeepL Pro pour éditer ce document.  
Visitez www.DeepL.com/Pro pour en savoir plus.

# Apprentissage automatique

Louis Fippo Fitime

12 novembre 2020

I L'ère de l'automatisation et des systèmes intelligents

I Un potentiel énorme dans tous les secteurs et toutes les entreprises I Les systèmes intelligents et les organisations axées sur les données deviennent une réalité.

I Les progrès des outils et des techniques ne font que l'aider à se développer davantage.

Les données sont le nouveau pétrole et l'apprentissage automatique est un concept et un cadre puissants pour en tirer le meilleur parti.

L'idée principale est de vous donner suffisamment d'informations sur les raisons pour lesquelles nous avons besoin de l'apprentissage automatique,

I Les blocs de construction fondamentaux de l'apprentissage automatique, I Ce que l'apprentissage automatique nous offre actuellement.

Cela vous permettra d'apprendre comment exploiter au mieux l'apprentissage automatique pour tirer le maximum de vos données.

Je comprends les définitions formelles,

I Concepts, fondements en ce qui concerne les algorithmes d'apprentissage,

I Gestion des données,

Construction, évaluation et déploiement de modèles.

Aspect pratique du cours I Cas d'utilisation spécifiques,

I Problème spécifique, cas concret.

La nécessité de l'apprentissage automatique

Pourquoi faire apprendre la machine ?

I L'absence d'une expertise humaine suffisante dans un domaine,

I Les scénarios et les comportements peuvent évoluer dans le temps, I Les humains ont une expertise suffisante dans le domaine mais il est extrêmement difficile d'expliquer ou de traduire formellement cette expertise en tâches informatiques,

I Résoudre des problèmes spécifiques à un domaine à l'échelle avec d'énormes volumes de données.

Paradigme de programmation traditionnel

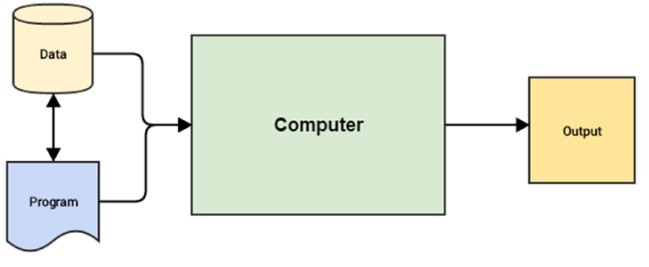


Figure 1 : Paradigme de la programmation traditionnelle

Pourquoi l'apprentissage automatique ?

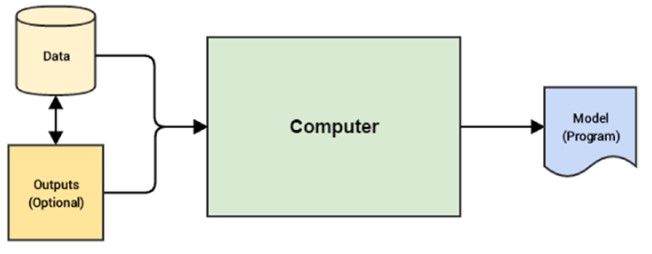


Figure 2 : Paradigme de programmation ML

Résoudre un problème d'apprentissage automatique

I Exploiter les données et les journaux des appareils,

I Décidez des attributs de données clés qui pourraient être utiles pour construire un modèle,

J'observe et je saisis les attributs des appareils et leur comportement sur diverses périodes de temps. Je soumets ces paires d'entrée et de sortie à un algorithme d'apprentissage automatique spécifique,

Je déploie ce modèle de manière à ce que, pour les nouvelles valeurs des attributs des appareils, il puisse prédire si un appareil spécifique se comporte normalement ou s'il risque de provoquer une panne potentielle.

Définition générale

La nécessité de l'apprentissage automatique

I Prendre des décisions fondées sur des données

I Efficacité et échelle

L'apprentissage automatique est le domaine d'étude qui donne aux ordinateurs la capacité d'apprendre sans être explicitement programmés. *Arthur*

*Samuel, 1959*

Définition formelle

*On dit d'un programme informatique qu'il apprend de l'expérience E en ce qui concerne une certaine classe de tâches T et une mesure de performance P, si sa performance aux tâches de T, telle que mesurée par P, s'améliore avec l'expérience E.*

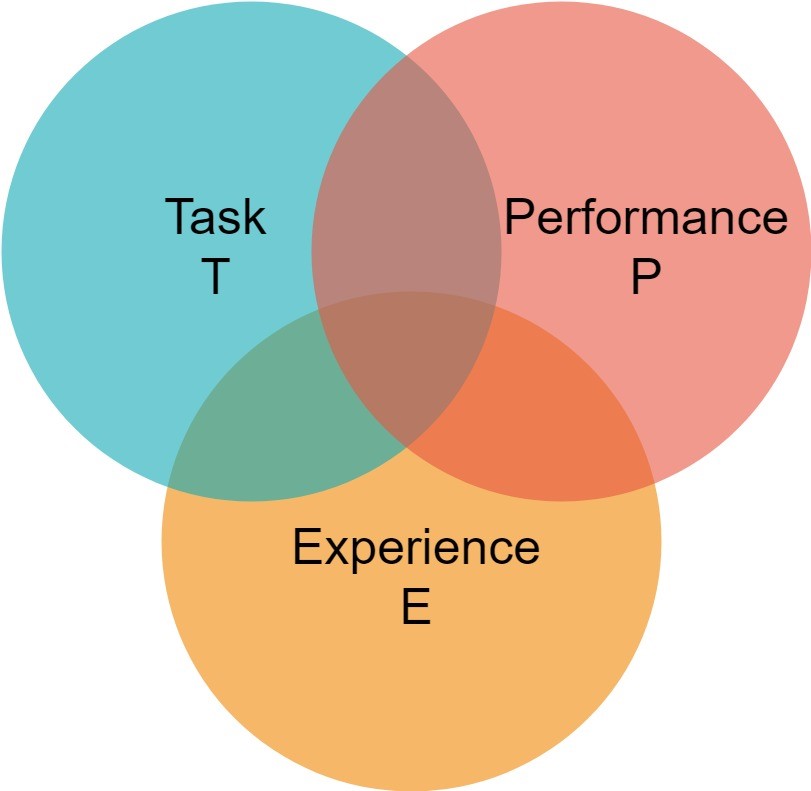


Figure 3 : Diagramme ML Ven

# Tâche, expérience et performance

Une tâche*,* T*, peut généralement être définie comme une tâche d'apprentissage automatique basée sur le processus ou le flux de travail que le système doit suivre pour opérer sur des points de données ou des échantillons.*

I Classification ou catégorisation

I Régression

I Détection d'anomalies

I Annotation structurée

I Traduction

I Regroupement ou groupement

I Transcriptions

# Tâche, expérience et performance

Le processus de consommation d'un ensemble de données constitué d'échantillons de données ou de points de données de telle sorte qu'un algorithme d'apprentissage ou un modèle apprenne des modèles inhérents est défini comme l'expérience, E, acquise par l'algorithme d'apprentissage.

La performance, P, est généralement une mesure quantitative ou une métrique utilisée pour voir dans quelle mesure l'algorithme ou le modèle exécute la tâche, T, avec l'expérience, E.

# L'apprentissage automatique : un véritable domaine pluridisciplinaire

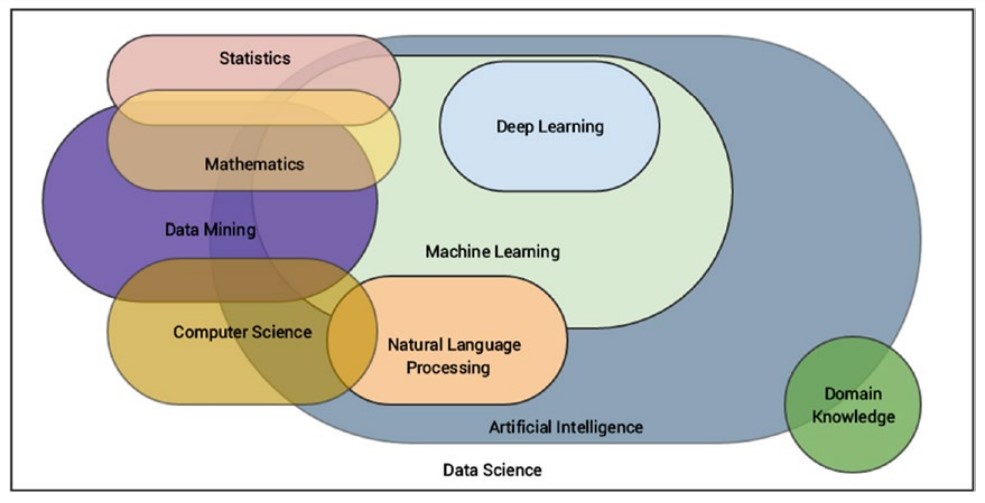


Figure 4 : Diagramme Discipline Ven

# (Intelligence artificielle)

L'art, la science et l'ingénierie de la création d'agents, de machines et de programmes intelligents.

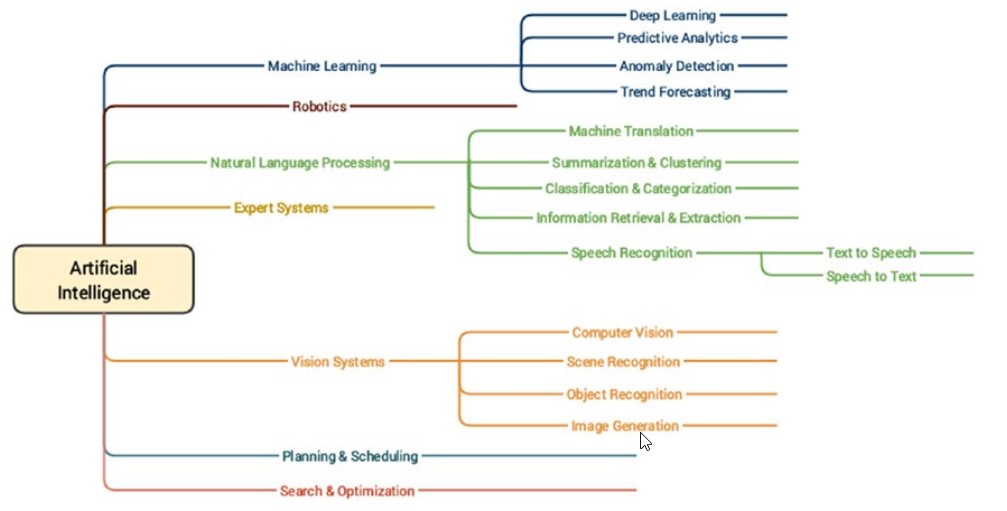


Figure 5 : Diverses facettes majeures sous l'égide de l'IA

Apprentissage profond

L'approche basée sur l'apprentissage profond tente de construire l'intelligence de la machine en représentant les données comme une hiérarchie de concepts en couches, où chaque couche de concepts est construite à partir d'autres couches plus simples.

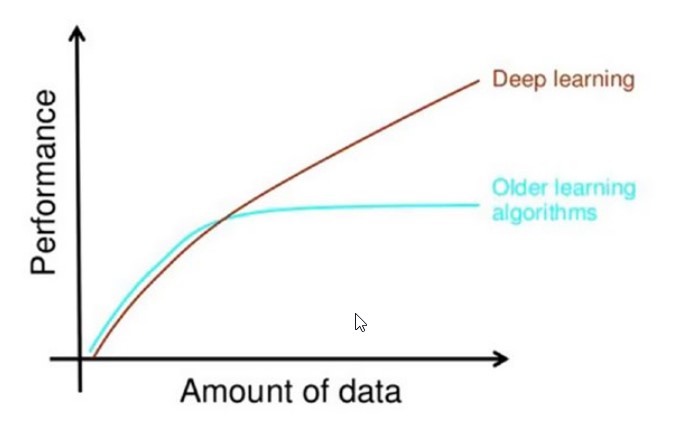


Figure 6 : Comparaison des performances de l'apprentissage profond et de l'apprentissage traditionnel.

Apprentissage automatique par Andrew Ng

Concepts importants de DL

I Réseaux neuronaux artificiels

I Backpropagation

I Perceptrons multicouches

I Réseaux neuronaux convolutifs

I Réseaux neuronaux récurrents

I Réseaux de mémoire à long terme

I Auto-codeurs

Illustration des réseaux neuronaux artificiels

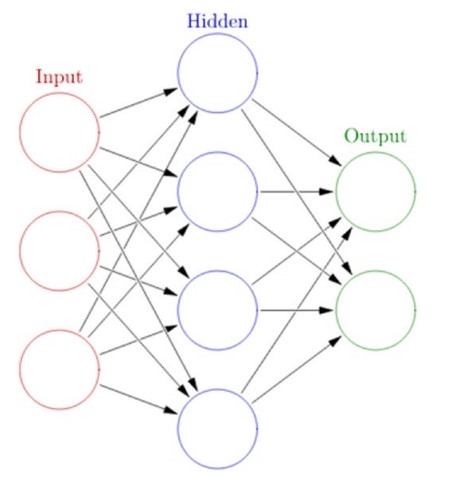


Figure 7 : Un réseau neuronal artificiel typique

# Méthodes d'apprentissage automatique

1. Méthodes basées sur le degré de supervision humaine dans le processus d'apprentissage I Apprentissage supervisé I Apprentissage non supervisé
   1. Apprentissage semi-supervisé

I Apprentissage par renforcement

1. Méthodes basées sur la capacité d'apprendre à partir d'échantillons de données incrémentiels I Apprentissage par lots I Apprentissage en ligne
2. Méthodes basées sur leur approche de la généralisation à partir d'échantillons de données
   1. Apprentissage basé sur les instances I Apprentissage basé sur les modèles

# Apprentissage supervisé

Les méthodes ou algorithmes d'apprentissage supervisé comprennent des algorithmes d'apprentissage qui prennent des échantillons de données (appelés données de formation) et des sorties associées (appelées étiquettes ou réponses) à chaque échantillon de données pendant le processus de formation du modèle. L'objectif principal est de

apprendre une correspondance ou une association entre les échantillons de données d'entrée x et leurs sorties correspondantes y, sur la base de multiples instances de données d'apprentissage. Principales méthodes d'apprentissage supervisé :

I Classification

I Régression

# Apprentissage supervisé : Classification

prédire des étiquettes de sortie, des classes ou des réponses de nature catégorielle pour les données d'entrée, en fonction de ce que le modèle a appris pendant la phase d'apprentissage.

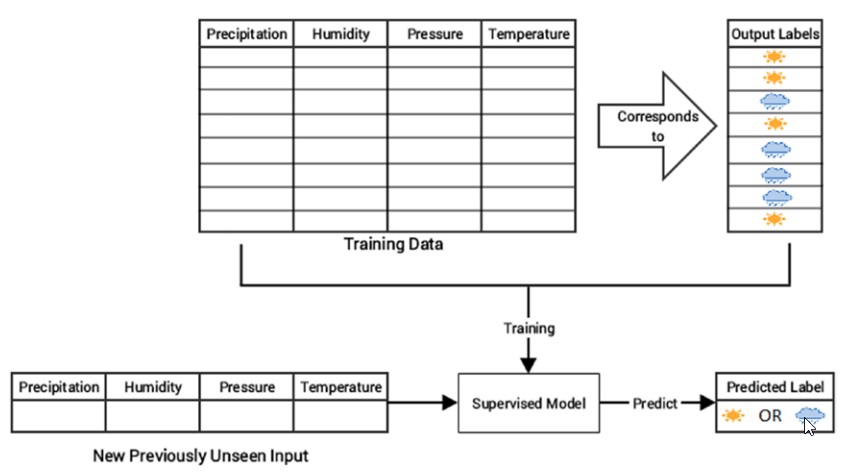


Figure 8 : Illustration de la classification

# Apprentissage supervisé : Régression

Les tâches d'apprentissage automatique dont l'objectif principal est l'estimation de la valeur peuvent être qualifiées de *tâches de régression*.

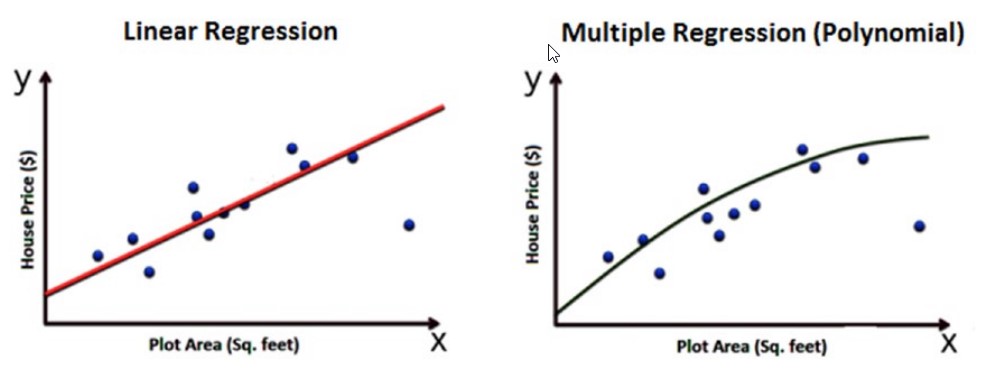


Figure 9 : Apprentissage supervisé : modèles de régression pour la prédiction du prix des maisons

# Apprentissage non supervisé

le modèle ou l'algorithme tente d'apprendre les structures latentes inhérentes, les modèles et les relations à partir de données données données sans aucune aide ou supervision. Les méthodes d'apprentissage non supervisé peuvent être classées dans les grands domaines de tâches ML suivants

I Clustering

I Réduction de la dimensionnalité

I Détection d'anomalie

I L'exploration des règles d'association

# Apprentissage non supervisé : clustering

Les méthodes de regroupement sont des méthodes d'apprentissage automatique qui tentent de trouver des modèles de similarité et de relations entre les échantillons de données dans notre ensemble de données, puis de regrouper ces échantillons en divers groupes, de sorte que chaque groupe ou groupe d'échantillons de données présente une certaine similarité, sur la base des attributs ou caractéristiques inhérents.

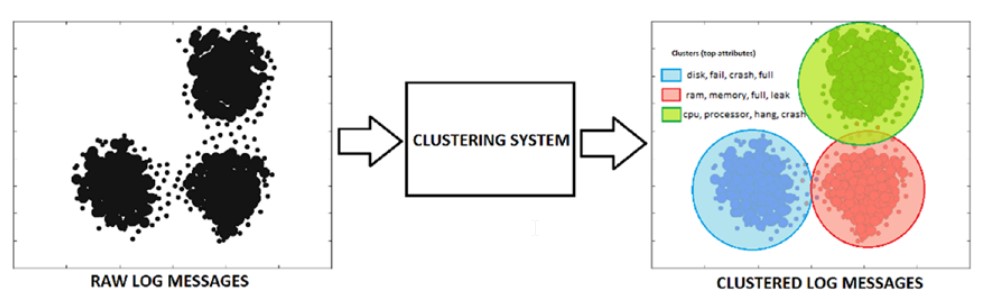


Figure 10 : Apprentissage non supervisé : regroupement de messages de journal

# Apprentissage non supervisé : Réduction de la dimensionnalité

Ces méthodes réduisent le nombre de variables caractéristiques en extrayant ou en sélectionnant un ensemble de caractéristiques principales ou représentatives.

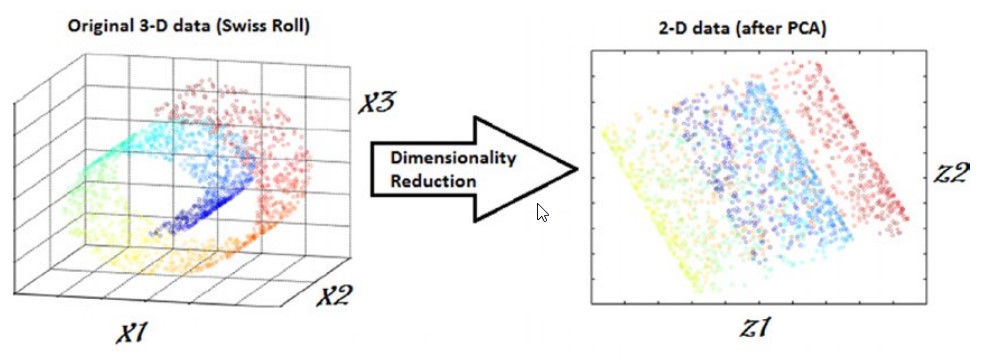


Figure 11 : Apprentissage non supervisé : réduction de la dimensionnalité

# Apprentissage non supervisé : Détection d'anomalies

Nous sommes intéressés par la découverte d'occurrences d'événements rares ou d'observations qui ne se produisent pas normalement sur la base d'échantillons de données historiques.

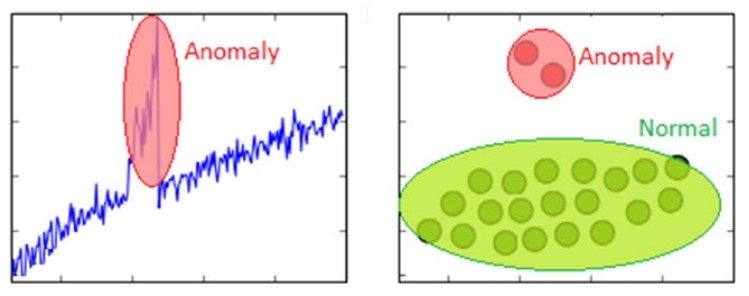


Figure 12 : Apprentissage non supervisé : détection d'anomalies

# Apprentissage non supervisé : Usinage de règles d'association

En règle générale, l'extraction de règles d'association est une méthode d'extraction de données utilisée pour examiner et analyser de grands ensembles de données transactionnelles afin de trouver des modèles et des règles d'intérêt.

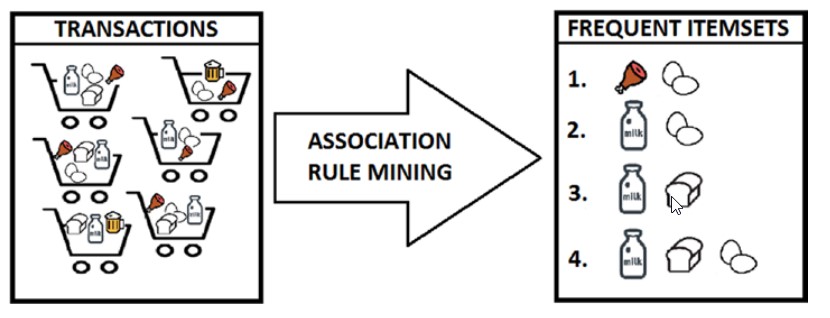


Figure 13 : Apprentissage non supervisé : exploration des règles d'association

# Apprentissage semi-supervisé : Apprentissage par renforcement

Nous avons un agent que nous voulons entraîner sur une période donnée à interagir avec un environnement spécifique et à améliorer ses performances sur une période donnée en ce qui concerne le type d'actions qu'il effectue sur l'environnement.

1. Préparer l'agent avec un ensemble de politiques et de stratégies initiales
2. Observer l'environnement et l'état actuel
3. Sélectionner la politique optimale et effectuer l'action
4. Obtenir la récompense (ou la pénalité) correspondante
5. Mettre à jour les politiques si nécessaire
6. Répétez les étapes 2 à 5 de manière itérative jusqu'à ce que l'agent apprenne les politiques les plus optimales.

# Apprentissage semi-supervisé : Apprentissage par renforcement

Nous avons un agent que nous voulons entraîner sur une période donnée à interagir avec un environnement spécifique et à améliorer ses performances sur une période donnée en ce qui concerne le type d'actions qu'il effectue sur l'environnement.

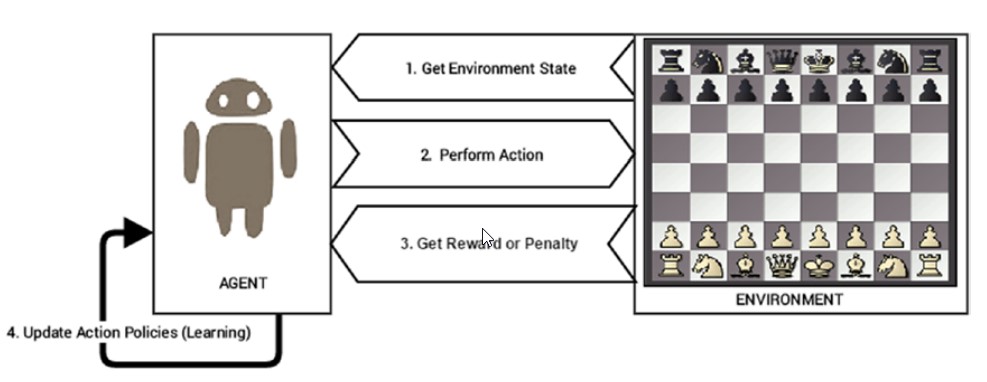


Figure 14 : Apprentissage par renforcement : former un robot à jouer aux échecs

# Aperçu de l'apprentissage automatique

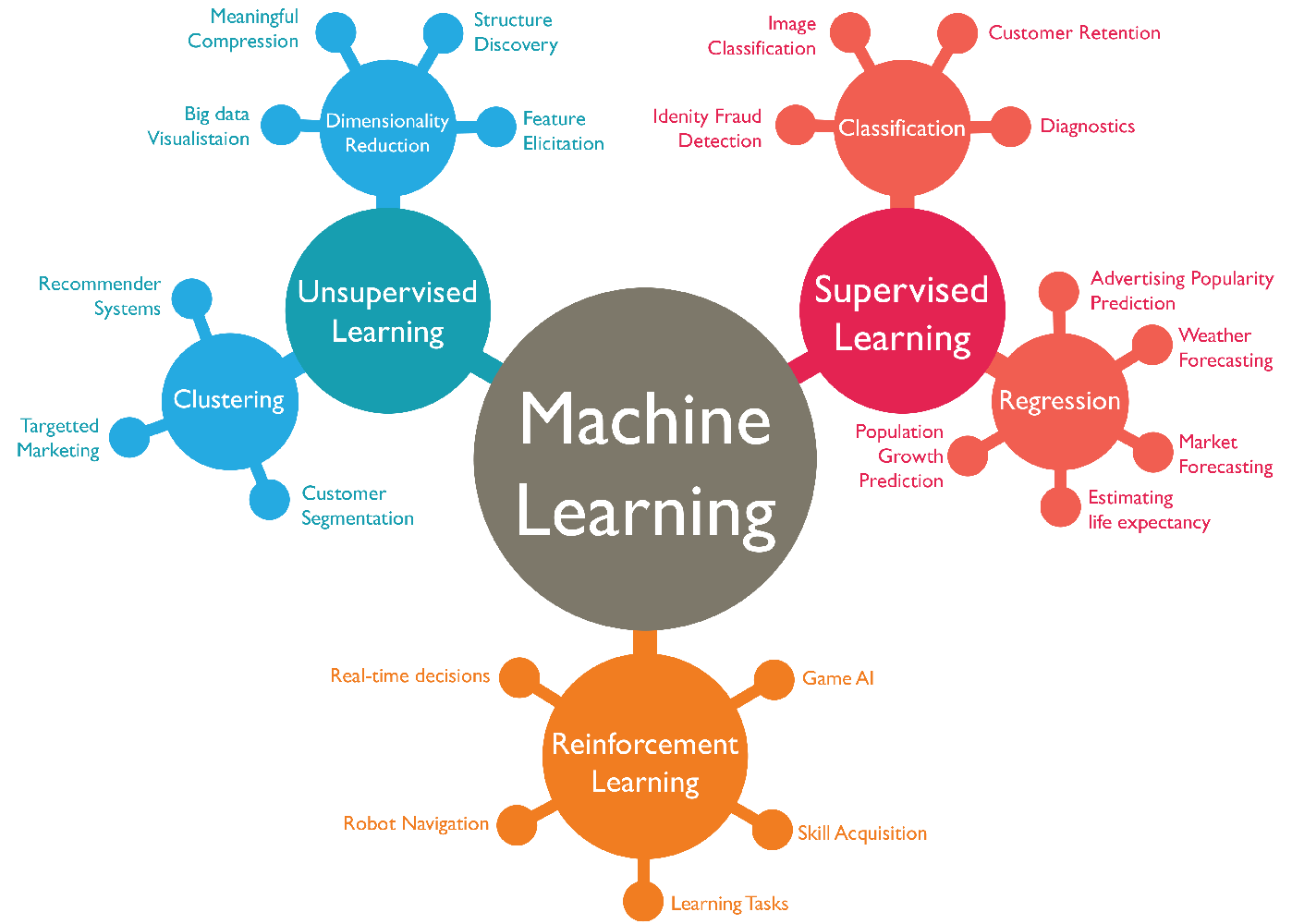


Figure 15 : Aperçu de l'apprentissage automatique

# Pipeline d'apprentissage automatique

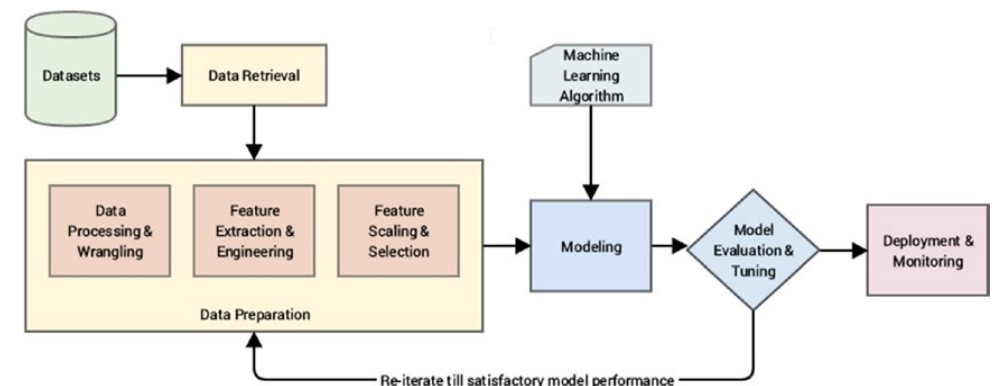


Figure 16 : pipeline d'apprentissage automatique

# Défis de l'apprentissage automatique (1)

L'acquisition, l'extraction et la récupération des données est un processus extrêmement fastidieux et long,

I Le manque de données d'entraînement de bonne qualité et suffisantes dans de nombreux scénarios,

Formuler clairement les problèmes de l'entreprise avec des buts et des objectifs bien définis,

# Défis de l'apprentissage automatique (2)

I Extraction et ingénierie des caractéristiques, en particulier l'élaboration manuelle des caractéristiques,

I Modèles surajustés ou sous-ajustés,

I La malédiction de la dimensionnalité : trop de caractéristiques peuvent être un véritable obstacle,

I Les modèles complexes peuvent être difficiles à déployer dans le monde réel.

Je recommande des produits sur les plateformes d'achat en ligne,

I Analyse des sentiments et des émotions,

I Détection d'anomalie,

I Détection et prévention des fraudes,

Recommandation de contenu (actualités, musique, films, etc.)

I Prévisions météorologiques

I Prévisions boursières

I Analyse du panier de la ménagère I Segmentation de la clientèle

Reconnaissance des objets et des scènes dans les images et les vidéos

I Reconnaissance vocale

I Analyse du taux de désabonnement

I Prédiction des clics I Détection et prévention des défaillances/défauts

Filtrage du courrier électronique non sollicité

Définition

Python est un langage interprété, ce qui signifie que le code source d'un programme Python est converti en bytecode, qui est ensuite exécuté par la machine virtuelle Python.

I Le code Python est rapide à développer.

I Le code Python n'est pas aussi rapide en exécution.

Points forts

I Facile à apprendre

Prise en charge de plusieurs paradigmes de programmation

I Extensible

I Communauté open source active

Configuration d'un environnement Python

J'installe Python et les bibliothèques nécessaires individuellement J'utilise une distribution Python pré-packagée qui est fournie avec les bibliothèques nécessaires, par exemple Anaconda

Configuration d'Anaconda Env

I Installation des paquets

Jupyter notebook

Je numpy

I Pandas

I Scikit-learn

I Theano

I Tensorflow

I Keras

I NLTK (The Natural Language Tool Kit)

# Ingénierie des données

Intro

I Collecte de données : Comprendre les différents mécanismes de récupération des données pour différents types de données.

I Description des données : Comprendre les différents attributs et propriétés des données collectées.

I Le traitement des données : Préparer les données en vue de leur utilisation dans les étapes de modélisation.

I Visualisation des données : Visualiser différents attributs pour partager les résultats, mieux comprendre, etc.

Intro

## Problème

Vous voulez former un modèle qui représente une relation linéaire entre la caractéristique et le vecteur cible.

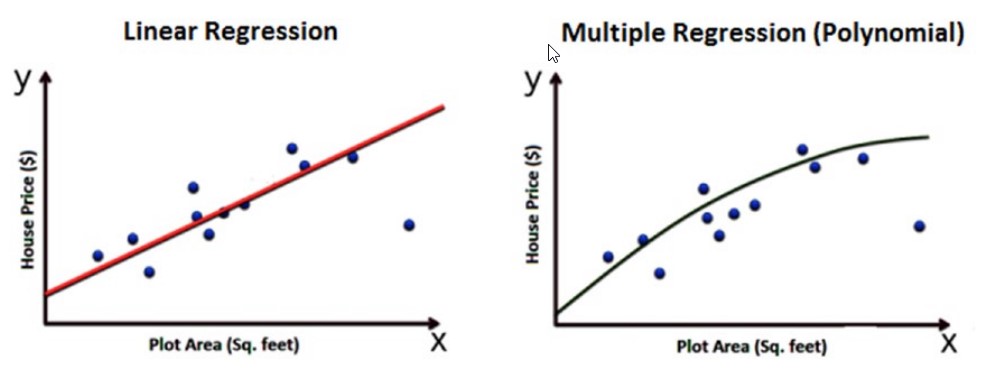


Figure 17 : Apprentissage supervisé : modèles de régression pour la prédiction du prix des maisons

Couvre

I Visualisation des données

I Implémentation d'une régression linéaire simple utilisant des équations normales (en utilisant Numpy)

I Optimisation par descente de gradient (en utilisant numpy)

Mise en œuvre de R-squared à l'aide de numpy

I Analyse du graphique résiduel

I Comparaison du modèle utilisant l'équation normale avec l'implémentation de scikit

Les étapes de base

J'importe les paquets et les classes dont vous avez besoin.

Je fournis des données pour travailler avec et éventuellement faire les transformations appropriées.

Je crée un modèle de régression et je l'adapte aux données existantes. Je vérifie les résultats de l'ajustement du modèle pour savoir si le modèle est satisfaisant.

I Appliquer le modèle pour les prédictions.

Couvre

I Récapitulation des modèles linéaires et discussion des méthodes d'optimisation stochastique

I Compréhension de base des réseaux neuronaux modernes et de leurs applications en matière de vision par ordinateur et de compréhension du langage naturel,

Je comprends les principales tendances technologiques qui sous-tendent l'apprentissage profond. Je suis capable de construire, d'entraîner et d'appliquer des réseaux neuronaux profonds entièrement connectés,

Je sais comment mettre en œuvre des réseaux neuronaux efficaces (vectorisés), Je comprends les paramètres clés de l'architecture d'un réseau neuronal. Ce cours vous apprend également comment fonctionne réellement l'apprentissage profond, au lieu de ne présenter qu'une description superficielle.

Exemple d'introduction

Recommandation visuelle de produits

I Acheter de nouveaux articles (chaussures, vêtements,...)

Je cherche un animal (chien, chat, ...)

### Problème

J'ai trop d'options en ligne...

I Les mots-clés ne sont pas très utiles

Les fonctionnalités sont la clé de l'apprentissage automatique

I Comment pouvons-nous représenter les caractéristiques ?

I Pb : apprentissage de caractéristiques très non linéaires

Classificateurs linéaires

Que peut représenter un classificateur linéaire ?

I OR

I ET

I Comment pouvons-nous représenter les caractéristiques ?

I Pb : apprentissage de caractéristiques très non linéaires

Classificateurs linéaires

Ce qu'un classificateur linéaire ne peut pas représenter ?

I XOR

I Comment pouvons-nous représenter les caractéristiques ?

I Pb : apprentissage de caractéristiques très non linéaires

Extraction de caractéristiques à partir d'une image I Comment représenter les caractéristiques ?

I Pb : apprentissage de caractéristiques très non linéaires



Figure 18 : extraction de caractéristiques