

Introduction au Machine Learning

■ Définition

- **Théorique**

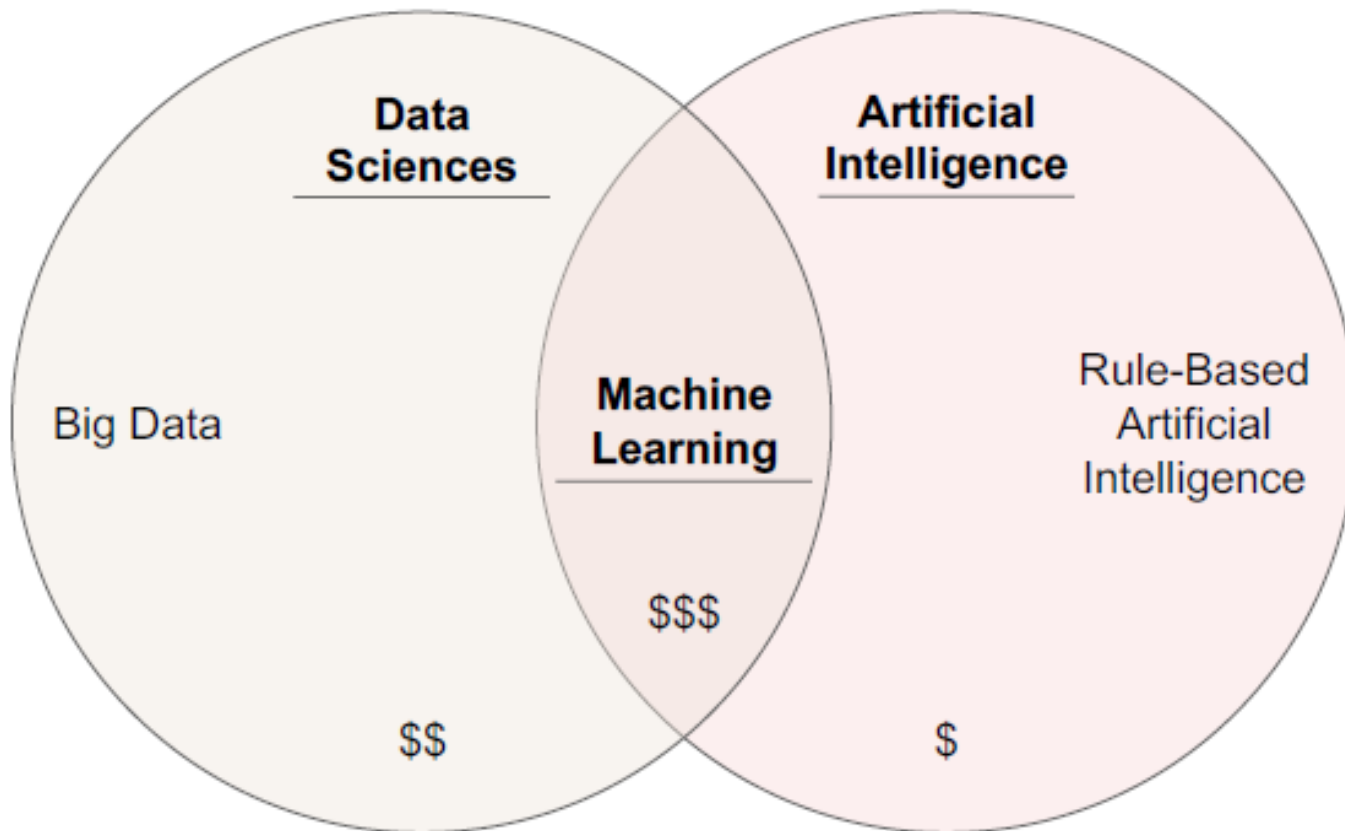
- ✓ L'apprentissage automatique peut-être vu comme l'ensemble des techniques permettant à une machine d'apprendre à réaliser une tâche sans avoir à la programmer explicitement pour cela. Arthur Samuel

- **En d'autres termes**

- ✓ « technique permettant à la machine de résoudre des problèmes pour les humains. »
- ✓ « Donner la possibilité à l'ordinateur d'apprendre sans être programmé »

Introduction au Machine Learning

■ Où se situe le Machine Learning



Introduction au Machine Learning

■ Types de problèmes en Machine Learning

- 4 types de problèmes en ML
 - ✓ Supervised Learning (Apprentissage Supervisé)
 - Répond à des problématiques de classification ou régression
 - Idée : associer un label à des données dont on possède des mesures
 - Si les labels sont discrets (libellés ou valeurs finies) alors on parlera de classification
 - Au contraire si labels sont continus (ensemble de nombres réels) on parlera de régression
 - Classification
 - Ensemble de données connues déjà classées (photos, plantes, individus,...)
 - Souhaiter à partir de cette 1^{ère} classification, dite connaissance, classer de nouveaux éléments

Introduction au Machine Learning

■ Types de problèmes en Machine Learning

- 4 types de problèmes en ML
 - ✓ Supervised Learning (Apprentissage Supervisé)
 - Régression
 - fournir aux algorithmes d'apprentissages un jeu de données d'apprentissage (Training Set) sous forme de (\mathbf{X}, \mathbf{Y})
 - avec \mathbf{X} les variables prédictives,
 - et \mathbf{Y} le résultat de l'observation.
 - En se basant sur le *Training Set*, l'algorithme va trouver **une fonction mathématique** qui permet de transformer (au mieux) \mathbf{X} vers \mathbf{Y}
 - Exemple : Prédire la valeur d'une maison en fonction de sa superficie et de sa localisation

Introduction au Machine Learning

■ Types de problèmes en Machine Learning

- 4 types de problèmes en ML
 - ✓ Unsupervised Learning (Apprentissage Non supervisé)
 - Répond à un besoin de classification de données, mais on dispose pas de données déjà classées, on parle de **clustering**
 - Clustering
 - Fournir à l'algorithme des données, éventuellement **non structurées** et lui demander de les répartir en N groupes
 - Exemple : En disposant des données d'achats des internautes dans un site e-commerce, l'algorithme de clustering va trouver les produits qui se vendent le mieux ensemble.

Introduction au Machine Learning

■ Types de problèmes en Machine Learning

- 4 types de problèmes en ML
 - ✓ Apprentissage par renforcement
 - s'applique plus à des problèmes d'optimisations.
 - idée : faire prendre des décisions à un système pour obtenir un résultat qui soit le meilleur possible.
 - ✓ Apprentissage par transfert
 - L'apprentissage par transfert vise à utiliser les connaissances d'un jeu de tâches sources pour non seulement influencer l'apprentissage mais aussi améliorer les performances sur une autre tâche cible.

Introduction au Machine Learning

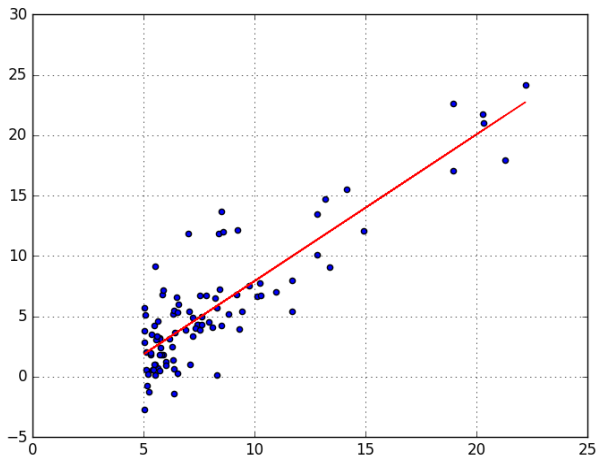
■ Différents types d'algorithmes en ML

- Chaque algorithme a ses spécificités mathématiques et algorithmiques
 - ✓ Régression linéaire (Linear regression)
 - Les algorithmes de régression linéaire modélisent **la relation entre des variables prédictives et une variable cible**.
 - La relation est modélisée par une fonction mathématique de prédiction. Le cas le plus simple est la régression linéaire univariée. Elle va trouver une **fonction sous forme de droite** pour estimer la relation.
 - La régression linéaire multivariée intervient quand **plusieurs variables explicatives** interviennent dans la fonction de prédiction.
 - Et finalement, la régression polynomiale permet de modéliser des relations complexes qui ne sont pas forcément linéaires.

Introduction au Machine Learning

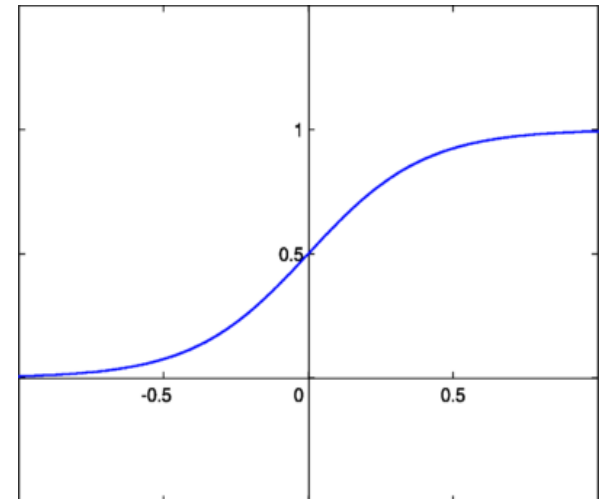
■ Différents types d'algorithmes en ML

- ✓ Régression logistique (Logistic regression)
 - Méthode statistique pour effectuer des classifications binaires.
 - Elle prend en entrée des variables prédictives qualitatives et/ou ordinales et mesure la probabilité de la valeur de sortie en utilisant la fonction sigmoïd. Répond aux question du genre : est-ce que le client est solvable pour lui accorder un crédit?



← linéaire

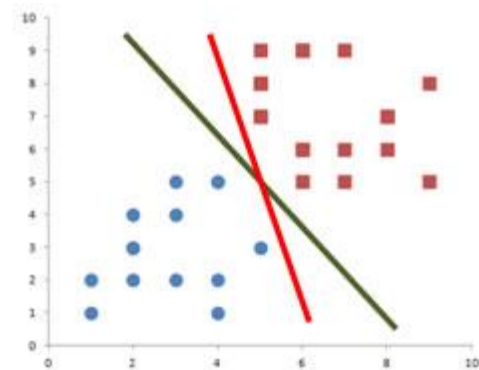
logistique →



Introduction au Machine Learning

■ Différents types d'algorithmes en ML

- ✓ Support Vector Machine (SVM)
 - Algorithme de classification binaire tout comme la régression logistique.
 - Exemple de l'image ci-dessus, on a 2 deux classes (mails Spam en rouge et les non spam sont en bleu). L
 - a régression Logistique pourra séparer ces deux classes en définissant le trait en rouge. le SVM va opter à séparer les deux classes par le trait vert.



Introduction au Machine Learning

■ Différents types d'algorithmes en ML

✓ Naïve Bayes

- basé sur le **théorème de Bayes des probabilités conditionnelles**.
formule du théorème de Bayes : $P(A|B) = P(B|A)P(A)/P(B)$
- Naïve Bayes assume une **hypothèse forte** (naïve) : on suppose que **les variables sont indépendantes entre elles**. Cela permet de simplifier le calcul des probabilités.
- Généralement, le Naïve Bayes est utilisé pour **les classifications de texte** (en se basant sur le nombre d'occurrences de mots).

✓ Anomaly detection

- algorithme de Machine Learning pour détecter des patterns anormaux.
- très utile pour la détection de fraudes dans les transactions bancaires, et les détections d'intrusions

Introduction au Machine Learning

■ Différents types d'algorithmes en ML

✓ Decision Trees

- Algorithme qui se basant sur **un modèle de graphe (les arbres)** pour définir la décision finale.
 - Chaque nœud comporte une condition, et les branchements sont en fonction de cette condition (Vrai ou Faux).
 - Plus on descend dans l'arbre, plus on cumule les conditions.

✓ K-Means

- K-Means est un algorithme de clustering en Unsupervised Learning.
- On donne un ensemble d'éléments (des données), et un nombre de groupes K.
- K-means va segmenter en K groupes les éléments. Le groupement s'effectue en minimisant la distance euclidienne entre le centre du cluster et un élément donné

Introduction au Machine Learning

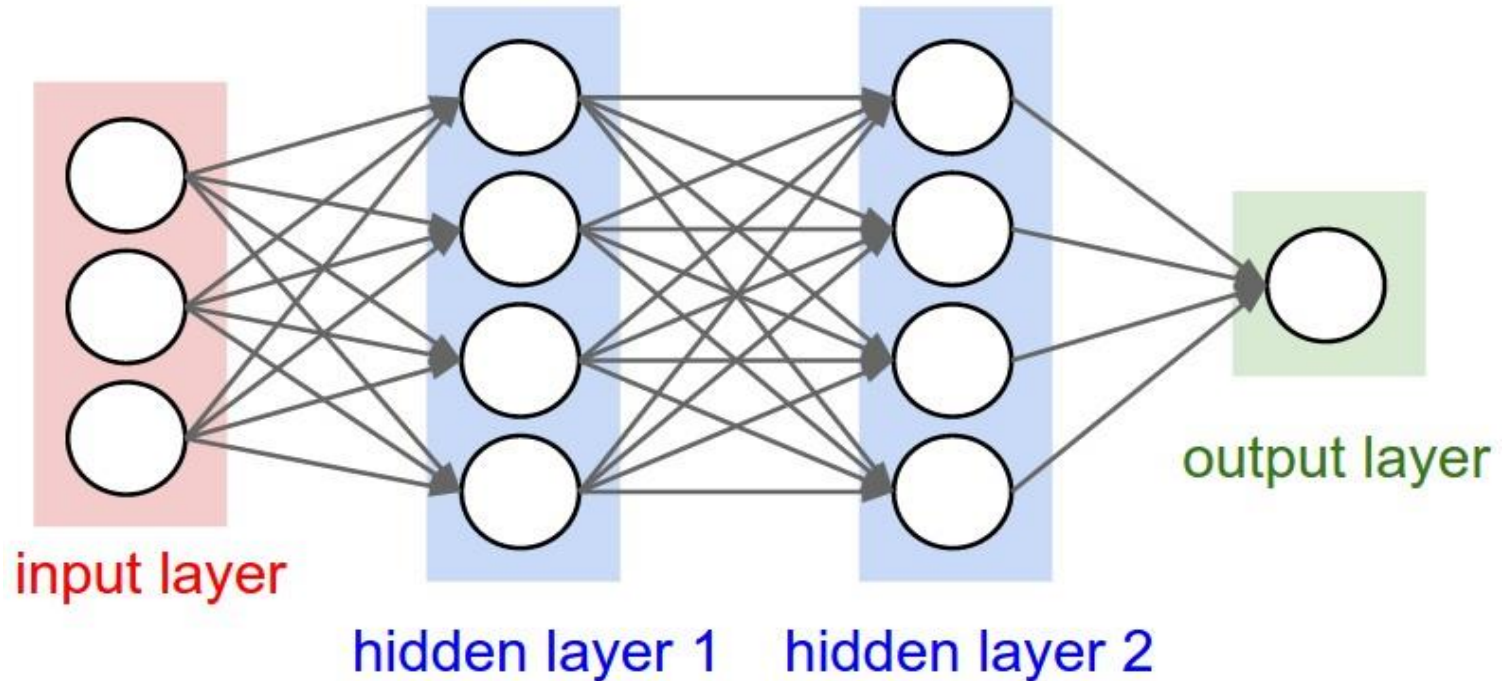
■ Différents types d'algorithmes en ML

- ✓ **Neurals Networks (Réseaux de neurones)**
 - Inspirés des neurones du système nerveux humains.
 - Permettent de trouver des patterns complexes dans les données.
 - Apprennent tâche spécifique en fonction des données d'entraînement.
 - se composent de nœuds (les cercles dans l'image). Dans ces réseaux,
 - Le tiers d'entrée (Input Layer) va recevoir les données d'entrées.
 - L'Input Layer va propager les données par la suite aux tiers cachés (Hidden Layers).
 - Finalement le Tiers de sortie (le plus à droite) permet de produire le résultat de classification.
 - Chaque tiers du réseau de neurones est un ensemble d'interconnexions des nœuds d'un tiers avec ceux des autres tiers.

Introduction au Machine Learning

■ Différents types d'algorithmes en ML

- ✓ Neurons Networks (Réseaux de neurones)

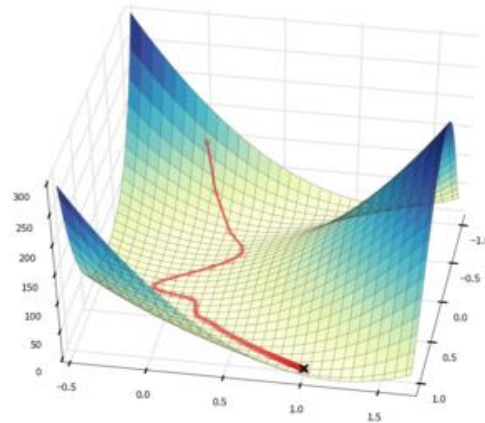


Introduction au Machine Learning

■ Différents types d'algorithmes en ML

✓ Gradient Descent

- Pas vraiment un algorithme de Machine Learning, mais très important
- Algorithme itératif de minimisation de fonction de coût.
- Minimisation servira à produire des modèles prédictifs comme la régression logistique et la régression linéaire



Introduction au Machine Learning

■ Aborder problèmes de Machine Learning

- Processus en 5 étapes

1. Définition du problème

- Que cherche-t-on à résoudre ?

- Décrire le problème de façon informelle

- Donner une description formelle

- Pourquoi c'est utile de le résoudre ?

- Donner les motivations pour résoudre le problème

2. Préparation des données

- Recenser les types de données utiles

- Définir la difficulté pour les récupérer

- Sources de données et difficultés rencontrées

- Définir les traitements préparatoires de ces données

Introduction au Machine Learning

■ Aborder problèmes de Machine Learning

- Processus en 5 étapes
 1. Définition du problème
 2. Préparation des données
 3. Choix des bons algorithmes
 - Faire le bon choix d'algorithme en fonction du nombre de caractéristiques, la quantité de données, ...
 4. Entraînement du modèle
 5. Évaluation du modèle



Introduction au Machine Learning

■ Aborder problèmes de Machine Learning

- Processus en 5 étapes
 4. Optimisation des résultats
 - Après données d'entraînement(Training set) et prédictions avec jeu de test (Test set) – évaluation de la performance de l'algorithme
 - Sur-apprentissage (overfitting) – pas de bonnes prédictions sur situations nouvelles
 - Sous-apprentissage (underfitting) – prédiction sans rapport avec la réalité
 - Affiner modèle de prédiction pour qu'il puisse se généraliser

Introduction au Machine Learning

■ Aborder problèmes de Machine Learning

- Processus en 5 étapes

5. Présentations des résultats finaux

- **Le contexte** : Soulever le contexte de la problématique et les motivations de sa résolution
- **Le problème** : Décrire de façon concise le problème qu'on cherche à résoudre
- **La solution** : Décrire la solution apportée en terme d'architecture, moyen d'exploitation de la solution etc...
- **Limitations** :
 - Dire si la solution n'est pas universelle et en lister les limitations
- **Conclusion** : Revisiter d'une façon rapide la description du problème ainsi que de la solution et les bénéfices tirées de cette dernière.