

Apprentissage Statistique

« Machine Learning »

ENSAH – 2^{ème} année Génie Informatique
– 2020/2021–

Pr. Routaib Hayat
hroutaib@uae.ac.ma



Téléchargement des vidéos Explicatives

1. Connectez-vous sur le site :

<https://my.pcloud.com/#page=login>

2. Login : gi2.enssah@gmail.com

3. Mot de passe : [azerty123@ENSAH](#)



Plan

- **Introduction**
- **La modélisation du Machine Learning**
- **Éléments de statistiques**
- **Le Réseaux Bayésien**
- **La chaine du Markov**
- **Le Modèle Markovien Caché**
- **La méthode de Monte-Carlo**
- **Modèles Statiques & Modèles Dynamiques**
- **Éléments de Théorie D'Apprentissage Statistique**
- **La Reconnaissance des Formes**

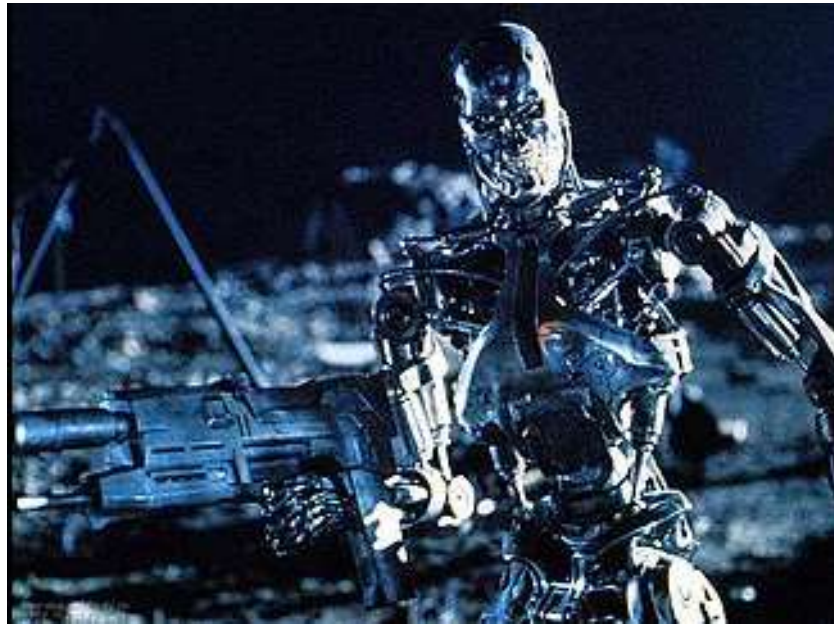


Machine Learning

- ✓ C'est Quoi ?
- ✓ Pourquoi ?
- ✓ Comment ?

Machine Learning

- Lorsqu'on entend parler de **machine learning** – ou plus généralement de **l'intelligence artificielle**, dont lequel le **machine learning** est un sous-domaine – on pense généralement à ça :



- Mais les experts du domaine sont formels « malgré toutes les inquiétudes évoquées dans les médias », ***le machine learning, bien évidemment l'intelligence artificielle, ne constituent pas une réelle menace.***

⇒ *En réalité, on est vraiment loin d'avoir atteint un niveau élevé d'intelligence chez les machines pour avoir de quoi s'inquiéter.*



Introduction

- Une des **tâches** essentielles **du cerveau** consiste à **transformer des informations pour faire connaissance** afin de :

1. Identifier les lettres qui constituent un texte.
2. Les assembler en mots et en phrases, pour extraire un sens.

⇒ **Ceux-ci sont des activités qui nous paraissent naturelles une fois que l'apprentissage nécessaire est accompli avec succès.**

- L'objectif de **l'apprentissage statistique** en anglais « **Machine Learning** » est **d'imiter**, à l'aide des algorithmes exécutés par des ordinateurs, la capacité que les êtres vivants ont à *apprendre par l'exemple*.



Introduction

A. Exemple réel d'apprentissage:

Pour apprendre à un enfant la lecture des lettres ou des chiffres, on lui présente **des exemples** de ceux-ci, écrits dans des styles et avec des polices différents. On ne fournit généralement pas à l'enfant une description analytique et discursive de la forme et de la topologie des caractères => **on se contente de lui montrer des exemples.**

À la fin de l'apprentissage : on attend de l'enfant qu'il soit capable de lire non seulement tous les chiffres et lettres qui lui ont été présentés durant son apprentissage, mais aussi tous les chiffres et lettres qu'il est susceptible de rencontrer.

=> on attend de lui qu'il ait une capacité de généralisation à partir des exemples qui lui ont été présentés.

De même, à l'issue d'apprentissage d'un modèle statistique à partir d'exemples, celui-ci doit être capable de **généraliser**, c'est-à-dire de fournir un résultat correct, dans des situations qu'il n'a pas connues pendant l'apprentissage.



Introduction

B. Exemple de tri postal:

La lecture automatique des codes postaux, et des autres éléments de l'adresse des lettres et paquets, est fréquemment effectuée à l'aide de modèles obtenus par **apprentissage statistique**, à partir des exemples de chacune des classes de chiffres. Il s'agit là d'un **problème de classification** : chaque chiffre inconnu doit être attribué à une classe parmi les 10 classes de chiffres possibles (ou être attribué à une classe dite « **de rejet** » si le chiffre est trop mal écrit pour être reconnu par la machine.

=> dans ce cas l'objet postal doit alors être traité manuellement.

Ce problème ne peut pas être résolu par une application de connaissances existante. Il n'existe pas d'équation mathématique qui décrive les propriétés topologiques des chiffres manuscrits. C'est dans telles conditions que le recours à **l'apprentissage statistique** à partir des exemples se révèle très fructueux.



La modélisation du Machine Learning

- Le **machine learning** constitue une manière de modéliser des phénomènes, dans le but de prendre des décisions stratégiques.
- Le problème de **machine learning** constitue l'étape suivante et permet à un ordinateur de modéliser les données qui lui sont fournies.
- "**Modéliser**" signifie dans ce cas à représenter **le comportement d'un phénomène**, afin de pouvoir **aider à la résolution d'un problème**.
- En **machine learning**, l'algorithme se construit une "**représentation interne**" afin de pouvoir effectuer la tâche qui lui est demandée (*prédiction, identification, etc.*).

⇒ Pour cela, il va d'abord falloir lui entrer un jeu de données d'exemples afin qu'il puisse **s'entraîner** et s'améliorer, d'où le mot **apprentissage**. Ce jeu de données s'appelle le **training set**. On peut appeler une entrée dans le jeu de données une **instance** ou une **observation**.



La modélisation du Machine Learning

A. Exemple :

Voici un jeu de données images classiques qui permet d'entraîner un modèle de machine learning. Chaque image constitue une observation du data set.

airplane



automobile



bird



cat



deer



dog



frog



horse



ship



truck





La modélisation du Machine Learning

A. Exemple :

- ✓ Vous allez donc être responsable dans une première phase du choix et de l'entraînement de **l'algorithme d'apprentissage** du modèle, mais le traitement de la tâche spécifique sera appris à partir du **training set** et ensuite effectué par l'algorithme lui-même dans une seconde phase.

B. le processus du machine learning :

- ✓ Le schéma suivant, représente les différentes étapes qui interviennent dans l'utilisation d'un algorithme de **machine learning**. Vous serez responsable **de la phase 1** et vous pourrez ensuite simplement passer les données dans le modèle durant **la phase 2**. Il y'a des allers-retours entre apprentissage et analyse d'erreur (voir le schéma suivant)

La modélisation du Machine Learning

Phase 1 Apprentissage

Données
D'entraînement

Modèle
Statistique

Analyse
d'Erreur

Phase 2 Déploiement

Nouvelles
Données

Modèle
Statistique
entraîné

Prédiction /
recommandation

Application





La modélisation du Machine Learning

C. Résultat :

- En résumé, le travail en machine learning consiste à :
 - 1) Sélectionner les bonnes **données de test**,
 - 2) Choisir et entraîner le **bon algorithme**,
 - 3) Vérification de la performance du modèle grâce à **l'analyse d'erreur**.
- ⇒ Si les performances s'améliorent lorsqu'on lui fournit les données d'entraînement, on dit alors que la machine **"apprend"**
- ⇒ Une fois le modèle est correctement paramétré sur les données d'entraînement, on peut le déployer par la suite afin qu'il traite de nouvelles données, **pour accomplir la tâche spécifique** poursuivie (prédiction, recommandation, décision.).



La modélisation du Machine Learning

D. Les éléments du machine learning :

Un problème d'apprentissage machine comporte ainsi différents éléments spécifiques :

- **Les données** (les données d'entraînement mais aussi les nouvelles données) ;
- **La tâche spécifique** à accomplir (prédire, recommander, décider quelque chose, etc.)
- **L'algorithme d'apprentissage** en lui-même ;
- **L'analyse d'erreur** (ou **mesure des performances** du modèle).

D.1 Les données :

Voici l'ensemble du type de données habituelles rencontrées en machine learning.

- Les bases de données :

Les bases de données peuvent comprendre différents types d'information, une bonne partie généralement spécifique à l'activité de l'entreprise ou de client.



La modélisation du Machine Learning

D. Les éléments du machine learning :

À titre d'exemple et de manière non exhaustive :

- ☐ Les logs d'un serveur web ;
- ☐ Le catalogue produits d'un site de e-commerce ;
- ☐ Les transactions bancaires ;
- ☐ Les comportements des utilisateurs d'un site.

- Les données brutes:

Ces données, souvent sont plus complexes et nécessitent des pré-traitements spécifiques pour les rendre manipulables par les algorithmes, elles peuvent servir des sources pour un problème de modélisation.



Le **deep learning**, dont vous avez peut-être déjà entendu parler, regroupe les algorithmes et modèles assez gros et complexes pour pouvoir traiter les données brutes directement, et sans pré-traitement.

La modélisation du Machine Learning

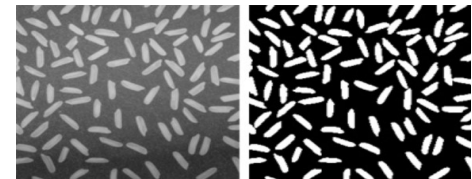
D. Les éléments du machine learning :

- Le texte :

Le **texte libre**, rédigé en **langage naturel (humain)**, est ainsi une autre source de données principale pour le travail du **machine learning**. Cela comporte tous les types de texte auxquels on peut penser naturellement (*articles, livres, messages, etc*), mais aussi d'autres types de textes tels que du code HTML ou encore des séquences d'ADN.. Le traitement du texte (appelé **NLP** comme **Natural Language Processing**)

- Les images (ou vidéos) :

Les images sont aussi une des sources de captation de l'environnement souvent utile sur des problématiques d'entreprise. Beaucoup d'entreprises ont des banques d'images à traiter pour les classer par type ou autre.





La modélisation du Machine Learning

D. Les éléments du machine learning :

- IoT data :

Les objets connectés sont une autre source de données brutes, qui récupèrent un grand nombre de données grâce à leurs capteurs.

D.2 Les tâches à accomplir :

La tâche spécifique à accomplir correspond au problème qu'on cherche à résoudre grâce à la modélisation du phénomène. On peut distinguer un certain nombre de cas qui reviennent souvent dans un environnement business, tels que les **recommandations** de produits par exemple, l'**identification** de transactions frauduleuses, la **prédiction** de l'impact d'une campagne marketing sur le taux de conversion ou la **prédiction du prix** optimal d'un produit pour maximiser le nombre de ventes.

=> Chaque tâche se traduira différemment et nécessitera, bien sûr, le choix d'algorithmes différents.



La modélisation du Machine Learning

D. Les éléments du machine learning :

D.3 L'algorithme d'apprentissage :

L'algorithme d'apprentissage constitue la méthode avec laquelle le modèle statistique va se paramétrer à partir des données d'exemple. Il existe plusieurs algorithmes différents ! On choisira un type d'algorithme particulier en fonction du type de tâche que l'on souhaite accomplir et du type de données dont on dispose.

Voici Quelques exemples d'algorithmes de machine learning :

- ✓ La régression linéaire ;
- ✓ K-means ;
- ✓ Les Support Vector Machine (SVM) ;
- ✓ Les réseaux de neurones ;
- ✓ Les random forests.



La modélisation du Machine Learning

D. Les éléments du machine learning :

D.4 La mesure des performances :

Il fait partie intégrante du travail de modélisation. Il faut en général déterminer une mesure principale, souvent spécifique à la tâche à accomplir. Le choix de cette métrique est très important ! .

Exemple :

Imaginez que vous voulez créer un algorithme de détection de fraudes bancaires. Vous voulez mesurer à quel point votre programme est performant. Une manière de faire serait de mesurer la proportion totale de transactions détectées comme fraude. Cependant, on compte ici les transactions qui ne sont pas des fraudes et qui ont quand même été notées comme en étant (appelé "faux positifs").



La modélisation du Machine Learning

D. Les éléments du machine learning :

D.4 La mesure des performances :

Donc, avec ce genre de métrique, on n'est pas exigeant sur ce type d'erreur produit par notre algorithme. Il faut peut-être utiliser une autre métrique plus pertinente. Par exemple préciser la proportion de "vraies fraudes" détectées par rapport au total de transactions détectées comme frauduleuses.



Le résumé en une phrase, un ordinateur **apprend** à partir de **données** pour résoudre une **tâche** en faisant attention à **mesurer les performances**. S'il **améliore** les performances sur cette tâche lorsqu'on lui fournit les **données d'entraînement**, on dit alors qu'il **apprend**.