



Les système d'apprentissage en Machine Learning



Introduction

Définition Machine Learning :

- Donner à une machine la capacité d'apprendre sans la programmer de façon explicite
- Science moderne permettant de découvrir des patterns et d'effectuer des prédictions à partir de données en se basant sur des statistiques, sur du forage de données, sur la reconnaissance de patterns et sur les analyses prédictives
- Extraire de la valeur à travers différents types d'algorithmes

Apprentissage supervisé

- Dataset

Exemple de Dataset sur des appartements

Target y	Features		
	x_1	x_2	x_3
Prix	Surface	Qualité	Adresse postale
313,000	90	3	95000
720,000	110	5	93000
250,000	40	4	44500
290,000	60	3	67000
190,000	50	3	59300
...

Par convention:
 m : nombre d'exemples
 n : nombre de features

Par convention, on note:
 $x_{feature}^{(exemple)}$

Dataset (x, y)

y	x_1	x_2	x_3	...	x_n
$y^{(1)}$	$x_1^{(1)}$	$x_2^{(1)}$	$x_3^{(1)}$...	$x_n^{(1)}$
$y^{(2)}$	$x_1^{(2)}$	$x_2^{(2)}$	$x_3^{(2)}$...	$x_n^{(2)}$
$y^{(3)}$	$x_1^{(3)}$	$x_2^{(3)}$	$x_3^{(3)}$...	$x_n^{(3)}$
...
$y^{(m)}$	$x_1^{(m)}$	$x_2^{(m)}$	$x_3^{(m)}$...	$x_n^{(m)}$

vecteur target $y \in \mathbb{R}^{m \times 1}$

matrice features $X \in \mathbb{R}^{m \times n}$

$$y = \begin{pmatrix} y^{(1)} \\ y^{(2)} \\ \vdots \\ y^{(m)} \end{pmatrix}$$

$$X = \begin{pmatrix} x_1^{(1)} & \dots & x_n^{(1)} \\ \vdots & \ddots & \vdots \\ x_1^{(m)} & \dots & x_n^{(m)} \end{pmatrix}$$

- Modèles et ses paramètres

A nous de choisir quels modèles la machine doit prendre en compte

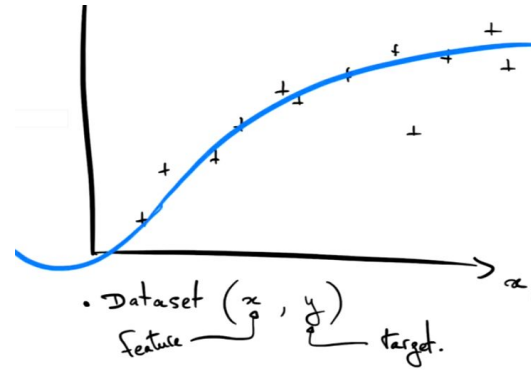
La machine doit apprendre les paramètres de ce modèles

- Fonction coût

Assemblage d'erreurs en fonction de $f(x)$ que l'on retrouve dans le dataset

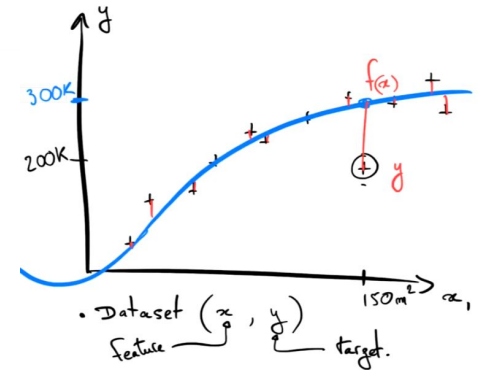
- Notion d'algorithme d'apprentissage

Développer un stratégie qui minimise la fonction coût, soit les erreurs



$$f(x) = \underline{a}x^2 + \underline{b}x^3 + \underline{c}$$

paramètres



$$f(x) = \underline{a}x^2 + \underline{b}x^3 + \underline{c}$$

paramètres



Apprentissage non supervisé

	Apprentissage supervisé	Apprentissage non supervisé
Données d'entrée	Données connues en entrée	Données inconnues en entrée
Complexité informatique	Complexe	Moins complexe
Domaines d'activités	Classification et régression	Exploitation de règles de <u>clustering</u> et d'association
Précision	Produit des résultats précis	Génère des résultats modérés

- Opérer à partir d'exemples non annotés / **données non étiquetées**
- Pas d'exemple : l'apprentissage par la machine se fait de manière entièrement indépendante



Apprentissage semi supervisé

Utilise un ensemble de données étiquetées et non-étiquetées.

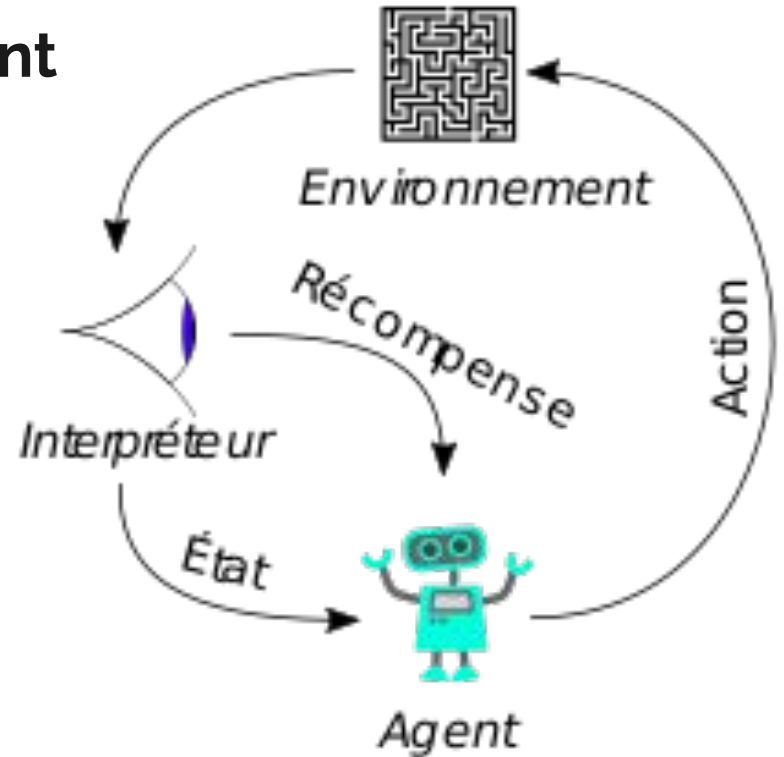
Objectif : améliorer significativement la qualité de l'apprentissage en utilisant l'ensemble du dataset

- Labeled / unlabeled data
- Pseudo-labeling: utiliser les données étiquetées pour classer celles qui ne le sont pas à travers un réseaux de neurones

Apprentissage par renforcement

Apprendre à partir d'essais et d'erreurs afin de maximiser les gains

Récompenser l'ordinateur pour lui indiquer qu'il doit ou pas continuer





Conclusion

Définir le modèle en fonction :

- des données à disposition
- des objectifs en terme d'analyse