

Apprentissage automatique (Machine Learning)

ProgFest 2022

Maxence Larose

7 mars 2022





$$\bar{x}_1 = \frac{1+3+3+6+8+9}{6} = 5$$

$$\bar{x}_2 = \frac{2+4+4+8+12}{5} = 30$$

$$\bar{x}_3 = \frac{4+7+1+6}{4} = 18$$

$$\log_b b^x = x$$

$$\log_a x = \frac{\log_b x}{\log_b a}$$

$$\log_b (x^r) = r \log_b x$$

$$\log_b (xy) = \log_b x + \log_b y$$

$$\log_b \left(\frac{x}{y} \right) = \log_b x - \log_b y$$



$$\begin{aligned} x(2x+3) &= 90 \\ 2x^2+3x-90 &= 0 \\ (2x+15)(x-6) &= 0 \end{aligned}$$

$$\begin{aligned} a^2 + b^2 &= c^2 \\ a &= \sqrt{c^2 - b^2} \end{aligned}$$

$$ab+ac = a(b+c)$$

$$\frac{a(\frac{b}{c})}{\frac{a}{c}} = \frac{ab}{c}$$

$$\frac{(\frac{a}{b})}{\frac{a}{c}} = \frac{a}{bc}$$

$$\frac{a}{(\frac{b}{c})} = \frac{ac}{b}$$

$$\frac{\frac{a}{b} + \frac{c}{d}}{\frac{a}{b} + \frac{c}{d}} = \frac{ad+bc}{bd}$$

$$\frac{a}{b} + \frac{c}{d} = \frac{ad+bc}{bd}$$

$$\frac{a}{b} + \frac{c}{d} = \frac{ad+bc}{bd}$$

$$\frac{a}{b} + \frac{c}{d} = \frac{ad+bc}{bd}$$

$$\frac{a}{b} + \frac{c}{d} = \frac{ad+bc}{bd}$$

$$\frac{a}{b} + \frac{c}{d} = \frac{ad+bc}{bd}$$

$$\frac{a}{b} + \frac{c}{d} = \frac{ad+bc}{bd}$$

$$\frac{a}{b} + \frac{c}{d} = \frac{ad+bc}{bd}$$

$$\frac{a}{b} + \frac{c}{d} = \frac{ad+bc}{bd}$$

$$\frac{a}{b} + \frac{c}{d} = \frac{ad+bc}{bd}$$

$$\frac{a}{b} + \frac{c}{d} = \frac{ad+bc}{bd}$$

$$\frac{a}{b} + \frac{c}{d} = \frac{ad+bc}{bd}$$

$$\frac{a}{b} + \frac{c}{d} = \frac{ad+bc}{bd}$$

$$\frac{a}{b} + \frac{c}{d} = \frac{ad+bc}{bd}$$

$$\frac{a}{b} + \frac{c}{d} = \frac{ad+bc}{bd}$$

$$\frac{a}{b} + \frac{c}{d} = \frac{ad+bc}{bd}$$

$$f(x) \leq 5$$

$$X^2 - 4X + 5 \leq 5$$

$$X^2 - 4X \leq 0$$

$$n(B \cap C) = 22$$

$$n(B) = 68$$

$$n(C) = 84$$

$$n(B \cup C) = n(B) + n(C) - n(B \cap C)$$

$$\begin{aligned} \text{He} &= 4.002602 \\ \text{Na} &= 22.989769 \\ \text{Ar} &= 39.948 \end{aligned}$$

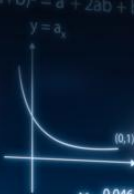


$$\begin{aligned} (100^2)a + 100b \\ 10000a + 100b - 5 \end{aligned}$$

$$\begin{aligned} a_n &= \frac{1}{2^{n-1}} = \\ &= \frac{1}{2^9} = \end{aligned}$$

$$y = ax + b$$

$$AB + BC = AC$$



$$M = \frac{0.046765}{3.0L}$$

$$a + 2ab + b^2$$

$$\frac{a-b}{c-d} = \frac{b-a}{bd}$$

$$\frac{a+b}{c-d} = \frac{b-a}{bd}$$

$$\frac{a+b}{c-d} = \frac{b-a}{bd}$$

$$\frac{a+b}{c-d} = \frac{b-a}{bd}$$

$$\frac{a+b}{c-d} = \frac{b-a}{bd}$$

$$\frac{a+b}{c-d} = \frac{b-a}{bd}$$

$$\frac{a+b}{c-d} = \frac{b-a}{bd}$$

$$\frac{a+b}{c-d} = \frac{b-a}{bd}$$

$$\frac{a+b}{c-d} = \frac{b-a}{bd}$$

$$\frac{a+b}{c-d} = \frac{b-a}{bd}$$

$$\frac{a+b}{c-d} = \frac{b-a}{bd}$$

$$\frac{a+b}{c-d} = \frac{b-a}{bd}$$

$$\frac{a+b}{c-d} = \frac{b-a}{bd}$$

$$\frac{a+b}{c-d} = \frac{b-a}{bd}$$

$$\frac{a+b}{c-d} = \frac{b-a}{bd}$$

$$\frac{a+b}{c-d} = \frac{b-a}{bd}$$

$$\frac{a+b}{c-d} = \frac{b-a}{bd}$$

$$\frac{a+b}{c-d} = \frac{b-a}{bd}$$

$$\frac{a+b}{c-d} = \frac{b-a}{bd}$$

$$\sin B = \frac{4\sqrt{3}}{5}$$

$$\sin 60^\circ = \frac{4\sqrt{3}}{5}$$

$$\sin 60^\circ = \frac{4\sqrt{3}}{5}$$

$$\sin 60^\circ = \frac{4\sqrt{3}}{5}$$

$$\sin 60^\circ = \frac{4\sqrt{3}}{5}$$

$$\sin 60^\circ = \frac{4\sqrt{3}}{5}$$

$$\sin 60^\circ = \frac{4\sqrt{3}}{5}$$

$$\sin 60^\circ = \frac{4\sqrt{3}}{5}$$

$$\sin 60^\circ = \frac{4\sqrt{3}}{5}$$

$$\sin 60^\circ = \frac{4\sqrt{3}}{5}$$

$$\sin 60^\circ = \frac{4\sqrt{3}}{5}$$

$$\sin 60^\circ = \frac{4\sqrt{3}}{5}$$

$$\sin 60^\circ = \frac{4\sqrt{3}}{5}$$

$$\sin 60^\circ = \frac{4\sqrt{3}}{5}$$

$$\sin 60^\circ = \frac{4\sqrt{3}}{5}$$

$$\sin 60^\circ = \frac{4\sqrt{3}}{5}$$

$$\sin 60^\circ = \frac{4\sqrt{3}}{5}$$

$$\sin 60^\circ = \frac{4\sqrt{3}}{5}$$

$$\sin 60^\circ = \frac{4\sqrt{3}}{5}$$

$$\sin 60^\circ = \frac{4\sqrt{3}}{5}$$

$$\sin B = \frac{4\sqrt{3}}{5}$$

$$\sin 60^\circ = \frac{4\sqrt{3}}{5}$$

$$\sin 60^\circ = \frac{4\sqrt{3}}{5}$$

$$\sin 60^\circ = \frac{4\sqrt{3}}{5}$$

$$\sin 60^\circ = \frac{4\sqrt{3}}{5}$$

$$\sin 60^\circ = \frac{4\sqrt{3}}{5}$$

$$\sin 60^\circ = \frac{4\sqrt{3}}{5}$$

$$\sin 60^\circ = \frac{4\sqrt{3}}{5}$$

$$\sin 60^\circ = \frac{4\sqrt{3}}{5}$$

$$\sin 60^\circ = \frac{4\sqrt{3}}{5}$$

$$\sin 60^\circ = \frac{4\sqrt{3}}{5}$$

$$\sin 60^\circ = \frac{4\sqrt{3}}{5}$$

$$\sin 60^\circ = \frac{4\sqrt{3}}{5}$$

$$\sin 60^\circ = \frac{4\sqrt{3}}{5}$$

$$\sin 60^\circ = \frac{4\sqrt{3}}{5}$$

$$\sin 60^\circ = \frac{4\sqrt{3}}{5}$$

$$\sin 60^\circ = \frac{4\sqrt{3}}{5}$$

$$\sin 60^\circ = \frac{4\sqrt{3}}{5}$$

$$\sin 60^\circ = \frac{4\sqrt{3}}{5}$$

$$\sin 60^\circ = \frac{4\sqrt{3}}{5}$$

$$\sin B = \frac{4\sqrt{3}}{5}$$

$$\sin 60^\circ = \frac{4\sqrt{3}}{5}$$

$$\sin 60^\circ = \frac{4\sqrt{3}}{5}$$

$$\sin 60^\circ = \frac{4\sqrt{3}}{5}$$

$$\sin 60^\circ = \frac{4\sqrt{3}}{5}$$

$$\sin 60^\circ = \frac{4\sqrt{3}}{5}$$

$$\sin 60^\circ = \frac{4\sqrt{3}}{5}$$

$$\sin 60^\circ = \frac{4\sqrt{3}}{5}$$

$$\sin 60^\circ = \frac{4\sqrt{3}}{5}$$

$$\sin 60^\circ = \frac{4\sqrt{3}}{5}$$

$$\sin 60^\circ = \frac{4\sqrt{3}}{5}$$

$$\sin 60^\circ = \frac{4\sqrt{3}}{5}$$

$$\sin 60^\circ = \frac{4\sqrt{3}}{5}$$

$$\sin 60^\circ = \frac{4\sqrt{3}}{5}$$

$$\sin 60^\circ = \frac{4\sqrt{3}}{5}$$

$$\sin 60^\circ = \frac{4\sqrt{3}}{5}$$

$$\sin 60^\circ = \frac{4\sqrt{3}}{5}$$

$$\sin 60^\circ = \frac{4\sqrt{3}}{5}$$

$$\sin 60^\circ = \frac{4\sqrt{3}}{5}$$

$$\sin 60^\circ = \frac{4\sqrt{3}}{5}$$

$$\sin B = \frac{4\sqrt{3}}{5}$$

$$\sin 60^\circ = \frac{4\sqrt{3}}{5}$$

$$\sin 60^\circ = \frac{4\sqrt{3}}{5}$$

$$\sin 60^\circ = \frac{4\sqrt{3}}{5}$$

$$\sin 60^\circ = \frac{4\sqrt{3}}{5}$$

$$\sin 60^\circ = \frac{4\sqrt{3}}{5}$$

$$\sin 60^\circ = \frac{4\sqrt{3}}{5}$$

$$\sin 60^\circ = \frac{4\sqrt{3}}{5}$$

$$\sin 60^\circ = \frac{4\sqrt{3}}{5}$$

$$\sin 60^\circ = \frac{4\sqrt{3}}{5}$$

$$\sin 60^\circ = \frac{4\sqrt{3}}{5}$$

$$\sin 60^\circ = \frac{4\sqrt{3}}{5}$$

$$\sin 60^\circ = \frac{4\sqrt{3}}{5}$$

$$\sin 60^\circ = \frac{4\sqrt{3}}{5}$$

$$\sin 60^\circ = \frac{4\sqrt{3}}{5}$$

$$\sin 60^\circ = \frac{4\sqrt{3}}{5}$$

$$\sin 60^\circ = \frac{4\sqrt{3}}{5}$$

$$\sin 60^\circ = \frac{4\sqrt{3}}{5}$$

$$\sin 60^\circ = \frac{4\sqrt{3}}{5}$$

$$\sin 60^\circ = \frac{4\sqrt{3}}{5}$$

Pourquoi est-ce si populaire ?

- Disponibilités de très grands jeux de données (big data)

Pourquoi est-ce si populaire ?

- Disponibilités de très grands jeux de données (big data)
- Augmentation de la capacité de calcul (GPU)



Pourquoi est-ce si populaire ?

- Disponibilités de très grands jeux de données (big data)
- Augmentation de la capacité de calcul (GPU)
- Nouveaux modèles d'apprentissage très flexibles

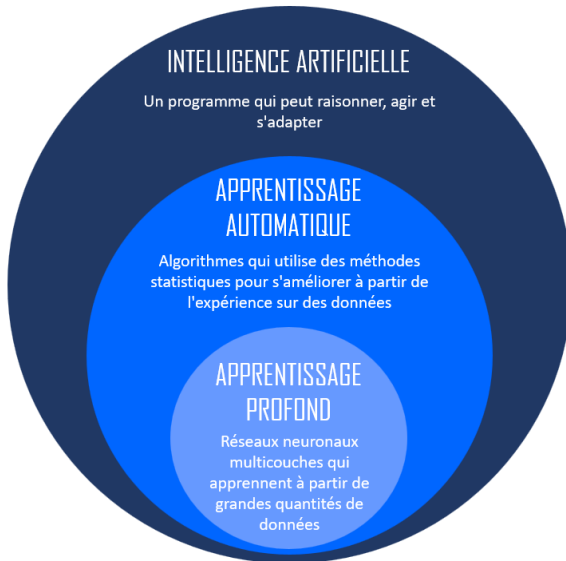


Pourquoi est-ce si populaire ?

- Disponibilités de très grands jeux de données (big data)
- Augmentation de la capacité de calcul (GPU)
- Nouveaux modèles d'apprentissage très flexibles



Ceci explique l'avènement de l'apprentissage profond. On discutera plutôt ici de l'apprentissage automatique. C'est quoi la différence entre les deux ?



Définition de l'apprentissage automatique

- L'apprentissage automatique consiste à utiliser des ordinateurs pour **optimiser** un **modèle de traitement** de l'information selon certains **critères de performance** à partir d'**observations**, que ce soit des données-exemples ou des expériences passées.
 - Quand on parle de modèle, on parle d'un objet approximant une fonction complexe $f(x)$. On cherche donc une fonction $h(x) \approx f(x)$.

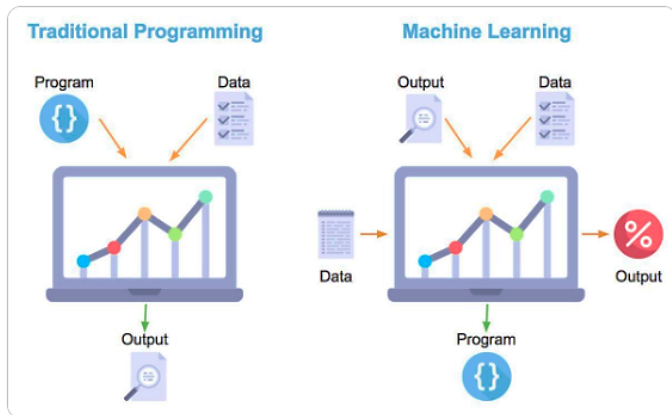
Définition de l'apprentissage automatique

- L'apprentissage automatique consiste à utiliser des ordinateurs pour **optimiser** un **modèle de traitement** de l'information selon certains **critères de performance** à partir d'**observations**, que ce soit des données-exemples ou des expériences passées.
 - Quand on parle de modèle, on parle d'un objet approximant une fonction complexe $f(x)$. On cherche donc une fonction $h(x) \approx f(x)$.
- Lorsque l'on connaît le bon modèle de traitement à utiliser, il n'est pas nécessaire de faire de l'apprentissage...

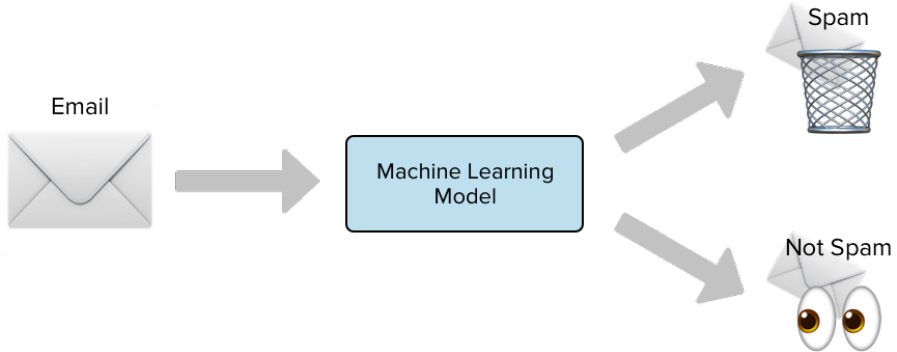
Définition de l'apprentissage automatique

- L'apprentissage automatique consiste à utiliser des ordinateurs pour **optimiser** un **modèle de traitement** de l'information selon certains **critères de performance** à partir d'**observations**, que ce soit des données-exemples ou des expériences passées.
 - Quand on parle de modèle, on parle d'un objet approximant une fonction complexe $f(x)$. On cherche donc une fonction $h(x) \approx f(x)$.
- Lorsque l'on connaît le bon modèle de traitement à utiliser, il n'est pas nécessaire de faire de l'apprentissage...
- Exemples où l'apprentissage automatique est utile :
 - On n'a pas d'expertise sur le problème (ex. robot navigant sur Mars)
 - On a une expertise, mais on ne sait pas comment l'expliquer (ex. reconnaissance d'objets)
 - Les solutions doivent être personnalisées (ex. biométrie)

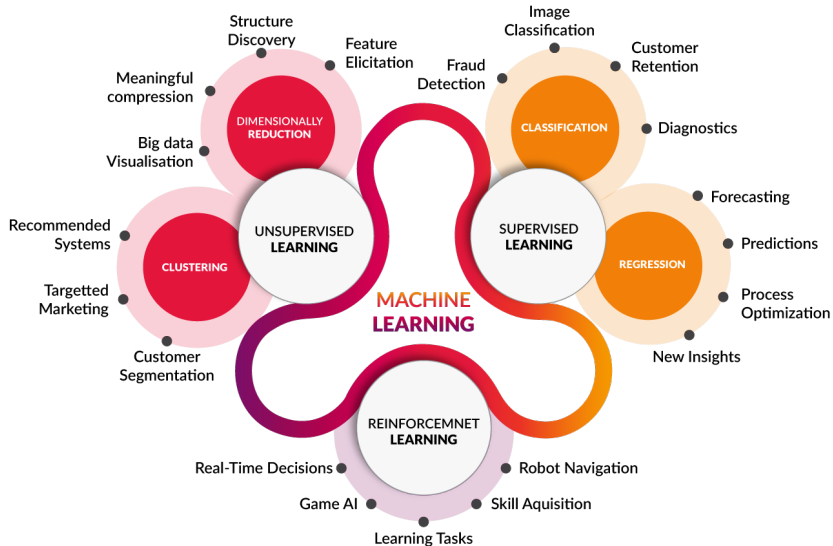
Programmation traditionnelle vs. Apprentissage automatique



Example

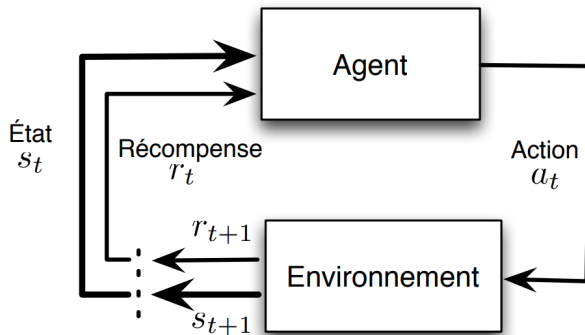


Types d'apprentissage

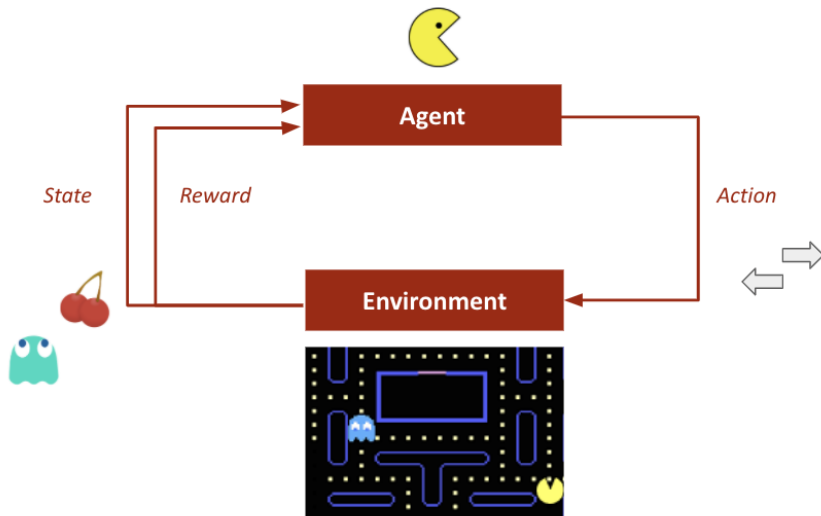


Apprentissage par renforcement

But : Apprendre une politique, une séquences d'actions, pour maximiser les récompenses obtenues pendant un épisode.



Apprentissage par renforcement : Exemple



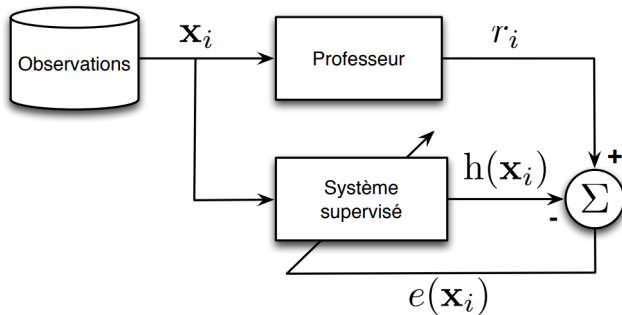
But : Apprendre une projection entre des observations X en entrée et des valeurs associées Y en sortie.

Modélisation mathématique :

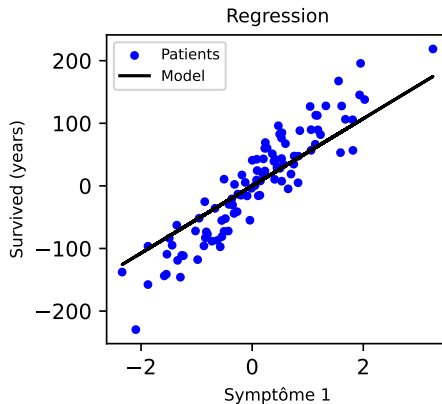
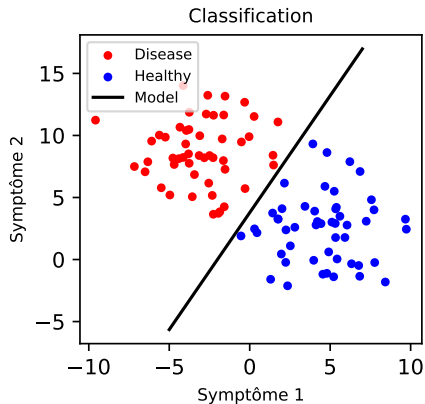
$$y = h_{\theta}(x)$$

où $h(\cdot)$ est une fonction générale du modèle et θ sont les paramètres du modèle.

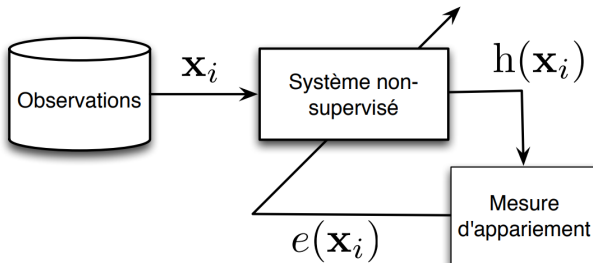
Apprentissage supervisé : Schéma



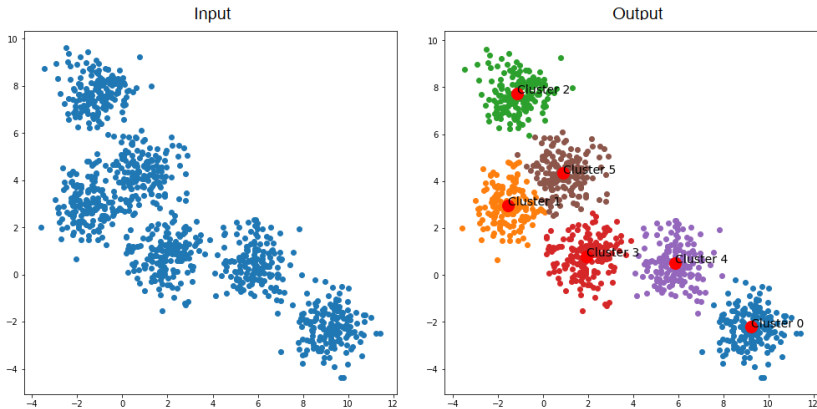
Apprentissage supervisé : Classification et régression



But : Découvrir des régularités dans des observations X sans étiquette Y .

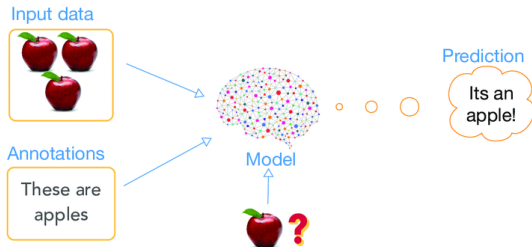


Apprentissage non supervisé : Application au clustering

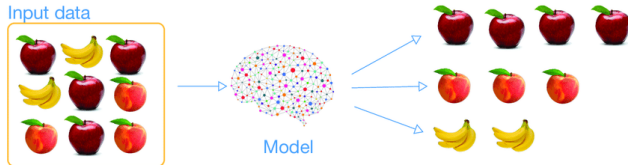


Apprentissage supervisé vs. non supervisé

supervised learning



unsupervised learning

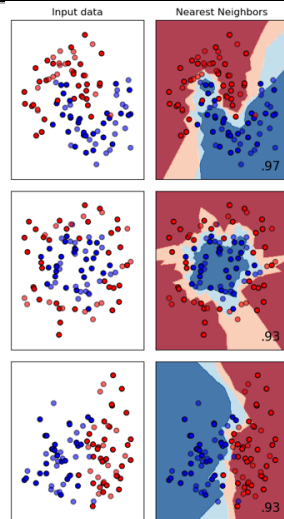


- K - plus proche voisin
- Arbre de décision
- Processus Gaussien
- Machine à vecteurs de support
- Réseau de neurones

K - Plus proches voisins

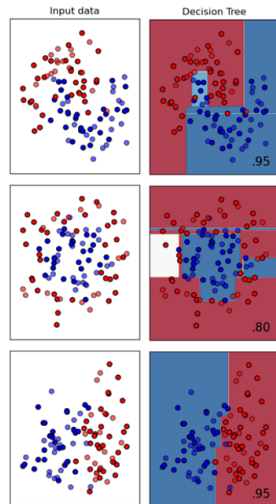
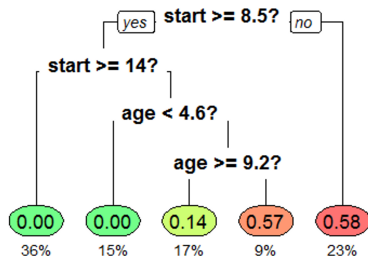
$$c_i = \max_c \{ \text{Count}(\mathcal{C}) \}$$

$$\mathcal{C} = \operatorname{argmin}_c^k \{ d(x_i, x_j) \} \quad \forall j \neq i$$



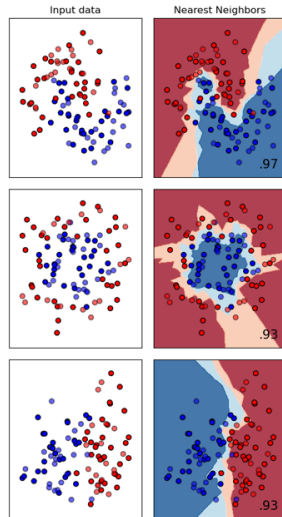
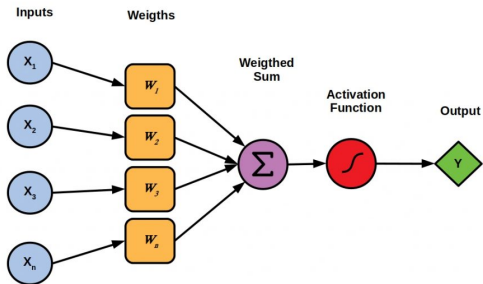
Arbre de décision

- Séparation hiérarchique (récursive) de l'espace d'entrée.
- Chaque nœud de l'arbre est un test sur valeur avec issues discrète.
- Effectue une division de plus en plus fine de l'espace d'entrée



Réseau de neurones

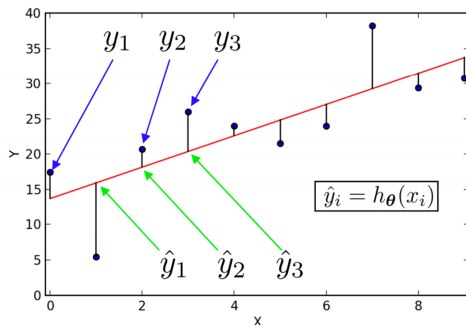
$$h_{\theta}(x) = \sigma \left(\sum_{i=1}^d \theta_i x_i + b \right)$$



Optimisation

Objectif : Optimiser une fonction $h_{\theta}(x)$ qui approxime une fonction complexe inconnue $f(x)$. Ce sont les paramètres θ qui sont optimisés lors de l'apprentissage.

L'erreur empirique $J_{\theta}(\mathcal{X})$ correspond à la moyenne de la perte calculée à chaque point avec une fonction de perte choisie $\mathcal{L}(y, \hat{y})$ où \hat{y} est la prédiction donnée par notre fonction $h_{\theta}(x)$ et $\mathcal{X} = \{x_i, y_i\}_{i=1}^N$.



Objectif : Elle sert à connaître la performance du modèle pour une certaine entrée x .

$$J_{\theta}(\mathcal{X}) = \frac{1}{N} \sum_{i=1}^N \mathcal{L}(y_i, \hat{y}_i)$$

- **Erreur absolue** : $\mathcal{L}(y_i, \hat{y}_i) = |y_i - \hat{y}_i| \rightarrow$ régression

Objectif : Elle sert à connaître la performance du modèle pour une certaine entrée x .

$$J_{\theta}(\mathcal{X}) = \frac{1}{N} \sum_{i=1}^N \mathcal{L}(y_i, \hat{y}_i)$$

- **Erreur absolue** : $\mathcal{L}(y_i, \hat{y}_i) = |y_i - \hat{y}_i| \rightarrow$ régression
- **Erreur quadratique** : $\mathcal{L}(y_i, \hat{y}_i) = (y_i - \hat{y}_i)^2 \rightarrow$ régression

Objectif : Elle sert à connaître la performance du modèle pour une certaine entrée x .

$$J_{\theta}(\mathcal{X}) = \frac{1}{N} \sum_{i=1}^N \mathcal{L}(y_i, \hat{y}_i)$$

- **Erreur absolue** : $\mathcal{L}(y_i, \hat{y}_i) = |y_i - \hat{y}_i| \rightarrow$ régression
- **Erreur quadratique** : $\mathcal{L}(y_i, \hat{y}_i) = (y_i - \hat{y}_i)^2 \rightarrow$ régression
- **Entropie croisée** : $\mathcal{L}(y_i, \hat{y}_i) = - \sum_{c=1}^M y_c \log(\hat{y}_c) \rightarrow$ classification

Comment minimiser l'erreur empirique ?

Il existe plusieurs méthodes pour minimiser l'erreur empirique :

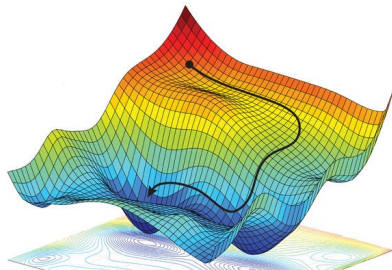
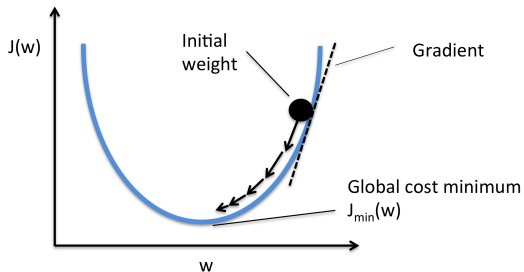
- Maximum de vraisemblance
- Descente du gradient
- Programmation quadratique
- ...

Descente du gradient

$$\text{Gradient : } \nabla J_{\theta}(\mathcal{X}) = \left[\frac{\partial J}{\partial \theta_1} \cdots \frac{\partial J}{\partial \theta_d} \right]$$

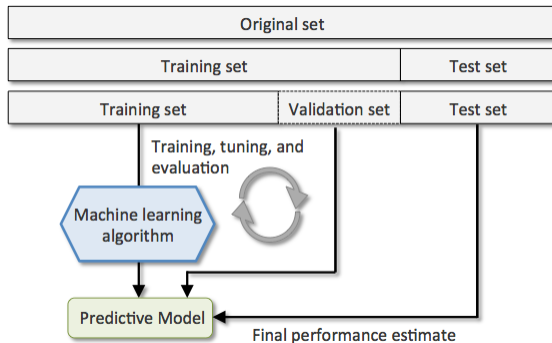
$$\text{Incrémentation : } \theta = \theta - \eta \nabla J_{\theta}(\mathcal{X})$$

$$\text{Solution : } \theta^* = \operatorname{argmin}_{\nabla \theta_i} J_{\theta}(\mathcal{X})$$



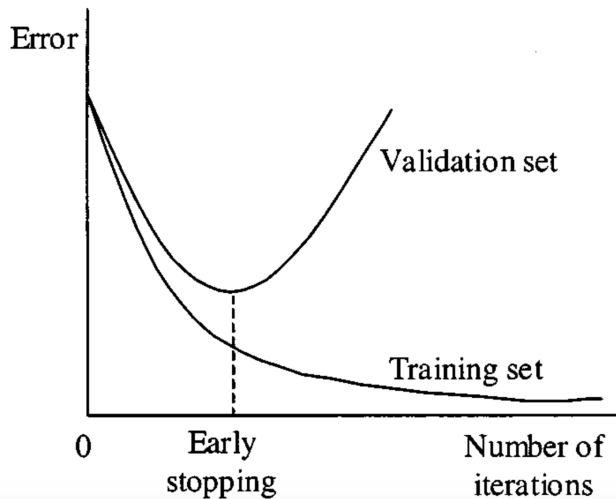
Jeu d'entraînement, de validation et de test

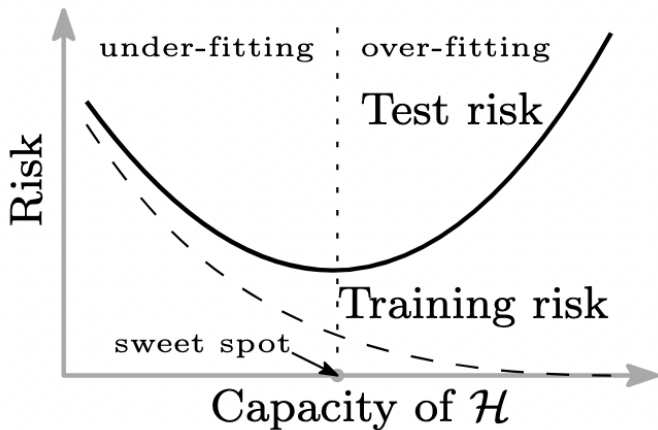
Pour estimer l'erreur de généralisation, on doit utiliser des données non vues durant l'entraînement. Approche classique, partitionner le jeu d'exemples (Entraînement (60%), validation (20%), test (20%)).



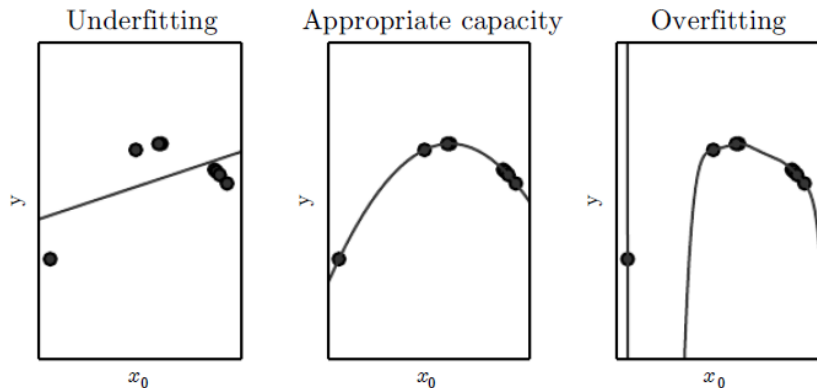
Il existe différents types d'erreur et le contexte du problème détermine lesquelles doivent être contrôlées.

		Predicted condition	
		Positive (PP)	Negative (PN)
Actual condition	Total population = P + N		
	Positive (P)	True positive (TP), hit	False negative (FN), type II error, miss, underestimation
	Negative (N)	False positive (FP), type I error, false alarm, overestimation	True negative (TN), correct rejection





Sous-apprentissage et sur-apprentissage



But : Introduire une pénalité dans la fonction optimisée afin de minimiser la complexité du modèle et ainsi combattre le sur-apprentissage.

Forme courante :

$$J_{\theta}(\mathcal{X}) = \frac{1}{N} \sum_{i=1}^N \mathcal{L}(y_i, \hat{y}_i) + \lambda C(\theta)$$

où λ est la pondération relative entre l'erreur empirique J et la complexité $C(\theta)$.

Exemples courants de complexité $C(\theta)$:

- Nombre de paramètres utilisés
- Norme des valeurs de paramètres

Exemple de programmation scikit-learn : Classification

Exemple de classification du jeu de données Iris avec un K - plus proche voisin. Voir le jupyter notebook : "".

- Il est possible d'approximer des fonctions complexes avec des modèles simples.
- Les paramètres des modèles doivent être optimisés en minimisant l'erreur empirique sur un jeu de validation, par exemple en utilisant une descente de gradient.
- Le module scikit-learn offre une approche facile et rapide.
- L'apprentissage profond est basé sur le perceptron.