



Eléments de Théorie D'Apprentissage Statistique / Machine Learning

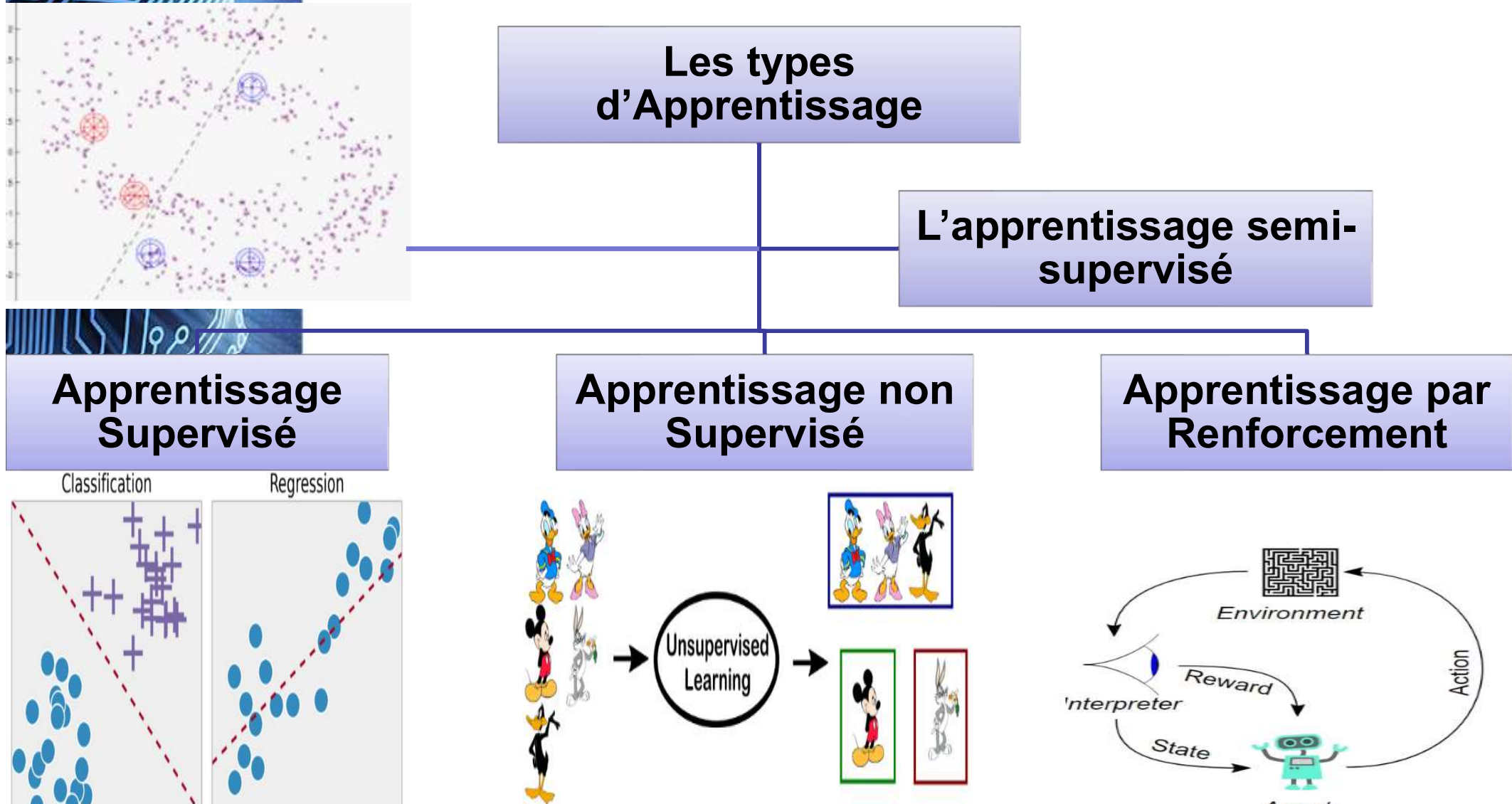
II. Les types d'Apprentissage du Machine Learning:

Les algorithmes d'apprentissage servent à apprendre à agir d'une manière à maximiser les futures récompenses dans un environnement variable où il y'a d'autre apprenants aussi en même temps

La modélisation prédictive vise principalement à réduire au minimum l'erreur d'un modèle ou à rendre les prévisions les plus précises possibles, au détriment de l'explicabilité.

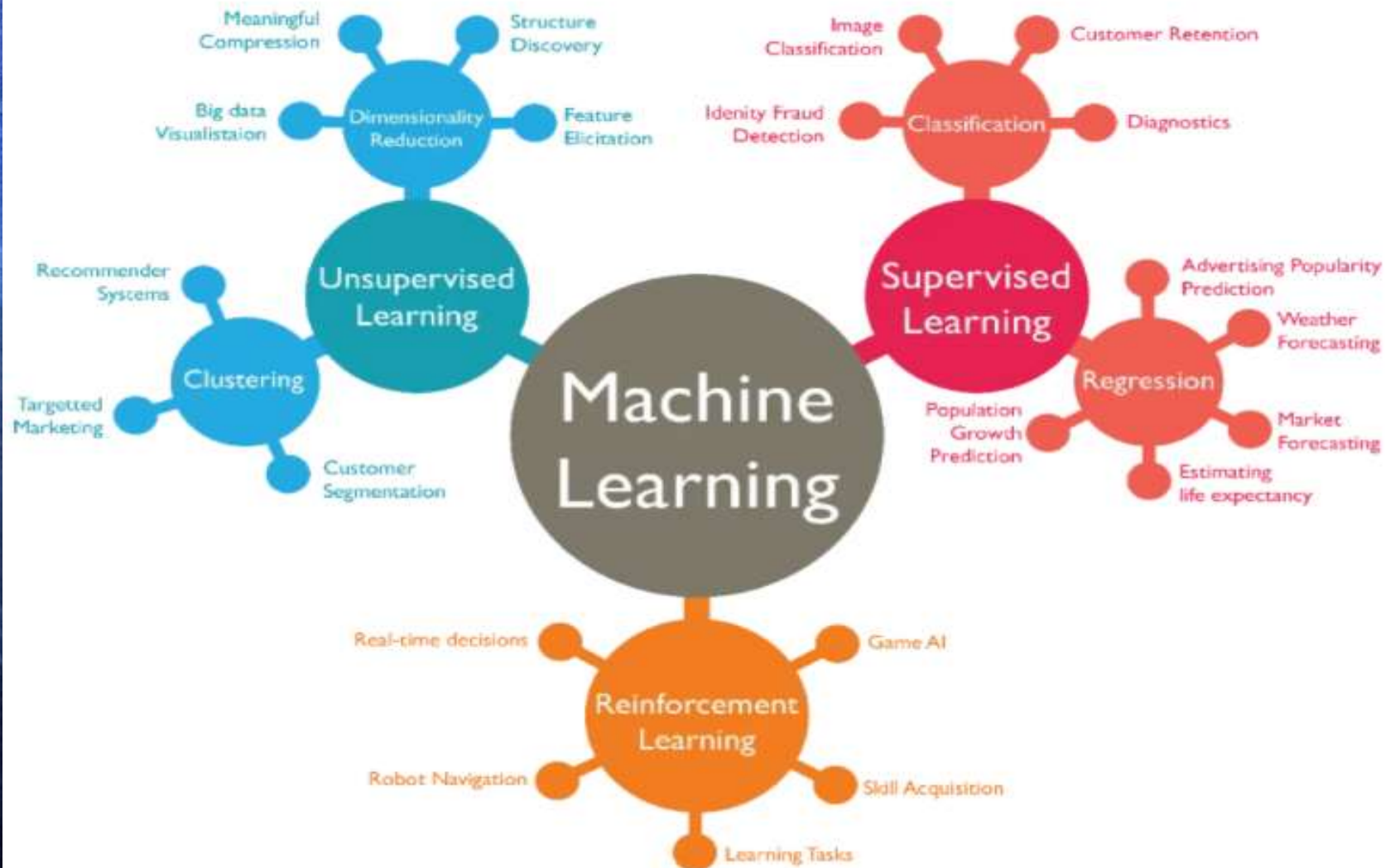
Eléments de Théorie D'Apprentissage Statistique / Machine Learning

II. Les types d'Apprentissage du Machine Learning:



Eléments de Théorie D'Apprentissage Statistique / Machine Learning

II. Les types d'Apprentissage du Machine Learning:





Eléments de Théorie D'Apprentissage Statistique / Machine Learning

II. Les types d'Apprentissage du Machine Learning :

Apprentissage semi-supervisé & Reinforcement Learning

Apprentissage semi-supervisé

- ❑ Il prend en entrée certaines données annotées et d'autres non. Ce sont des méthodes très intéressantes qui tirent parti des deux mondes (supervisé et non supervisé), mais ils apportent leur lot de difficultés.

Reinforcement Learning

- ❑ Il se base sur un cycle **d'expérience / récompense** et améliore les performances à chaque itération. Une analogie souvent citée est celle du cycle de dopamine : une **"bonne"** expérience augmente la dopamine et donc augmente la probabilité que l'agent répète l'expérience.

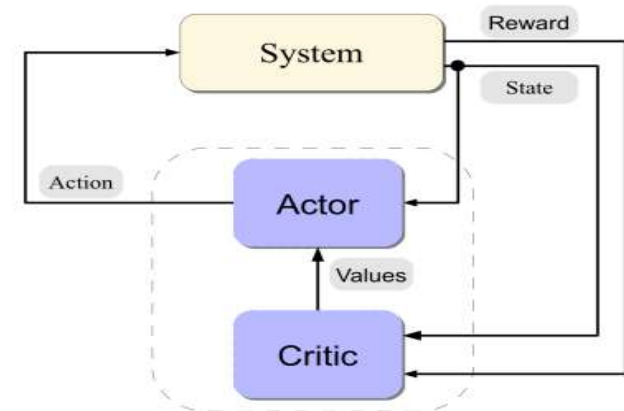
Eléments de Théorie D'Apprentissage Statistique / Machine Learning

II. Les types d'Apprentissage du Machine Learning :

Dans cet apprentissage, l'acteur effectue une action sur le système, ce dernier va générer cette action sous forme des récompenses et de nouveaux états qui vont être envoyés par la suite au critic. Le critic va envoyer après son interprétation à propos des nouvelles valeurs à l'acteur et ainsi de suite

Reinforcement Learning

Actor-Critic Learning



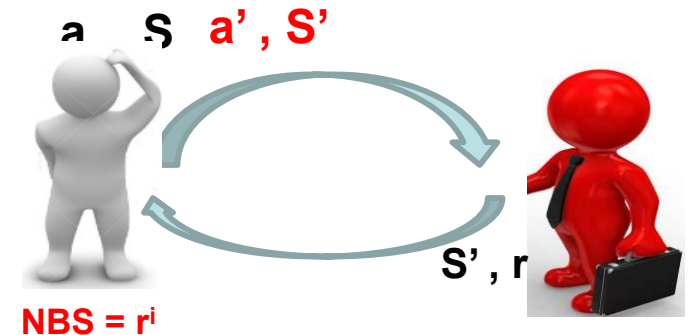
Eléments de Théorie D'Apprentissage Statistique / Machine Learning

II. Les types d'Apprentissage du Machine Learning :

Dans cet apprentissage, l'agent envoie une action dont le but est de recevoir une récompense de l'autre agent et un nouvel état dont lequel il va le prendre en considération pour maximiser la prochaine récompense et va envoyer par la suite selon son nouvel état s' une nouvelle action a' pour recevoir une nouvelle récompense r' de telle manière $r' > r$. Ce scénario sera répété en boucle jusqu'à ce que la récompense finale r^i convergera vers **Nash Bargaining Solution (NBS)** => on utilise cet algorithme pour montrer la convergence vers le NBS

Reinforcement Learning

Nash Bargaining Learning



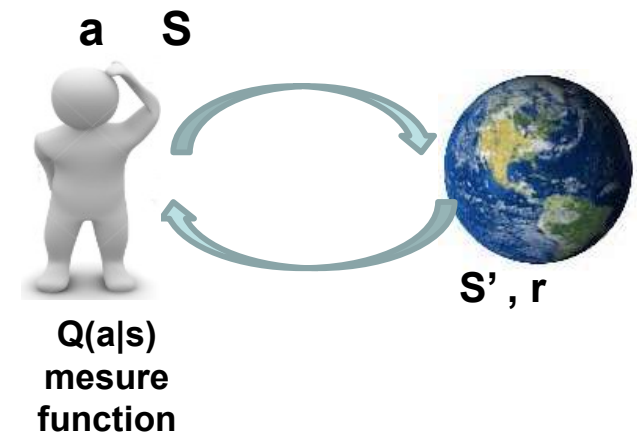
Eléments de Théorie D'Apprentissage Statistique / Machine Learning

II. Les types d'Apprentissage du Machine Learning :

Dans le Q-learning, l'agent exécute une action **a** comme un signal en fonction de son état **s** sur l'environnement et la fonction **Q(a|s)** qui désigne la qualité de l'action **a** exécutée dans l'état **s**. il reçoit le nouvel état **s'** et une récompense **r** de l'environnement ensuite il met à jour sa fonction **Q**

Reinforcement Learning

Q- Learning



Eléments de Théorie D'Apprentissage Statistique / Machine Learning

II. Les types d'Apprentissage du Machine Learning :

A. Apprentissage supervisé :

On reprend cet exemple qui se base sur cette photo :

En **apprentissage supervisé**, est un algorithme qui s'appuie sur des données d'entrée étiquetées pour apprendre une fonction qui produit une sortie appropriée lorsqu'elle reçoit de nouvelles données non étiquetées. Une fois entraîné, de prédire cette cible sur de nouvelles données non annotées. Dans notre exemple, les données d'entrée seraient des images et la cible (ou *target* en anglais) la catégorie de photos que vous voulez.

airplane

automobile

bird

cat

deer

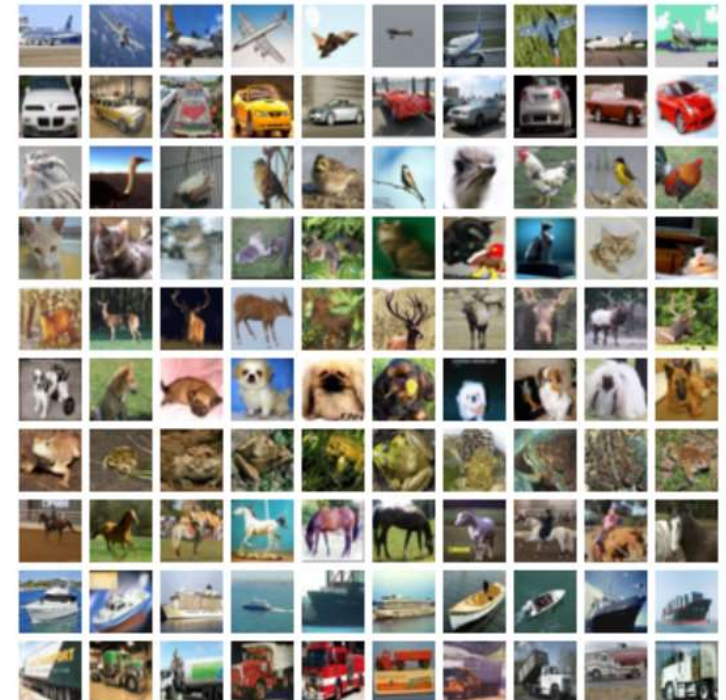
dog

frog

horse

ship

truck



Eléments de Théorie D'Apprentissage Statistique / Machine Learning

II. Les types d'Apprentissage du Machine Learning :

A. Apprentissage supervisé :

Exemple « Apprentissage supervisé » :

Imaginez un ordinateur est un enfant,

Quand on voit un ours, on crie « **ours !** »
Quand ce n'est pas un ours, on crie « **non, C'est pas un ours !** » Après avoir fait ceci plusieurs fois avec l'enfant, on lui montre une photo et on lui demande « **est-ce que c'est un ours ?** » et il nous répond correctement dans (la plupart du temps) « **ours !** » ou « **non, C'est pas un ours !** » selon ce que l'image est. **C'est l'apprentissage automatique supervisé.** Ces algorithmes sont utilisés pour résoudre des problèmes de *classification ou de régression.*



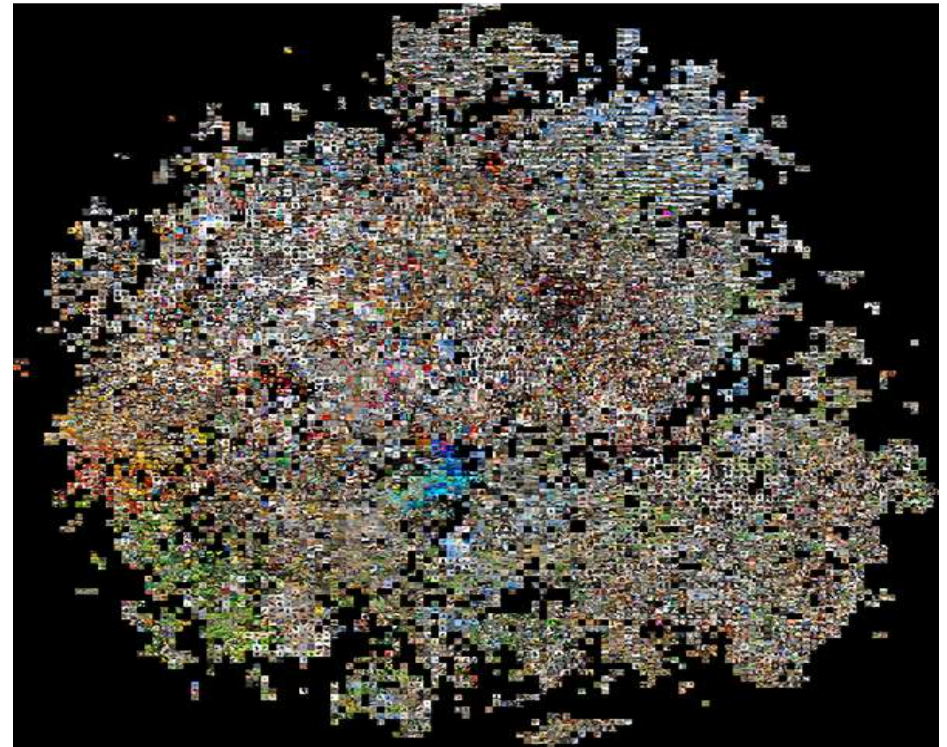
Eléments de Théorie D'Apprentissage Statistique / Machine Learning

II. Les types d'Apprentissage du Machine Learning :

A. Apprentissage non supervisé :

Un autre exemple des images qui ne sont pas étiquetées :

En apprentissage non supervisé, les données d'entrées ne sont pas annotées. Dans ce cas, l'algorithme d'entraînement s'applique à trouver seul les similarités et distinctions au sein de ces données, et à regrouper ensemble celles qui partagent des caractéristiques communes. Dans notre exemple, les photos similaires seraient ainsi regroupées automatiquement au sein d'une même catégorie.





Eléments de Théorie D'Apprentissage Statistique / Machine Learning

II. Les types d'Apprentissage du Machine Learning :

A. Apprentissage non supervisé :

- ❑ Un algorithme d'apprentissage automatique non supervisé utilise des données d'entrée sans aucune étiquette - en d'autres termes, aucun enseignant (étiquette) ne dit à l'enfant (ordinateur) quand il a raison ou quand il a fait une erreur pour qu'il puisse s'auto-corriger.
- ❑ Contrairement à l'apprentissage supervisé qui tente d'apprendre une fonction qui nous permettra de faire des prédictions compte tenu de certaines nouvelles données non étiquetées, l'apprentissage non supervisé tente d'apprendre la structure de base des données pour nous donner plus d'informations sur les données.



Eléments de Théorie D'Apprentissage Statistique / Machine Learning

II. Les types d'Apprentissage du Machine Learning :

Apprentissage supervisé vs non supervisé

Apprentissage supervisé

- ❑ Représentation Mathématique
on reçoit des données d'exemple annotées : $(x_1, y_1), (x_2, y_2), (x_3, y_3), \dots$ et on espère prédire la sortie sur de nouvelles observations : $x^* \rightarrow y^*$.
- ❑ Cas d'utilisation
Dans le cas où vous avez un problème et vous pouvez annoter précisément pour chaque observation la cible que vous voulez en sortie.

Apprentissage non supervisé

- ❑ Représentation Mathématique
on reçoit uniquement des observations brutes de variables aléatoires : $x_1, x_2, x_3, x_4, \dots$ et on espère découvrir la relation avec des variables latentes structurelles : $x_i \rightarrow y_i$.
- ❑ Cas d'utilisation
Dans le cas où vous essayez de mieux comprendre votre dataset ou d'identifier des comportements intéressants



Eléments de Théorie D'Apprentissage Statistique / Machine Learning

III. Les types d'Apprentissage supervisé :

- ❑ Une question typique posée au début, face à une grande variété d'algorithmes d'apprentissage automatique, est la suivante : « [Quel algorithme dois-je utiliser?](#) ».

=> La réponse à la question varie en fonction de nombreux facteurs, notamment :

1. La taille, la qualité et la nature. de données;
2. Le temps de calcul disponible;
3. L'urgence de la tâche;
4. Et qu'est-ce que vous voulez faire avec les données ?

- ❑ Même un scientifique expérimenté ne peut pas dire quel algorithme donnera les meilleurs résultats avant d'essayer différents algorithmes. Bien qu'il existe de nombreux autres algorithmes d'apprentissage automatique, ceux qui sont les plus populaires.

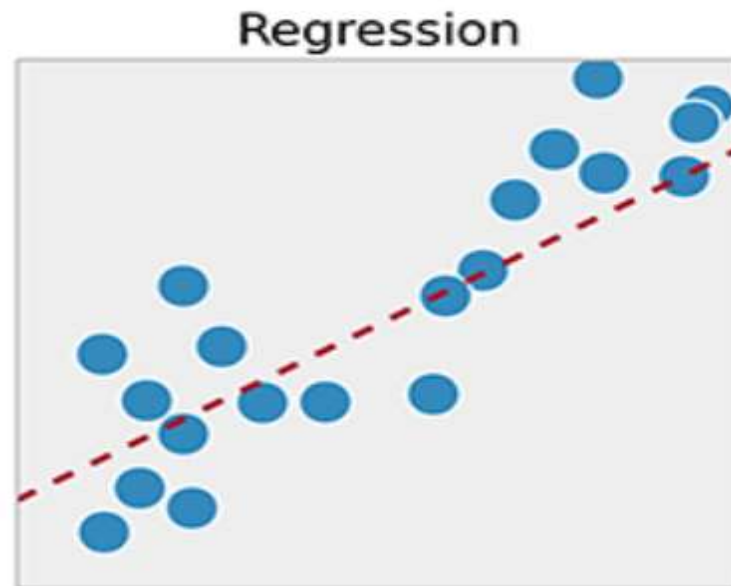
Eléments de Théorie D'Apprentissage Statistique / Machine Learning

III. Les types d'Apprentissage supervisé :

- ✓ Si le type des variables de sortie est quantitatif un nombre => on utilisera l'algorithme **de régression linéaire**
- ✓ Si le type des variables de sortie est qualitatif une catégorie => on utilisera l'algorithme de **classification**

A. La régression linéaire :

Par exemple, si je veux déterminer le coût par clic d'une publicité web, j'effectue **une régression**.





Eléments de Théorie D'Apprentissage Statistique / Machine Learning

III. Les types d'Apprentissage supervisé :

A. La régression linéaire :

La construction de la régression repose d'une part sur une modélisation des variables statistiques par **des variables aléatoires (réelles ou non)**. Elle consiste alors à formuler **un indicateur sur les valeurs de la variable expliquée dépendant uniquement des valeurs des variables explicatives**. On distingue essentiellement la nature de la variable expliquée si elle est quantitative, représentée ici par une variable aléatoire Y . Les variables explicatives seront notées par X_1, \dots, X_n . La fonction de régression est définie par un indicateur de la loi de Y conditionnellement aux valeurs des autres variables X_i sous forme:

$$f(x_1, \dots, x_n) = \mathbb{E}(Y \mid X_i = x_i)$$

Les algorithmes de régression prédisent les valeurs de sortie en fonction des caractéristiques d'entrée à partir des données introduites dans le système.

Eléments de Théorie D'Apprentissage Statistique / Machine Learning

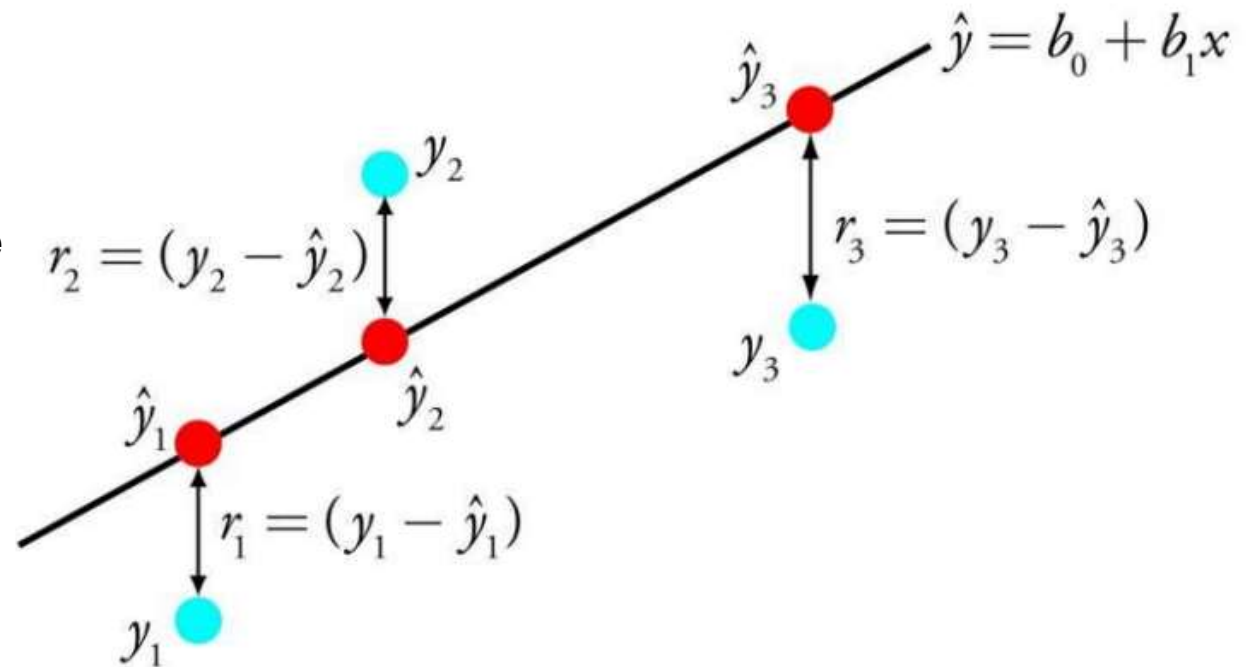
III. Les types d'Apprentissage supervisé :

A. La régression linéaire :

La représentation de la régression linéaire est une équation qui décrit une ligne qui correspond le mieux à la relation entre les variables d'entrée (x) et les variables de sortie (y), en recherchant des pondérations spécifiques pour les variables d'entrée appelées coefficients (b_i). En respectant la propriété du **moindre carrée** : $\min \|y_i - \hat{y}_i\|_2^2$

\hat{y}_i : variable à prédire

y_i : variable sortie





Eléments de Théorie D'Apprentissage Statistique / Machine Learning

III. Les types d'Apprentissage supervisé :

A. La régression linéaire :

Par exemple: $y = B_0 + B_1 * x$

- Nous allons prédire y étant donné l'entrée x et l'objectif de l'algorithme d'apprentissage par régression linéaire est de trouver les valeurs des coefficients B_0 et B_1 .
- Différentes techniques peuvent être utilisées pour apprendre le modèle de régression linéaire à partir de données, telles qu'une solution d'algèbre linéaire pour l'optimisation *des moindres carrés ordinaires* et *la descente de gradient*.
- La régression linéaire existe depuis plus de 200 ans et a fait l'objet de nombreuses études. L'utilisation de cette technique a pour règle de supprimer les variables très *similaires (corrélées)* et *d'éliminer le bruit de vos données*, si possible. **C'est une technique simple et rapide et un bon premier algorithme à essayer.**



Eléments de Théorie D'Apprentissage Statistique / Machine Learning

III. Les types d'Apprentissage supervisé :

B. La régression logistique « la classification » :

C'est la méthode de référence pour les problèmes de classification binaire (*problèmes avec deux valeurs de classe*).

La régression logistique s'apparente à la régression linéaire en ce sens que l'objectif est de trouver les valeurs des coefficients qui pondèrent chaque variable d'entrée. Contrairement

à la régression linéaire, la prédiction pour la sortie est transformée à l'aide d'une fonction

non linéaire => *La fonction Sigmoid*e appelée fonction logistique $f(x) : f(x) = \frac{1}{1 + e^{-x}}$

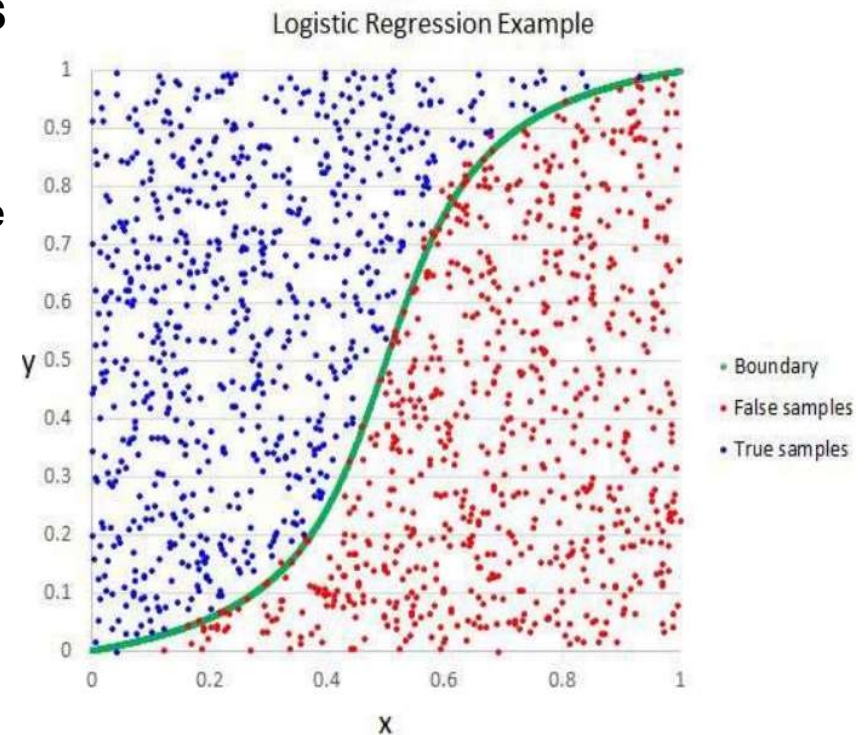
Eléments de Théorie D'Apprentissage Statistique / Machine Learning

III. Les types d'Apprentissage supervisé :

B. La régression logistique « la classification » :

La fonction logistique ressemble à un grand **S** et transformera toute valeur entre **0 et 1**.

Ceci est utile car nous pouvons appliquer une règle à la sortie de la fonction logistique pour aligner les valeurs sur **0 et 1** (par exemple, IF inférieur à 0,5, puis sortie 1) et prédire une valeur de classe.



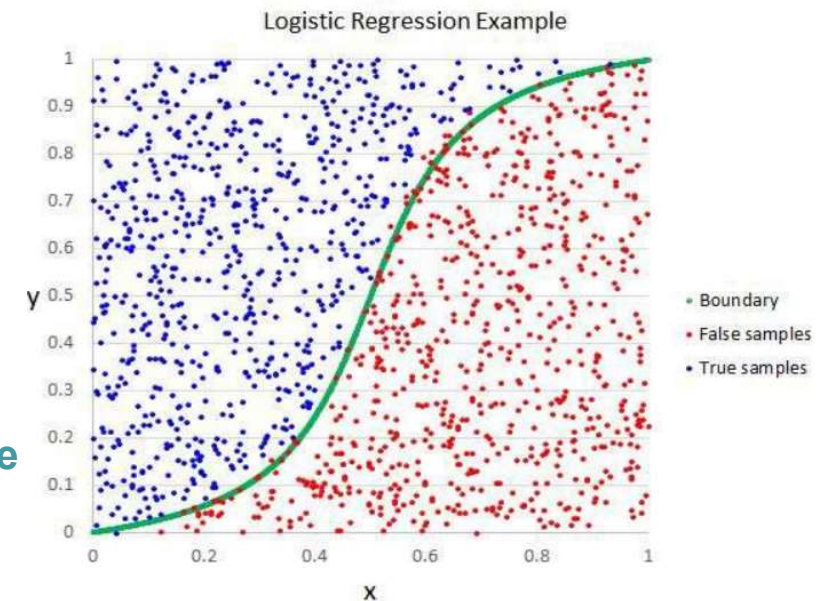
Eléments de Théorie D'Apprentissage Statistique / Machine Learning

III. Les types d'Apprentissage supervisé :

B. La régression logistique « la classification » :

En raison de la manière dont le modèle est appris, les prédictions faites par régression logistique peuvent également être utilisées comme la probabilité qu'une instance de données appartienne à **la classe 0** ou à **la classe 1**. Cela peut être utile pour les problèmes pour lesquels vous devez justifier davantage une prédiction.

Comme la régression linéaire, la régression logistique fonctionne mieux lorsque vous supprimez des attributs qui ne sont pas liés à la variable de sortie, ainsi que des attributs très similaires (**corrélés**) les uns aux autres. C'est un modèle **rapide à apprendre** et **efficace sur les problèmes de classification binaire**.





Eléments de Théorie D'Apprentissage Statistique / Machine Learning

III. Les types d'Apprentissage supervisé :

B. La régression logistique « la classification » :

Avantages	Inconvénients
Rapide à entrainer et prévoir	Pas très précis
Bon pour les petits problèmes de données de classification	Ne pas utiliser pour des données non linéaires
Facile à comprendre	Pas flexible pour adapter à des données complexes
	Le modèle finit parfois par sur-adapter



Eléments de Théorie D'Apprentissage Statistique / Machine Learning

III. Les types d'Apprentissage supervisé :

C. Analyse Discriminante Linéaire «LDA» :

- ❑ La régression logistique est un algorithme de classification traditionnellement limité aux problèmes de classification à deux classes. Si vous avez plus de deux classes, l'algorithme d'analyse linéaire discriminante est la technique de classification linéaire préférée.
- ❑ La représentation de LDA est assez simple. Pour une seule variable d'entrée, elle consiste des propriétés statistiques de vos données, calculées pour chaque classe comme:
 - ✓ La valeur moyenne pour chaque classe.
 - ✓ La variance calculée dans toutes les classes.

Figure 1 shows flow cytometry plots for the isolation of CD4⁺ Treg cells. The figure is divided into four panels. The top-left panel shows a plot of CD4 (y-axis, 0-4000) vs. CD25 (x-axis, 0-4000) with three gates labeled 1 (13.9%), 2 (18.9%), and 3 (36.3%). The bottom-left panel shows a plot of CD4 (y-axis, 0-4000) vs. CD25 (x-axis, 0-4000) with three gates labeled 1 (13.9%), 2 (18.9%), and 3 (36.3%). The top-right panel shows a plot of CD4 (y-axis, 0-4000) vs. CD25 (x-axis, 0-4000) with two gates labeled 1 (13.9%) and 2 (18.9%). The bottom-right panel shows a plot of CD4 (y-axis, 0-4000) vs. CD25 (x-axis, 0-4000) with two gates labeled 1 (13.9%) and 2 (18.9%).

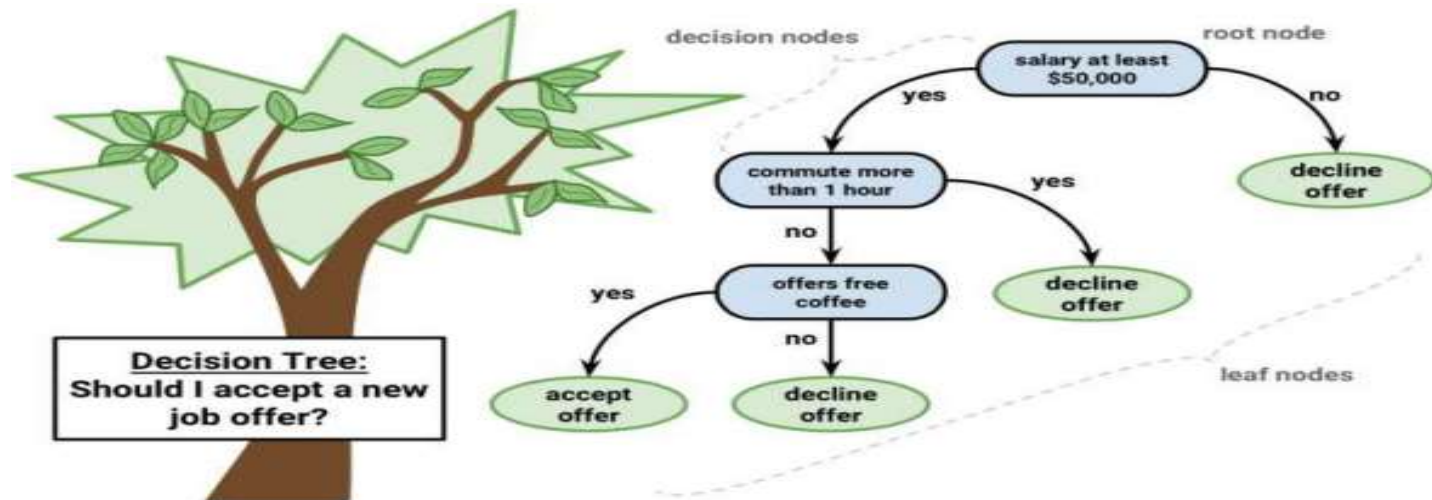
Eléments de Théorie D'Apprentissage Statistique / Machine Learning

III. Les types d'Apprentissage supervisé :

D. Arbres de classification et de régression :

Les arbres de décision constituent un type important d'algorithme pour la modélisation prédictive.

- ❑ La représentation du modèle d'arbre de décision est un arbre binaire. Ceci est votre arbre binaire d'algorithmes et de structures de données.
- ❑ Chaque nœud représente une variable d'entrée unique (x) et un point de partage sur cette variable (en supposant que la variable est numérique).





Eléments de Théorie D'Apprentissage Statistique / Machine Learning

III. Les types d'Apprentissage supervisé :

D. Arbres de classification et de régression :

- ❑ Les nœuds feuille de l'arbre contiennent une variable de sortie (y) utilisée pour effectuer une prédiction. Les prédictions sont effectuées en parcourant les divisions de l'arbre jusqu'à arriver à un nœud feuille et en sortie la valeur de classe à ce nœud feuille.
- ❑ Les arbres sont rapides à apprendre et très rapides pour faire des prédictions. En outre, ils sont souvent exacts pour un large éventail de problèmes et ne nécessitent aucune préparation particulière pour vos données.

Eléments de Théorie D'Apprentissage Statistique / Machine Learning

III. Les types d'Apprentissage supervisé :

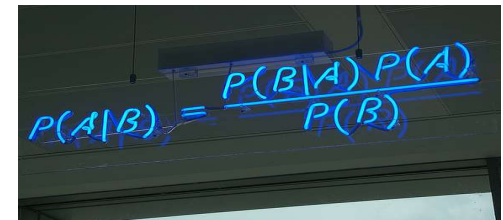
E. Naïve Bayes :

- ❑ Naïve Bayes est un algorithme simple mais étonnamment puissant pour la modélisation prédictive. C'est un **classifieur** assez **intuitif à comprendre**. Il se base sur le [théorème de Bayes des probabilités conditionnelles](#). Généralement, le **Naïve Bayes** est utilisé pour les classifications de texte (en se basant sur le nombre d'occurrences de mots).
- ❑ **Naïve Bayes** assume une hypothèse forte (**naïve**). En effet, il suppose que les variables sont indépendantes entre elles. Cela permet de simplifier le calcul des probabilités. Le modèle comprend deux types de probabilités qui peuvent être calculées directement à partir de vos données d'entraînement:

1) La probabilité de chaque classe;

et

2) La probabilité conditionnelle pour chaque classe étant donnée chaque valeur x.


$$P(A|B) = \frac{P(B|A) P(A)}{P(B)}$$

Une fois calculé, le modèle de probabilité peut être utilisé pour établir des prévisions pour les nouvelles données à l'aide du théorème de Bayes. Lorsque vos données ont une valeur réelle, il est courant de supposer [une distribution gaussienne \(courbe de cloche\)](#) afin de pouvoir estimer facilement ces probabilités.

Eléments de Théorie D'Apprentissage Statistique / Machine Learning

III. Les types d'Apprentissage supervisé :

E. Naïve Bayes :

Exemple :

Supposons qu'on ait un jeu de données sur 1000 fruits. On dispose de trois types : **Banane, Orange, et "autre"**. Pour chaque fruit, on a 3 caractéristiques :

- ❖ Si le fruit est **long** ou non
- ❖ S'il est **sucré** ou non
- ❖ Si sa couleur est **jaune** ou non

Notre jeu de données se présente comme suit : le symbole \neg signifie la négation (\neg froid = chaud)

Type	Long	petit (\neg long)	sucré	\neg sucré	Jaune	Pas jaune	Total
Banane	400	100	350	150	450	50	500
Orange	0	300	150	150	300	0	300
Autre fruit	100	100	150	50	50	150	200
Total	500	500	650	350	800	200	1000



Eléments de Théorie D'Apprentissage Statistique / Machine Learning

III. Les types d'Apprentissage supervisé :

E. Naïve Bayes :

Exemple :

L'idée du jeu est de prédire le type d'un fruit (orange, banane ou autre) qu'on n'a pas encore vu. Ceci en se basant sur ses caractéristiques.

Supposons que quelqu'un nous demande de lui donner le type d'un fruit qu'il a. Ses caractéristiques sont les suivantes :

- ☐ Il est jaune
- ☐ Il est long
- ☐ Il est sucré

Pour savoir s'il s'agit d'une banane, ou d'une orange ou d'un autre fruit, il faut qu'on calcule les trois probabilités suivantes :

- $P(\text{Banane} \mid \text{long, jaune, sucré})$: La probabilité qu'il s'agisse d'une banane sachant que le fruit est long, jaune et sucré
- $P(\text{Orange} \mid \text{long, jaune, sucré})$: La probabilité qu'il s'agisse d'une orange sachant que le fruit est long, jaune et sucré
- $P(\text{Autre fruit} \mid \text{long, jaune, sucré})$: La probabilité qu'il s'agisse d'un autre fruit sachant que ce dernier est long, jaune et sucré



Eléments de Théorie D'Apprentissage Statistique / Machine Learning

III. Les types d'Apprentissage supervisé :

E. Naïve Bayes :

Exemple :

Le type du fruit “inconnu” qu’on cherche à classifier sera celui où **on a la plus grande probabilité**. Selon la formule de Bayes, on fait ce calcul pour le fruit de Banane : (De la même manière pour les autres fruits) $P(\text{Orange} \mid \text{long, jaune, sucré})$ et $P(\text{Autre fruit} \mid \text{long, jaune, sucré})$

$$\begin{aligned} \blacksquare P(\text{Banane} \mid \text{long, jaune, sucré}) = \\ \frac{P(\text{Long} \mid \text{Banane}) * P(\text{Sucre} \mid \text{Banane}) * P(\text{Jaune} \mid \text{Banane}) * P(\text{Banane})}{P(\text{Long}) * P(\text{Sucre}) * P(\text{Jaune})} \end{aligned}$$

$$P(\text{Banane}) = \frac{\text{cardinal}(\text{Banane})}{\text{cardinal}(\text{Tous les fruits})} = \frac{50}{100} = 0.5$$

De même :

- $P(\text{Orange}) = 0.3$
- $P(\text{Autre fruit}) = 0.2$
- $P(\text{Long}) = 0.5$
- $P(\text{Sucre}) = 0.65$
- $P(\text{Jaune}) = 0.8$

et on a aussi : vu que les variables sont indépendantes

$$P(\text{Long} \mid \text{Banane}) = \frac{\text{cardinal}(\text{Banane ET Long})}{\text{cardinal}(\text{Banane})} = \frac{400}{500} = 0.8$$

La même chose pour $P(\text{sucré} \mid \text{Banane})$ et $P(\text{jaune} \mid \text{Banane})$



Eléments de Théorie D'Apprentissage Statistique / Machine Learning

III. Les types d'Apprentissage supervisé :

E. Naïve Bayes :

Exemple :

D'après un calcul, on remarque que la probabilité $P(\text{Banane} \mid \text{long, jaune, sucré})$ que notre fruit soit une **banane est largement plus grande que celle des autres probabilités**. On classifie notre fruit **inconnu** comme étant une banane.

Avantage	Inconvénient
Très rapide pour la classification	Supposition l'indépendance des variables
Possible à le faire même avec un petit jeu de données	



Eléments de Théorie D'Apprentissage Statistique / Machine Learning

III. Les types d'Apprentissage supervisé :

F. K-Nearest Neighbor :

- ❑ L'algorithme des k-plus proches voisins (KNN) est un algorithme d'apprentissage automatique supervisé simple et facile à mettre en œuvre qui peut être utilisé pour résoudre des problèmes de classification et de régression.
- ❑ Les arbres sont rapides à apprendre et très rapides pour faire des prédictions. En outre, ils sont souvent exacts pour un large éventail de problèmes et ne nécessitent aucune préparation particulière pour vos données.
- ❑ L'algorithme KNN suppose que des choses similaires existent à proximité. En d'autres termes, des choses similaires sont proches les unes des autres.



Eléments de Théorie D'Apprentissage Statistique / Machine Learning

III. Les types d'Apprentissage supervisé :

F. K-Nearest Neighbor :

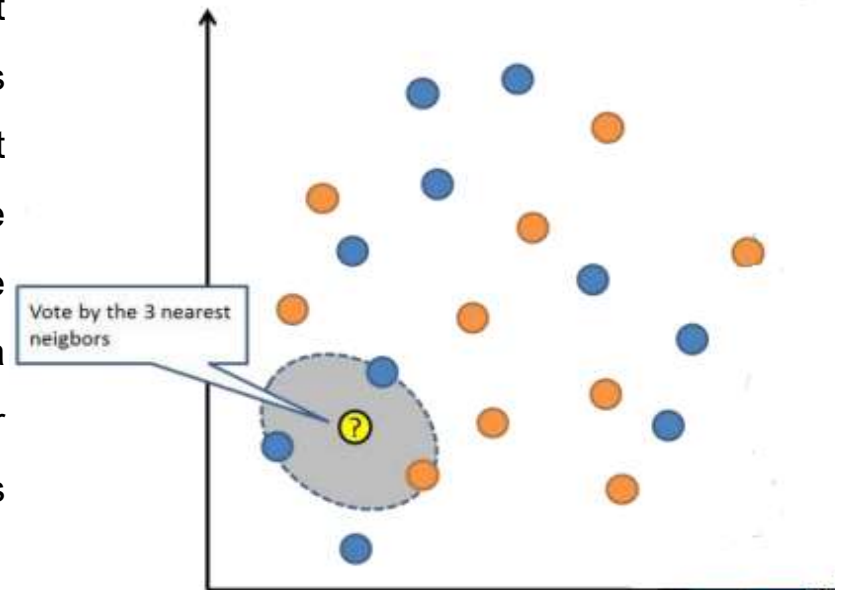
- ❑ Les prévisions sont établies pour un nouveau point de données en effectuant une recherche dans l'ensemble complet d'apprentissage pour les K instances les plus similaires (les voisins) et en résumant la variable de sortie pour ces K instances. Pour les problèmes de régression, il peut s'agir de la variable de sortie moyenne. Pour les problèmes de classification, il peut s'agir de la valeur de classe mode (ou la plus courante).
- ❑ KNN peut nécessiter beaucoup de mémoire ou d'espace pour stocker toutes les données, mais effectue uniquement un calcul (ou un apprentissage) lorsqu'une prévision est nécessaire, juste à temps. Vous pouvez également mettre à jour et gérer vos instances de formation au fil du temps pour que les prévisions restent exactes.

Eléments de Théorie D'Apprentissage Statistique / Machine Learning

III. Les types d'Apprentissage supervisé :

F. K-Nearest Neighbor :

- ❑ L'astuce consiste à déterminer la similarité entre les instances de données. La technique la plus simple si vos attributs ont tous la même échelle (tous en pouces, par exemple) consiste à utiliser la distance euclidienne, un nombre que vous pouvez calculer directement en fonction des différences entre chaque variable en entrée.
- ✓ L'idée de distance ou de proximité peut tomber en très grandes dimensions (beaucoup de variables d'entrée), ce qui peut affecter négativement les performances de l'algorithme sur votre problème. C'est ce qu'on appelle la malédiction de la dimensionnalité. Il vous suggère d'utiliser uniquement les variables d'entrée les plus pertinentes pour prédire la variable de sortie.





Eléments de Théorie D'Apprentissage Statistique / Machine Learning

III. Les types d'Apprentissage supervisé :

F. K-Nearest Neighbor :

Avantages	Inconvénients
Simple	Mémoire intensive coûteuse
Adaptable au problème	Toutes les données de formation peuvent être impliquées dans la prise de décision
Précis	Ralentissement des performances dû aux opérations IO
Facile à comprendre	Choisir une mauvaise mesure de distance peut produire des résultats inexacts
Utilise des arbres spatiaux pour améliorer les problèmes d'espace	



Eléments de Théorie D'Apprentissage Statistique / Machine Learning

III. Les types d'Apprentissage supervisé :

I. Random Forest :

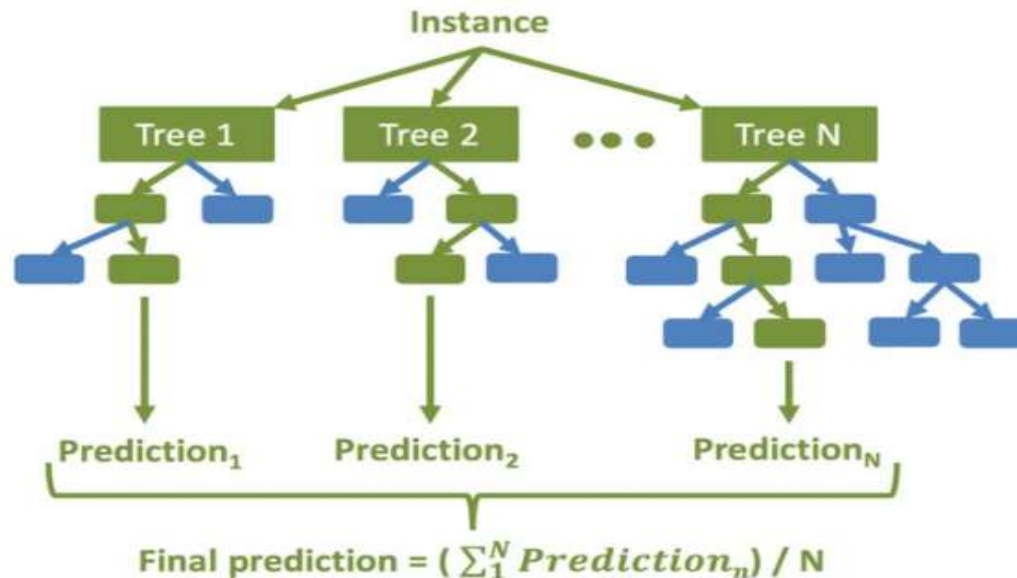
- **Random Forest** est l'un des algorithmes d'apprentissage automatique les plus populaires et les plus puissants. Il s'agit d'un type d'algorithme d'apprentissage automatique appelé « **Bootstrap Aggregation** » ou « **bagging** ».
- C'est un algorithme de classification qui réduit la variance des prévisions d'un arbre de décision seul, améliorant ainsi leurs performances. Pour cela, il combine de nombreux arbres de décisions construits aléatoirement et entraînés sur des sous-ensembles de données différents pour qu'il puisse effectuer un apprentissage en parallèle. Le nombre idéal d'arbres, est un paramètre important : qui peut aller jusqu'à plusieurs centaines voire plus, il est très variable et dépend du problème. Concrètement, chaque arbre de forêt aléatoire est entraîné sur un sous ensemble aléatoire de données selon le principe du **bagging**,

Eléments de Théorie D'Apprentissage Statistique / Machine Learning

III. Les types d'Apprentissage supervisé :

I. Random Forest :

- **Bagging** est une méthode statistique puissante permettant d'estimer une quantité à partir d'un échantillon de données. Comme un moyen. Vous prenez beaucoup d'échantillons de vos données, calculez la moyenne, puis vous faites la moyenne de toutes vos valeurs moyennes pour vous donner une meilleure estimation de la vraie valeur moyenne.



This figure was uploaded by [Tyler McCandless](#)



Eléments de Théorie D'Apprentissage Statistique / Machine Learning

III. Les types d'Apprentissage supervisé :

I. Random Forest :

- Dans la mise en sac, la même approche est utilisée, mais plutôt pour estimer des modèles statistiques entiers, le plus souvent des arbres de décision. Plusieurs échantillons de vos données d'entraînement sont prélevés, ensuite des modèles sont construits pour chaque échantillon de données. Lorsque vous devez effectuer une prévision pour les nouvelles données, chaque modèle en fait une prédiction et la moyenne des prédictions est calculée afin de fournir une meilleure estimation de la valeur de sortie réelle.



Eléments de Théorie D'Apprentissage Statistique / Machine Learning

III. Les types d'Apprentissage supervisé :

I. Random Forest :

Avantages	Inconvénients
Très précis	Lent à l'entraînement
bon point de départ pour résoudre un Problème	Sur-adapter
Flexible et peut s'adapter à une variété de données différentes	Ne convient pas aux petits échantillons
Rapide à exécuter	Petit changement dans les données d'entraînement => changement de modèle
Facile à utiliser	Parfois trop simple pour des problèmes très complexes
Utile pour les problèmes de régression et de classification	
Peut modéliser les valeurs manquantes	
Haute performance	

Les Réseaux de Neurones

I. Définition :

- ✓ Les réseaux de neurones artificiels sont des imitations simples des fonctions **d'un neurone dans le cerveau humain** pour résoudre des problématiques d'apprentissage de la machine (Machine Learning). Le neurone est une unité qui est exprimée généralement par une fonction sigmoïde :

$$f(x) = \frac{1}{1 + e^{-x}}$$

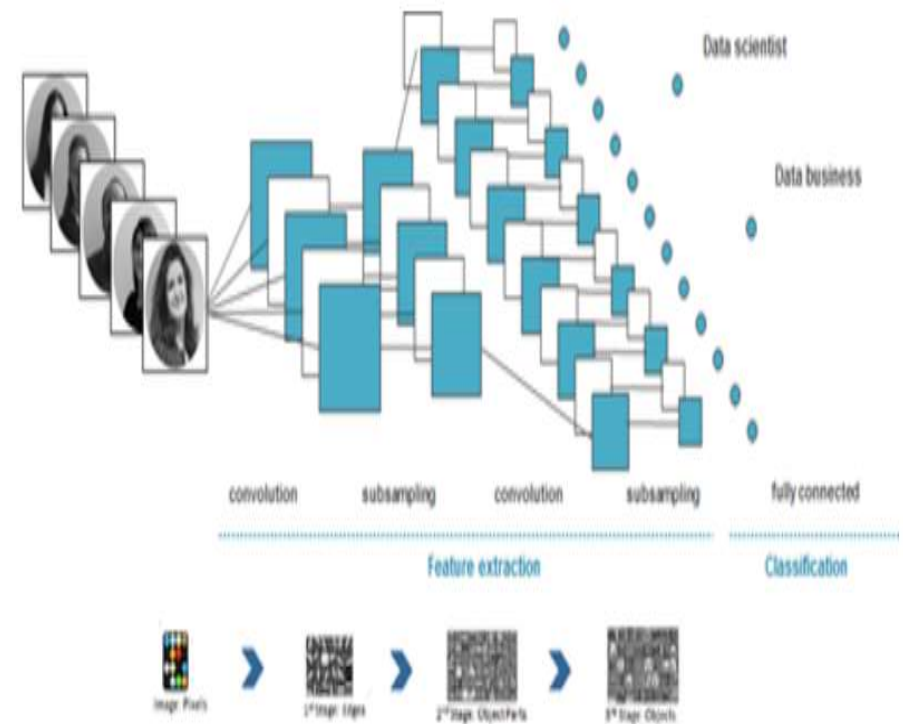
- ✓ ***Pourquoi nous utilisons des réseaux de neurones ?*** La réponse est plutôt simple dans le sens où les réseaux de neurones s'avèrent plus performants que les techniques de régressions pour des tâches de Machine Learning.

Les Réseaux de Neurones

I. Définition :

- ✓ Les domaines d'application des réseaux neuronaux sont souvent caractérisés par une relation entrée-sortie de la donnée d'information :

1. La reconnaissance d'image
2. Les classifications de textes ou d'images
3. Identification d'objets
4. Prédiction de données
5. Filtrage d'un set de données

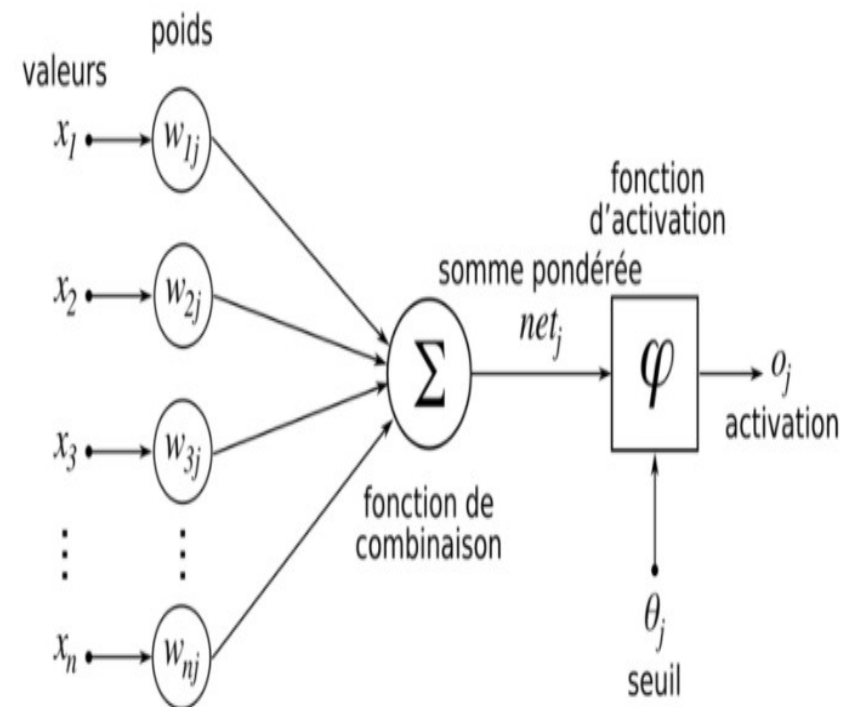


Les Réseaux de Neurones

II. L'Architecture d'un Réseau Neuronal :

- ✓ Un réseau de neurones peut prendre des formes différentes selon l'objet de la donnée qu'il traite et selon sa complexité et la méthode de traitement de la donnée.
- ✓ Les architectures ont leurs forces et faiblesses et peuvent être combinées pour optimiser les résultats. Le choix de l'architecture s'avère ainsi crucial et il est déterminé principalement par l'objectif.
- ✓ Un réseau de neurones est en général composé d'une succession de couches dont chacune prend ses entrées sur les sorties de la précédente. Chaque couche (i) est composée de N_i neurones, prenant leurs entrées sur les N_{i-1} neurones de la couche précédente.

À chaque synapse est associé un poids synaptique, de sorte que les N_{i-1} sont multipliés par ce poids, puis additionnés par les neurones de niveau i,

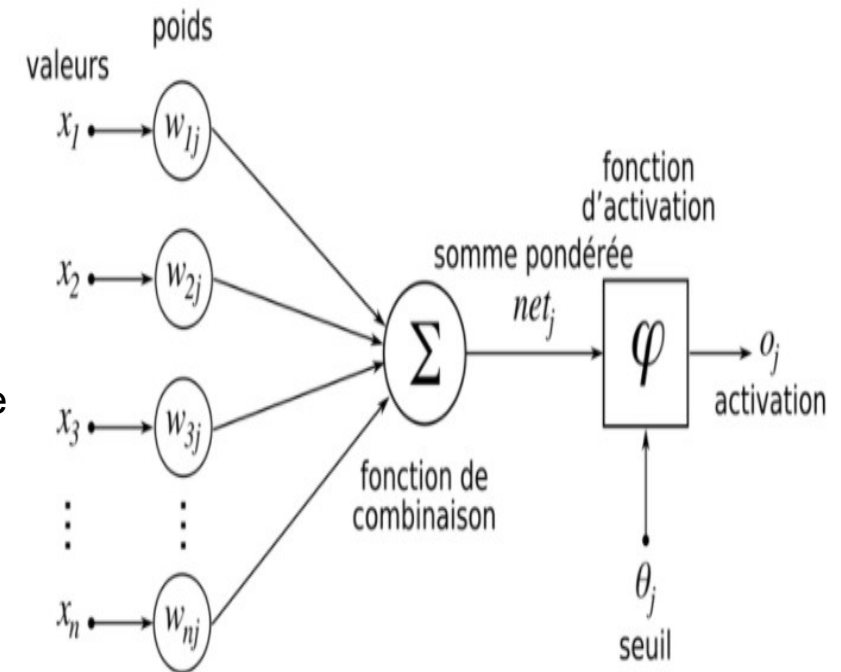


Les Réseaux de Neurones

II. L'Architecture d'un Réseau Neuronal :

Fonction de combinaison :

- ✓ Considérons un neurone quelconque. Il reçoit des neurones en amont un certain nombre de valeurs via ses connexions synaptiques, et il produit une certaine valeur en utilisant une fonction de combinaison. Cette fonction peut être formalisée comme :
- ✓ **Sous les réseaux de type MLP (multi-layer perceptron)** ils calculent une combinaison linéaire des entrées, c'est-à-dire que la fonction de combinaison renvoie le produit scalaire entre le vecteur des entrées et le vecteur des poids synaptiques.
- ✓ **Sous les réseaux de type RBF (radial basis function)** ils calculent la distance entre les entrées, c'est-à-dire que la fonction de combinaison renvoie la norme euclidienne du vecteur issu de la différence vectorielle entre les vecteurs d'entrées.



Les Réseaux de Neurones

II. L'Architecture d'un Réseau Neuronal :

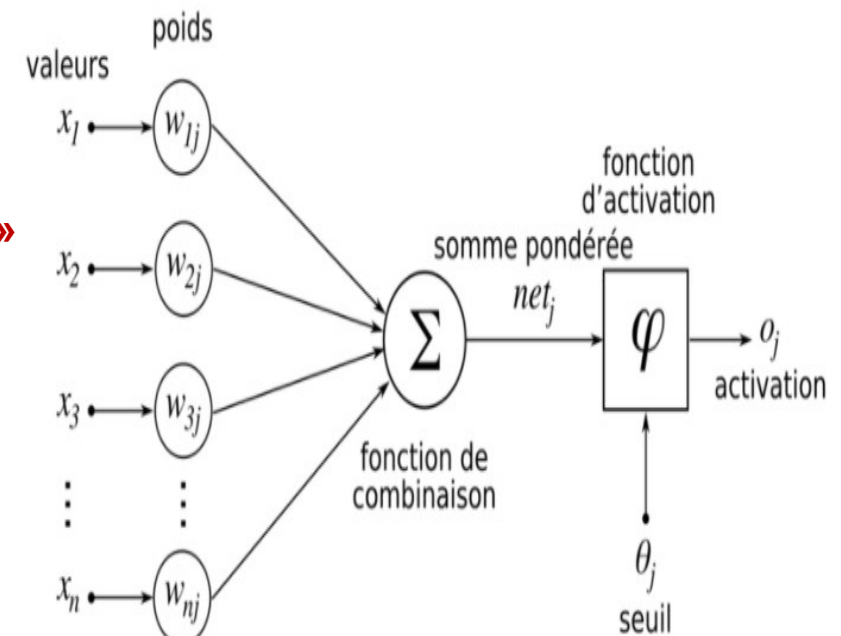
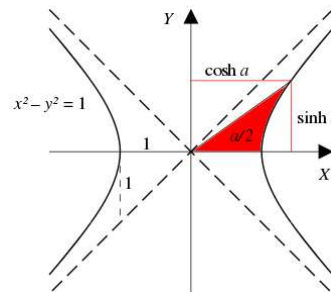
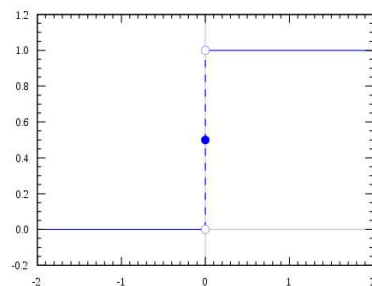
Fonction d'activation :

La fonction d'activation (ou fonction de transfert) sert à introduire une non-linéarité dans le fonctionnement du neurone. Les fonctions d'activation présentent généralement trois intervalles :

- ✓ En dessous du seuil, où la sortie vaut 0 ou -1 : le neurone est non-actif ;
- ✓ Aux alentours du seuil, une phase de transition ;
- ✓ Au-dessus du seuil, la sortie vaut 1 le neurone est actif

Des exemples classiques de fonctions d'activation sont :

- ✓ La fonction **sigmoïde** « la plus populaire »
- ✓ La fonction **tangente hyperbolique**.
- ✓ La fonction de **Heaviside**.





Les Réseaux de Neurones

II. L'Architecture d'un Réseau Neuronal :

- ✓ Selon le calcul de la fonction d'activation effectué, le neurone propage son nouvel état interne sur son axone. Dans un modèle simple, la fonction neuronale est simplement une fonction d'activation :

**1 si la somme pondérée dépasse un certain seuil ;
0 sinon.**

- ⇒ Dans un modèle plus riche, le neurone fonctionne avec des nombres réels (souvent compris dans l'intervalle $[0,1]$ ou $[-1,1]$). On dit que le réseau de neurones passe d'un état à un autre lorsque tous ses neurones recalculent en parallèle leur état interne, en fonction de leurs entrées.
- ✓ Un réseau de neurones peut prendre des formes différentes selon l'objet de la donnée qu'il traite et selon sa complexité et la méthode de traitement de la donnée. Les architectures ont leurs forces et faiblesses et peuvent être combinées pour optimiser les résultats. Le choix de l'architecture s'avère ainsi crucial et il est déterminé principalement par l'objectif.
- ✓ Les architectures de réseaux neuronaux peuvent être divisées en 4 grandes familles :

A. Réseaux de neurones Feed forwarded

B. Réseaux de neurones à résonance

C. Réseaux de neurones récurrent (RNN)

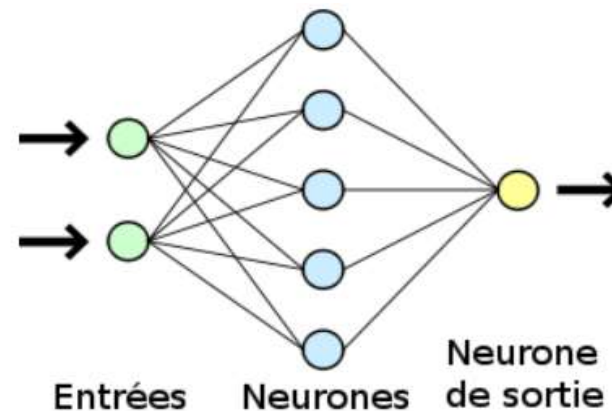
D. Réseaux de neurones auto-organisés

Les Réseaux de Neurones

III. Les Types de Réseaux Neuronaux :

A. Les réseaux de neurones feed-forwarded

- ✓ feed-forwarded (propagation avant) signifie tout simplement que la donnée traverse le réseau d'entrée à la sortie sans retour en arrière de l'information.
- ✓ Typiquement, dans la famille des réseaux à propagation avant, on distingue les réseaux monocouches (perceptron simple) et les réseaux multicouches (perceptron multicouche)
- ✓ Le perceptron simple est dit simple parce qu'il ne dispose que de deux couches ; la couche en entrée et la couche en sortie. Le réseau est déclenché par la réception d'une information en entrée. Le traitement de la donnée dans ce réseau se fait entre la couche d'entrée et la couche de sortie qui sont toutes reliées entre elles. Le réseau intégral ne dispose ainsi que d'une matrice de poids. Le fait de disposer d'une seule matrice de poids limite le perceptron simple à un classificateur linéaire permettant de diviser l'ensemble d'informations obtenues en deux catégories distinguées.

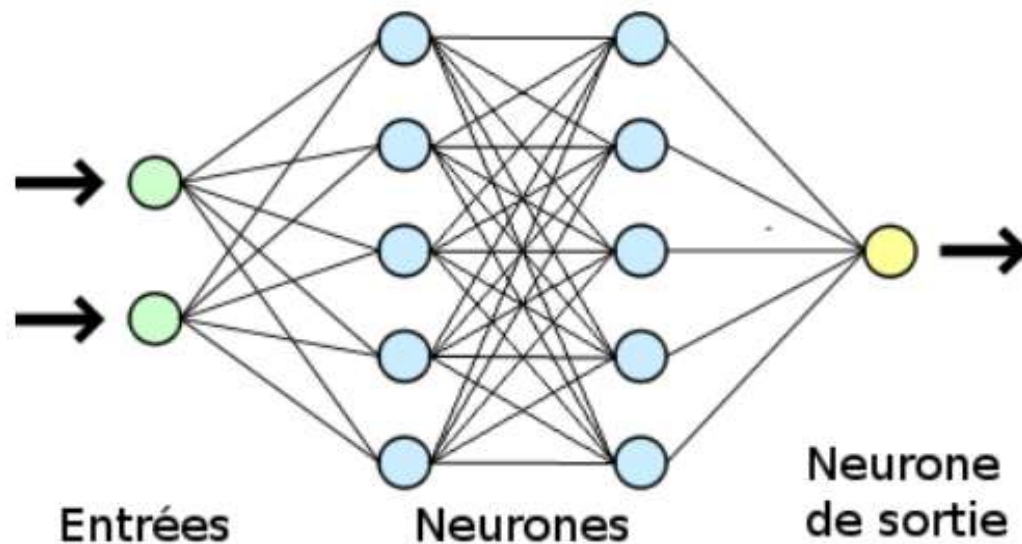


Les Réseaux de Neurones

III. Les Types de Réseaux Neuronaux :

A. Les réseaux de neurones feed-forwarded

- ✓ Le perceptron multicouche se structure de la même façon. L'information entre par une couche d'entrée et sort par une couche de sortie. À la différence du perceptron simple, le perceptron multicouche dispose entre la couche en entrée et la couche en sortie une ou plusieurs couches dites « **cachées** ». Le nombre de couches correspond aux nombres de matrices de poids dont disposent le réseau. Un perceptron multicouche est donc mieux adapté pour traiter les types de fonctions non-linéaires.





Les Réseaux de Neurones

III. Les Types de Réseaux Neuronaux :

A. Les réseaux de neurones feed-forwarded

Qu'est-ce que c'est que le Deep Learning ?

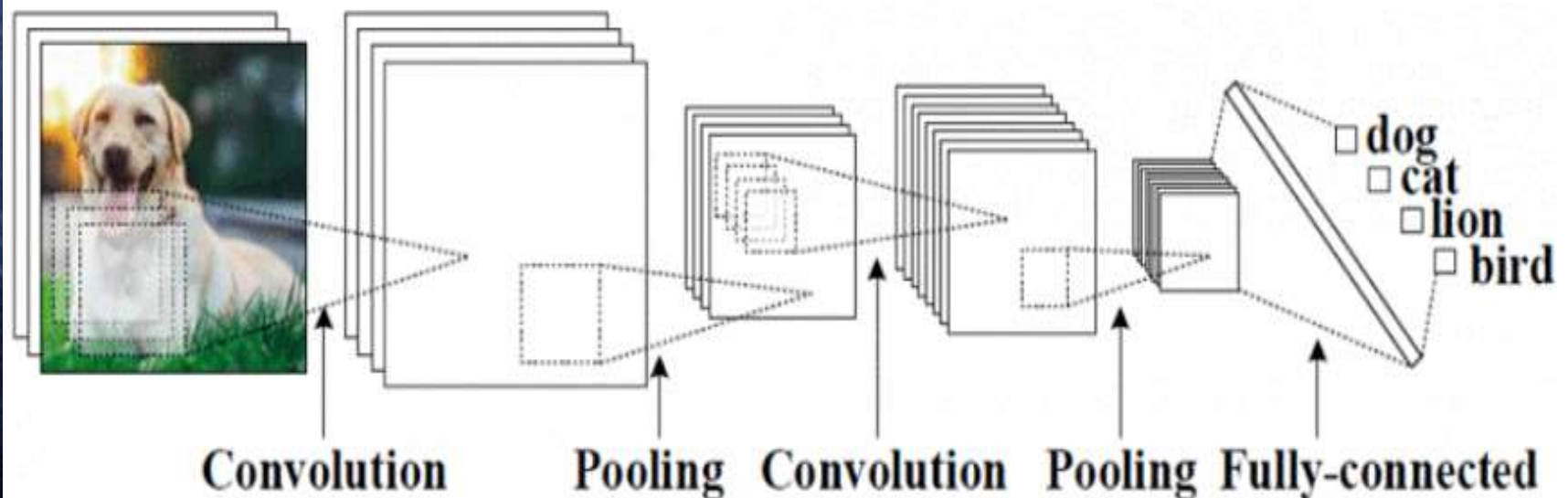
- ✓ **L'apprentissage profond (Deep Learning)** est une notion issue du fait que les réseaux de neurones disposaient de plus en plus **de couches cachées** et **que le nombre élevé de couches devenait une source de problèmes**. En effet, à partir d'un nombre de couches, le réseau neuronal n'était plus capable d'assimiler les informations et d'apprendre correctement.
- ✓ Des solutions ont été apportées à ces problèmes et les réseaux de neurones sont de plus en plus dotées de couches multiples et capables d'apprendre. Tous ces types de réseaux de neurones peuvent être regroupés sous la notion « **Deep Learning** ».

Les Réseaux de Neurones

III. Les Types de Réseaux Neuronaux :

A. Les réseaux de neurones feed-forwarded

- ✓ Pour le traitement d'informations complexes et très variées, il est envisageable de créer plusieurs réseaux de neurones distincts dédiés à traiter chacun une partie de l'information. Ces réseaux de neurones sont appelés des réseaux neuronaux convolutifs (**Convolutional Neural Networks**). Ces réseaux peuvent être imaginés comme une compilation d'un segment d'informations pour au final traiter l'ensemble de l'information (par exemple le traitement d'image, de vidéos, de textes).

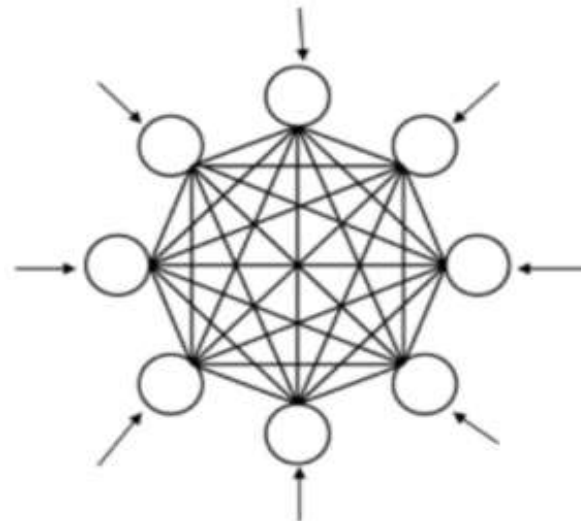


Les Réseaux de Neurones

III. Les Types de Réseaux Neuronaux :

B. Les réseaux de neurones récurrents :

- ✓ Les Réseaux de Neurones récurrents traitent l'information en cycle. Ces cycles permettent au réseau de traiter l'information plusieurs fois en la renvoyant à chaque fois au sein du réseau.
- ✓ La force des Réseaux de neurones récurrents réside dans leur capacité de prendre en compte des informations contextuelles suite à la récurrence du traitement de la même information. Cette dynamique auto-entretient le réseau.



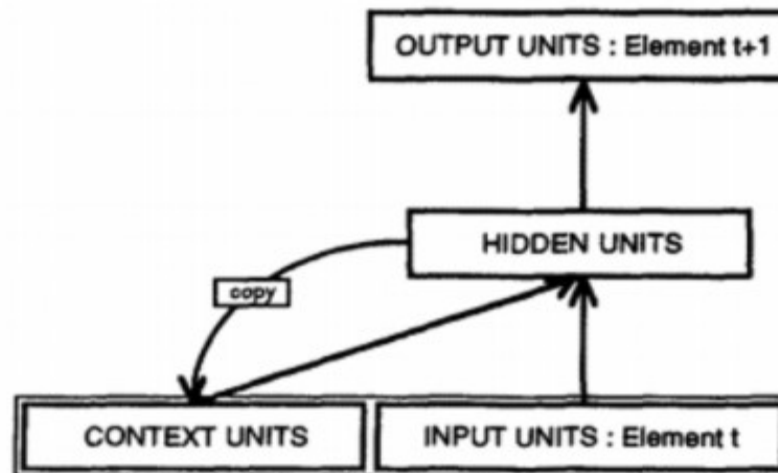
Ref :
http://www.scilog.fr/intelligence-mecanique/wp-content/blogs.dir/135/files/hopfield_Elman-1.png

Les Réseaux de Neurones

III. Les Types de Réseaux Neuronaux :

B. Les réseaux de neurones récurrents :

- ✓ Les Réseaux de neurones récurrents se composent d'une ou plusieurs couches. Le modèle de Hopfield (réseau temporel) est le réseau de neurones récurrent d'une seule couche le plus connu.
- ✓ Les Réseaux de neurones récurrents à couches multiples revendiquent quant à eux la particularité de posséder des couples (entrée/sortie) comme les perceptrons entre lesquels la donnée véhicule à la fois en propagation en avant et en rétro propagation



Ref :
http://www.scilog.fr/intelligence-mecanique/wp-content/blogs.dir/135/files/hopfield_Element-1.png



Les Réseaux de Neurones

III. Les Types de Réseaux Neuronaux :

C. Les réseaux de neurones à résonance :

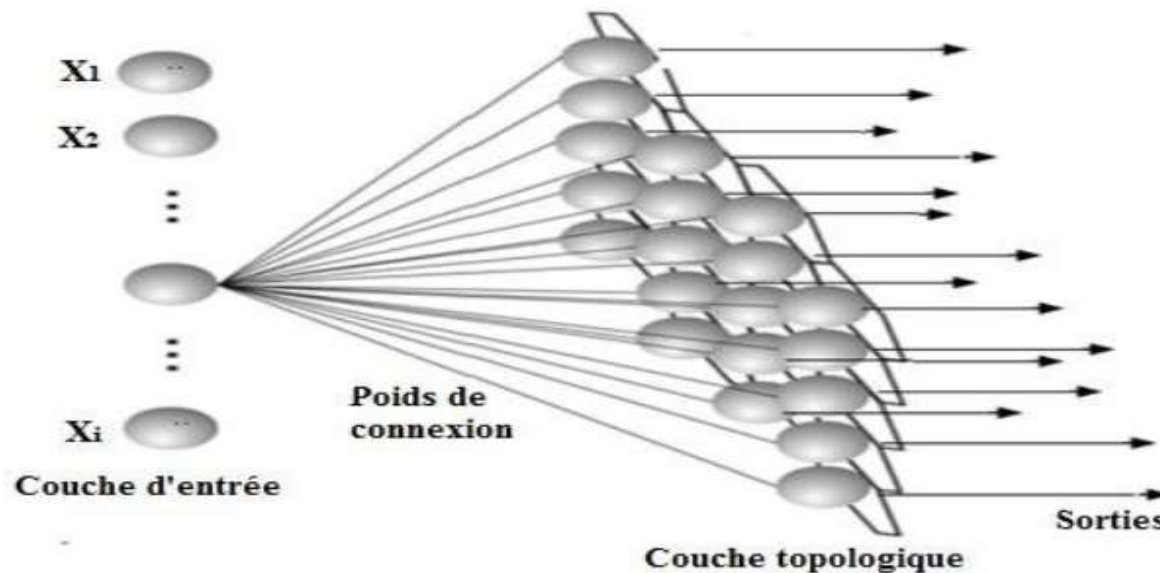
- ✓ Au sein des réseaux de neurones à résonance, l'activation de tous les neurones est renvoyée à tous les autres neurones au sein du système. Ce renvoi provoque des oscillations, d'où la raison du terme résonance.
- ✓ Ce réseau de neurones peut prendre différentes formes avec des degrés de complexité plutôt élevés. Prenons l'exemple d' **ART (Adaptative Resonance Theory)** c'est un réseau à apprentissage par compétition. Le problème majeur qui se pose dans ce type de réseaux est le dilemme « **stabilité/plasticité** ». En effet, dans un apprentissage par compétition, rien ne garantit que les catégories formées vont rester stables. La seule possibilité, qui existe, serait que le coefficient d'apprentissage tende vers zéro, **mais le réseau perdrait alors sa plasticité**. Dans les ART, les vecteurs de poids ne seront adaptés que si l'entrée fournie est suffisamment proche, d'un prototype déjà connu par le réseau d'où la notion de la résonance

Les Réseaux de Neurones

III. Les Types de Réseaux Neuronaux :

D. Les réseaux de neurones auto-organisés :

- ✓ Les Réseaux de neurones auto-organisés sont surtout adaptés pour le traitement de d'informations spatiales. Par des méthodes d'apprentissage non-supervisé, les réseaux neuronaux auto-organisés sont capables d'étudier la répartition de données dans des grands espaces comme par exemple pour des problématiques de clusterisation ou de classifications.
- ✓ Le modèle le plus connu de ce type de réseaux de neurones est sans doute la carte auto-organisatrice de Kohonen :



Ref :
https://www.researchgate.net/figure/Car-te-auto-organisatrice-de-Kohonen-Chaque-sphere-symbolise-un-neurone-de-la-couche_fig2_304559519



Les Réseaux de Neurones

IV. La nouvelle tendance du Réseaux de Neurones :

La 3^{ème} génération des réseaux de neurones « Spiking Neural Networks » :

- ✓ Elle vise à rapprocher la neuroscience et le Machine Learning. L'idée consiste à représenter les différentes procédures biologiques qui sont capables de traiter des événements ponctuels au lieu d'événements à valeur continue, c'est-à-dire, « lorsqu'un neurone dépasse une certaine valeur, il émet un signal et la valeur retombe à son potentiel initial ». A titre d'exemple : **le phénomène neurologique d'épilepsie**. C'est pour cette raison, que les réseaux de neurones de troisième génération veulent prendre en considération.
- ✓ Dès lors, les réseaux de neurones de 3^{ème} génération changent complètement l'objet de l'étude de Machine Learning qui se basait initialement sur une suite de valeur continue. Ceci mentionne, que « **Spiking Neural Networks** » serait synonyme de régression plutôt que de progrès. Le fait de saisir les pics de valeurs augmente la capacité du traitement des données spatio-temporelles. C'est pour cette raison il s'avère bien plus performante que les générations précédentes.



Les Réseaux de Neurones

IV. La nouvelle tendance du Réseaux de Neurones :

La 3^{ème} génération des réseaux de neurones « Spiking Neural Networks » :

- ✓ A cause de la méconnaissance initiale du fonctionnement de la méthode d'apprentissage du cerveau humain, l'élaboration d'une méthode d'apprentissage supervisé pour ce genre de réseau de neurones reste encore compliquée. Bien qu'on sache représenter le phénomène biologique pour saisir les pics des valeurs, les recherches n'ont pour l'instant pas pu reproduire le fonctionnement intrinsèque responsable de ces pics.
- ✓ Les « **Spike Neuronal Networks** » sont également très énergivores et ils demandent à simuler différents calculs d'équations. Vu que, la durée de vie de l'application des réseaux de neurones demeure très incertaine sachant qu'ils existent principalement que dans le discours théorique mais les applications pratiques se font rares.



La Reconnaissance des Formes (RdF)

I. Définition :

- ✓ Selon J.P. Haton : « *La RdF s'intéresse à la conception et à la réalisation de systèmes (matériels et logiciels) capables **de percevoir**, et dans une certaine mesure, **d'interpréter des signaux** captés dans le monde physique* »

Autrement dit :

- ✓ « **Reconnaissance des formes (pattern recognition)** = sert à identifier à quelle catégorie appartient une « **forme** » décrite par des données brutes »
- ✓ « **Forme** » = **observation** (ou partie d'une observation, ou ensemble d'observations)

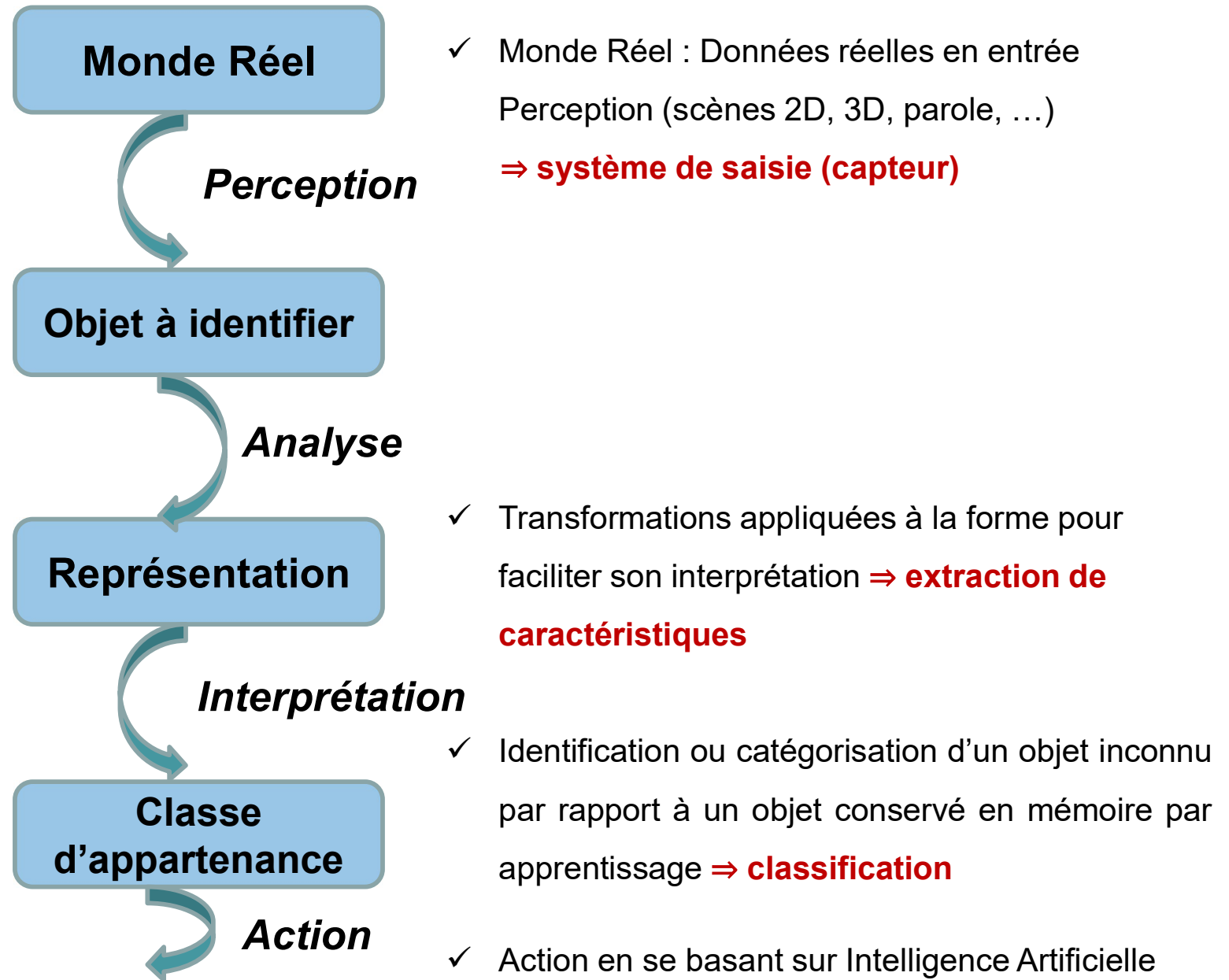
Exemples : des valeurs de variables décrivant un état clinique ; partie correspondant à un visage dans une image ; ensemble des valeurs prises par le cours d'une action sur une journée

Ses objectifs :

- ✓ Concevoir des systèmes automatiques ou semi-automatiques qui reconnaissent les formes qu'on leur présente pour avoir une compréhension automatique (IA)
- ✓ L'homme est le plus parfait des systèmes de RdF
- ✓ Reconstitution sur « machine » des fonctions typiquement humaines comme :
 - **La perception**
 - **L'analyse et représentation**
 - **L'interprétation**

La Reconnaissance des Formes (RdF)

II. Procéssus de la RdF :



La Reconnaissance des Formes (RdF)

II. Procéssus de la RdF :

+

Qualité
d'information

-

- ✓ Prétraitements :
binarisation, localisation, segmentation,
élimination du bruit, normalisation, ...
- ✓ Extraction des informations pertinentes :
 - *parole: fréquence fondamentale, harmoniques, énergie, ... quantité*
 - *mage: occlusions, concavités, contours, fins de trait, ... d'info.*
- ✓ Représentation de ces informations en vue de leur classification vecteur, graphe, chaîne
- ✓ Classification de la forme
apprentissage/décision, combinaison/fusion

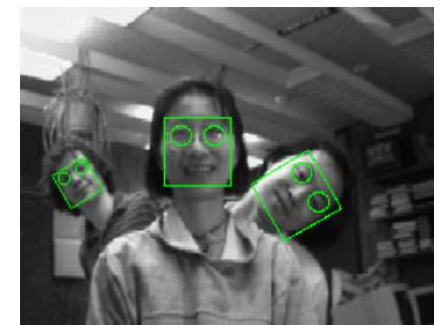
La Reconnaissance des Formes (RdF)

III. Système de RdF :

- ✓ Plusieurs Composantes d'un système de RdF :
 - Mécanique (satellite, bras d'un robot, dépileur, ...)
 - Saisie (caméra, barrette CCD, scanner, micro, ...)
 - Electronique (carte mémoire d'images, CPU, archivage, ...)
 - ALGORITHMIQUE
- ✓ Interaction de la RdF avec d'autres disciplines comme :
 - biologie / neuropsychologie
 - physique / mécanique / optique / capteurs
 - proba / stat / analyse de données
 - informatique / parallélisme / systèmes distribués
 - automatique / commande / robotique
 - traitement du signal et des images

➤ **Exemple : Détection des objets**

- ✓ Détecter des « objets » d'une certaine catégorie dans des images ou des vidéos



<http://vasc.ri.cmu.edu/NNFaceDetector/>

La Reconnaissance des Formes (RdF)

III. Application de RdF :

Applications

	Domaines d'étude	Applications
Traitement Du Signal	Reconnaissance de la parole	Bureautique, commande vocale, ordinateur sans clavier ni souris
	Reconnaissance du locuteur	Banques, commerces
	Électrocardiogramme, électro-encéphalogramme	Médical
Traitement Des Images	Reconnaissance des caractères, du manuscrit	Saisie de texte, bureautique, tri postal, compression télécopie, chèque
	Reconnaissance des signatures	Banques, commerces
	Reconnaissance des empreintes digitales, des visages	Banques, commerces, police
	Analyse de radiographies, échographies, reconnaissance chromosomes, comptage globules	Contrôles systématiques de santé
	Détection de défauts circuits intégrés, pièces métalliques, manufacturées	Contrôle de qualité industrielle
	Identification d'objets	Tri d'objets industriels, surveillance militaire
	Localisation d'objets	Guidage de robots industriels, guidage de missiles
	Analyse d'images de satellite	Météorologie, agriculture, ressources terrestres, surveillance militaire
	Analyse de photos aériennes	Agriculture surveillance militaire
	Analyse d'échos radar	Poursuite de cibles, pilotage missiles

La Reconnaissance des Formes (RdF)

IV. Application d'image en RdF :

Applications en Image

Domaines	Objets	Nature Image	Tâches	Références
Robotique	Scènes 3D	Luminence Rayons X	Identifier et décrire un objet dans une scène Tâches industrielles (peinture, métro, ...)	Modèles d'objets Modèles de réflexion de la lumière
Images aériennes	Terrains, bâtiments	Luminence Infra-rouge Radar	Restitution d'images, contrôle des cultures, météorologie, surveillance, repérage des cibles	Cartes Modèles géométriques
Astronomie	Etoiles, planètes	Luminence	Restitution de l'évolution Analyse chimique	Modèles géométriques
Médical	Organes du corps humain	Rayons X Ultra-sons Luminence ECG-EEG	Recherche d'anomalies Planification d'interventions Analyse ECG, EEG	Modèles anatomiques
Reconnaissance Caractères	Enveloppes, chèques, cartes, documents, textes libres	Luminence	Tri du courrier Traitement des chèques Indexation courriers Archivage	Modèles géométriques
Sismique	Cartes du sous-sol	Ultra-sons	Recherche du pétrole Tremblement de terre	Modèles souterrains



La Reconnaissance des Formes (RdF)

V. Les Descripteurs statistiques :

Un descripteur est une **quantité mesurable ou calculable** qui permet de décrire en partie un objet, un signal ou une donnée. Le **descripteur statistique** est toute information qui caractérise une ou un ensemble des variables aléatoires comme : **comme la moyenne, la variance, de même que les moments d'ordre plus élevé, apportent une information supplémentaire à faible coût**. Si on applique les descripteurs statistiques pour reconnaître la voix on a par exemple :

A. Le taux de passage par zéro (zero crossing rate) :

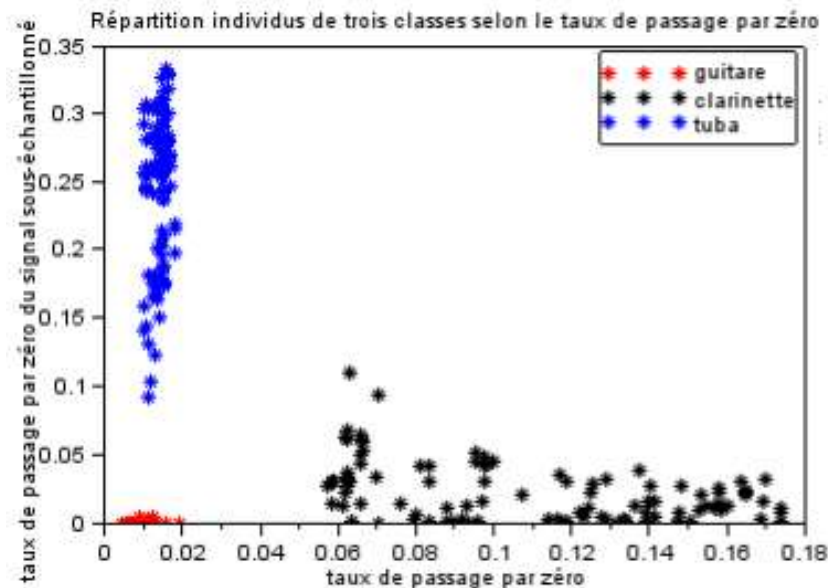
Le taux de passage par zéro ZCR est un descripteur intéressant. Il est évidemment corrélé à la fréquence fondamentale, mais donne aussi notamment une information sur le bruit présent dans le signal :

$$ZCR = \sum_{2 \leq i \leq N} \frac{|sign(i_i) - sign(i_{i-1})|}{2(N-1)}$$

La Reconnaissance des Formes (RdF)

V. Les Descripteurs statistiques :

A. Le taux de passage par zéro (zero crossing rate) :



Ces trois classes sont clairement séparées grâce à ces deux descripteurs. Les taux de passage par zéro des signaux de guitare sont très faibles à cause de modulations très basses fréquences qui décentrent localement le signal.

La Reconnaissance des Formes (RdF)

V. Les Descripteurs statistiques :

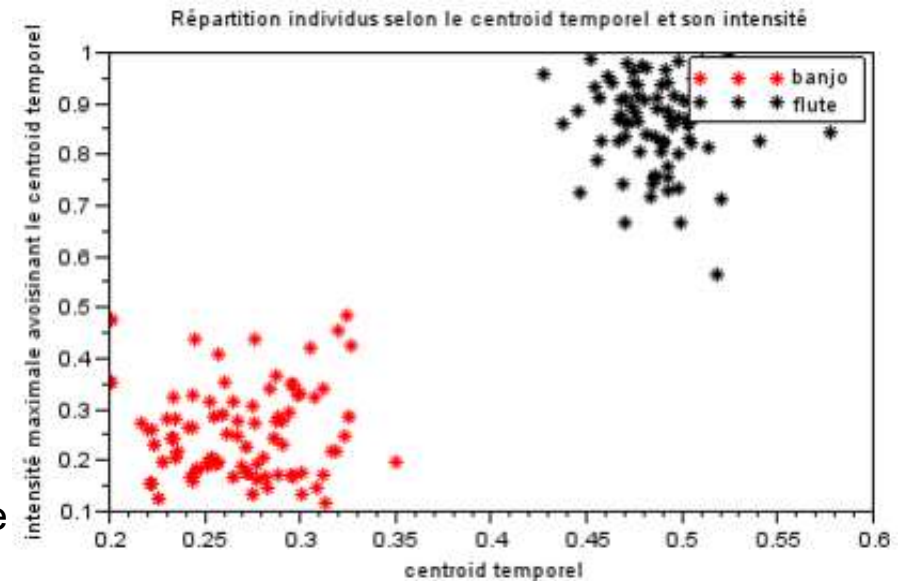
B. Centroïd (moment d'ordre1) : temporel et spectral:

Le centroïd, temporel ou spectral, est un des descripteurs statistiques les plus souvent utilisés. Le centroïd temporel permet de décrire l'enveloppe temporelle et le centroïd spectral, appelé brillance, Le centroïd est l'indice (temps ou fréquence) moyen :

$$C = \frac{\sum_{1 \leq i \leq N} t |i_t|}{\sum_{1 \leq i \leq N} |i_t|}$$

D'après cette classification on constate que

le banjo est un son percussif, avec une attaque forte et une décroissance rapide. Au contraire, la flute est un son tenu, avec une enveloppe temporelle de type plateau et un centroïd temporel plutôt au centre de l'échantillon sonore.





Fin de cours