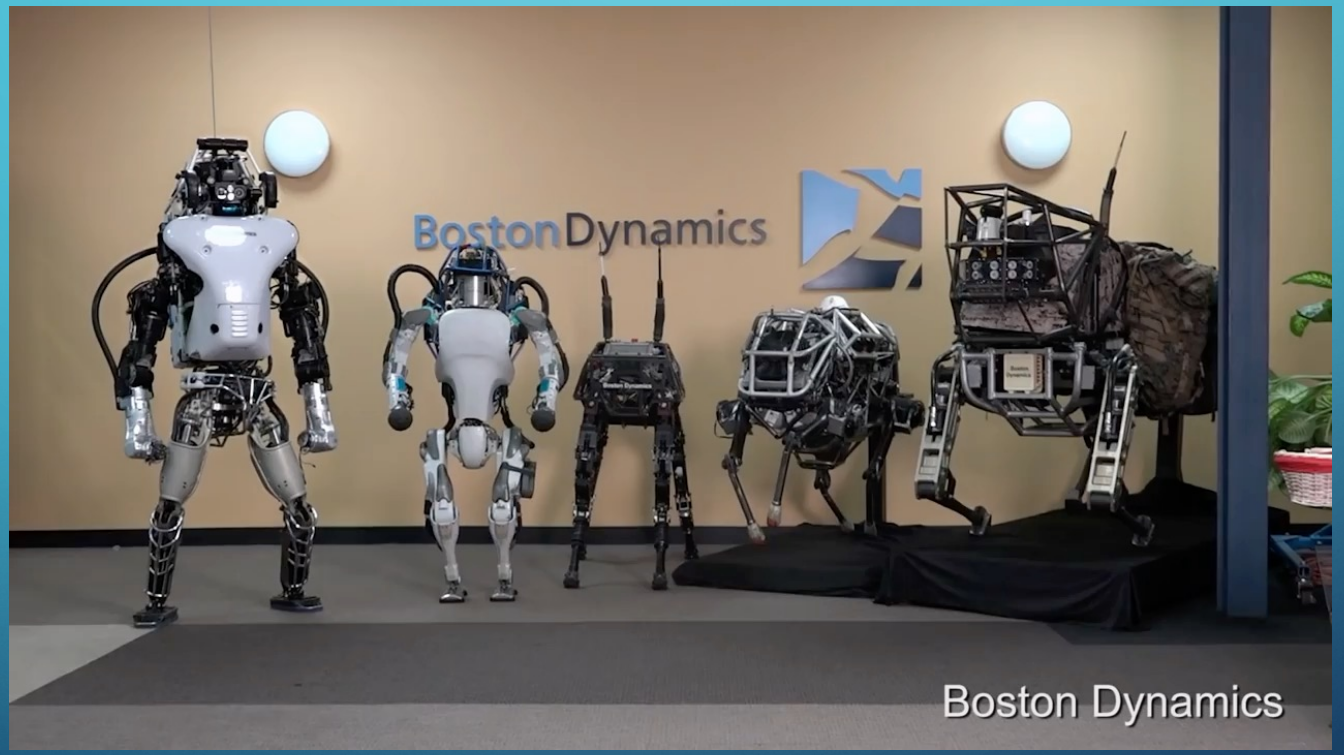
INTRODUCTION AU MACHINE LEARNING

FORMATION MACHINE LEARNING

VALÉRIE GAUTARD

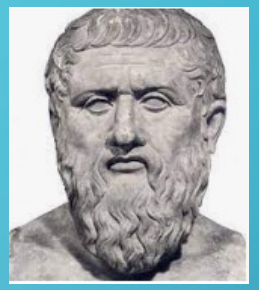
FASCINANTE IA



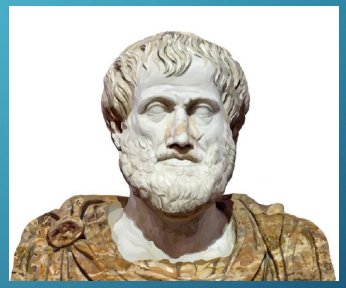
SOMMAIRE

- Historique
- Les bases du ML
- Une méthode non supervisée : Kmean
- Les arbres de décision
- Ensemble learning

UN BREF HISTORIQUE (1) : 300 BC

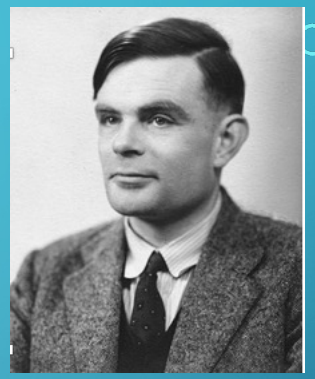


- 387 BC, Platon suggère que le cerveau contrôle nos processus mentaux
- 335 BC, Aristote
  - invente le raisonnement inductif, méthode formelle de représentation du raisonnement humain
  - Pour lui, tout part du cœur…

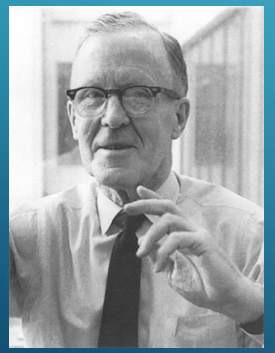


Tous les hommes sont mortels
Les athéniens sont des hommes
Donc les Athéniens sont mortels

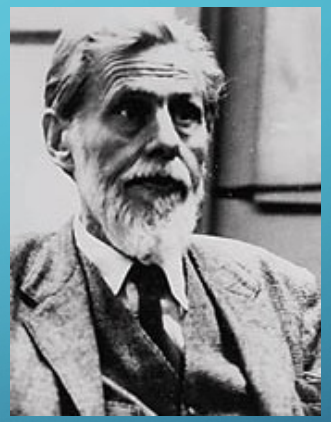
UN BREF HISTORIQUE (2) : LES ANNÉES 40



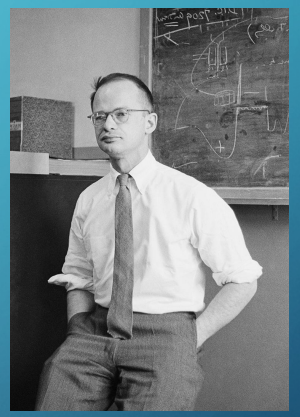
- 1942 : Alan Turing : Toute forme de raisonnement mathématique peut être implémenté sur une machine
- 1943 : Mac Culloch et Pitts : neurone formel
- 1949 : Donald Hebb :



- Mécanismes de plasticités synaptique
- Base de l'adaptation des neurones



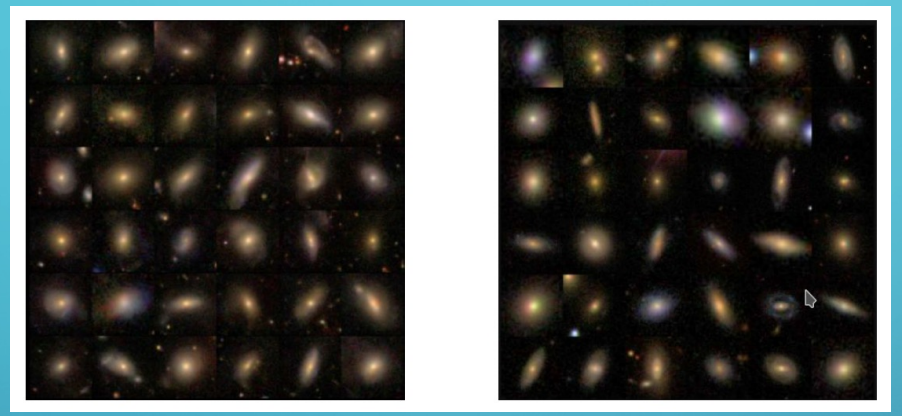
Neurophysiologist and cybernetician



Logician workingin the field of computational neuroscience

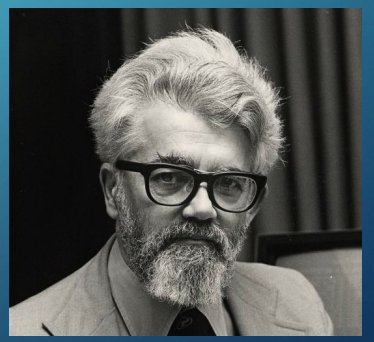
UN BREF HISTORIQUE (3) : LES ANNÉES 50

- 1950 :
  - Test de Turing pour vérifier si un système est « intelligent » , indistinguable d'un humain



Faux

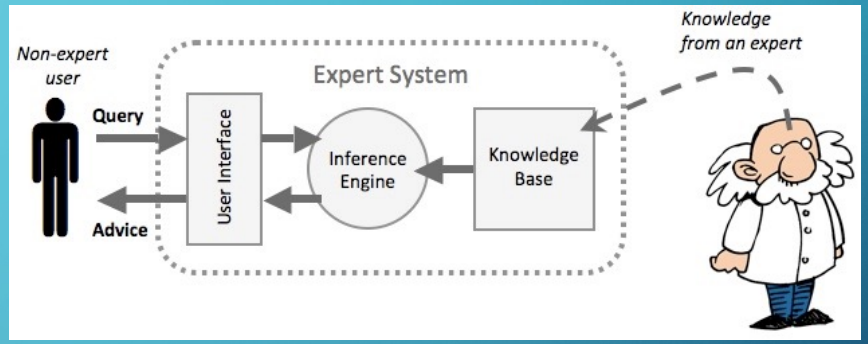
Vrai



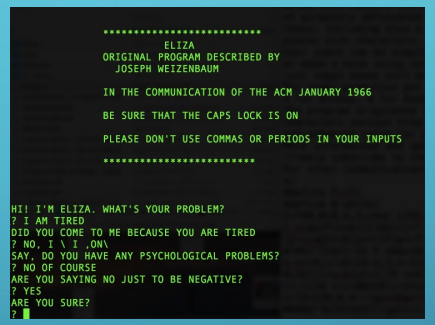
  - Isaac Asimov invente les 3 (4) lois de la robotique
- 1955 : McCarthy et Marvin Minsky : père de l'IA
- 1956 : Conférence de Darmouth -> terme d'Intelligence artificielle

UN BREF HISTORIQUE (4) : LES ANNÉES 60

- 1965 : systèmes expert, logique floue



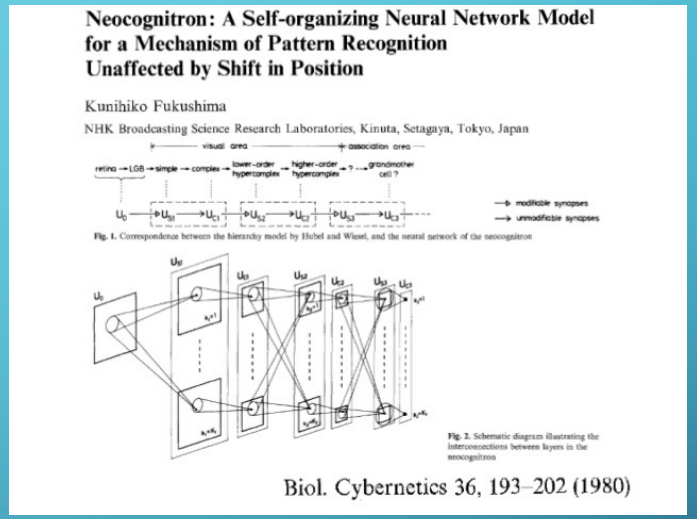
- 1966 : Chat bot Eliza



- 1969 : Premier hiver de l'IA après le livre de M. Minsky

UN BREF HISTORIQUE (5) : LES ANNÉES 80

- 1980 : K. Fukushima :
  - premier réseau de neurone profond
  - Inspiré de cortex visuel

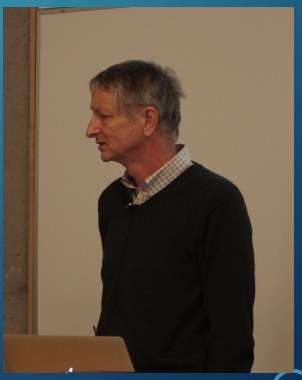


Neocognitron: A Self-organizing Neural Network Model for a Mechanism of Pattern Recognition Unaffected by Shift in Position

Kunihiko Fukushima

NHK Broadcasting Science Research Laboratories, Kinuta, Setagaya, Tokyo, Japan

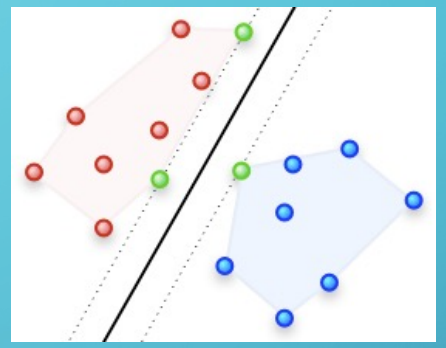
Biol. Cybernetics 36, 193–202 (1980)



- 1985 : Yann Le Cun
- 1986 : Hinton : back Propagation
- 1987 : 2ème hiver suite à l'article de Minsky & Papert

UN BREF HISTORIQUE (6) : LES ANNÉES 90

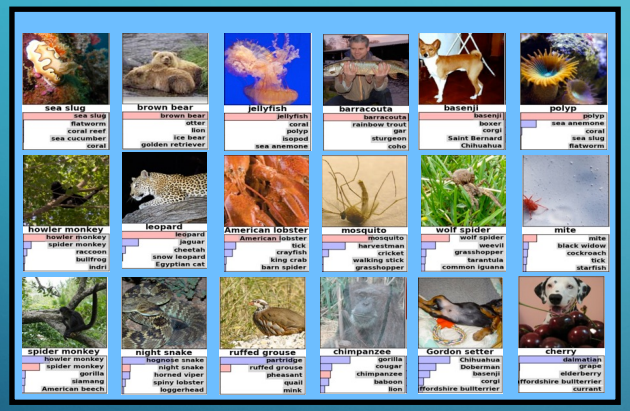
- 1995 : SVM:
  - SVM
  - 3eme hiver des réseaux de neurones

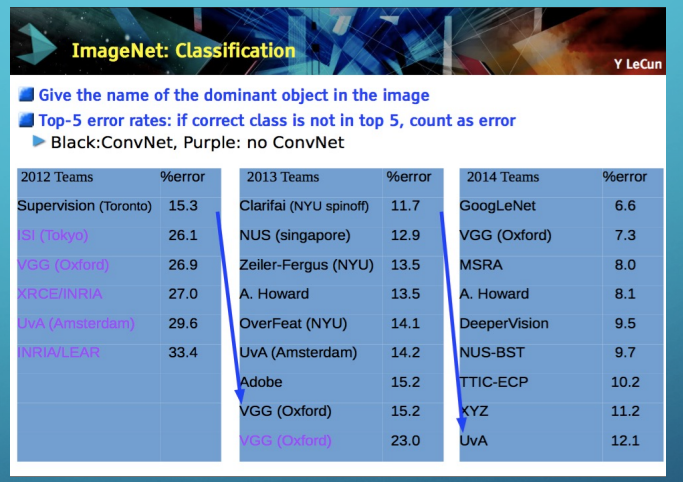


- 1997 : Deep Blue gagne G. Kaparov aux échecs

UN BREF HISTORIQUE (6) : LES ANNÉES 2010

- 2012 :
  - Retour des réseaux de neurones profonds (ImageNet)
  - Facebook : programme DeepFaceSVM



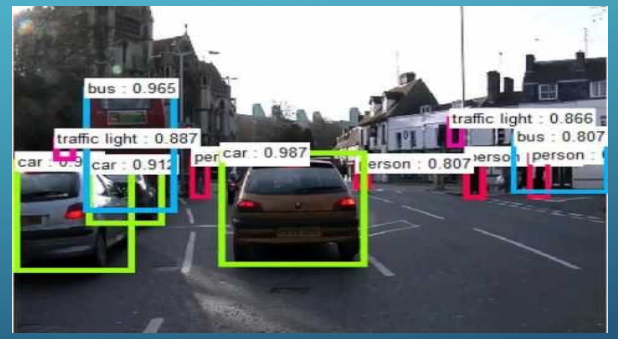


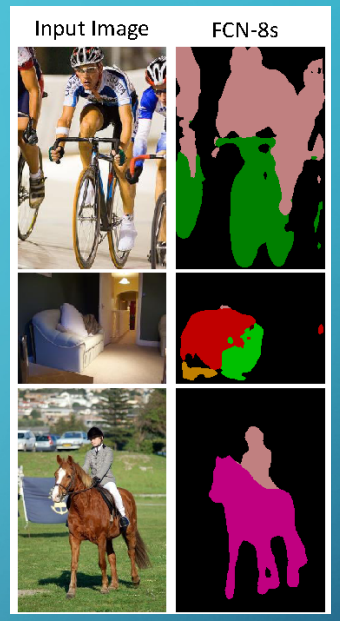
Name of the algorithm	Date	Error on test set
Supervision	2012	15.3%
Clarifai	2013	11.7%
GoogLeNet	2014	6.66%
Humain level (Adrej Karpathy)		**5%**
Microsoft	05/02/2015	4.94%
Google	02/03/2015	4.82%
Baidu/ Deep Image	10/05/2015	4.58%
Shenzhen Institutes of Advanced Technology, Chinese Academy of Sciences	10/12/2015 (le CNN a 152 couches!)	3.57%
Google Inception-v3 (Arxiv)	2015	3.5%
WMW (Momenta)	2017	2.2%
	Now	?

- 2018 : Prix Turing à Hinton (google), Y. LeCun (facebook) et Y Bengio

QUELQUES RÉSULTATS (1)

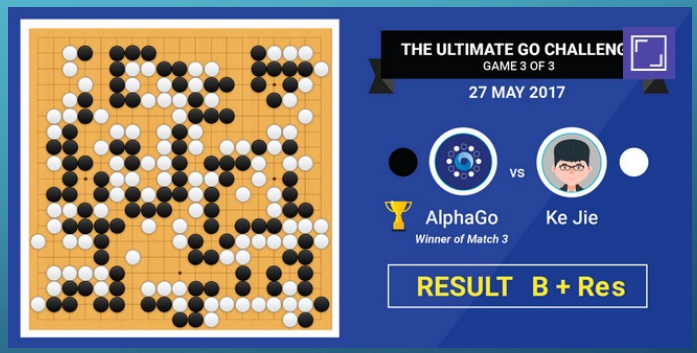
- Segmentation
- Extraction des zones d'interet
- Détection de cancer
  - Médecin spécialisé : 0.73
  - IA : 0.89

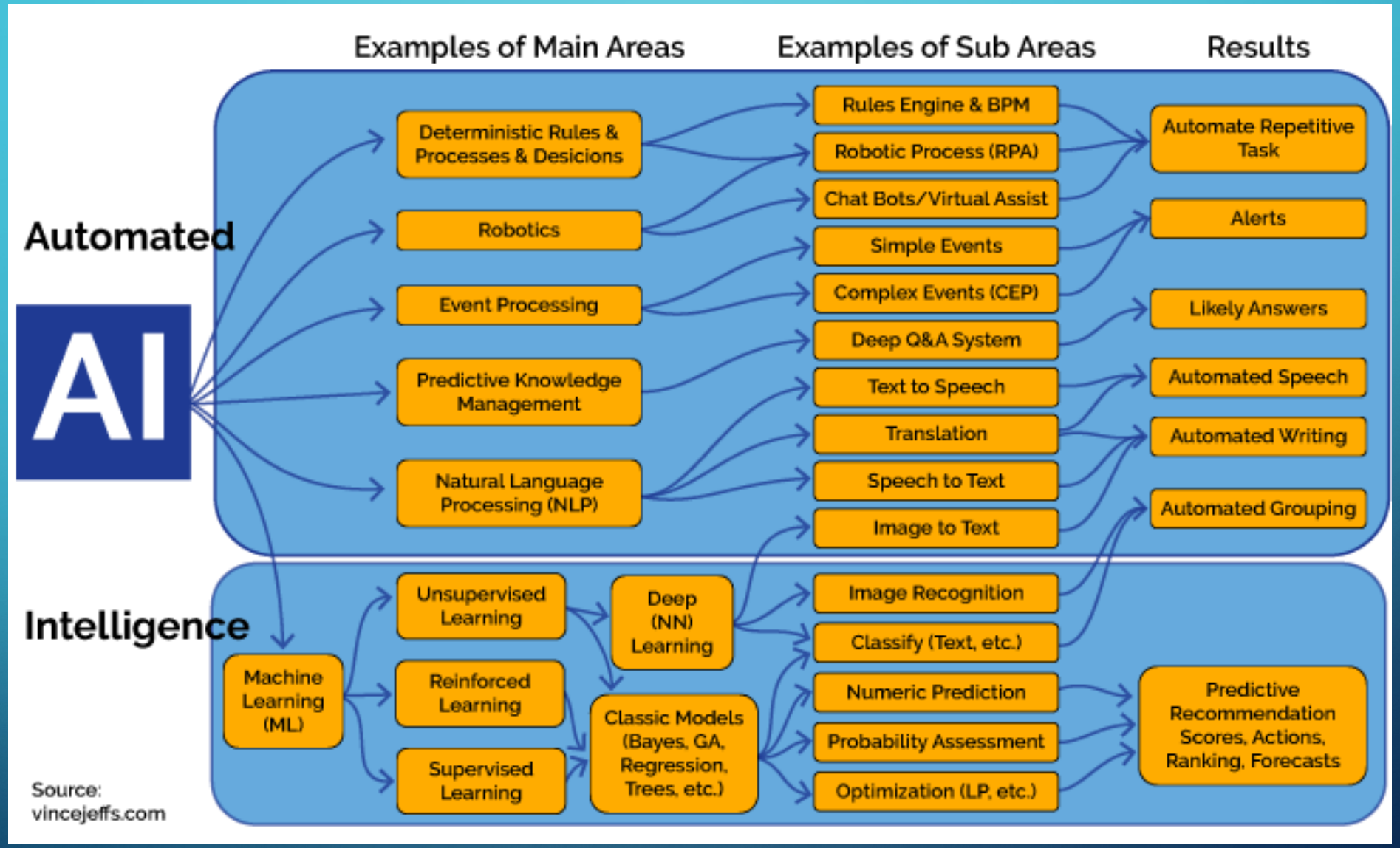




QUELQUES RÉSULTATS (2) : LE JEU DE GO

- AlphaGo
- AlphaGo Zero
  - Apprend avec les règles seulement
- AlphaZero




Examples of Main Areas
Examples of Sub Areas
Results
Automated
AI
Deterministic Rules & Processes & Desicions
Robotics
Event Processing
Predictive Knowledge Management
Natural Language Processing (NLP)
Rules Engine & BPM
Robotic Process (RPA)
Chat Bots/Virtual Assist
Simple Events
Complex Events (CEP)
Deep Q&A System
Text to Speech
Translation
Speech to Text
Image to Text
Automate Repetitive Task
Alerts
Likely Answers
Automated Speech
Automated Writing
Automated Grouping
Intelligence
Machine Learning (ML)
Unsupervised Learning
Reinforced Learning
Supervised Learning
Deep (NN) Learning
Classic Models (Bayes, GA, Regression, Trees, etc.)
Image Recognition
Classify (Text, etc.)
Numeric Prediction
Probability Assessment
Optimization (LP, etc.)
Predictive Recommendation Scores, Actions, Ranking, Forecasts
Source:
vincejeffs.com

SOMMAIRE

- Historique
- **Les bases du ML**
- Une méthode non supervisée : Kmean
- Les arbres de décision
- Ensemble learning

IA, UNE DÉFINITION (1)

« la construction de programmes informatiques qui s'adonnent à **des tâches qui sont, pour l'instant, accomplies de façon plus satisfaisante par des êtres humains** car elles demandent des processus mentaux de haut niveau tels que : l'apprentissage perceptuel, l'organisation de la mémoire et le raisonnement critique »

Marvin Minsky – 1927-2016

IA, UNE DÉFINITION (2)

- « On dit d'un programme informatique qu'il apprend une classe de tâches T de l'expérience E avec une mesure de performance P si sa performance sur les tâches T, telles que mesurées par P, s'améliore avec l'expérience E.''

- L'apprentissage automatique est le domaine d'étude qui donne aux ordinateurs la capacité d'apprendre sans être explicitement programmés

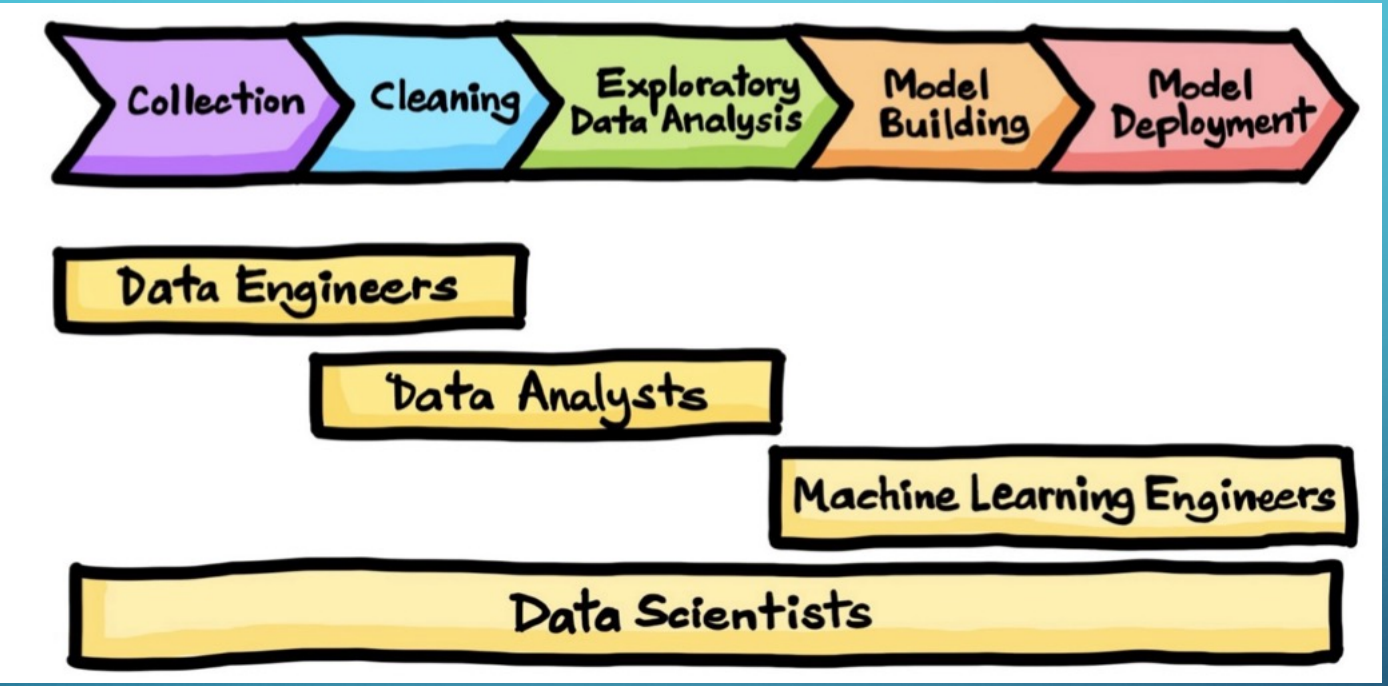
IA, UNE DÉFINITION (3)

L'étude scientifique des **algorithmes** et des **modèles statistiques** que les ordinateurs utilisent pour accomplir une tâche **sans instruction explicite**, mais plutôt en s'appuyant sur des motifs et de l'inférence.

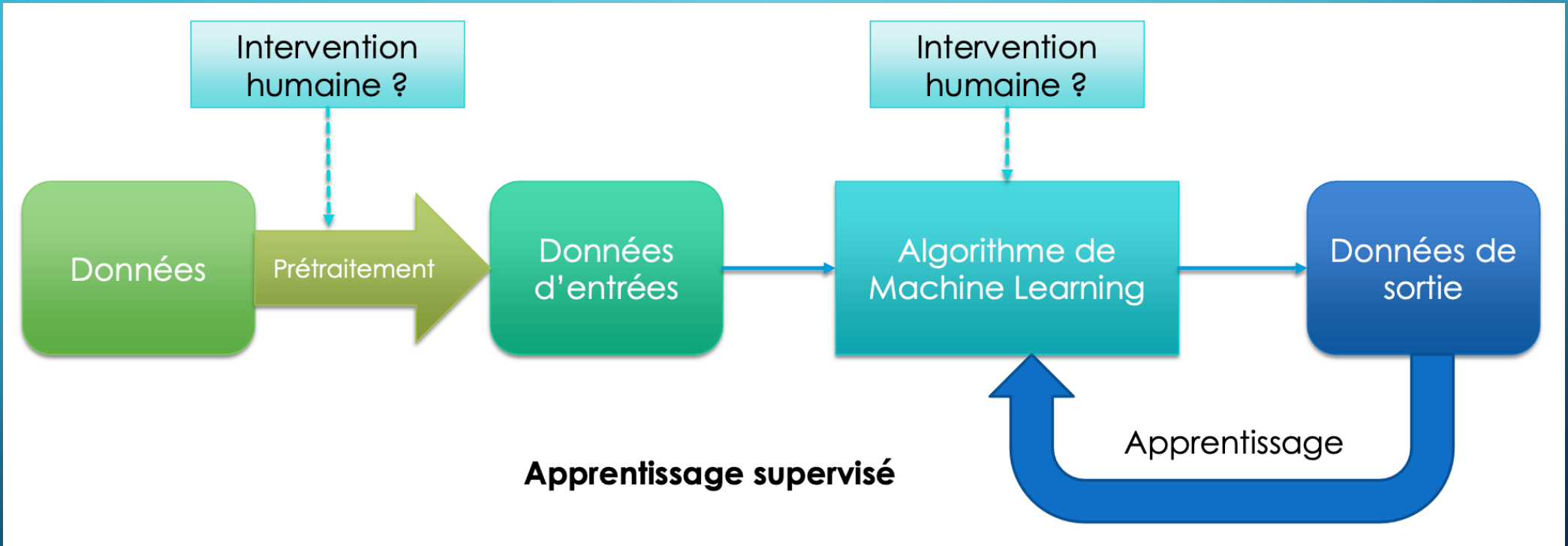
On ne sait pas toujours définir un algorithme explicite pour une tâche donnée …

… mais accès à des données massives à partir desquelles “apprendre” à résoudre une tâche

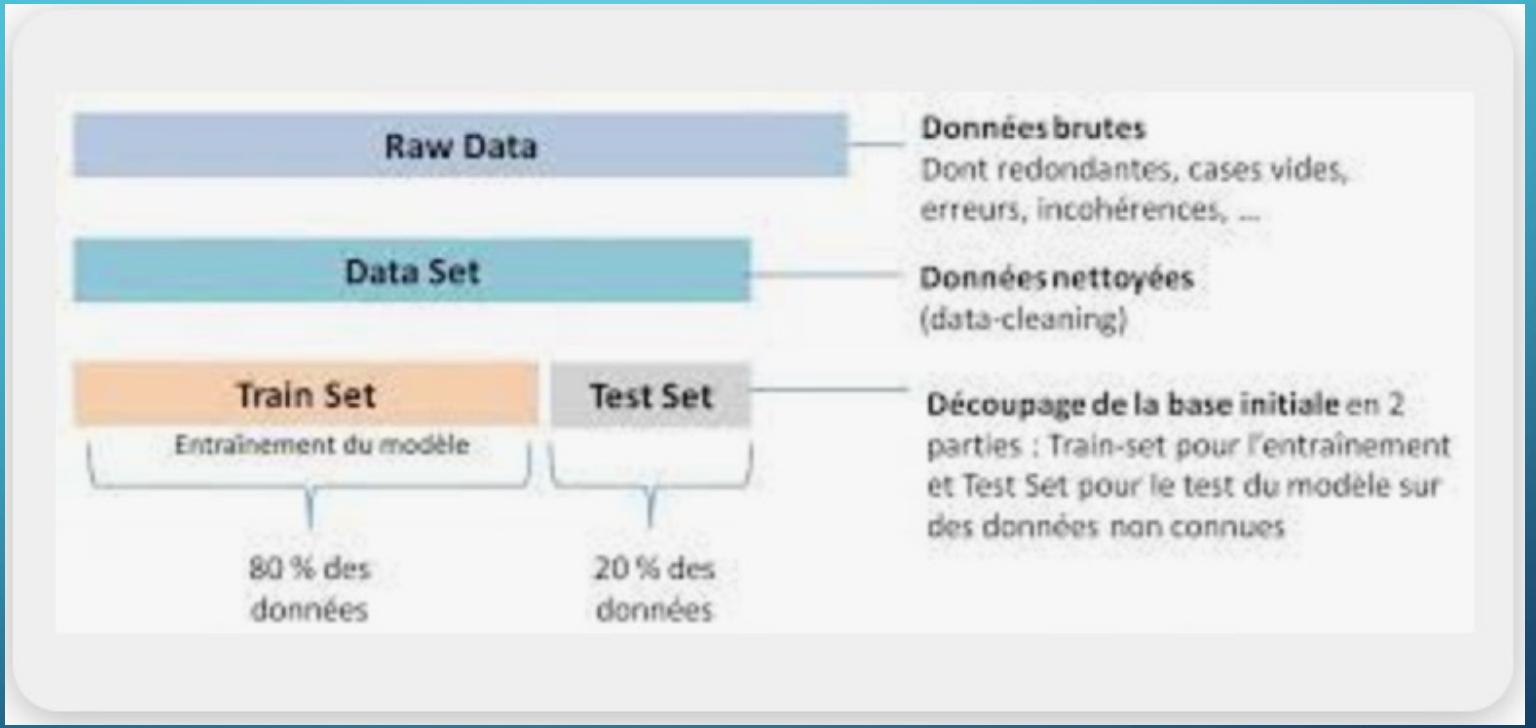
THE DATA SCIENCE PROCESS (1/3)



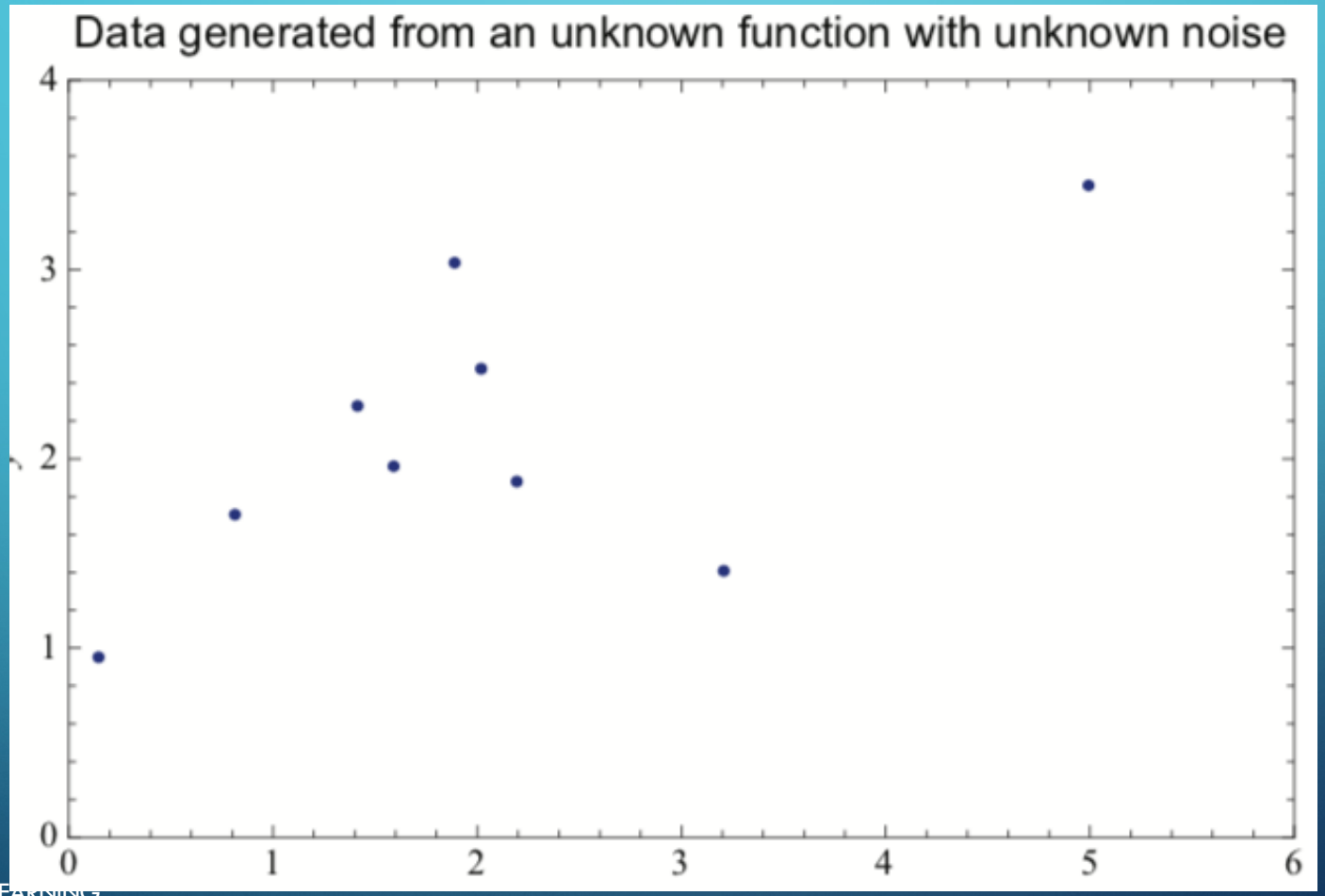
THE DATA SCIENCE PROCESS (2/3)



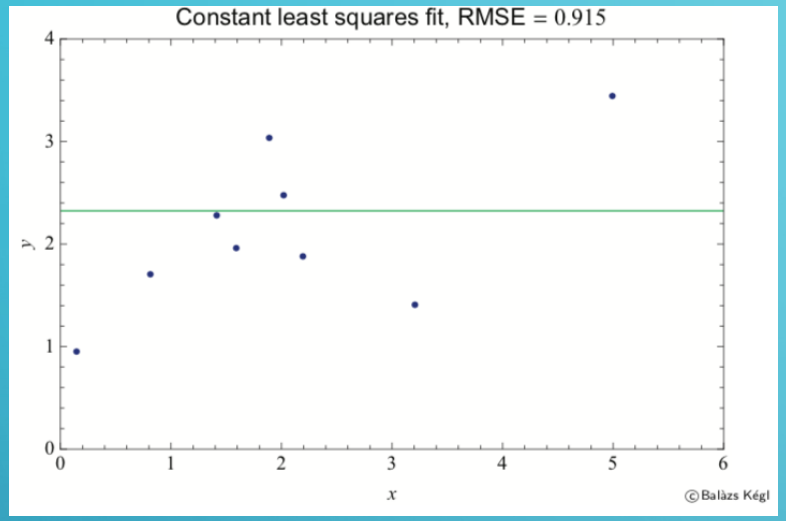
THE DATA SCIENCE PROCESS (3/3)

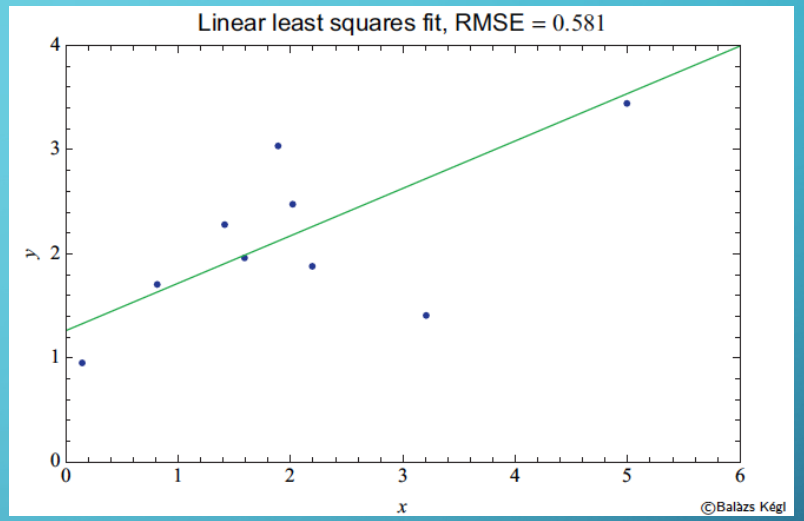


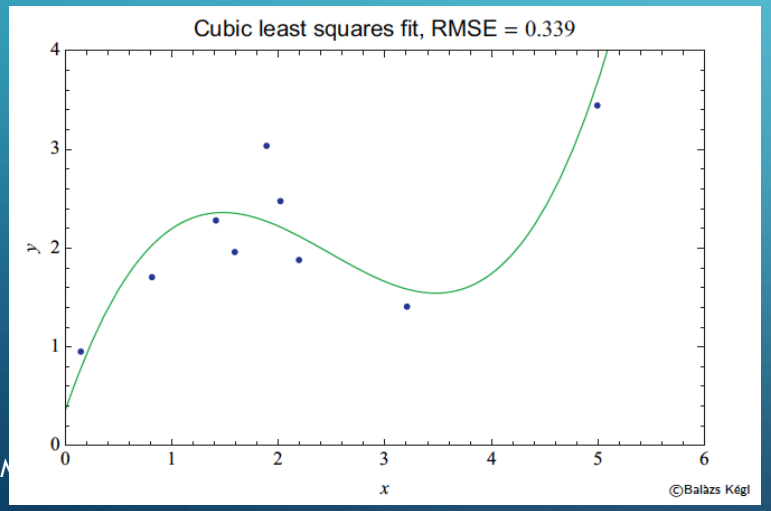
CHOIX DE LA CLASSE DE FONCTIONS

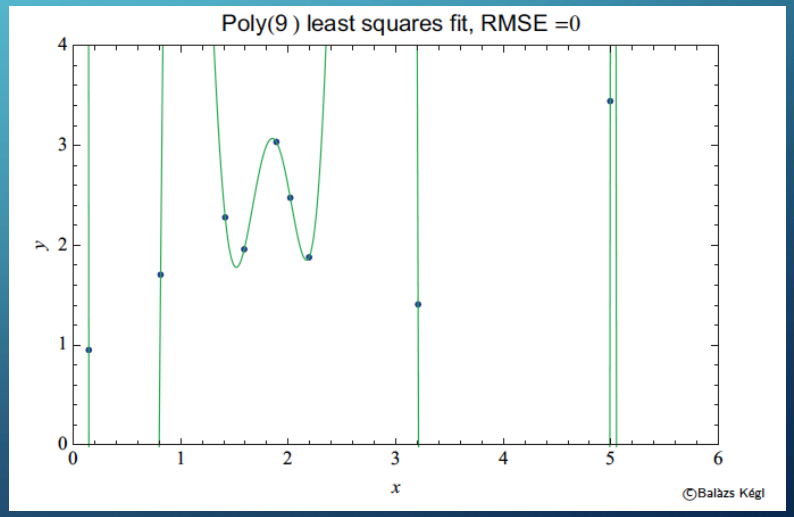


CHOIX DE LA CLASSE DE FONCTIONS

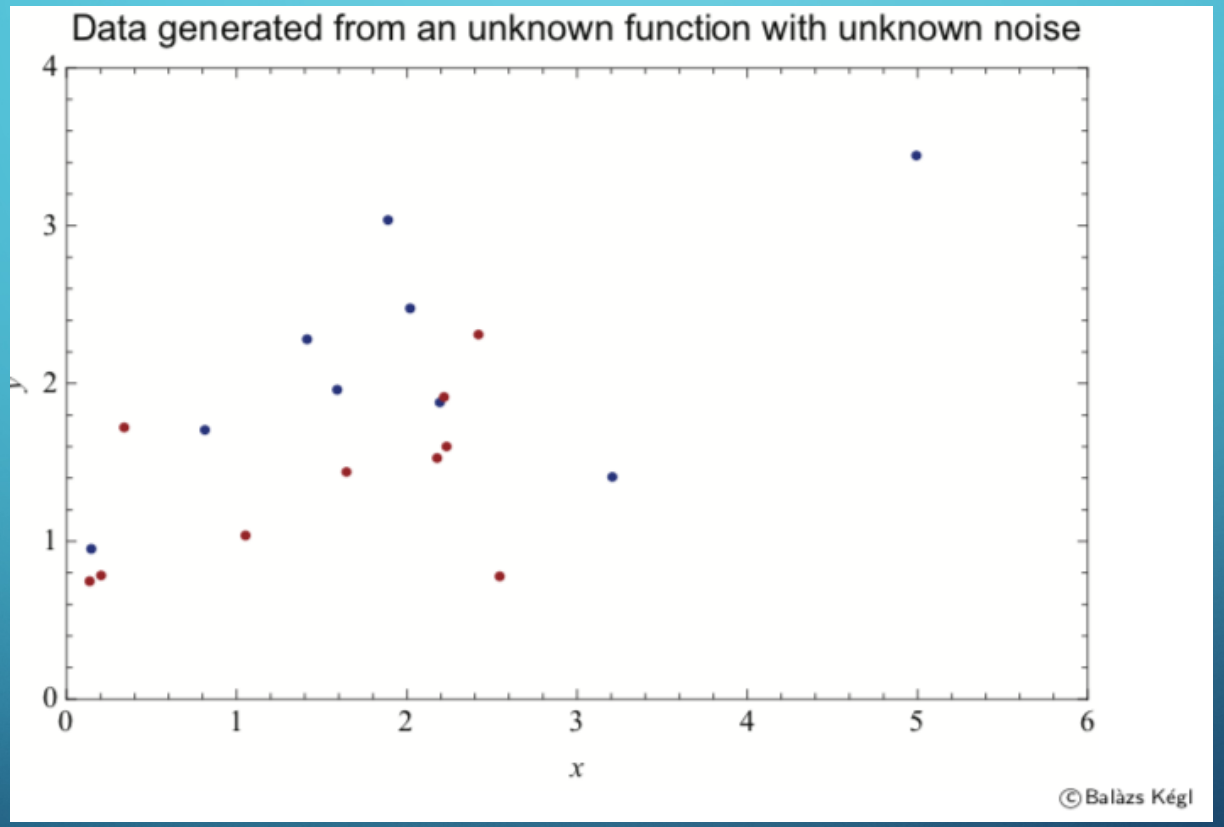




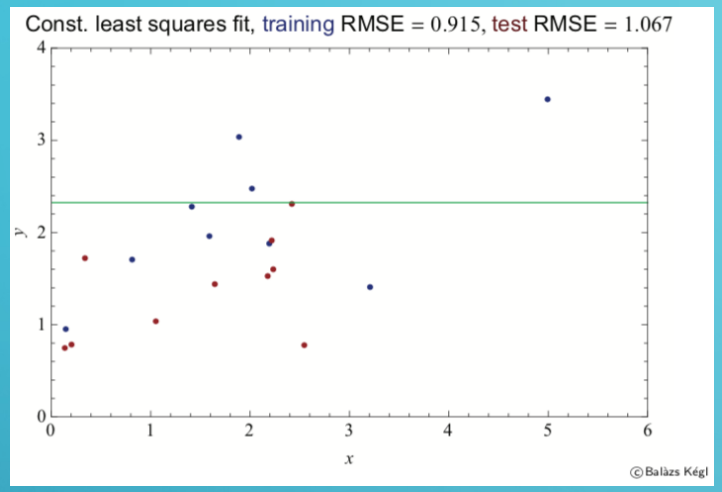


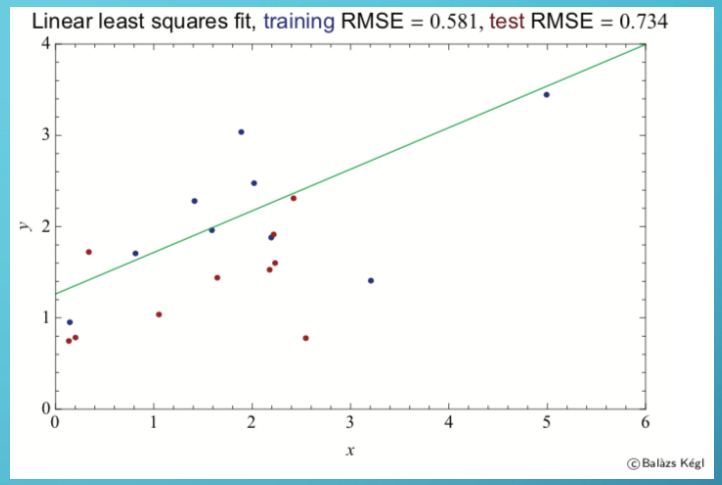


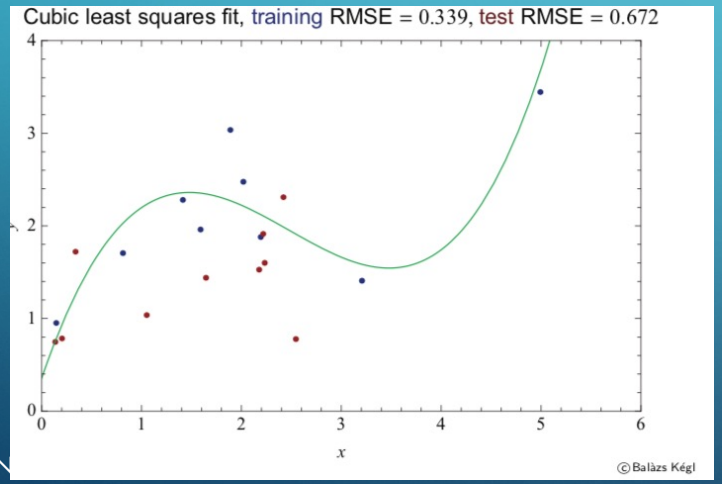
CHOIX DE LA CLASSE DE FONCTIONS

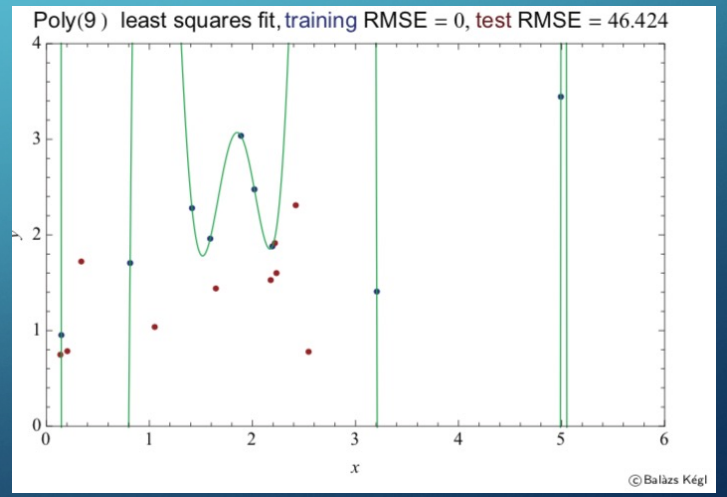


CHOIX DE LA CLASSE DE FONCTIONS

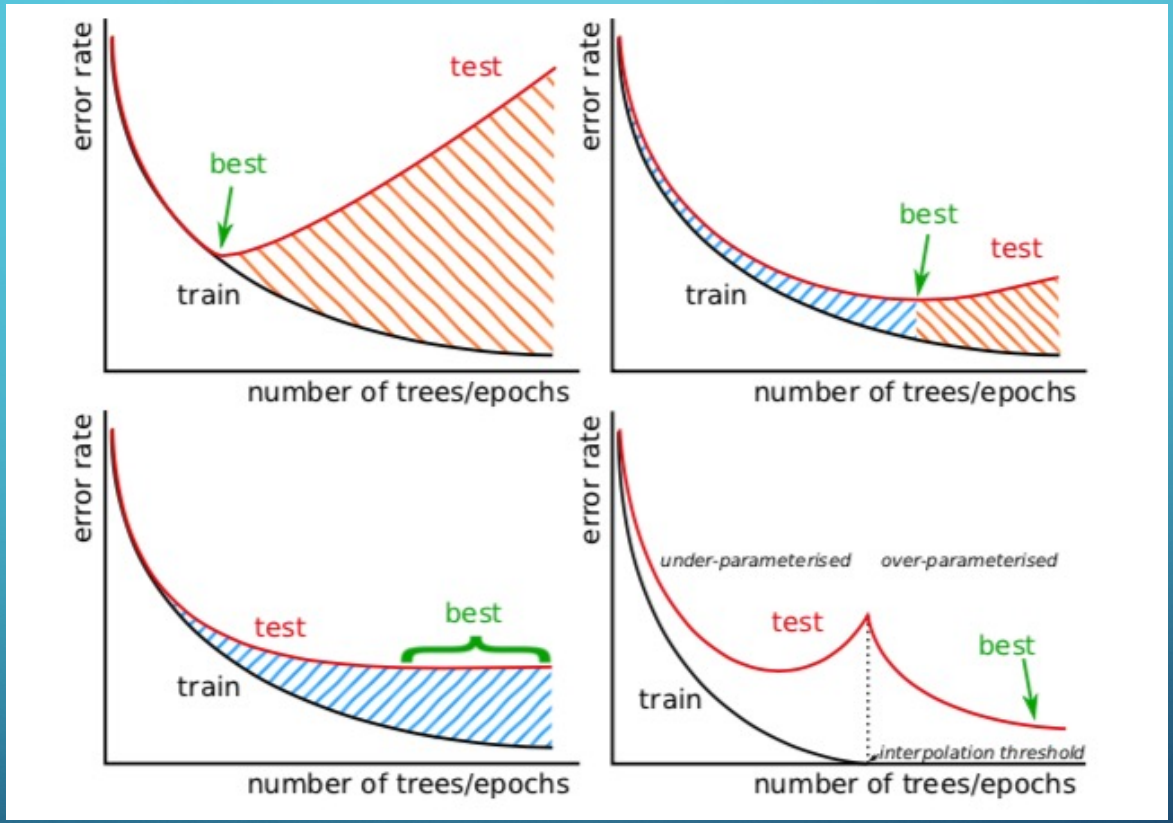






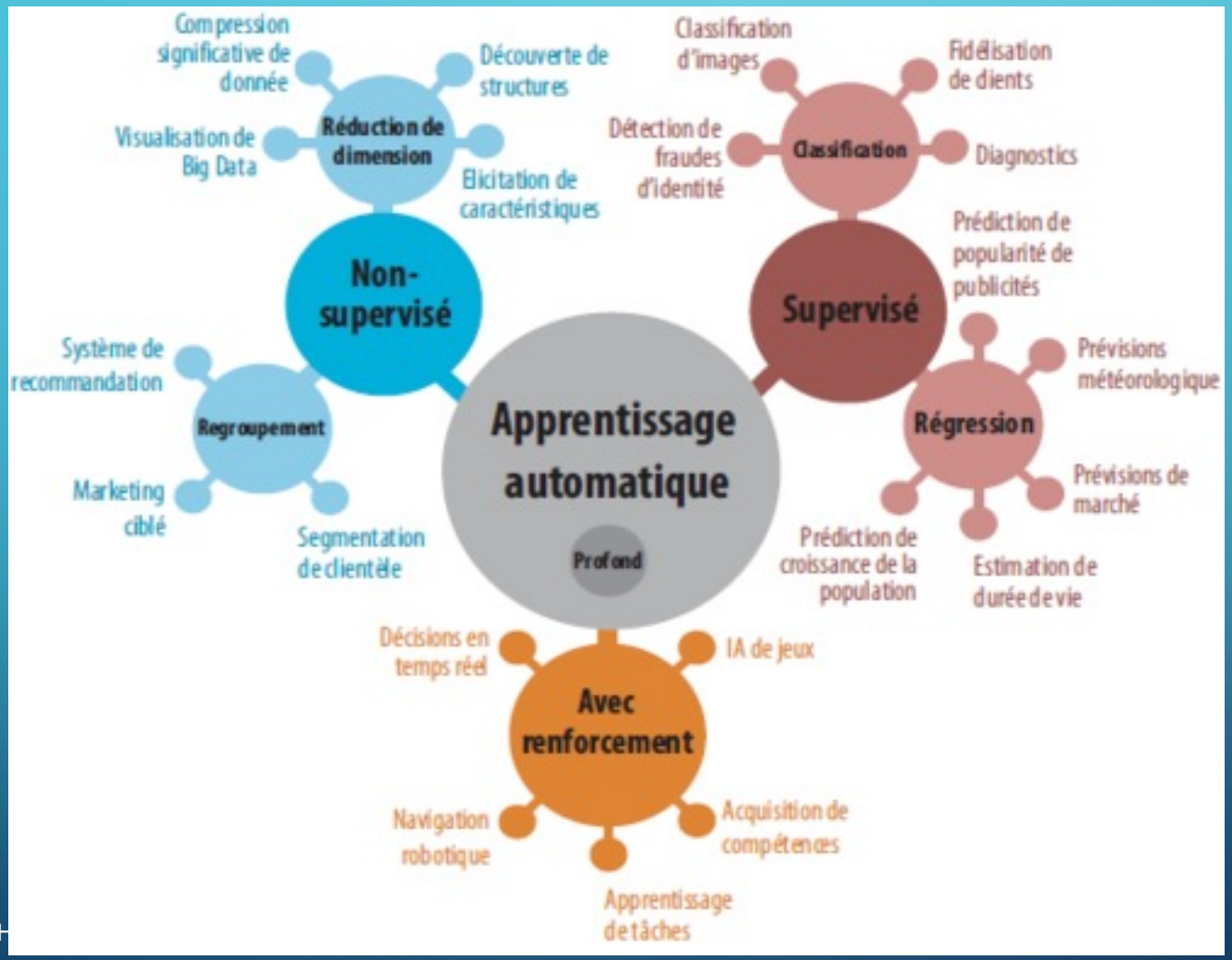


SUR-ENTRAINEMENT/SOUS-ENTRAINEMENT



Trop de sur-entrainement (overfitting) / bon sur-entrainement (encore sous entrainement)

DIVERSITÉ DU ML



DIVERSITÉ DU ML : APPRENTISSAGE SUPERVISÉ

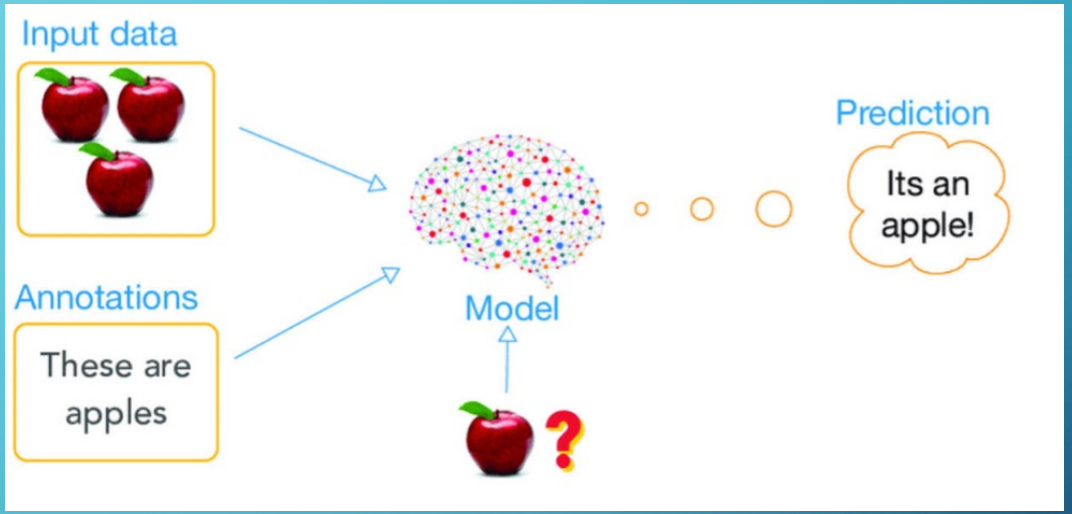
- Données avec label (étiquetées)
- On se donne des « data » :

  N exemples

  $(x,y)_1, (x,y)_2, \dots, (x,y)_N$

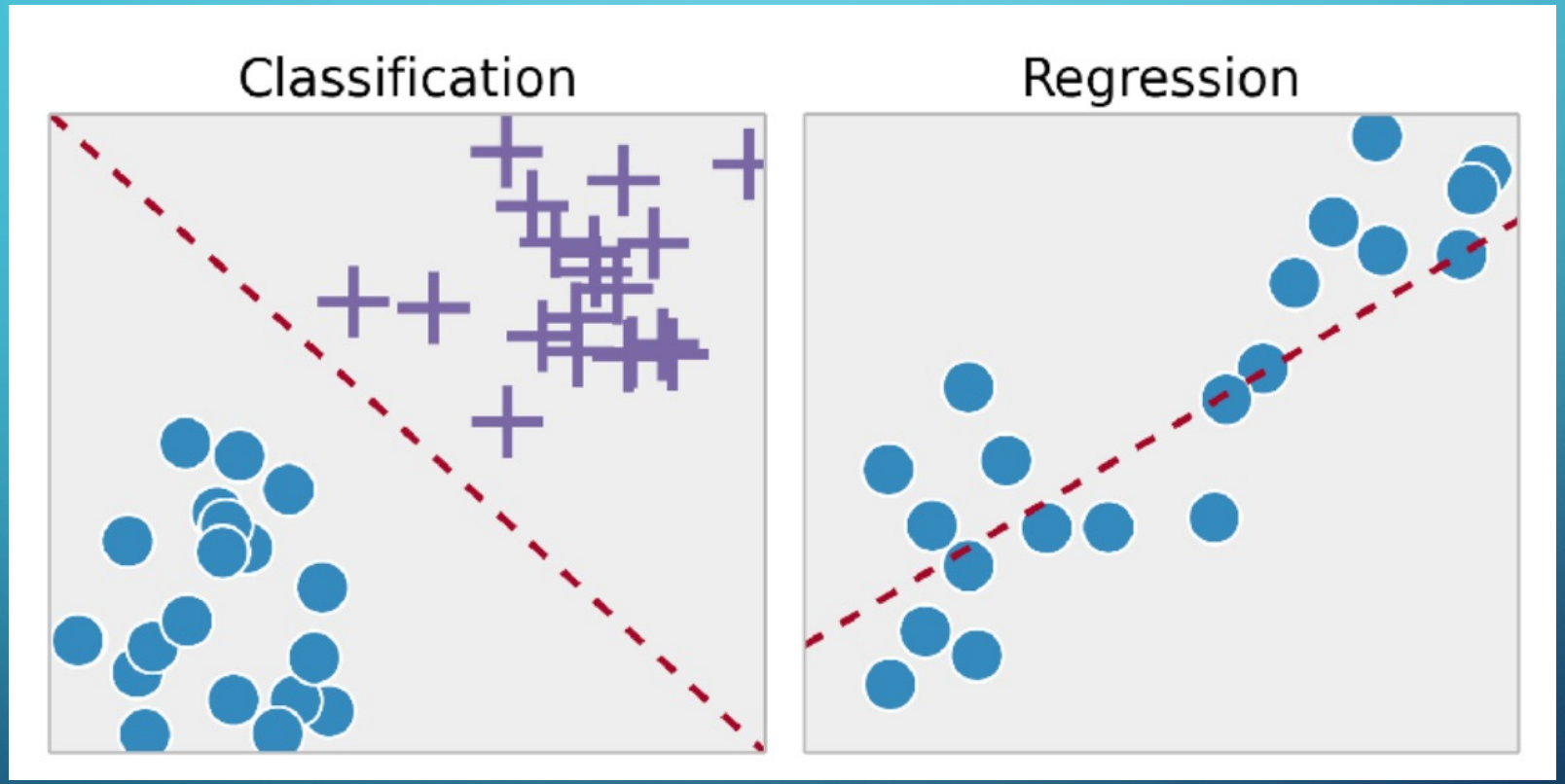
  x : feature

  y : label



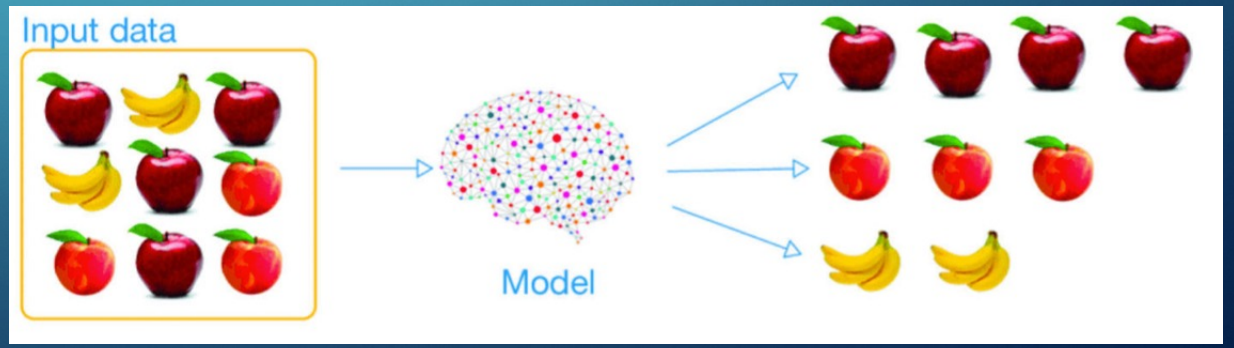
- On utilise des exemples avec des labels pour savoir si l'apprentissage est efficace
- Exemple :
  - Evaluer l'évolution du prix d'une maison
  - Accord ou non d'un prêt

DIVERSITÉ DU ML : APPRENTISSAGE SUPERVISÉ



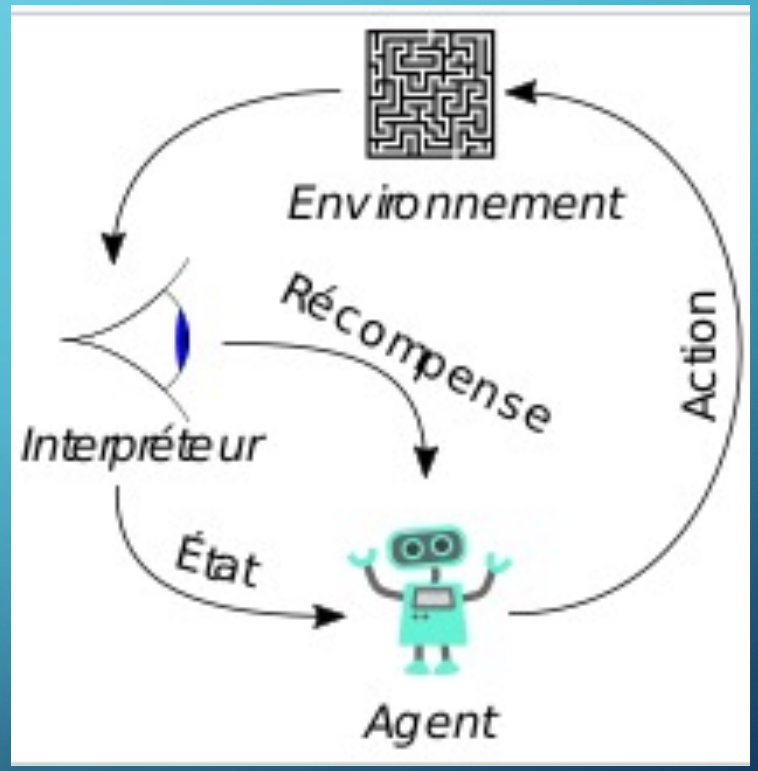
DIVERSITÉ DU ML : APPRENTISSAGE NON SUPERVISÉ

- Données non étiquetées
- score ne peut être calculé de manière certaine
- Découverte de la structure sous jacente
- Trouver des similarités sans avoir de catégories prédèfinies
- Exemple :
  - classer des mails en spam, non spam
  - classer des articles de journaux en rubrique

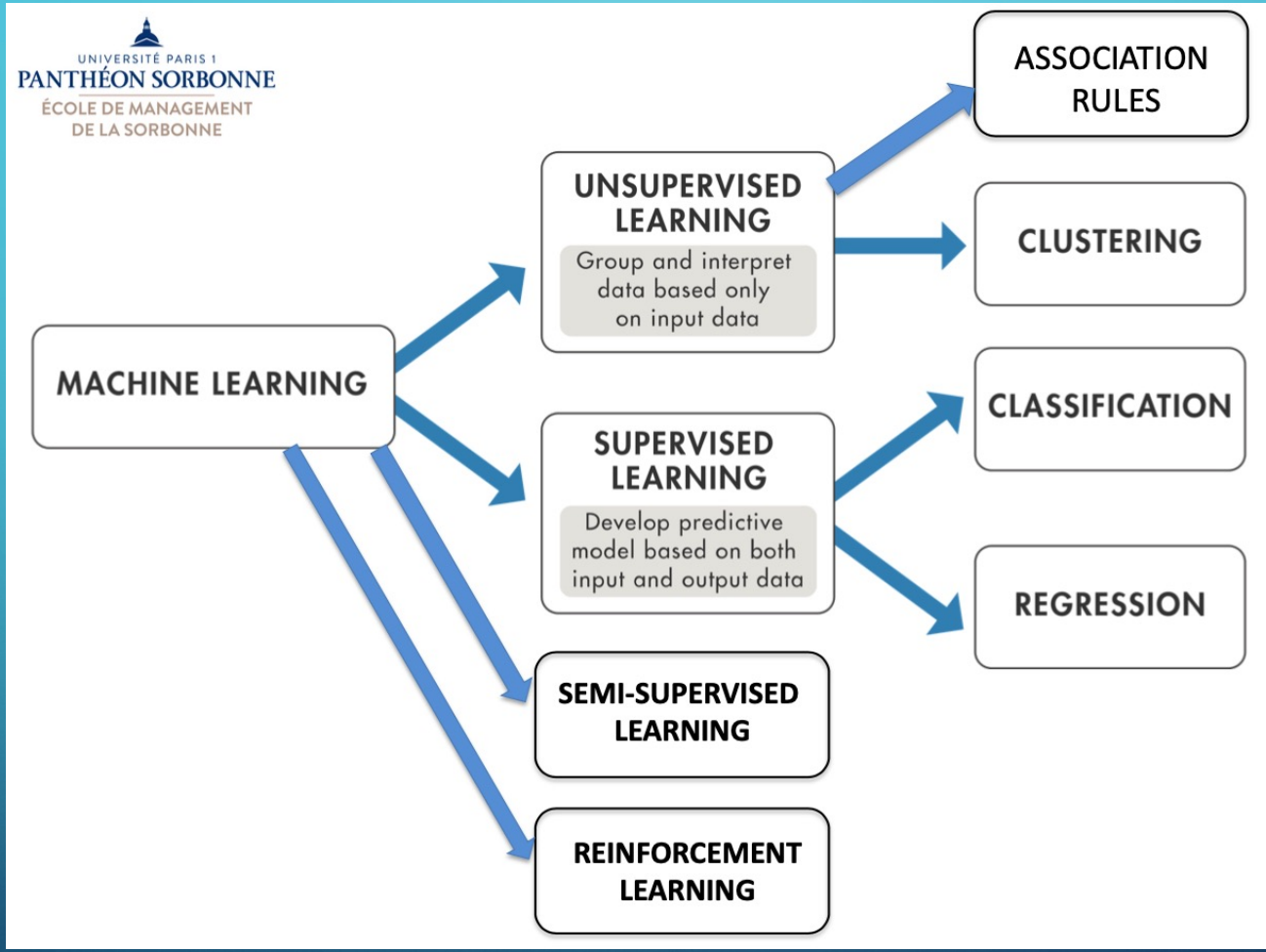


DIVERSITÉ DU ML : APPRENTISSAGE PAR RENFORCEMENT

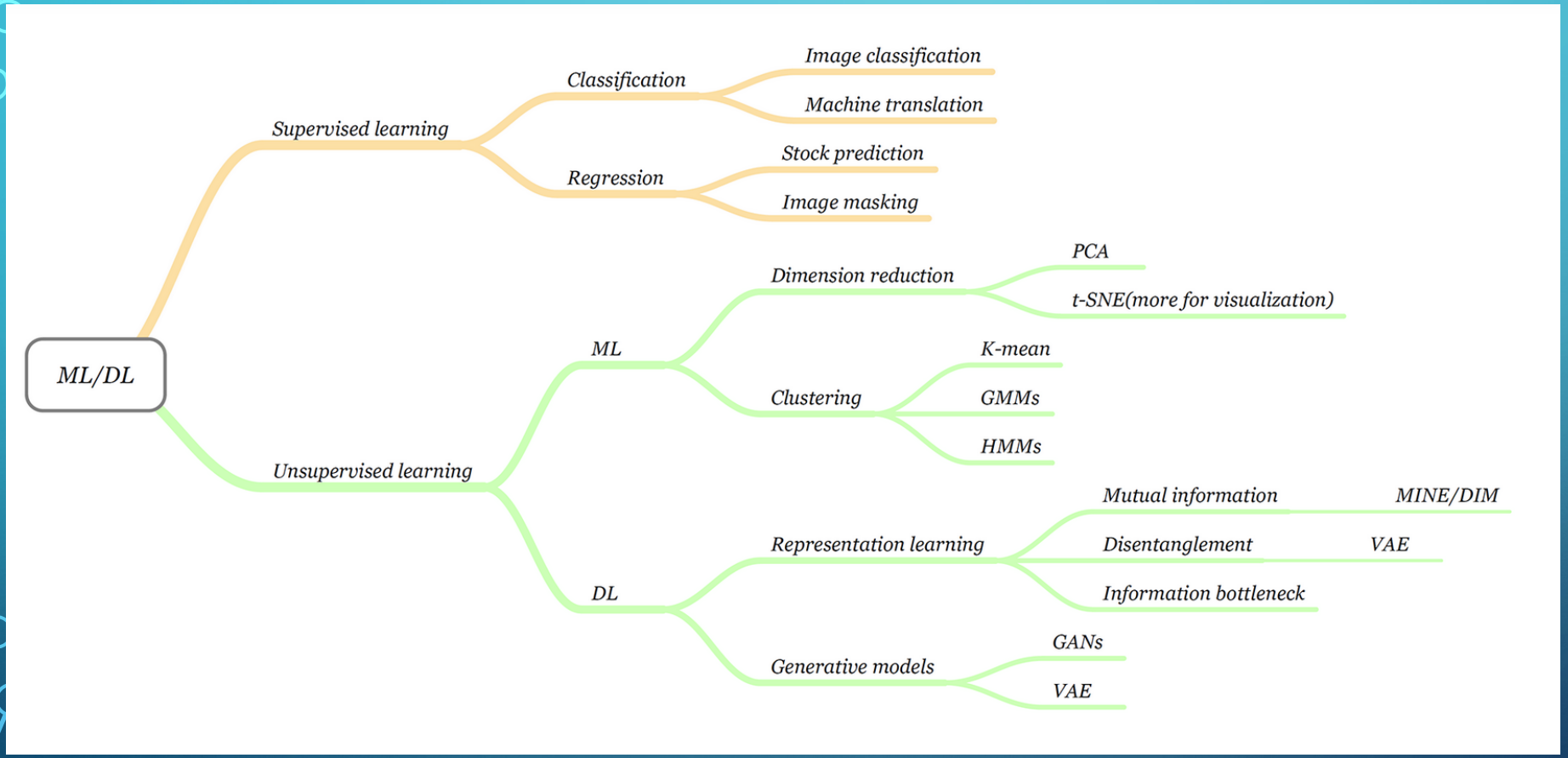
- Introduit en 1988 par Richard Sutton et en 19889 par Chris Watkins, publié en 1992
- Apprentissage en optimisant une récompense au cours du temps
- Processus de décision Markovien
- Décision prise en associant l'etat courant à l'état optimal
- Exemple :
  - IRobotique
  - Alpha Go zero



DIVERSITÉ DU ML : UN PAYSAGE VASTE

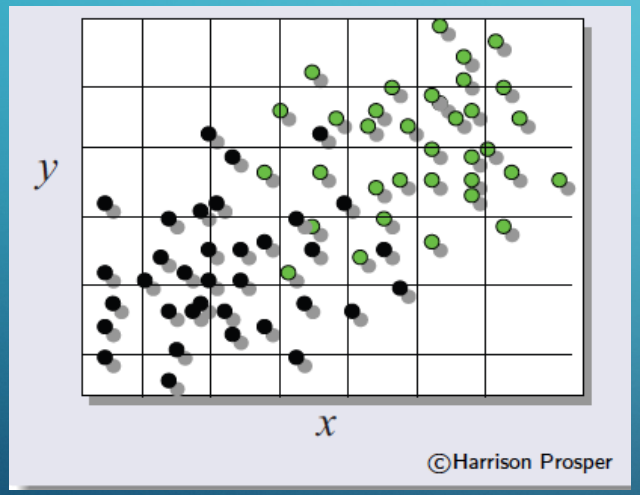


DIVERSITÉ DU ML : ET DE MULTIPLES MODELES



GRID SEARCH

- Annalyse basée sur des coupures
  - Utilisé pour optimiser les hyper paramètres
  - Approche simple basée sur des coupures autour de variables discriminantes
  - Difficulté : comment optimiser les coupures ?



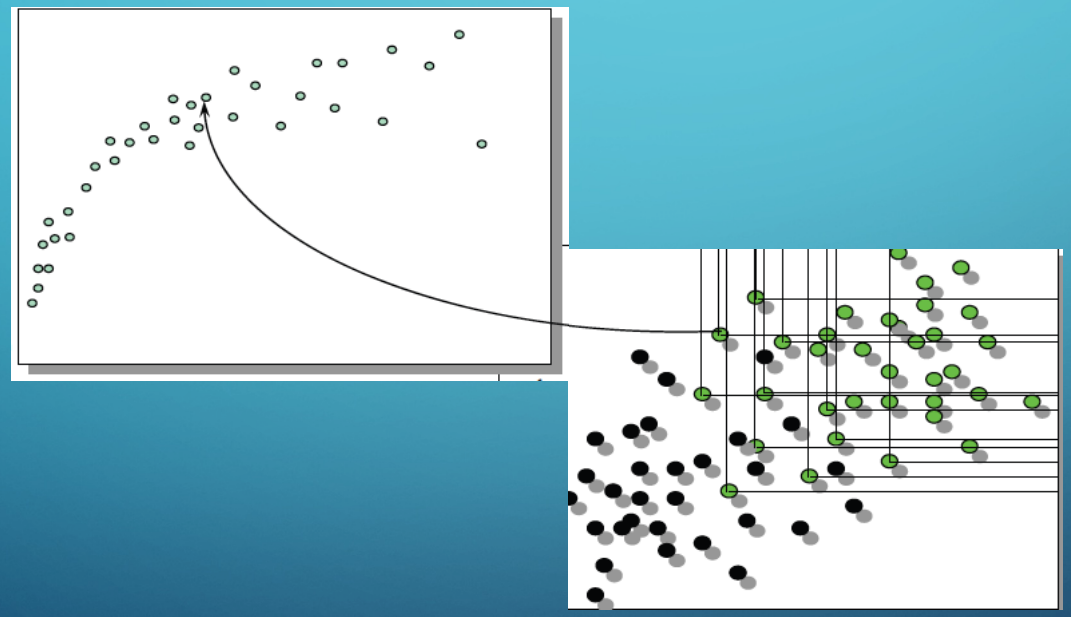
- Chaque variable est projeté dans chaque dimension
- On applique des coupures :

  $x > x_i, y > y_i$

- Nombre de points en fonction de la dimension

GRID SEARCH

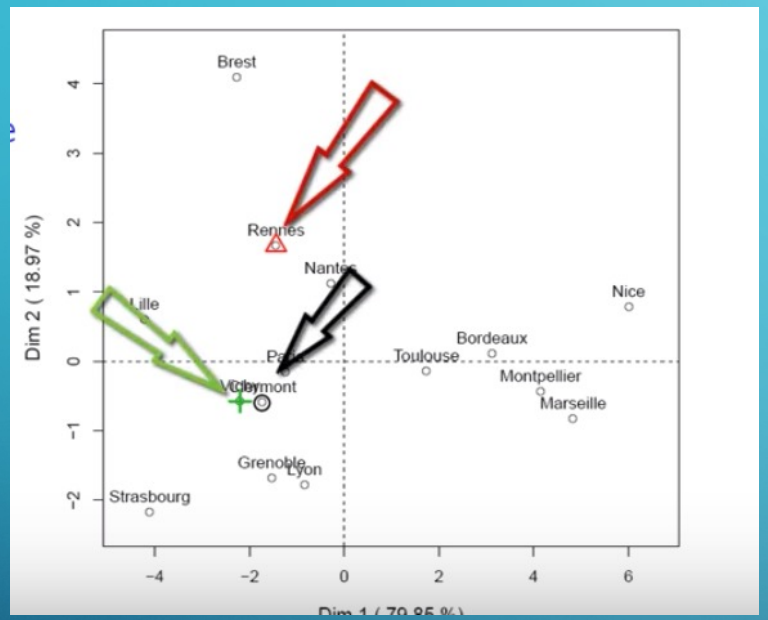
- Chaque point est vu comme un grille
  - Nombre de coupure est independant de la dimension
  - on s'affranchi de la malédiction de la dimension



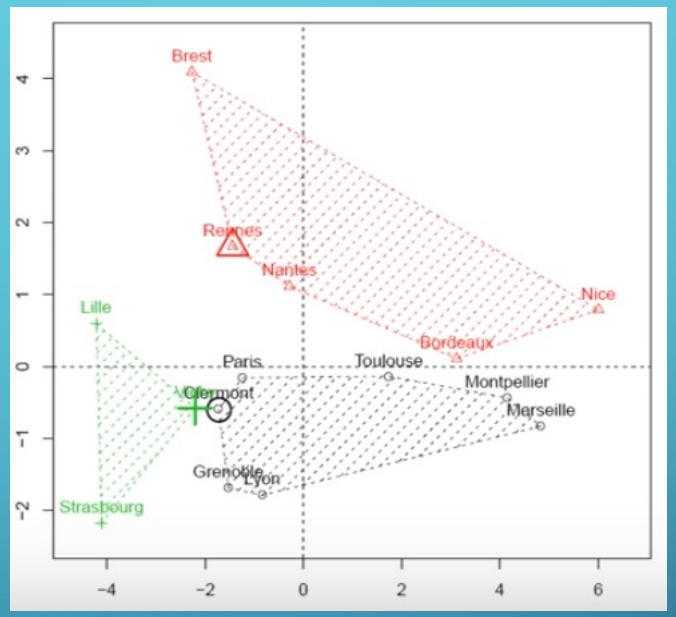
SOMMAIRE

- Historique
- **Les bases du ML**
- Une méthode non supervisée : Kmean
- Les arbres de décision
- Ensemble learning

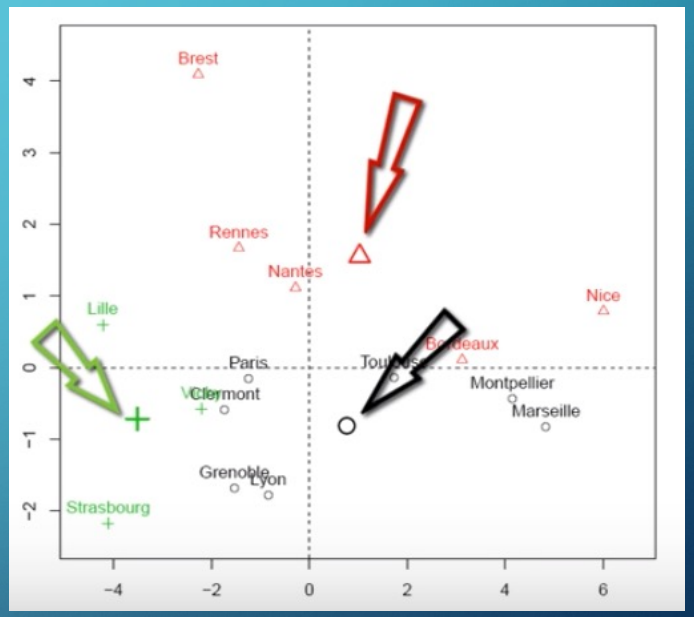
K-MEAN



On choisit 3 villes



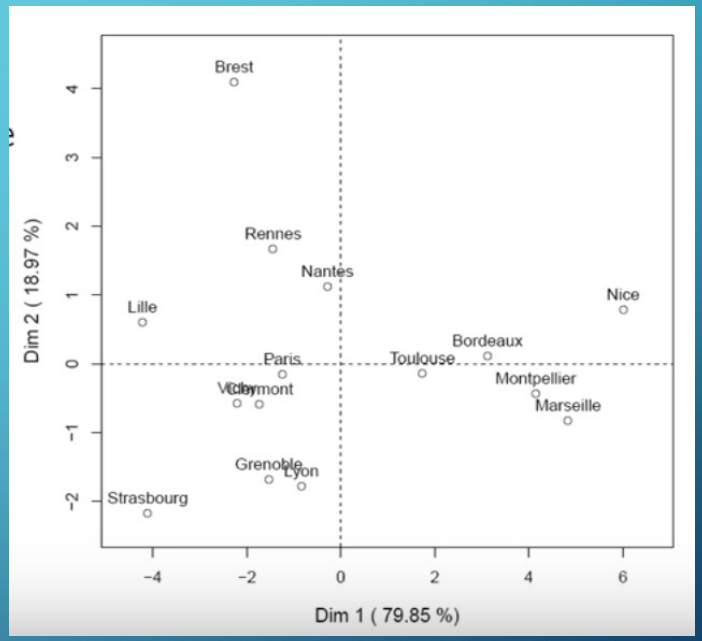
On affecte les villes
en minimisant la distance



On calcule le barycentre
Cette fois le centre des classe
n'est pas un individu

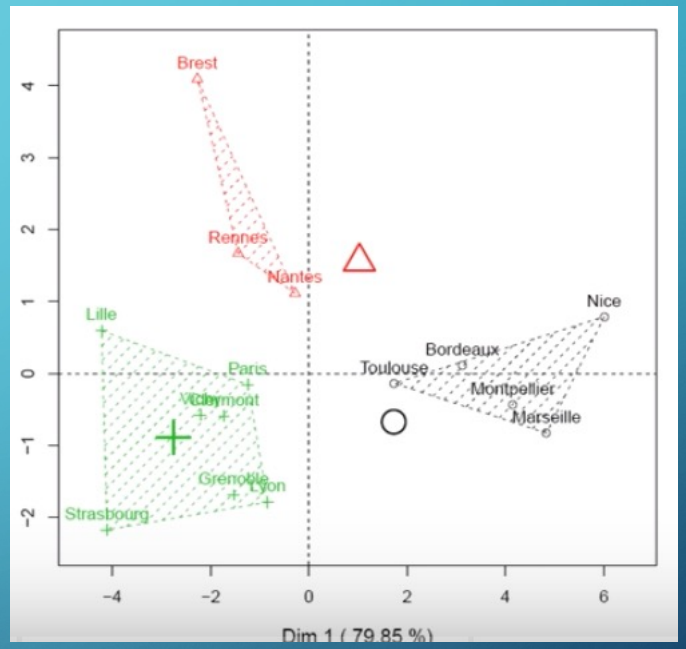
K-MEAN

- Méthode non supervisée
- Méthode de partitionnement des données (clustering)
  - Choisir Q centre de classe au hasard
  - Affecter les points au centre le plus proche
  - Calculer le centre de gravite de chaque classe



K-MEAN

- On itère :
  - Affecter les points au centre le plus proche
  - Calculer le centre de gravité de chaque classe
- On s'arrête quand les classes ne varient plus
- Avantage :
  - Méthode rapide
- Inconvénients
  - Il faut choisir à priori le nombre de classes
  - Résultat qui dépendent des centres choisis à la 1ère étape
  - Relancer plusieurs fois l'algo avec des centres différents

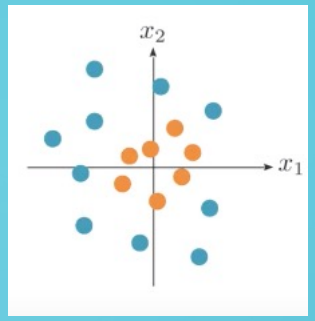


SOMMAIRE

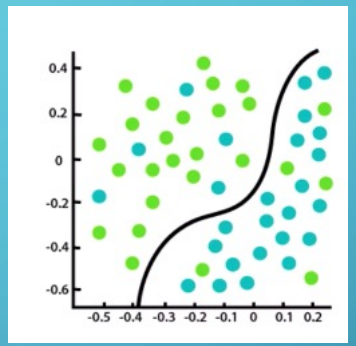
- Historique
- **Les bases du ML**
- Une méthode non supervisée : Kmean
- Les arbres de décision
- Ensemble learning

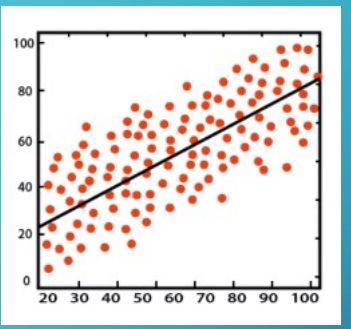
ARBRE DE DÉCISION

- Méthode supervisée
- Classification ou régression
- Algorithme transparent



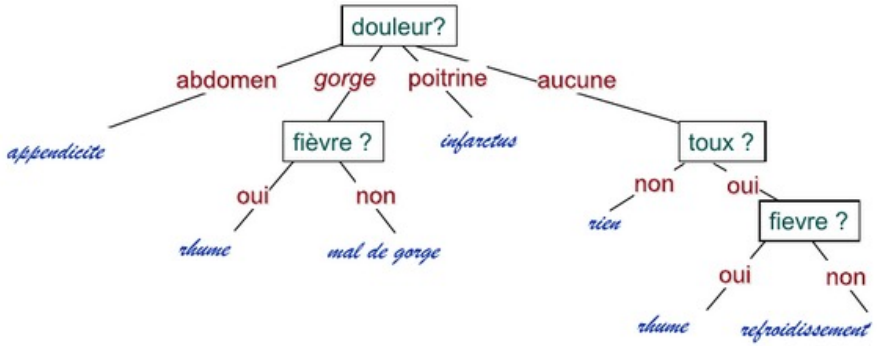
Classification





Régression

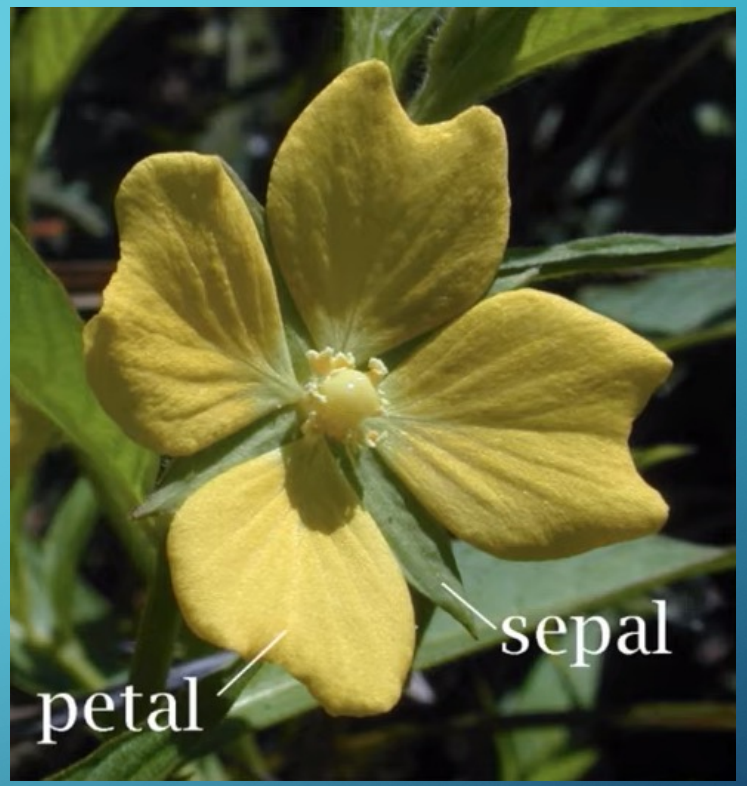
	Toux	Fièvre	Poids	Douleur
Marie	non	oui	normal	gorge
Fred	non	oui	normal	abdomen
Julie	oui	oui	maigre	aucune
Elvis	oui	non	obese	poitrine



ARBRE DE DÉCISION

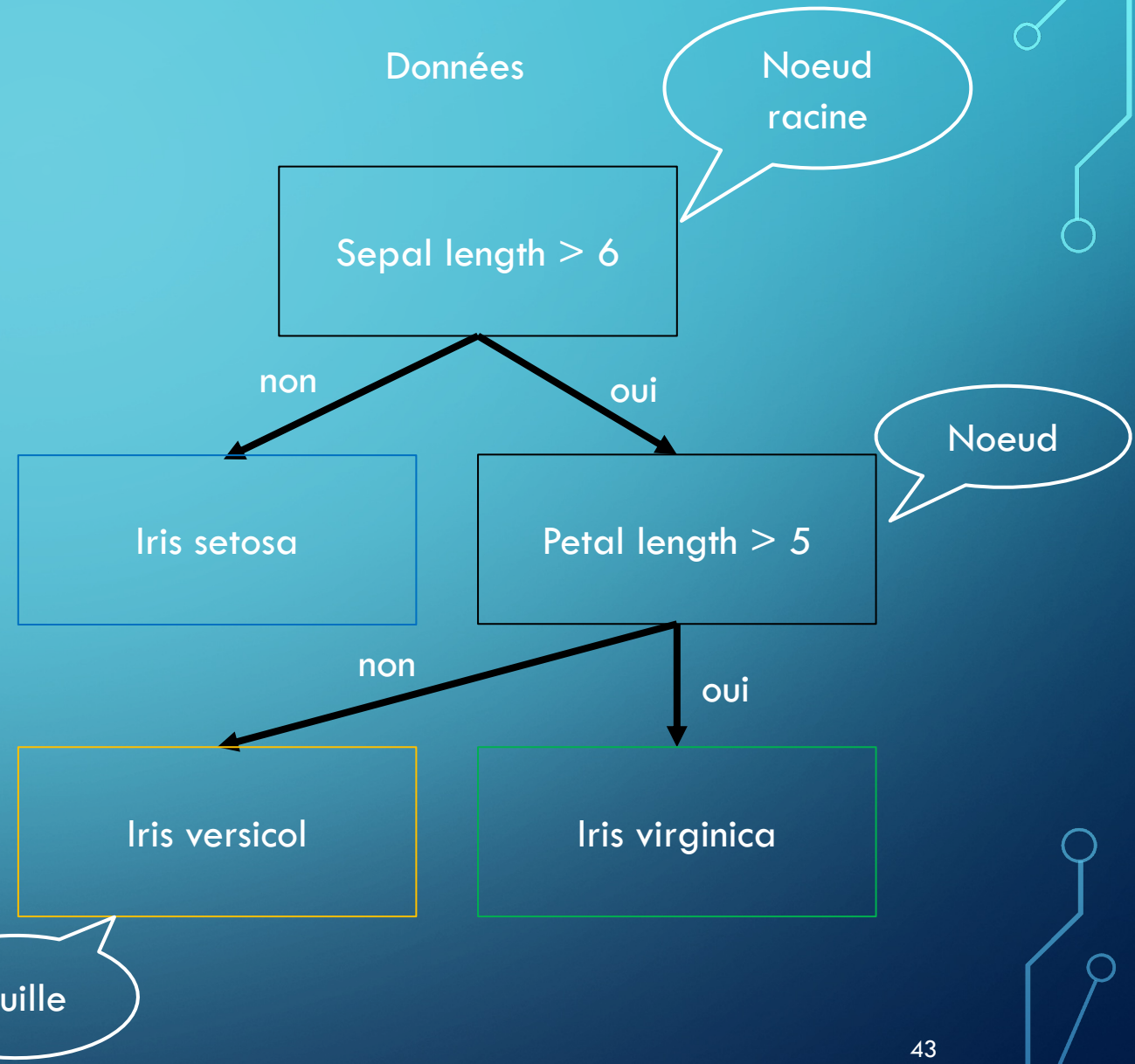
- Exemple avec la base de données Iris

Sepal Length	Sepal width	Petal Length	Petal width	Espèce
5.1	3.5	1.4	0.2	Iris setosa
7.0	3.2	4.7	1.4	Iris versicolor
6.3	3.3	6.0	2.5	Iris virginica

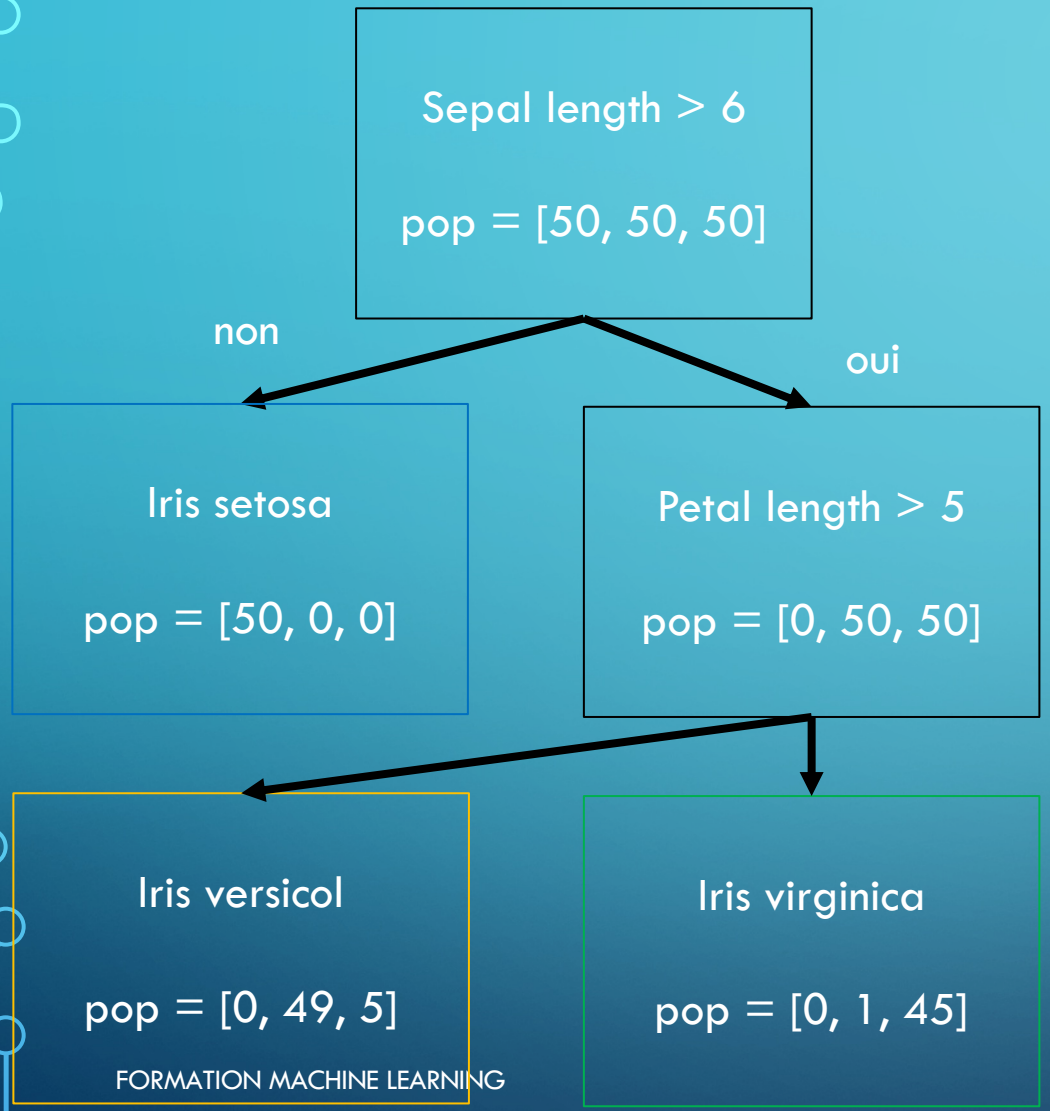


ARBRE DE DÉCISION

Sepal Length	Sepal width	Petal Length	Petal width	Espèce
5.1	3.5	1.4	0.2	Iris setosa
7.0	3.2	4.7	1.4	Iris versicolor
6.3	3.3	6.0	2.5	Iris virginica



ARBRE DE DÉCISION



probabilité d'appartenance à une classe

pop = [nb setosa, nb versicolor, nb virginica]

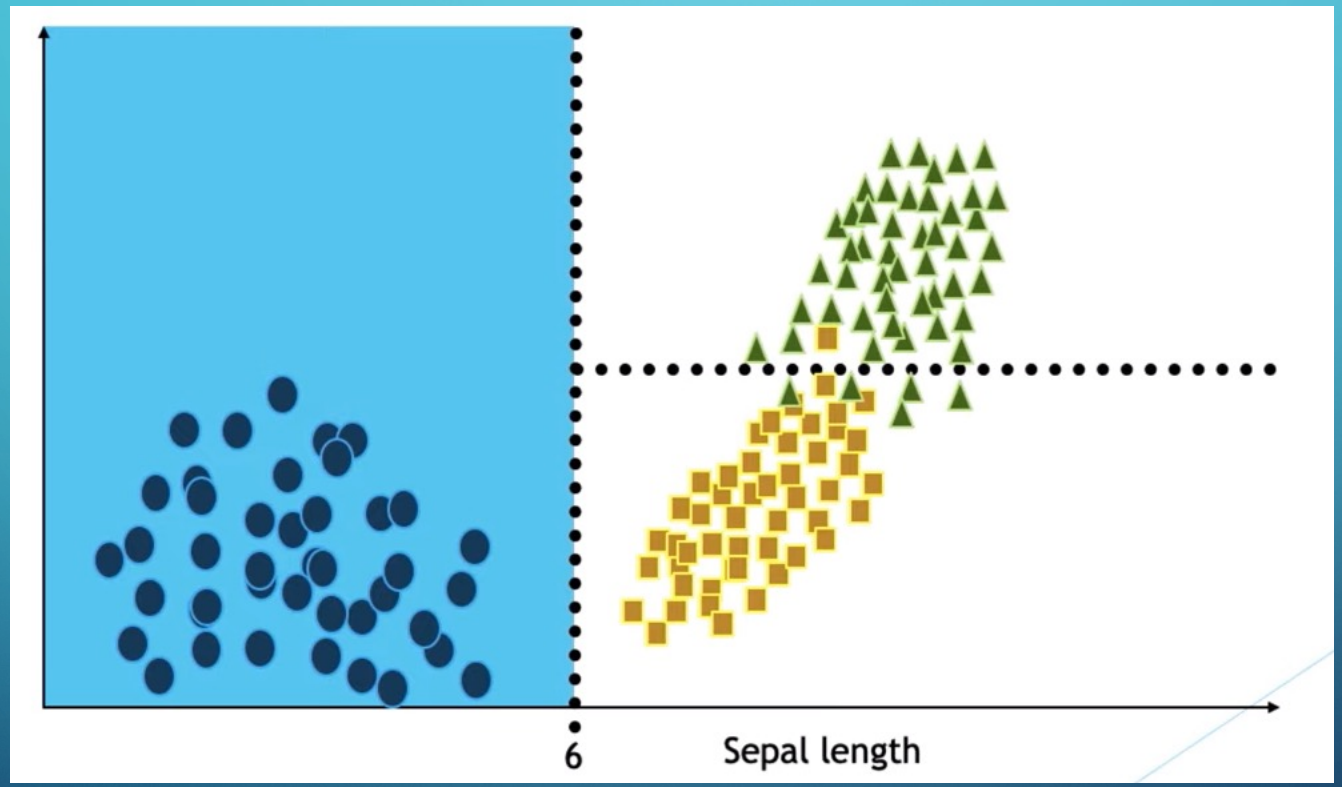
Par exemple :

Identifions une fleur avec un sépale de 6.5 cm et un pétale de 4 cm

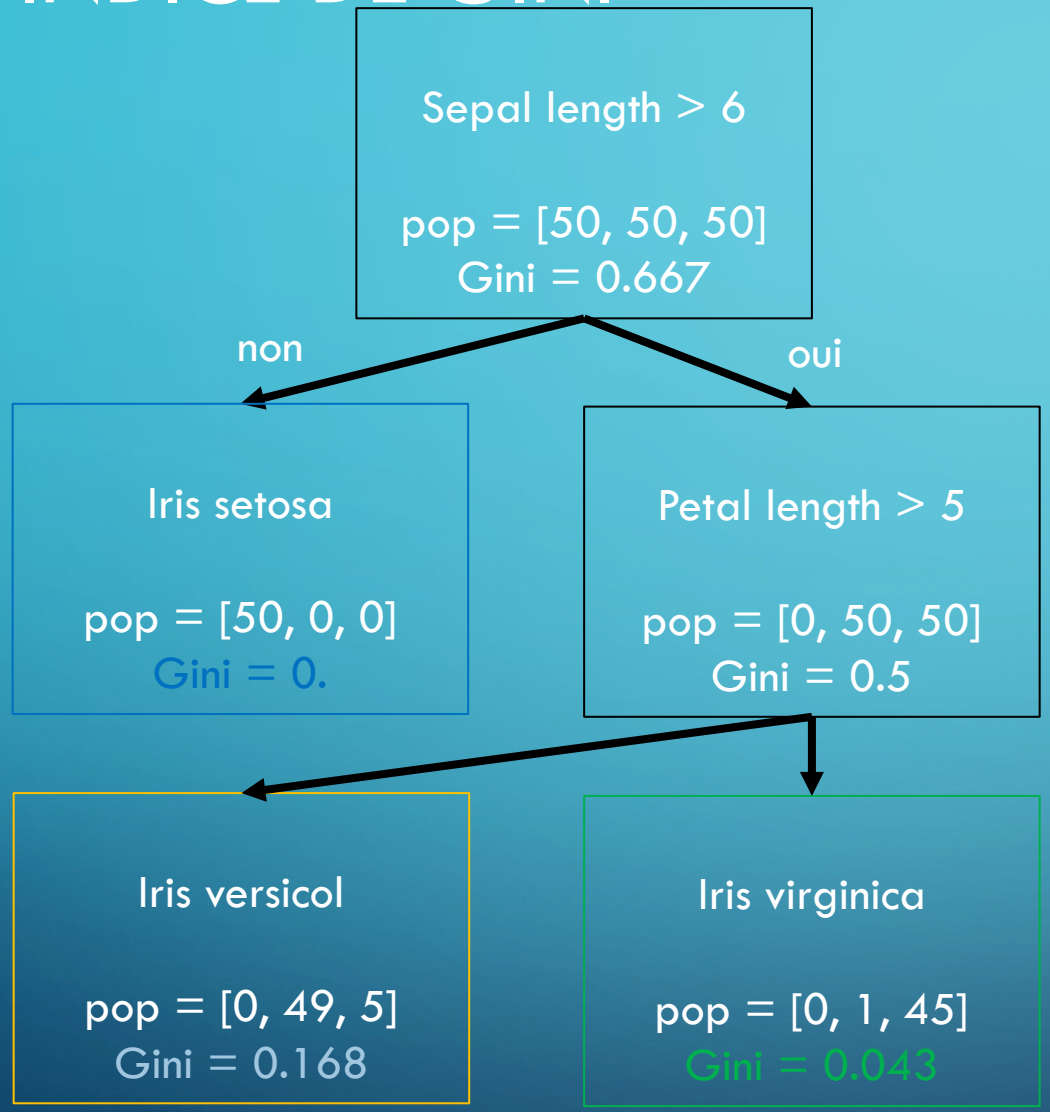
Setosa :	0/54 = 0
Versicolor :	49/54 = 0.91
virginica :	5/54 = 0.09

ARBRE DE DÉCISION

Frontière de décision



INDICE DE GINI



$$G_i = 1 - \sum_{k=1}^{n} {p_{i,k}}^2$$

$p_{i,k}$ est le ratio du nombre d'individus de la classe k parmi la population du $i^{ème}$ noeud.

$$\text{gini} = 1 - \left(\frac{50}{150}\right)^2 - \left(\frac{50}{150}\right)^2 - \left(\frac{50}{150}\right)^2 =$$

$$\text{gini} = 1 - \left(\frac{50}{50}\right)^2 - \frac{0}{50} - \frac{0}{50} = 0$$

$$\text{gini} = 1 - \frac{0}{50} - \left(\frac{50}{100}\right)^2 - \left(\frac{50}{100}\right)^2 = 0.5$$

$$\text{gini} = 1 - \frac{0}{54} - \left(\frac{49}{54}\right)^2 - \left(\frac{5}{54}\right)^2 = 0.168$$

$$\text{gini} = 1 - \frac{0}{46} - \left(\frac{1}{46}\right)^2 - \left(\frac{45}{46}\right)^2 = 0.043$$

ARBRE DE DÉCISION, PURETÉ

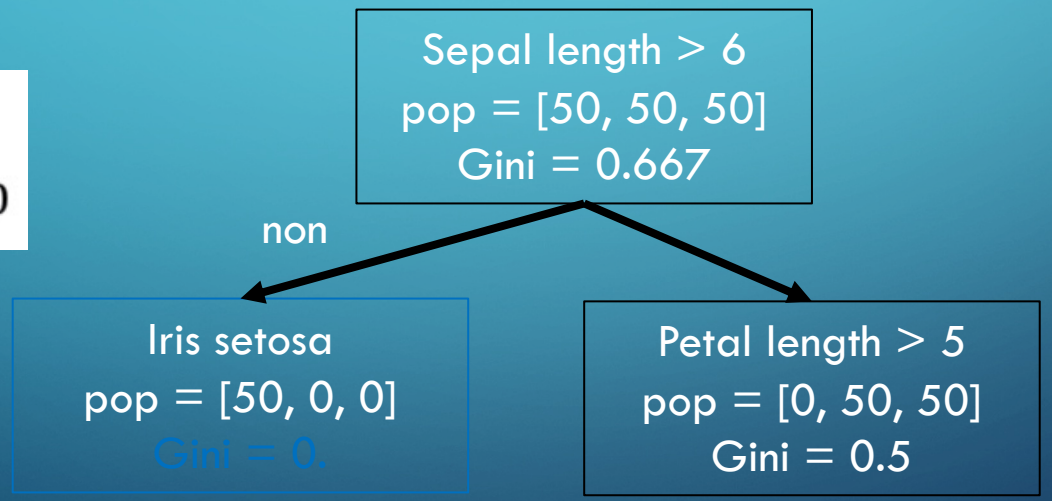
- La pureté représente le cout du noeud

$$J(k) = \frac{m_{gauche}}{m} G_{gauche} + \frac{m_{droite}}{m} G_{droite}$$

$$Ou \begin{cases} G_{gauche/droite}\ mesure\ l'impureté\ du\ sous\ ensemble\ droite/gauche \\ m_{gauche/droite}\ est\ la\ proportion\ de\ notre\ population\ du\ sous\ ensemble\ droite/gauche \end{cases}$$

$G_{gauche} = 0$

$\frac{m_{gauche}}{m} = 50/150$

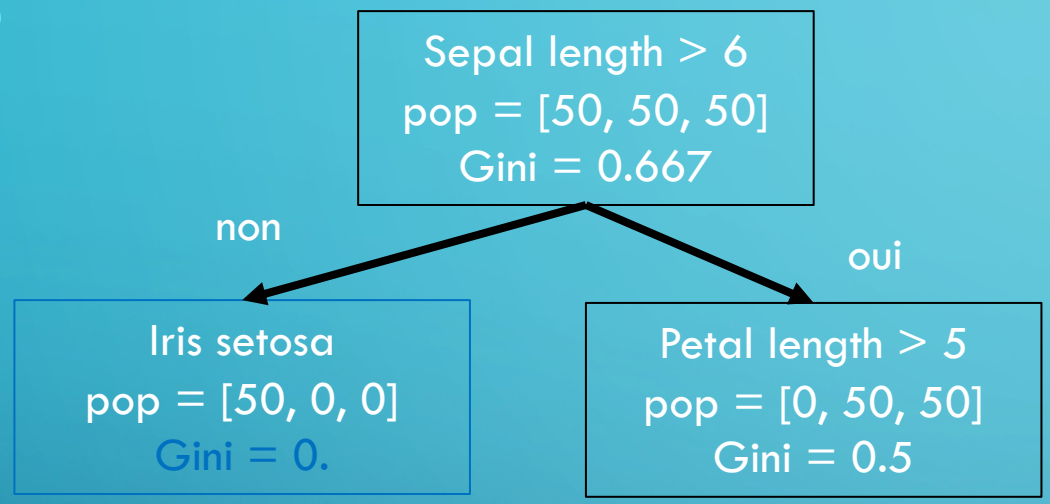


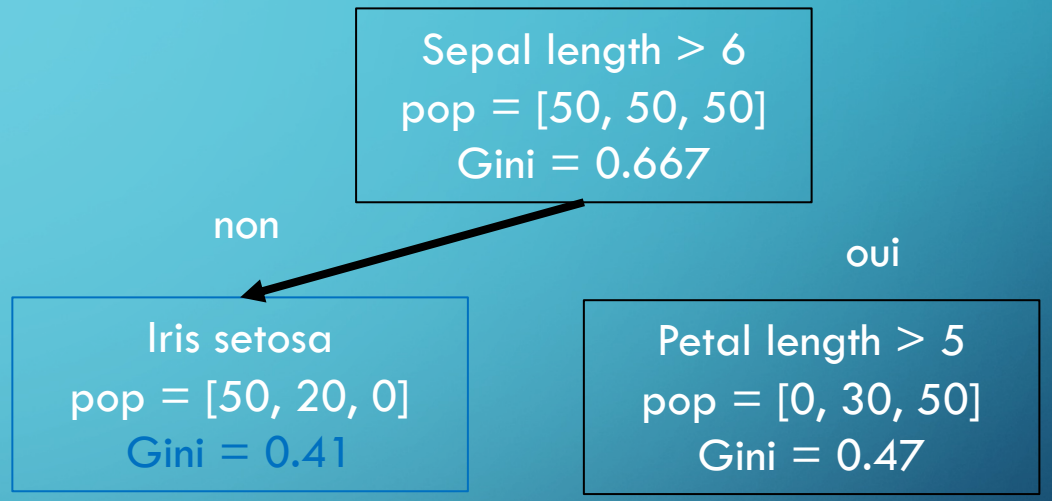
$G_{droite} = 0.5$

$\frac{m_{droite}}{m} = 100/150$

$$J(0) = \frac{50}{150}\ 0 + \frac{100}{150}\ 0.5 = 0.33$$

ARBRE DE DÉCISION, CHOIX DES NOEUDS





$$J(0) = \frac{50}{150}\,0 + \frac{100}{150}\,0.5 = 0.33$$

$$G_i = 1 - \sum_{k=1}^{n} {p_{i,k}}^2$$

$$J(0) = \frac{70}{150}\,0.41 + \frac{80}{150}\,0.47 =$$

ARBRE DE DÉCISION

- Avantage :
  - facile à entrainer
  - facile à utiliser
  - interprétable (algorithme transparent)

- Mais…
  - peu précis
  - généralisation qui manque de fiabilité

SOMMAIRE

- Historique
- **Les bases du ML**
- Une méthode non supervisée : Kmean
- Les arbres de décision
- Ensemble learning

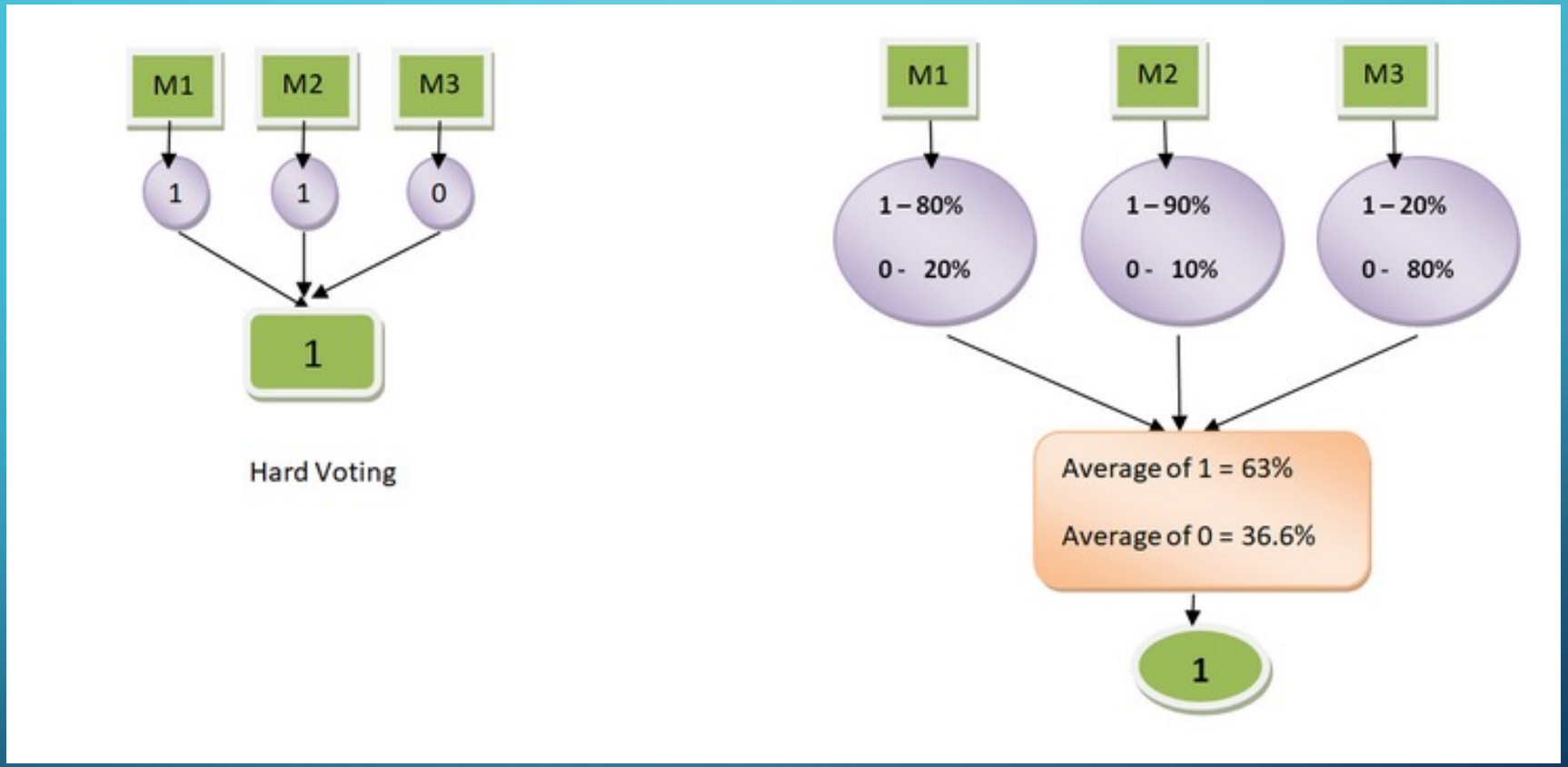
ENSEMBLE LEARNING

- Entrainement de plusieurs petits modèles de ML pour en faire un modèle plus performant
- Variabilité inter et intraopérateur
  - intra-operateur : variabilité pour un même opérateur
  - Inter-opérateur : variabilité pour des opérateurs différents
  - exemple
    - segmentation de tumeur
    - estimer le prix d'une maison
- Suppose que la résolution d'un problème est plus efficace par une foule que par un expert seul
- Hypothèse sur la foule :
  - diversité
  - indépendance
  - décentralisation : les jugements s'additionnent, pas d'autorité supérieure pour décider
- Les méthodes d'ensemble learning utilisent plusieurs algorithmes d'apprentissage et prennent en compte les résultats de ces modèles afin d'obtenir de meilleurs performances prédictives que les modèles pris séparément.

ENSEMBLE LEARNING, EXEMPLE

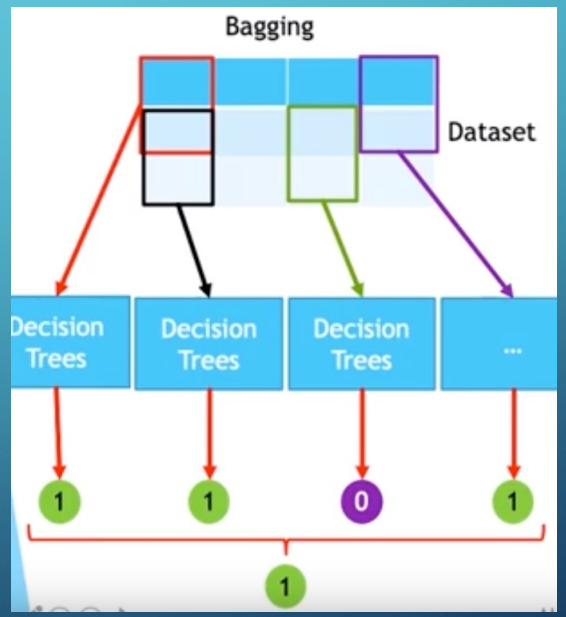
Index	Modèle 1	Modèle 2	Modèle 3	Modèle 4	Modèle 5	Mélange
1	1	1	0	0	1	1
2	1	1	1	0	0	1
3	0	0	1	1	1	1
4	0	1	1	1	0	1
5	1	0	1	0	1	1
	60%	60%	60%	60%	60%	100%

ENSEMBLE LEARNING



RANDOM FOREST

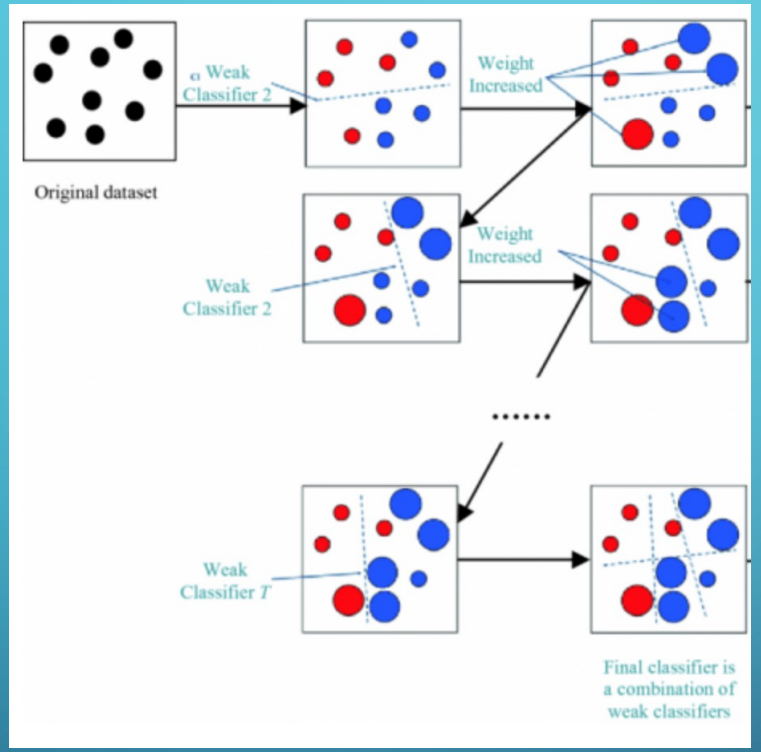
- Pour palier le problème de généralisation des arbres de décision, on utilise plusieurs arbres de décision : Random Forest
- Bagging : On entraine l'arbre de décision seulement sur
  - une partie des données
  - une partie des variables



RANDOM FOREST

- Les étapes
  - Répéter les étapes jusqu'à avoir le nombre d'arbres souhaité
    - Créer un jeu de données (avec le bagging on peut prendre plusieurs fois les mêmes données). La sélection aléatoire renforce la variabilité des arbres
    - Entrainement de l'arbre
  - Obtention d'une forêt d'arbres.
  - On moyenne les résultats de la forêt d'arbres

BOOSTED TREE



REMERCIEMENTS

- Yann Coadou – CPPM
- Geoffrey Daniel – CEA-Saclay
- Marc Duranton – CEA-Saclay