Informations

Mes coordonnées

Abdulrahman ALLOUCHE
Bâtiment Brillouin (2ème)
allouchear@univ-lyon1.fr

Enseignants

Cédric Ray ; cedric.ray@univ-lyon1.fr (3 premières séances de TP)
Abdulrahman Allouche ; allouchear@univ-lyon1.fr (CM+reste des TP)

Supports

moodle.univ-lyon1.fr
Mes cours actuels :
Intelligence Artificielle pour la physique
Tableau de bord

Informations

Prérequis

- Quelques notions en algèbre linéaire
- Avoir déjà programmé

Modalité de contrôle de connaissance :

- Contrôle partiel : Projet (binôme) début vers 8 décembre: 2semaines pour rendre le projet
- 1ère session : Examen individuel sur machine : 1h30
- 2ème session : Examen individuel sur machine : 1h00

Note Finale : 0,4 Projet + 0,6 Examen individuel

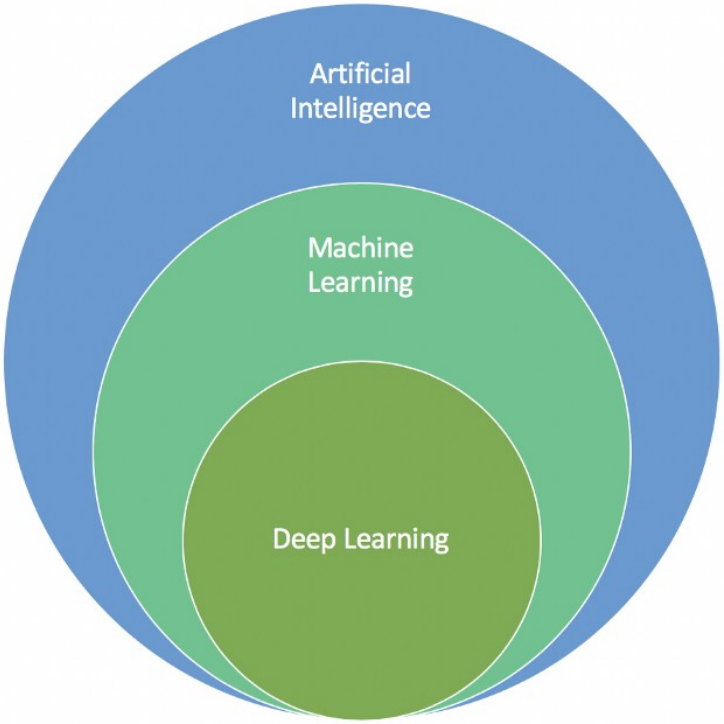
Logiciels nécessaires : Anaconda. 3 méthodes possibles :

- **jupyter.univ-lyon1.fr (serveur M2) :**
  machine personnelle ou machine en salle des TP
- **Machine dans les salles des TP. Tout est déjà installé**
- **Machine personnelle en local :**
  Installer Anacanda AVANT le début de TP
  Voir le fichier **InstallAnacanda.pdf** sur moodle (Students/Courses/Tutorials)

Some definitions

Artificial Intelligence (AI) :

A field of computer science thats aims to make computers achieve human-style intelligence.



Machine Learning (ML) :

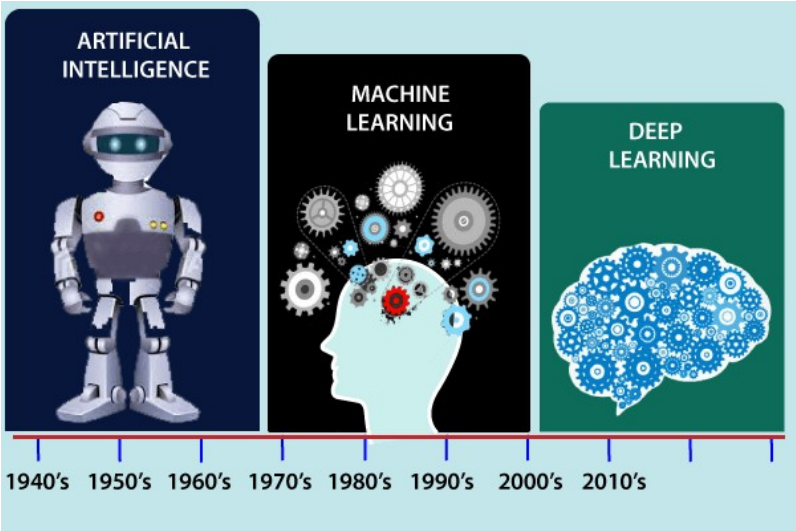
- Subset of AI.
- The science (art) of programming computers how to **learn from data.**

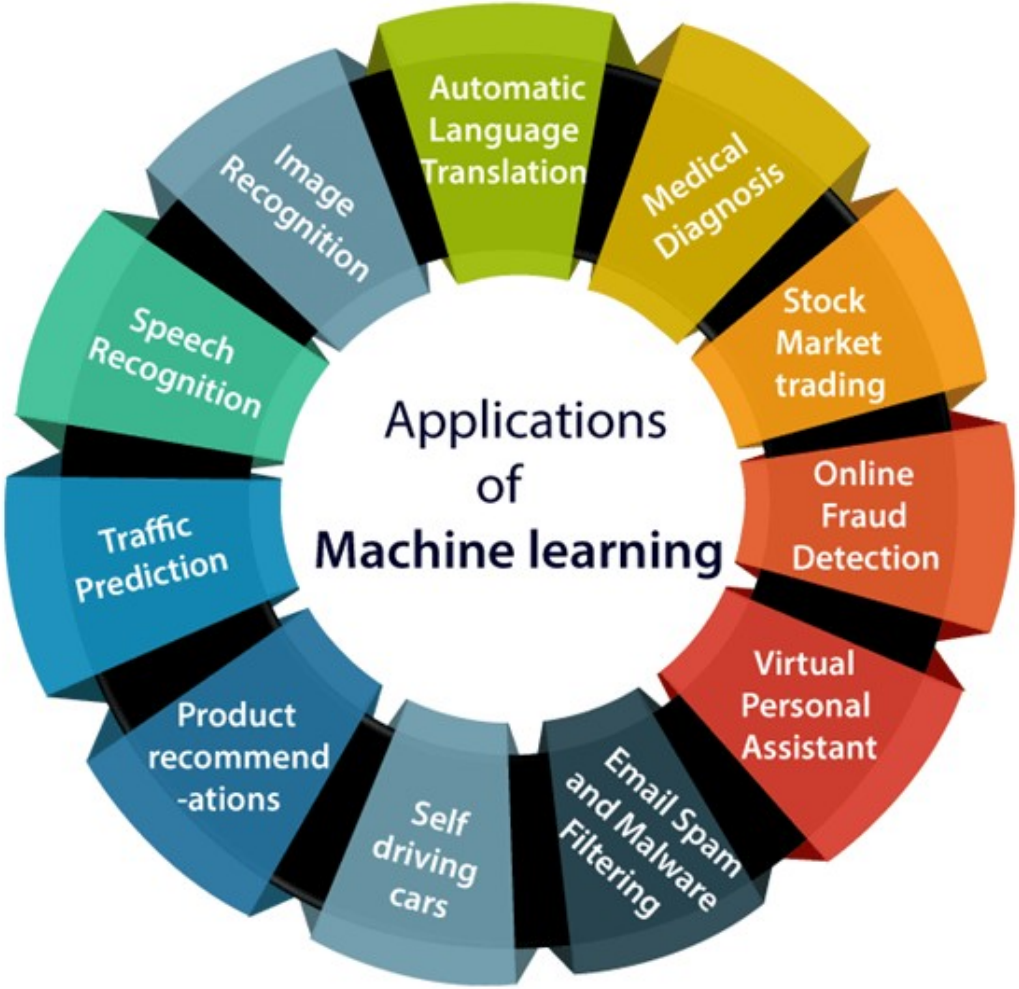
- ML algorithms are **not only "instructions"** to the computer but **dynamic models that analyze how well things are working**, using the information they get to improve. it changes and ouputs an algorithm as well as an ouput

Deep Learning :

- Subset of Machine Learning
- Use artificial neural networks réseau de neurone
- Analyzes data with a **logical structure similar to how a human** would draw conclusions

Applications





Applications

iLM

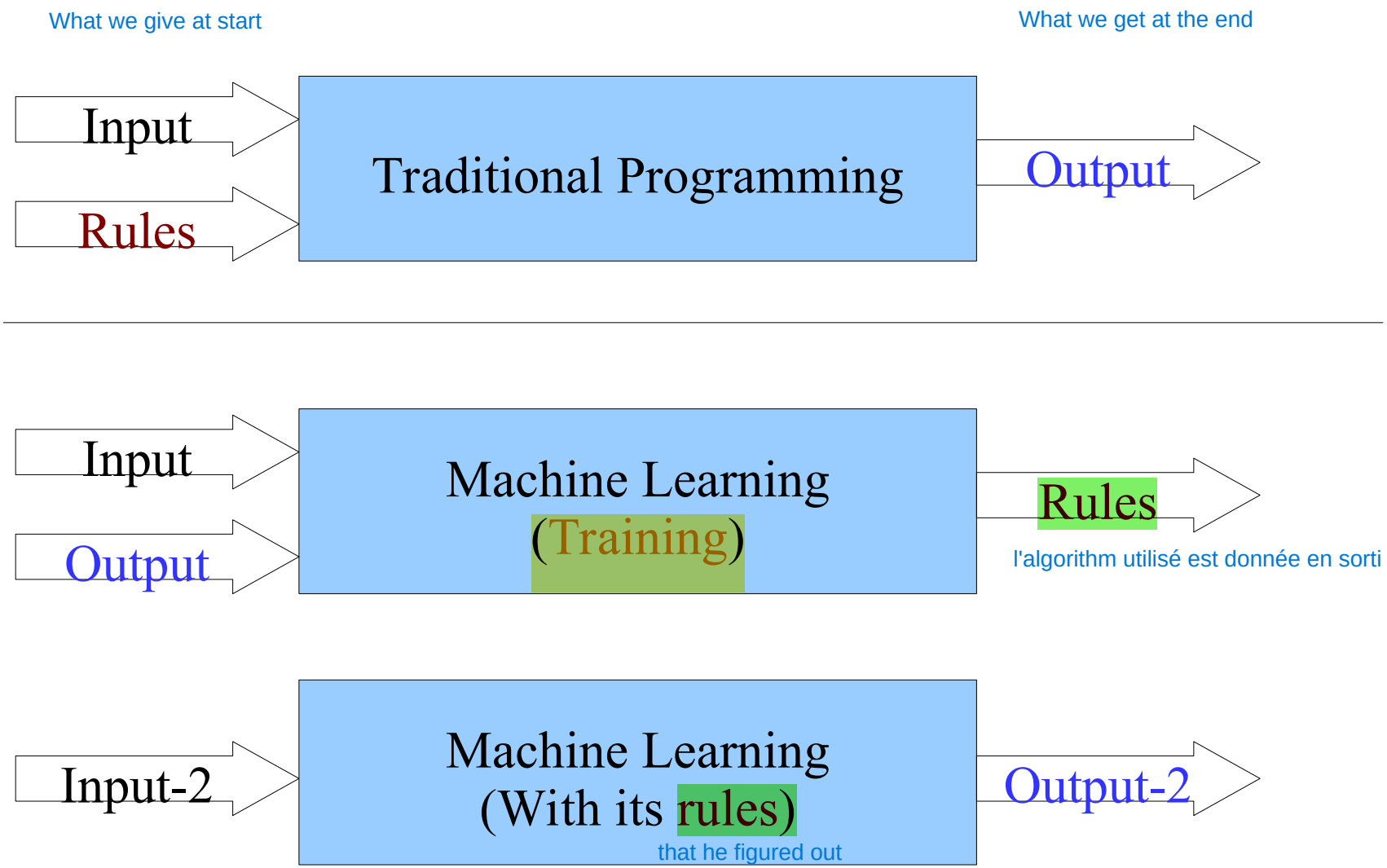
- Machine learning interatomic potential for transformation-induced plasticity of zirconia
- Repulsive potentials for the DFTB method from neural networks
- Control of non-equilibrium systems
- Biomedical image analysis with CNNs
- Imagerie LIBS et Intelligence Artificielle : une belle combinaison !
- Prediction of laboratory quakes with machine learning
- Imagerie LIBS et Intelligence Artificielle : Etude des mortiers archéologique
- Laboratory evaluation of the scattering matrix of ragweed, ash, birch and pine pollens towards pollen classification
- Quantum Mechanics and Machine-Learning Calculations for Vibrational Frequencies
- Neural network approach for a rapid prediction of metal-supported borophene properties
- ……...

TP

- Project : Prediction of atomization energies of molecules. ML methods without Schrodinger eq will be used, multiple approaches
- COVID-19 can be detected from the Raman spectrum of serum using machine learning. an exam exercice example
- *Model capable of classifying images of moles into benign and malignant*
- *Recurrent Neural Networks, forecast weather based on input features (T, P, ...)*

optional

Machine Learning vs. Traditional Programming



The methodology

Machine learning Life cycle

3. Build Model

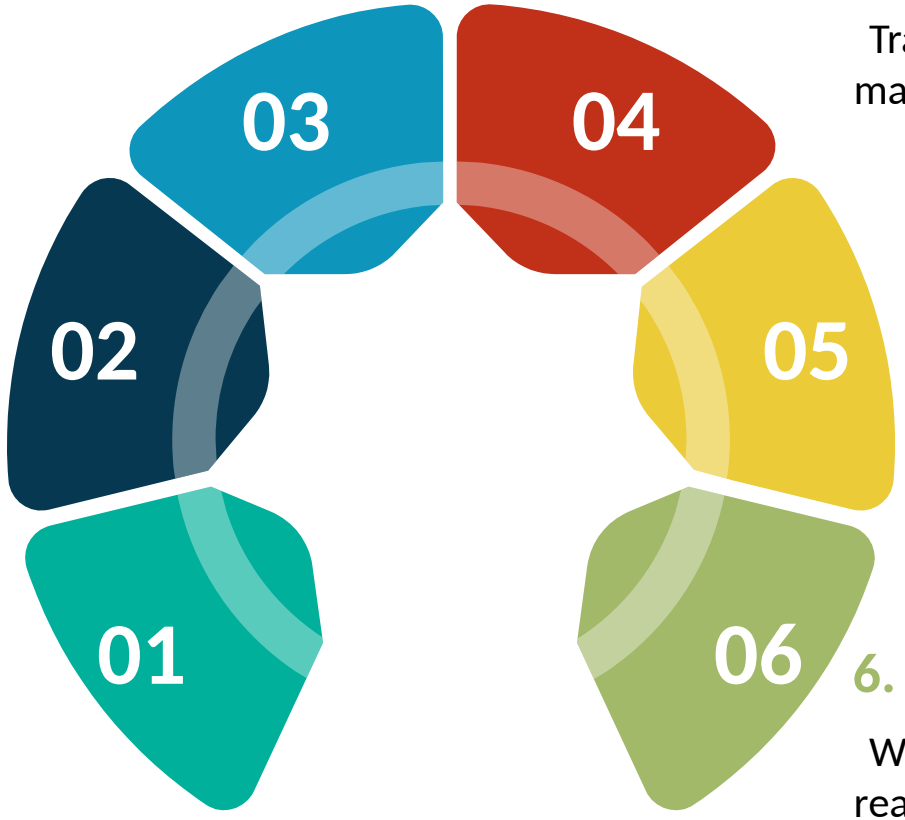
- Determination of the machine learning techniques
- Build the model

2. Data Preparation

- Data exploration: characteristics, format, quality
- Missing Values
- Duplicate data
- Invalid data
- Noise
- Data pre-processing: Standardization, Convert categorical data

1. Gathering Data

- Identify various data sources
- Collect data
- Integrate the data/different sources



4. Train Model

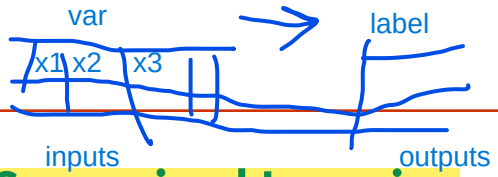
Train the model using various machine learning algorithms

5. Test Model

Check the accuracy of our model using a test dataset

6. Deployement

We deploy the model in the real-world system.

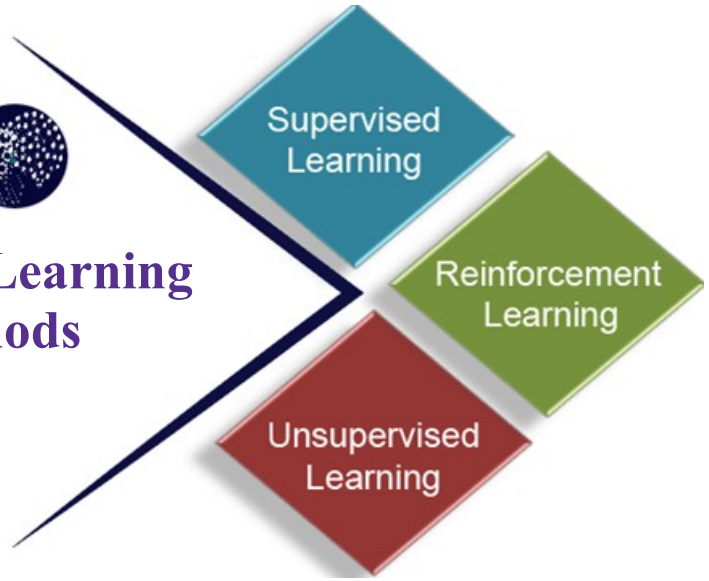


Supervised Learning and Unsupervised Learning will be treated

Machine Learning Methods

Supervised Learning

- Map input data with the output data.
- Supervision, as student learns in the supervision of the teacher.
- Example : spam filtering
- Can be grouped in two categories :
  - Classification un entier/ true false compte aussi
  - Regression les labels sont des reels ex spectre d'energie



Unsupervised Learning

- Data not labeled, classified, or categorized
- The algorithm needs to act on that data without any supervision.
- Can be grouped in two categories :
  - Clustering
  - Dimensionality Reduction

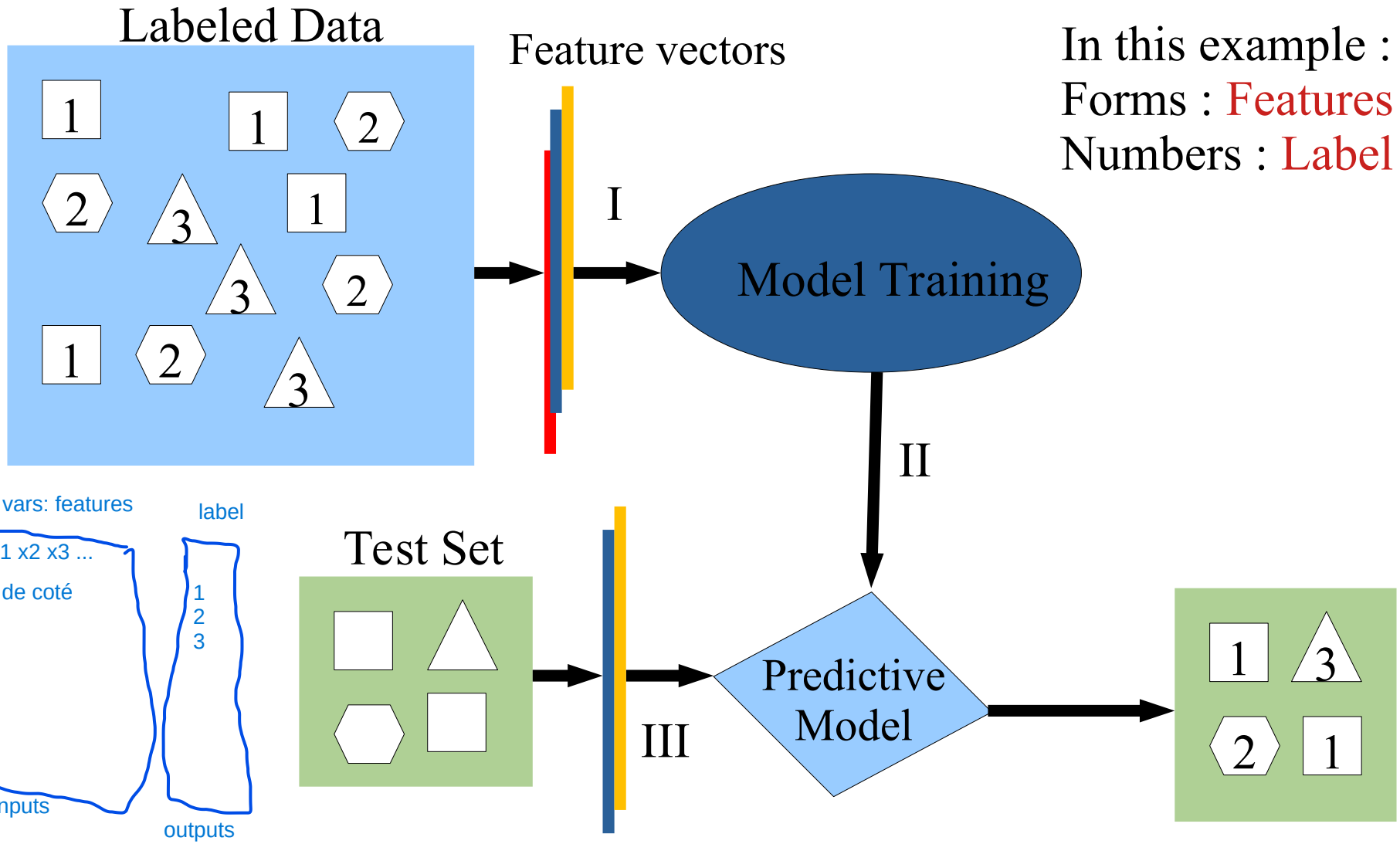
diminuer le nombre de varaible en eliminant des colonnes de var redondant ou tres correlée

Reinforcement Learning

- Feedback-based learning method
- Agent gets a reward for each right action and gets a penalty for each wrong action.
- Example : The robotic dog, which automatically learns the movement of his arms,

Exmaple: consider vars of forms as labeled data set

Supervised learning

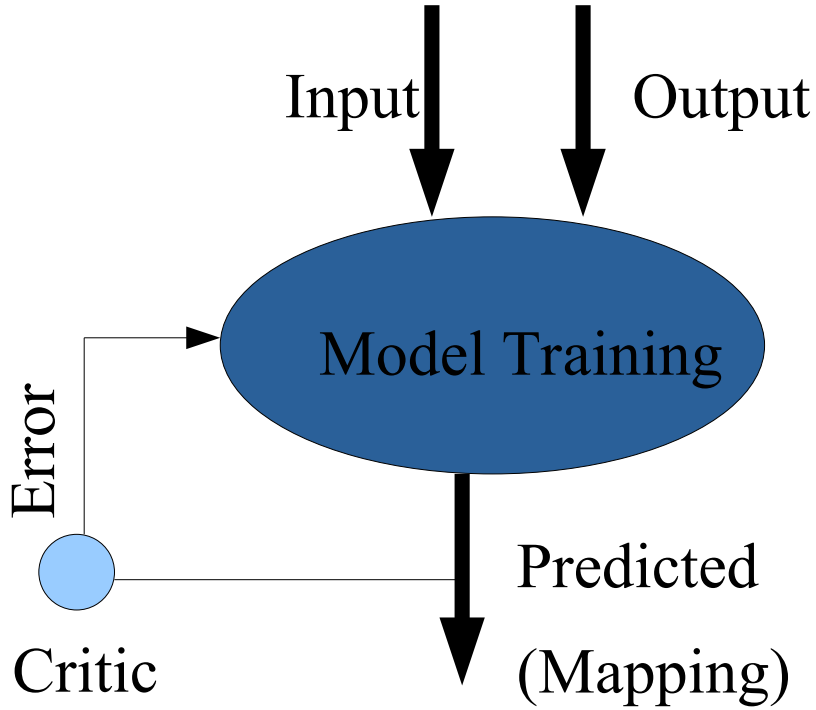


Supervised learning

• You have **input variables** (X) **and an output** variable (Y) and you use model to learn the **mapping function from the input to the output**.

$$Y = f(X)$$

• When you have new input data (X) that you can **predict** the output variables (Y) for that data.



Ouput - Predicted >=0
a minimiser par la machine

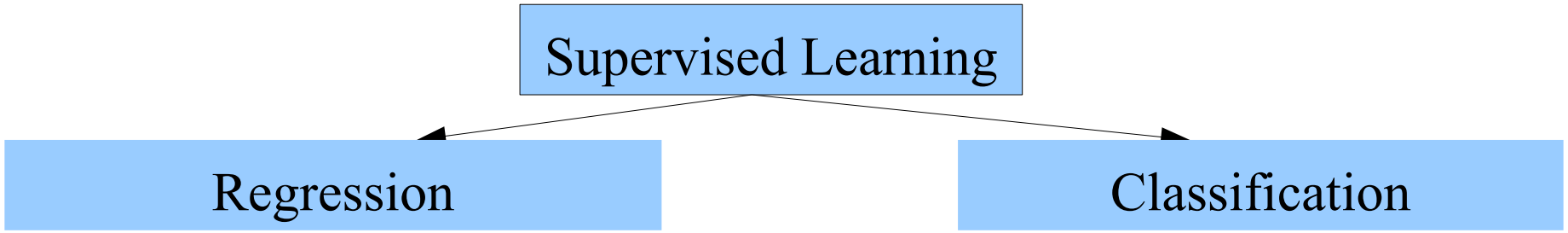
Model : Given algorithm (the form of the function)

on donne la forme de la fonction avec des parametre aleatoire initial que la machine va optimiser et trouver

Train : Search the optimal parameters of the function

There are two types of predetermined functions that are used in Supervised Learning

Supervised learning



The output variable is a real value such as “dollars” or “weight”

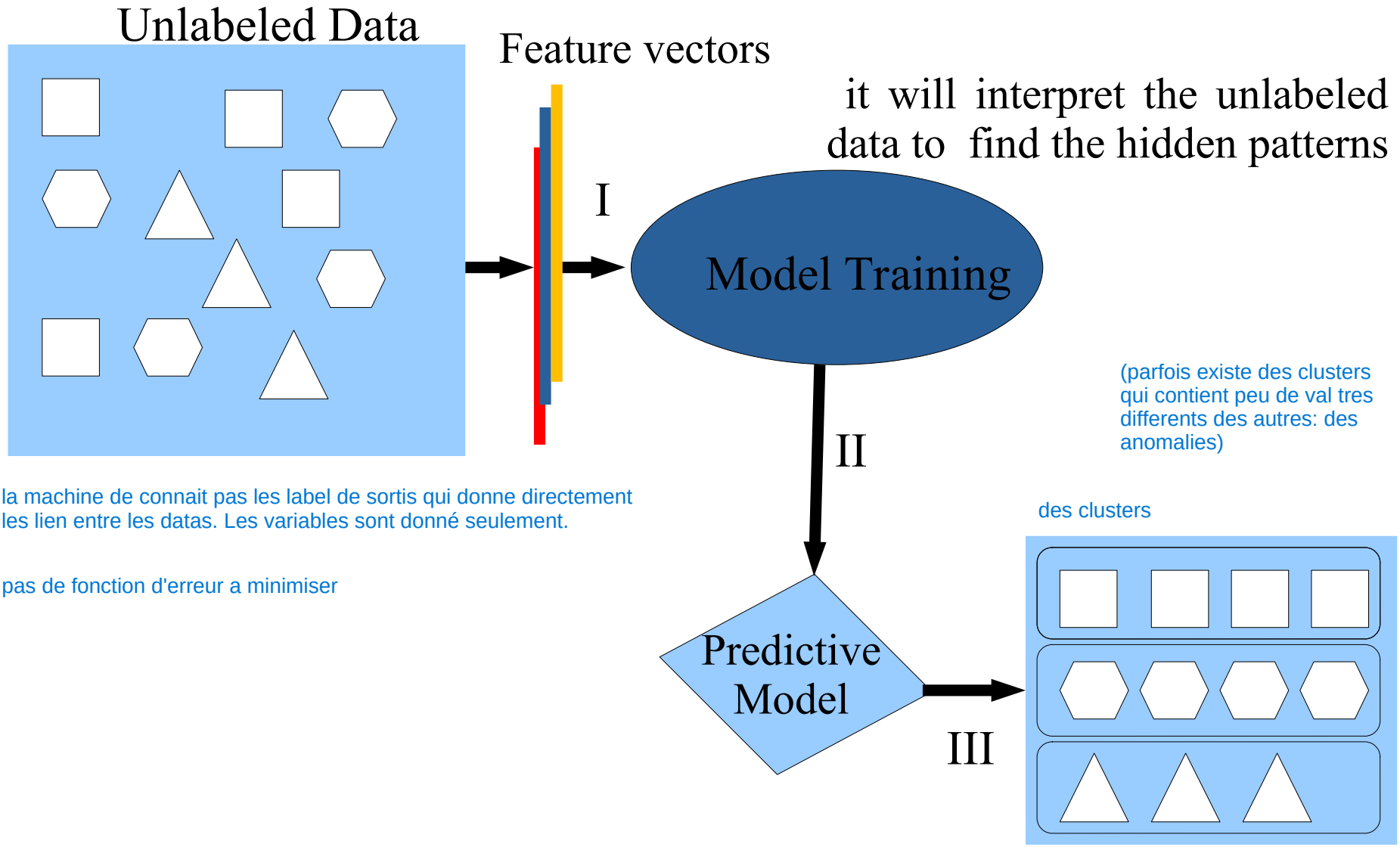
Linear regression
Polynomial Regression
Ridge Regression
Lasso Regression
Support vector Regression
Trees Regression
k-Nearest Neighbors
Random Forest Regression
Neural Networks

The output variable is an integer or a category, such as “red” or “blue”

Logistic Regression
Support vector Machines
Decision Trees
k-Nearest Neighbors
Random Forest

Neural Networks

Unsupervised learning



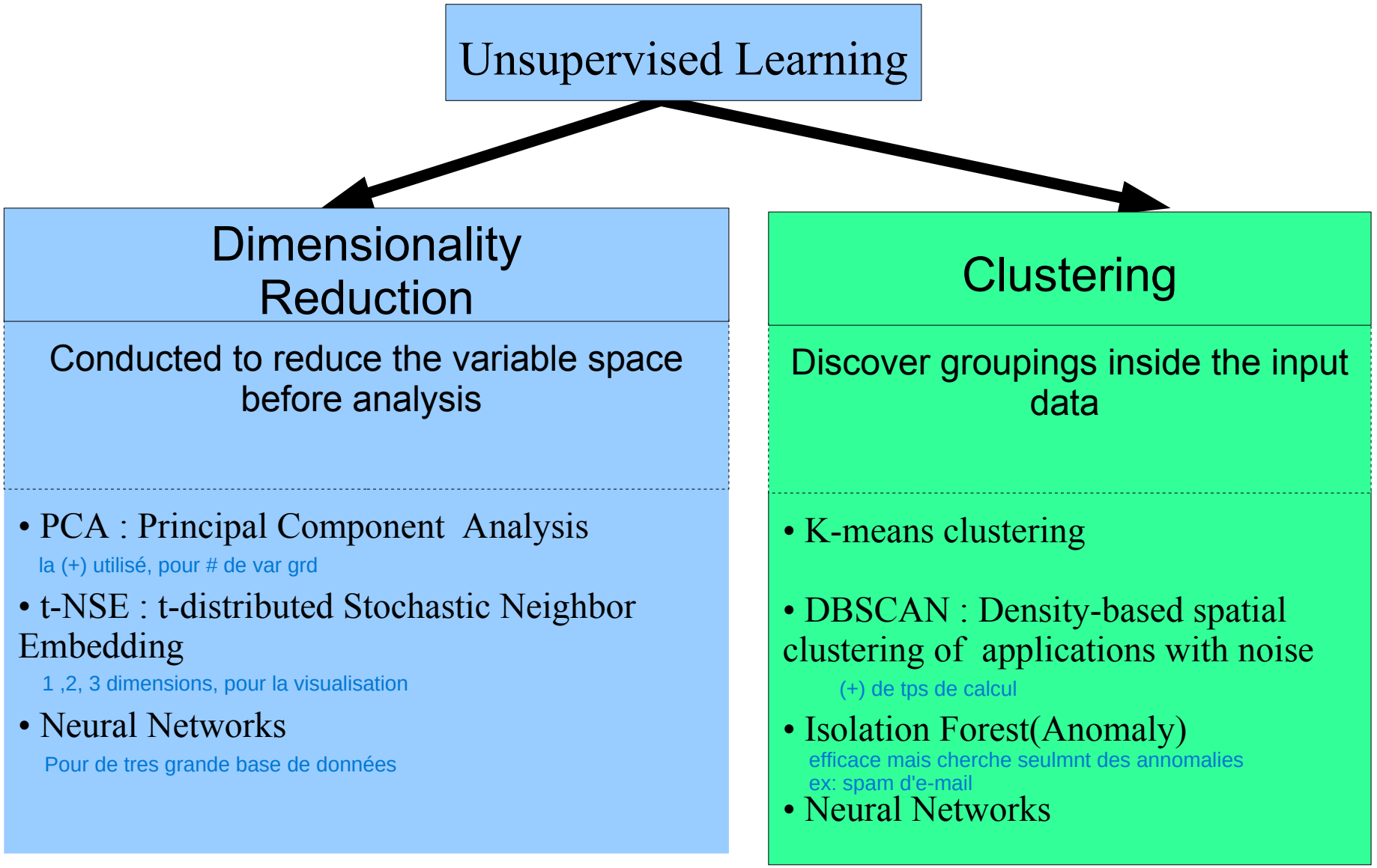
la machine de connait pas les label de sortis qui donne directement les lien entre les datas. Les variables sont donné seulement.

pas de fonction d'erreur a minimiser

(parfois existe des clusters qui contient peu de val tres differents des autres: des anomalies)

des clusters

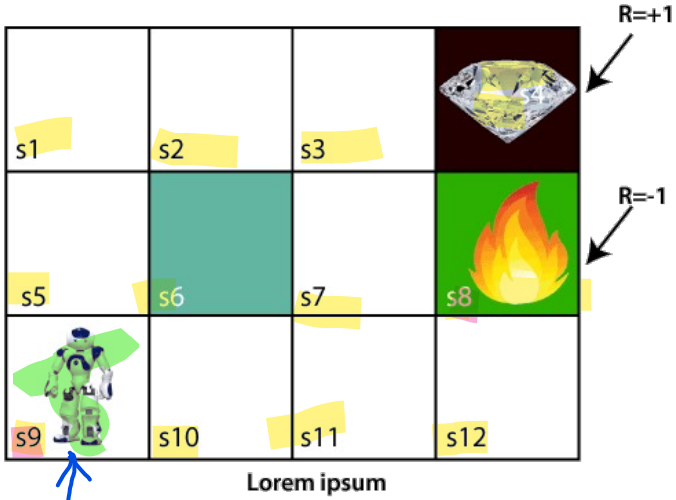
Unsupervised learning methods

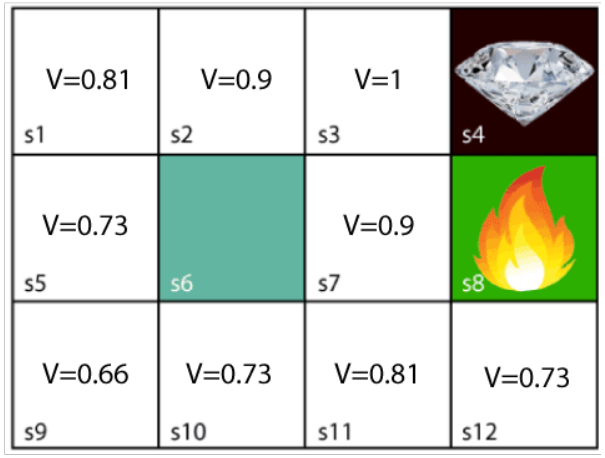


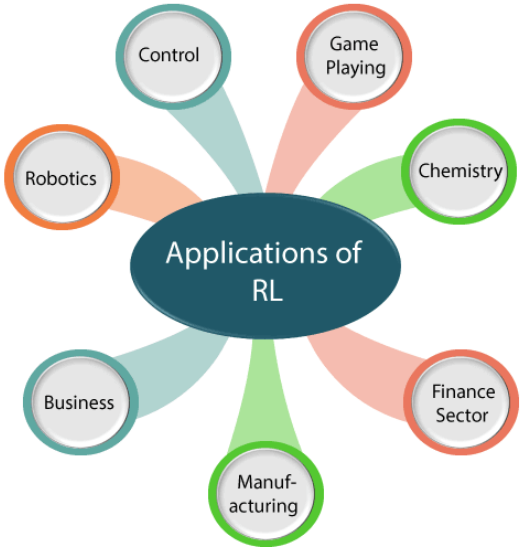
au depart: V=0 partout; on applique les rewards apres action, V des cases change apres le chemin parcouru, a la fin le best policy devient le choix de parcour qui maximise les V

Reinforcement Learning

The robot, diamond, and fire







- **Agent**(): An entity that can perceive/explore the environment and act upon it.
- **Environment**(): A situation in which an agent is present or surrounded by.
- **Action**(): Actions are the moves taken by an agent within the environment.
- **State**(): State is a situation returned by the environment after each action taken by the agent.
- **Reward**(): A feedback returned to the agent from the environment to evaluate the action of the agent.
- **Value**(): It is expected long-term retuned with the discount factor and opposite to the short-term reward.
- **Policy**(): Policy is a strategy applied by the agent for the next action based on the current state.

Tools needed to make Machine Learning

- **Anaconda :** Environment management for any language : Python, R, C/++, Fortran, ….
- **Python :** The reference language for Machine Learning.
- **NumPy** stands for Numerical Python
- **Matplotlib** is a comprehensive library for visualizations in Python.
- **Seaborn** is a Python data visualization library based on matplotlib. It provides a high-level interface for **drawing attractive and informative statistical graphics**.
- **Pandas** is a fast, powerful, flexible and easy to use open source **data analysis** and manipulation tool, built on top of the Python programming language.
- **Scikit-learn** (formerly scikits.learn and also known as sklearn) is a free software machine learning library for the Python programming language. It features **various** classification, regression and clustering **algorithms**
- **Tensorflow** : TensorFlow is the Google open source platform for machine learning. We use it for deep learning models.

See **InstallAnaconda.pdf** document to install all these tools (under anaconda)
(~ 10 GB without Tensorflow)

Course agenda

CM (9h00)

- General Introduction, 1h00
- Python Tutorial , 1h30
- NumPy Tutorial, 30 mn
- Pandas Tutorial, 30 mn
- Data Visualization Tutorial, 30 mn
- Supervised Learning Methods, 2h00
- Unsupervised Learning Methods, 1h00
- Introduction to Deep Methods, 2h00

Practical (13h30)

- Python, 1h30
- NumPy, 30mn
- Pandas & Visualization, 2h00
- Supervised & Unsupervised Methods, 3h30
- Project using ML Standard Methods, 1h30
- Deep Learning Methods, 3h00
- Project using Deep Learning, 1h30

To do at home

Before the first Practical session :

- Read InstallAnaconda.pdf
- Install Anaconda on your machine
- Read JupyterTutorial.pdf

- **Before the Visualization Tutorial session :**
  - Read CrashCoursesMathPlotLib.pdf