

AI2E : Workshop 12

RL : Q-Learning!

WHOA!

MADJDA ZERROUK
5th year student at ESI
Mobile developer at Yassir

01

From Classique ML to RL
A global View

02

What's RL
Analogy , elements,
mathematics, categories

03

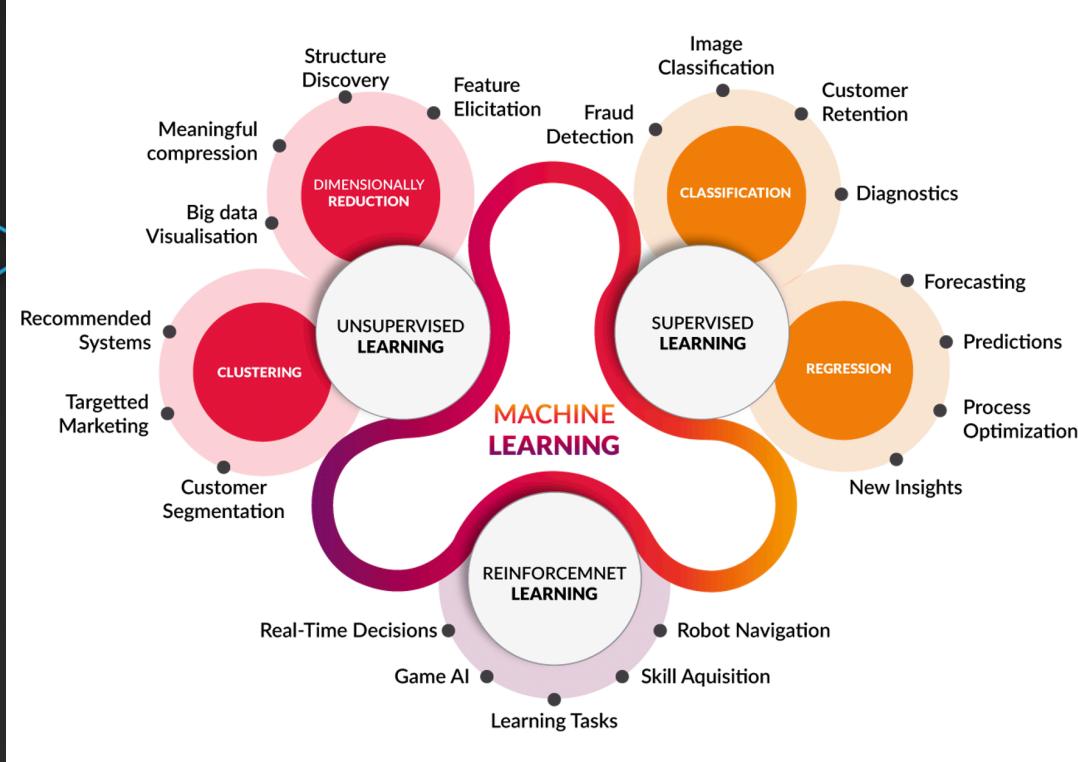
RL In The Practice
Use sectors , Challenges
and limitations

04

Q-Learning
Definition , process , a
practical case

01. From Classique ML to RL

A global View



RL In General

- Subclass of Machine Learning and an approach linked to Artificial Intelligence that pushes the machine to learn what to do in a certain state so as to maximize the total reward.
- Data in the form of actions & states
- Several algorithms and methods implement RL & speed up the learning process.

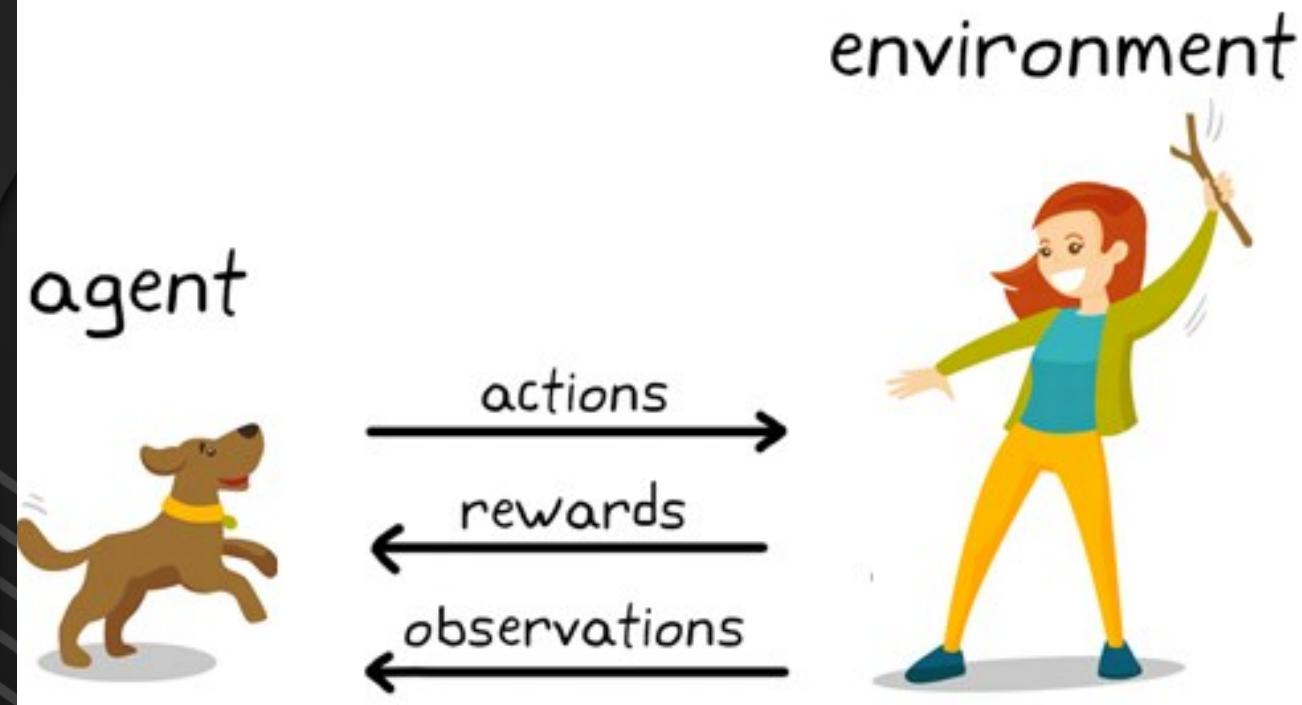
02. What's RL

Analogy , elements , mathematics ,
categories

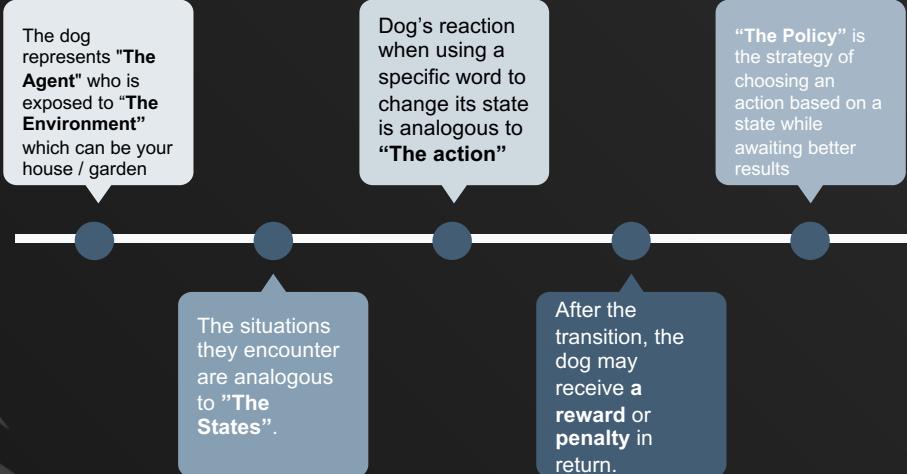


“a way of programming agents by reward and punishment without needing to specify how the task is to be achieved”
[Kaelbling, Littman, & Moore, 96]

RL's Analogy



RL's Elements



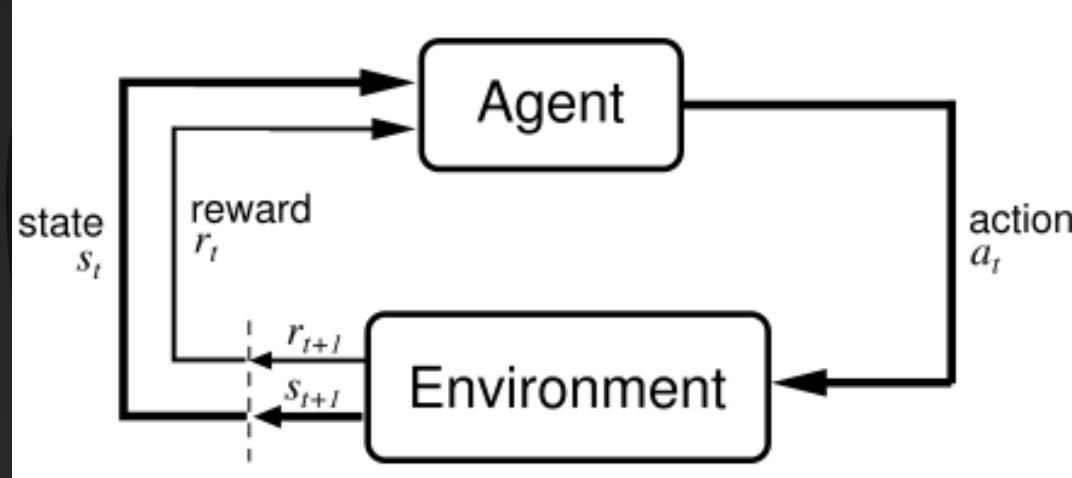


**More
formally ...**



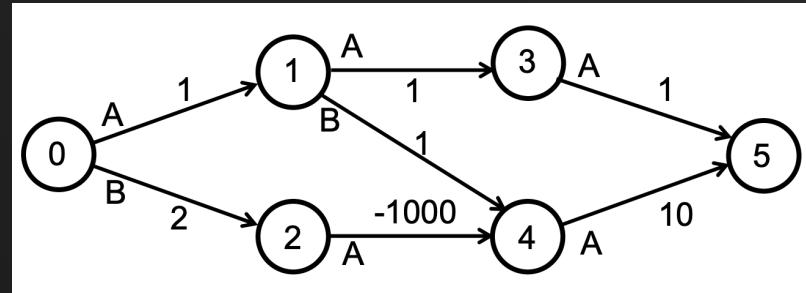
The standard model of reinforcement learning includes an agent that interacts with an environment through actions that generate a new state, the environment also responds with a reward: a numerical value that the agent aims to maximize.

More
formally ...



RL's mathematics

Markov Decision Processes



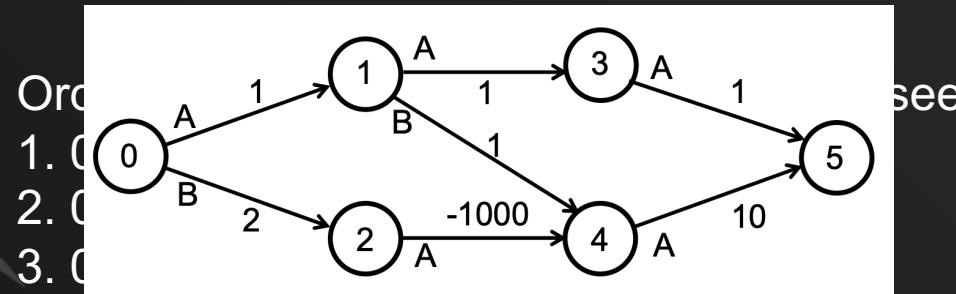
A set of states, $S = \{s_1, s_2, \dots, s_n\}$

A set of actions, $A = \{a_1, a_2, \dots, a_m\}$

A reward function, $R: S \times A \times S \rightarrow \mathcal{R}$

A transition function, $P : S \times A \times S \rightarrow [0, 1]$

We want to learn a policy, $\pi: S \rightarrow A$ that
maximize the sum of rewards



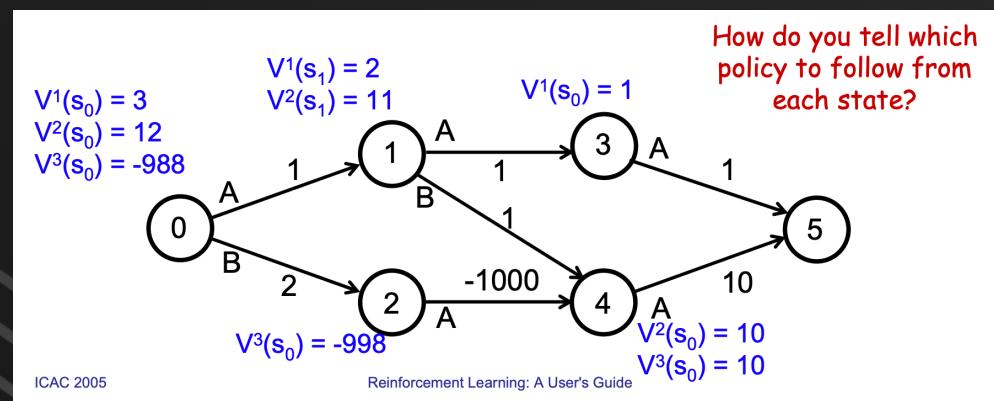
RL's mathematics

Policy

RL's mathematics

State value function V

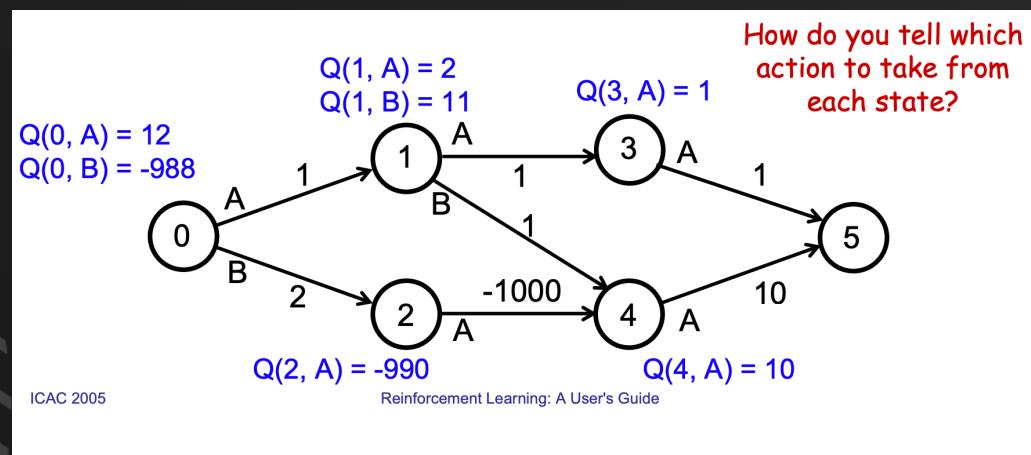
We can associate a value with each state
 For a fixed policy
 How good is it to run policy π from that state s
 This is the state value function, V



RL's mathematics

State-action value function, Q

We can define value without specifying the policy
 Specify the value of taking action a from state s
 and then performing optimally
 This is the state-action value function, Q



RL's categories

REINFORCEMENT
LEARNING

MODEL BASED

The agent learns from an environment-related model. This model is constructed from a reduced number of interactions with the real environment during the learning phase.

MODEL FREE

Value Based

The agent knows what to choose based on rewards (the strategy of exploration is well defined). The basis of RL, suitable for discreet environments & complex problems.

Policy Based

No exploration strategy is defined, self-decision. It works with continuous actions.

03. RL In The Practice

Use sectors , Challenges
and limitation

Use sectors



MANUFACTURING

Robots use RL to perform a certain mission



ENERGY SYSTEMS

RL is used to assess the security of electrical power systems and to develop control and protection systems.



FINANCE

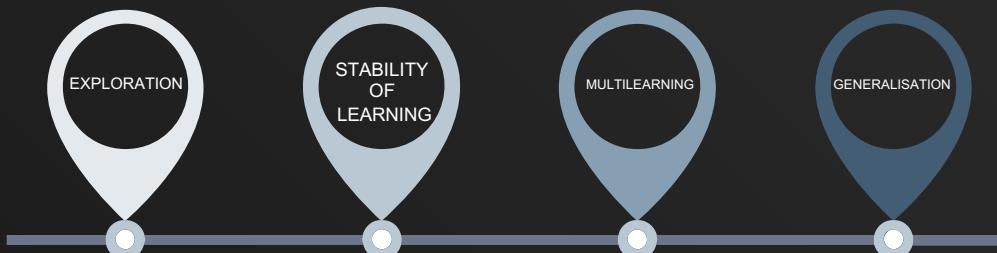
Stock management, evaluation of business decisions



GAMES SOLVING

The agent tries to beat his record every time Atari games .. Etc.

RL's challenges



04. Q-Learning

Definition , process ,
a practical case

Definition



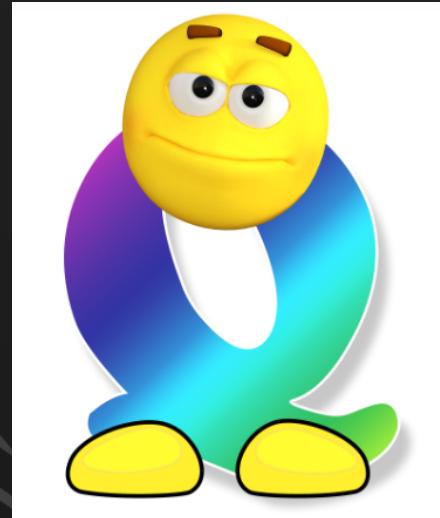
Q-learning is a non-political reinforcement learning algorithm (Free model value-based) It seeks to find the best action to take given the current state.

Q-learning iteratively approximates the state-action value function, Q

Keep an estimate of $Q(s, a)$ in a table

More specifically, Q-learning seeks to learn a policy that maximizes the total reward.

What does the
'Q' represent?



The Quality

=> How useful a given action is for obtaining a future reward.

Q-TABLE

When Q-table is initialized, it contains zero values for all [state, action] pairs. We then start training the Q-table over multiple episodes, allowing it to learn the optimal policy according to the reinforcement learning algorithm.

Initialized

Q-Table		Actions					
		South (0)	North (1)	East (2)	West (3)	Pickup (4)	Dropoff (5)
States	0	0	0	0	0	0	0
	327

	499	0	0	0	0	0	0

Training

Initialized

Q-Table		Actions					
		South (0)	North (1)	East (2)	West (3)	Pickup (4)	Dropoff (5)
States	0	0	0	0	0	0	0
	328	-2.30108105	-1.97092096	-2.30357004	-2.20591839	-10.3607344	-8.5583017

	499	9.96984239	4.02706992	12.96022777	29	3.32877873	3.38230603

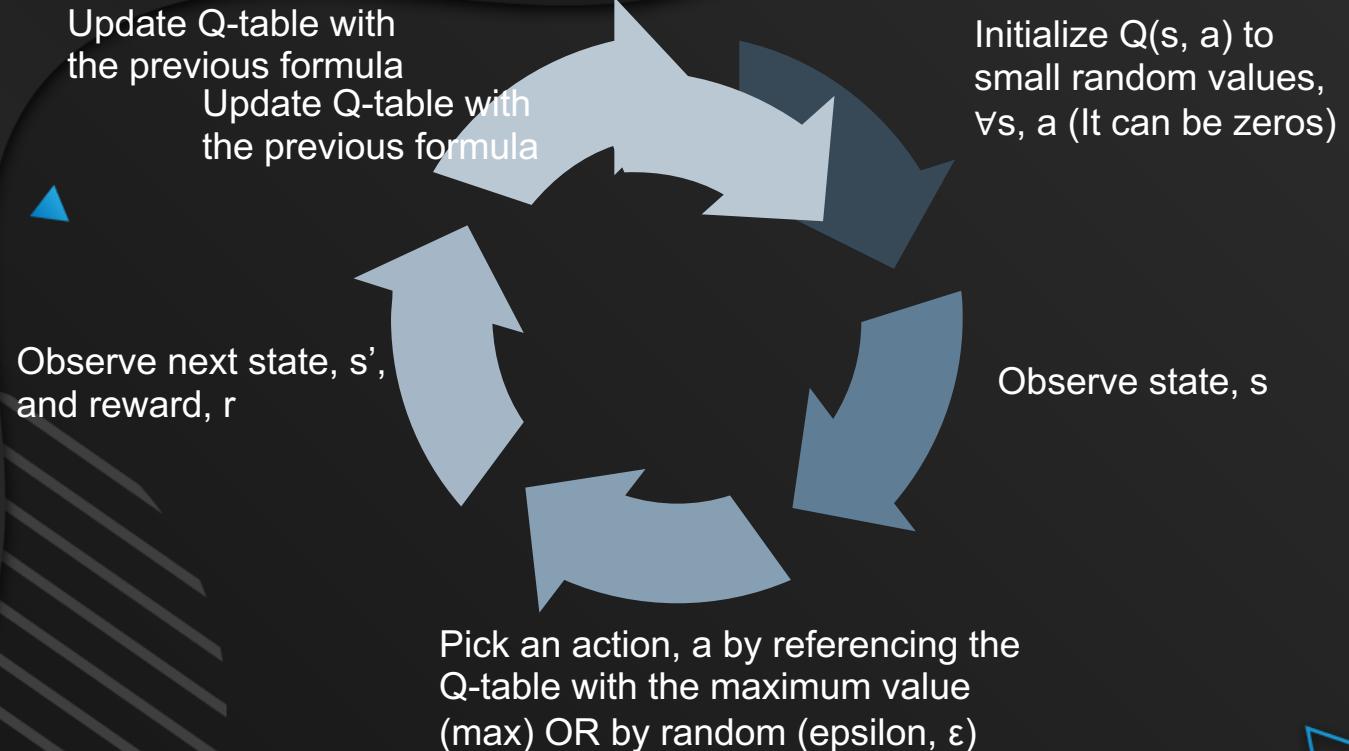
create what is called a Q-table. It is a matrix where rows represent states and columns represent actions. The entries in the table are initially set to zero. During training, the values are updated based on the rewards received during the episode. After training, the Q-table provides the optimal action for each state, which can be used to control the agent in the environment.

Update formula

$$Q(s, a) \leftarrow (1 - \alpha)Q(s, a) + \alpha(r + \gamma \max_{a'} Q(s', a'))$$

γ (gamma) is the discount factor ($0 \leq \gamma \leq 1$)
 α (alpha) represents the learning rate ($0 < \alpha < 1$)
As ~~is multiplying the future expected reward, it controls the importance of future rewards over past ones, and thus the extent to which Q-values are updated with less important future rewards are, and the Agent will tend to focus on actions which will yield immediate rewards only.~~

Q-Learning process





LET'S PRACTICE !

THANK YOU FOR ATTENDING

