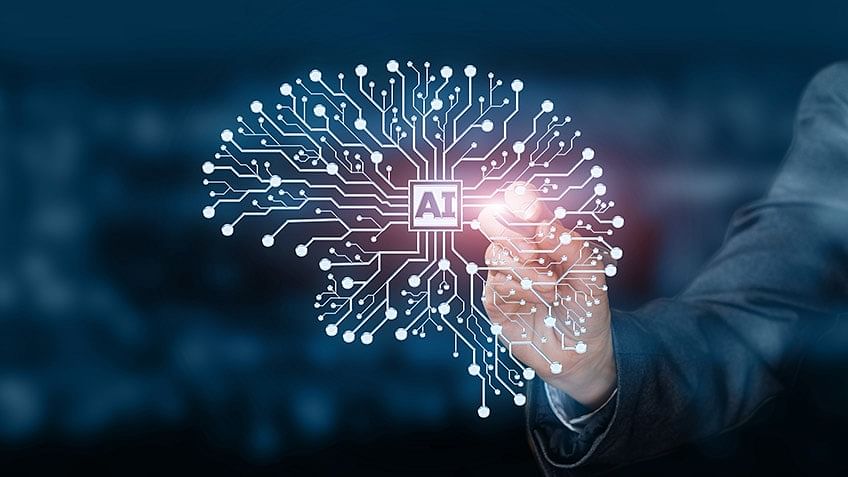
**Artificial Intelligence (AI)**



Artificial intelligence, or AI, is a field of study in computer science.

It was founded as an academic discipline in 1956.

Artificial intelligence has been growing exponentially since its first release, over the years many new techniques have been developed such as: q-learning, deep learning, etc...

Rosini Leonardo, 1/12/2023

**Fundamentals**

- Reinforcement learning

The "Reinforcement learning" is a type of system that has the  
objective of creating agents capable of solving certain problems once placed in a certain environment.

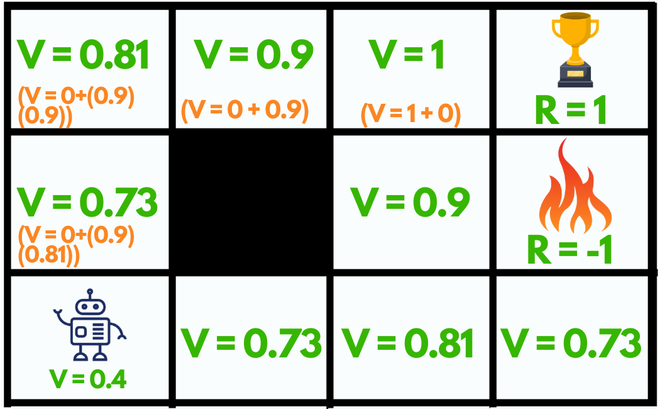
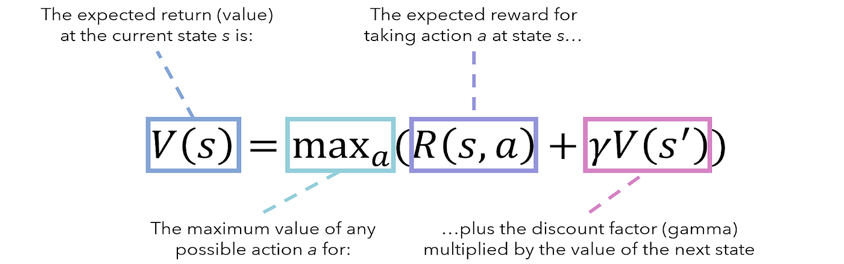
It is therefore distinguished from the other two models of "supervised learning" and "unsupervised learning".

Readings: [Simple Reinforcement Learning](https://medium.com/emergent-future/simple-reinforcement-learning-with-tensorflow-part-0-q-learning-with-tables-and-neural-networks-d195264329d0)

- The Bellman Equation

The bellman equation is simply the application of mathematical concepts on programming, as a result of [dynamic programming](https://en.wikipedia.org/wiki/Dynamic_programming).

It is a simple system of rewarding which the agent use to be oriented in the environment. There are several alterations in this formula, however they follow the same basic principle.



Concepts: s- state | a- action | R- reward | γ- discount

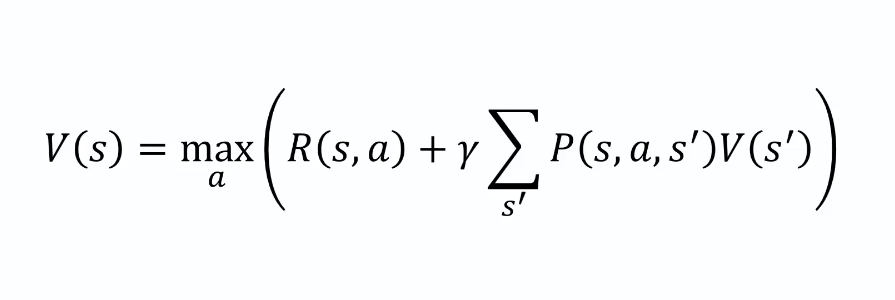
- Markov Decision Process (MDP)

This is basically a variant of the bellman equation.

The point of this one is to make a more realistic model that can be applied on a real-world basis.

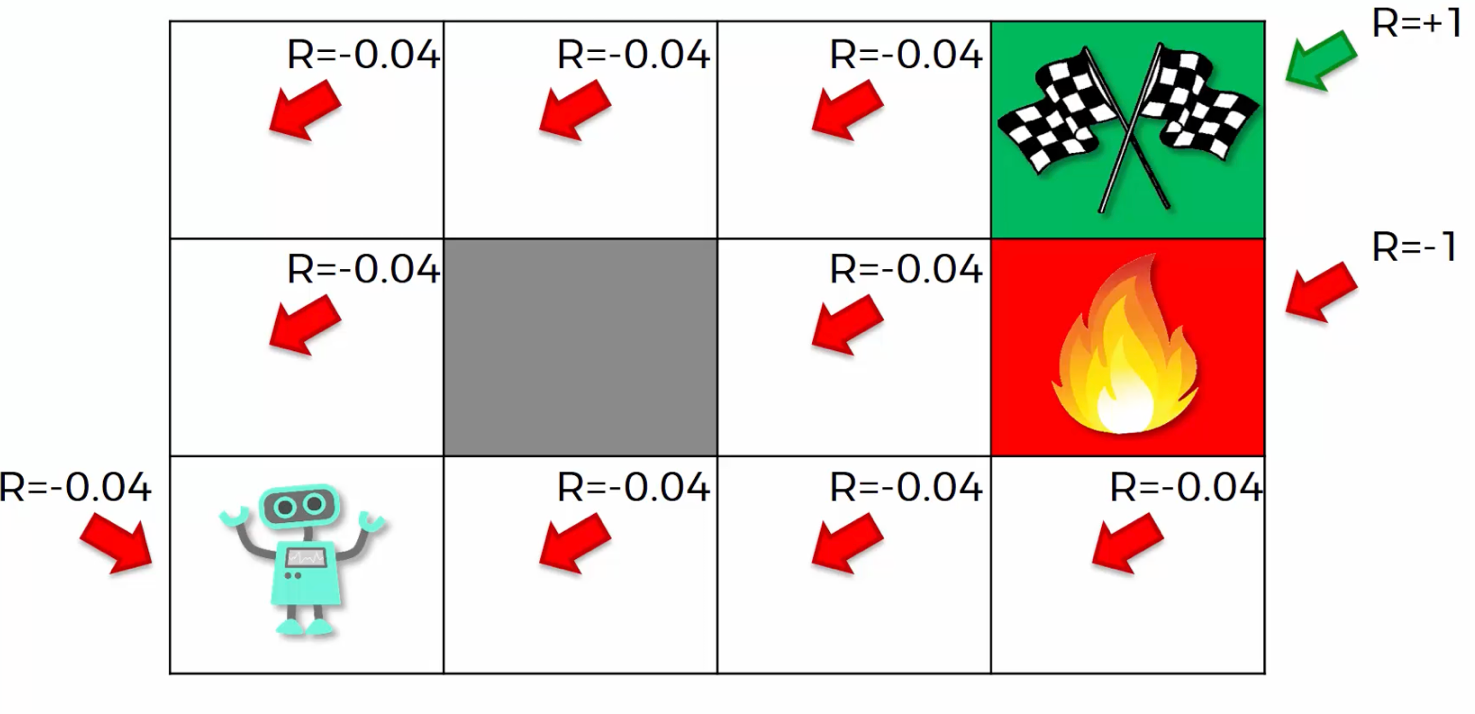
The basics behind this equation is to add some randomness because sometimes things do not work as expected.

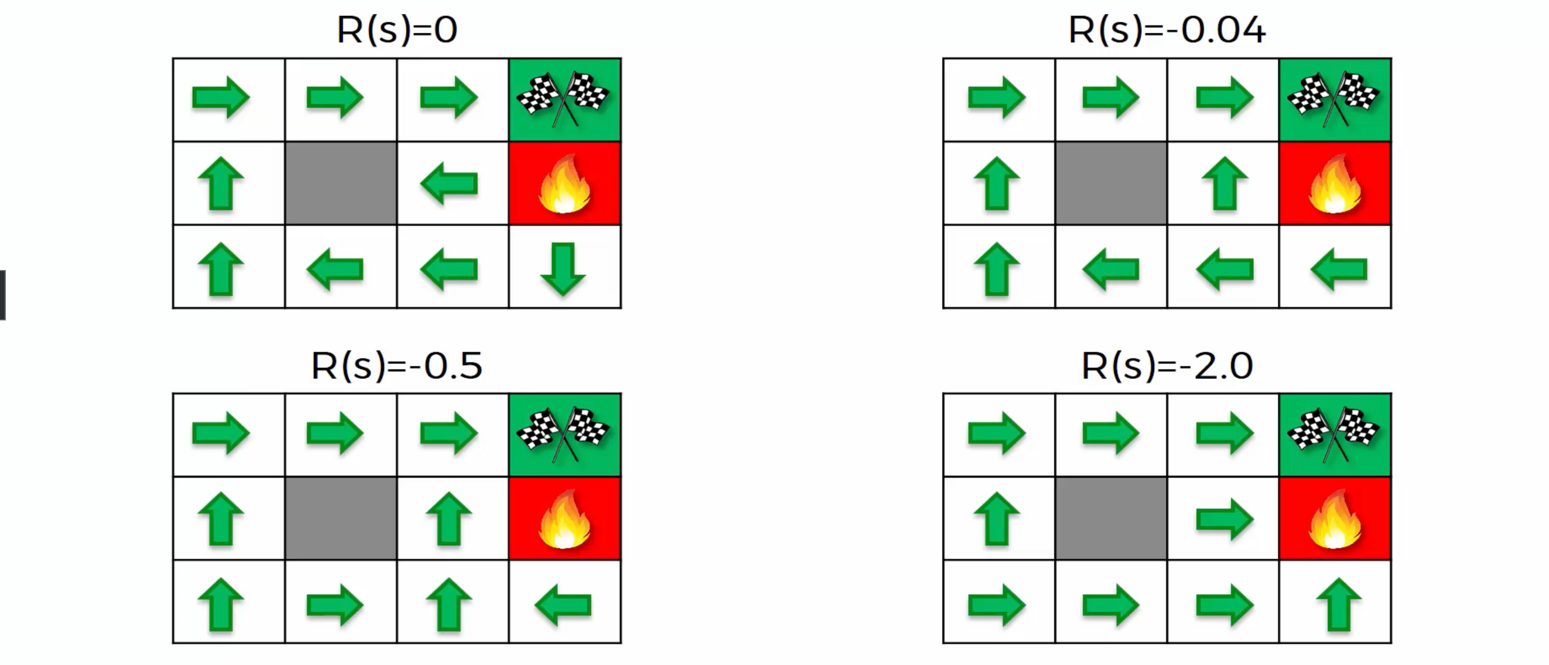
With the MDP we calculate the probability that the agent has to go on a certain spot, it does not matter the state before.



- Living penalty

The “living penalty” is a way which the agent is constantly losing points in every single spot of the environment, that until the game is end. Note that winning produce a positive reward (+1), losing is the max loss of points (-1) and single spots make you lose only a small quantity (like -0.04).  
This structure was built in a way that help the agent end his game faster.



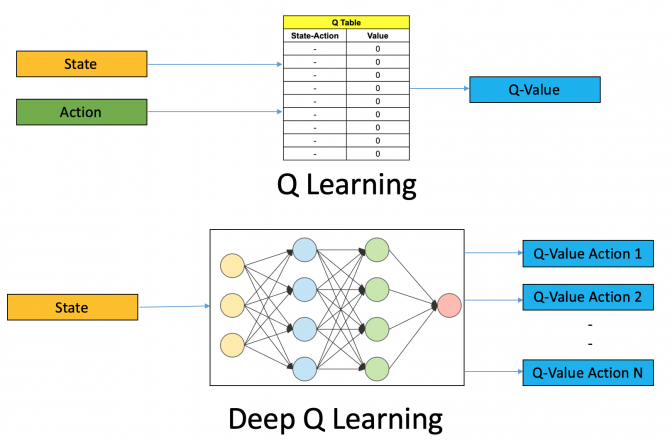


**Q-Learning intuition**

The Q-learning Intuition is still based on the Bellman equation, but there are some changes involved.

The Bellman equation is based on the Value for certain state of the agent *[V(S)]*, instead the Q-learning intuition is based on “Q” (some people speculate that means quality) which help us to get more quantifiable values based on actions and states (no more on the Value itself).

The formula for the Q-learning intuition is [Q(S0, An)].

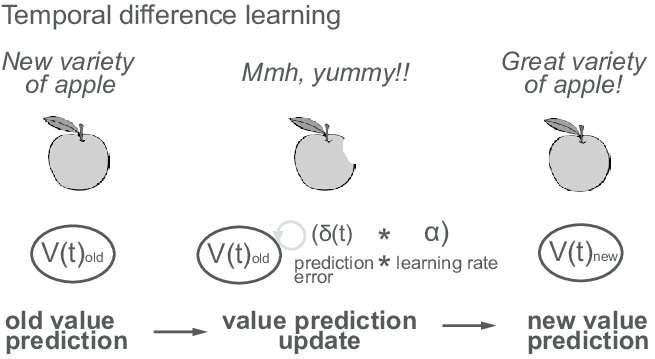


- Temporal Difference

The temporal difference learning (TD), it’s used to evaluate a

long-term pattern or behavior of the agent.

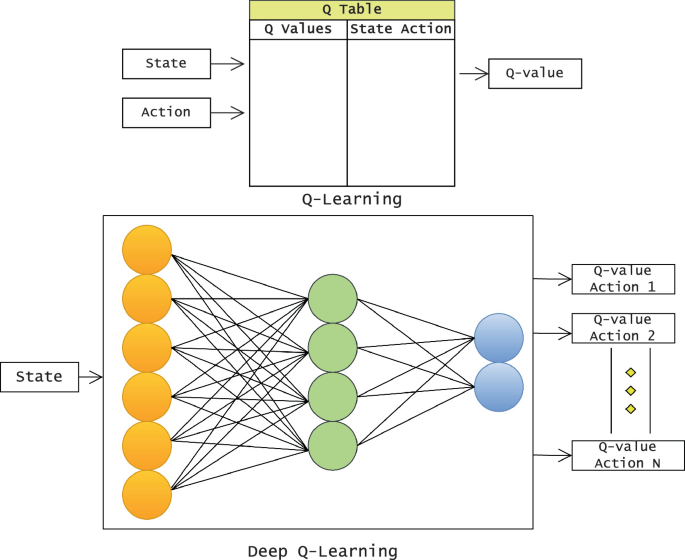
It is an unsupervised learning technique, mainly used to predict the total reward expected for the future.



Follow the [Article](https://link.springer.com/article/10.1007/BF00115009), to know more!

**Deep Q-Learning**

Deep Q-learning is one of the most used techniques in AI industries, that’s why it’s important to understand the basics of the previous one but not to focus only on that.



We use the “Loss formula” to value the loss in our environment, we want it to be as low as possible, and then the “[Softmax](https://developers.google.com/machine-learning/crash-course/multi-class-neural-networks/softmax?hl=it)” described in Annex 2 in order to value the best quality actions.

- Action selection policies

The action selection policies are used in order to filter the best actions made from the output layer.

The most common of these policies are:

*Softmax*, *ε-Greedy* and *ε-soft*.

Additional reading:   
<https://tokic.com/www/tokicm/publikationen/papers/AdaptiveEpsilonGreedyExploration.pdf>