Machine Learning

Session 7 - Reinforcement Learning

hadrien.salem@centralelille.fr

introduction-to-data-science

Introduction

What did we do last time?

Course outline

Machine learning course

Session 1: Regression

Session 2: Supervised classification

Session 3: Clustering

Session 4: Decision trees and ensemble methods

Session 5: Introduction to neural networks
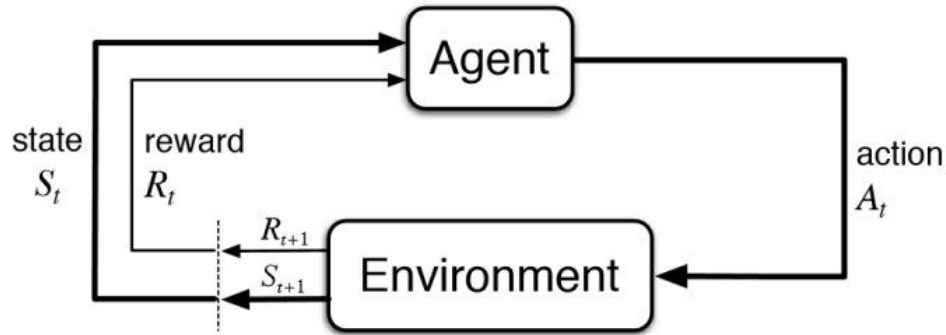
Session 6: Advanced neural networks

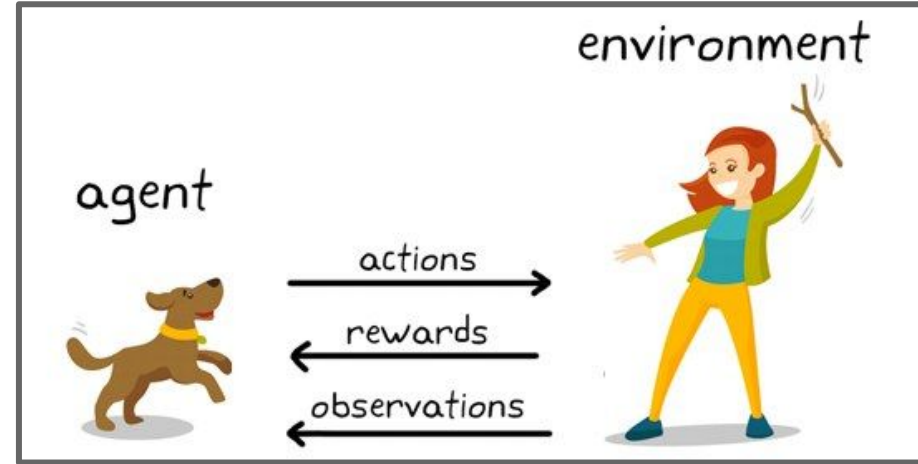**Session 7: Introduction to reinforcement learning**

Session 8: Reading science papers

Project

What is reinforcement learning?

Through trial and error, an agent keeps learning in an interactive environment (source)

Agents take actions in an environment aimed at maximizing their cumulative reward

Desired behaviors are rewards, undesired behaviors are punished. Through trial and error, the agent learns the best behavior.

Vocabulary

Agent - The decision maker and learner

Environment - A physical world where an agent learns and decides the actions to be performed

**Action –** Interaction of the agent with its environment

State – The current situation of the agent in the environment

Reward – Feedback (scalar value) given to the agent by the environment as a result of taking an action

Policy – A map of situations to actions, the strategy of the agent

Value Function – Function evaluating how good it is to be in a particular state or state-action pair

Applications of RL
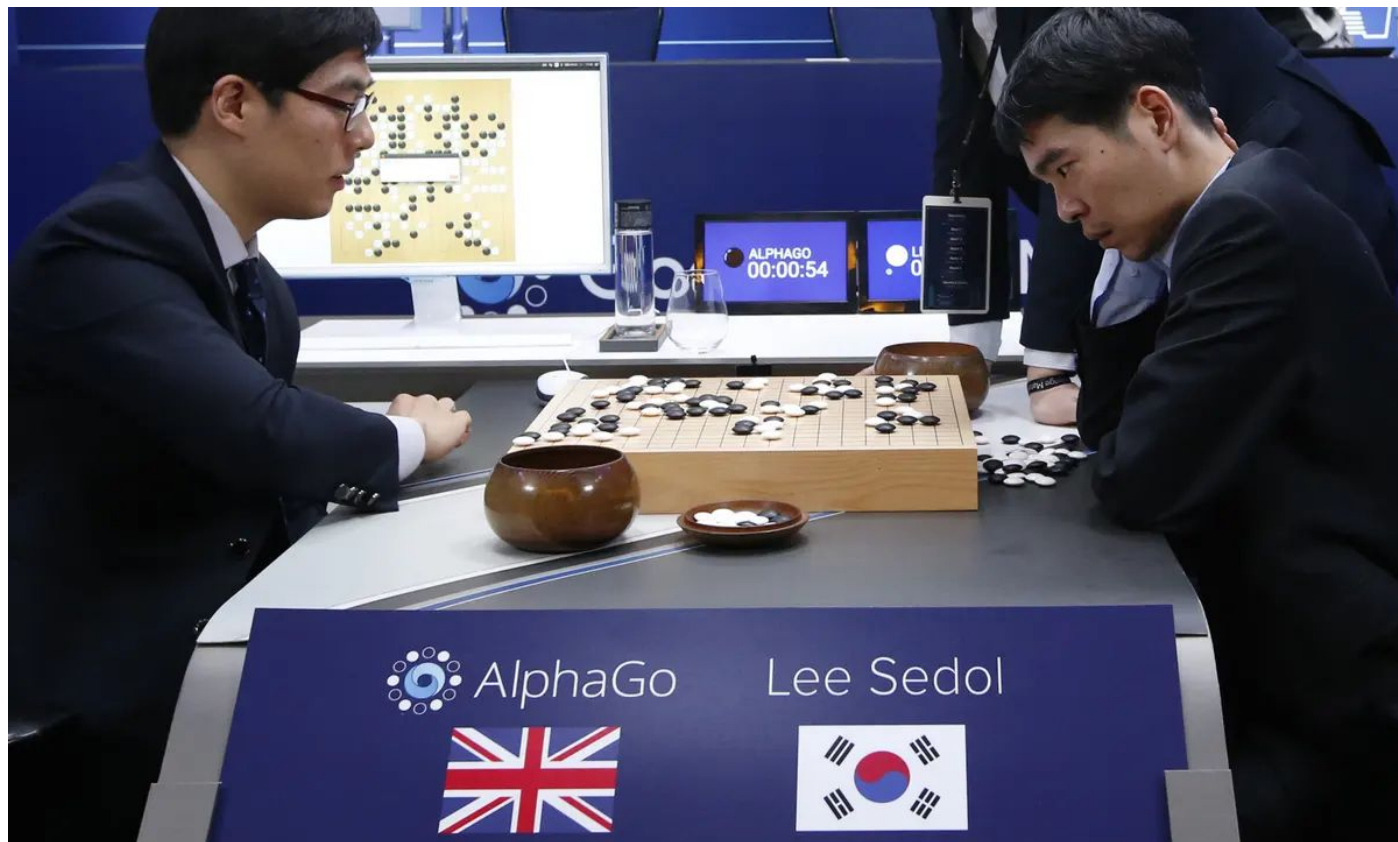
In 2016, AlphaGo beats professional Go player Lee Sedol

In 2019, AlphaStar attains GrandMaster in StarCraft 2

Applications of reinforcement learning

Game playing

Robotics - Robots learn independently how to perform certain tasks (walking, picking, placing)

Recommendation systems – Youtube videos, Spotify playlists, etc.

NLP – Chatbots learn from their past experiences

Education – Personalizing the students' learning experiences

Dynamic parameter optimization – Improve the performance of an algorithm as it is being used

How reinforcement learning works
The example of Q learning

Take action

⬇

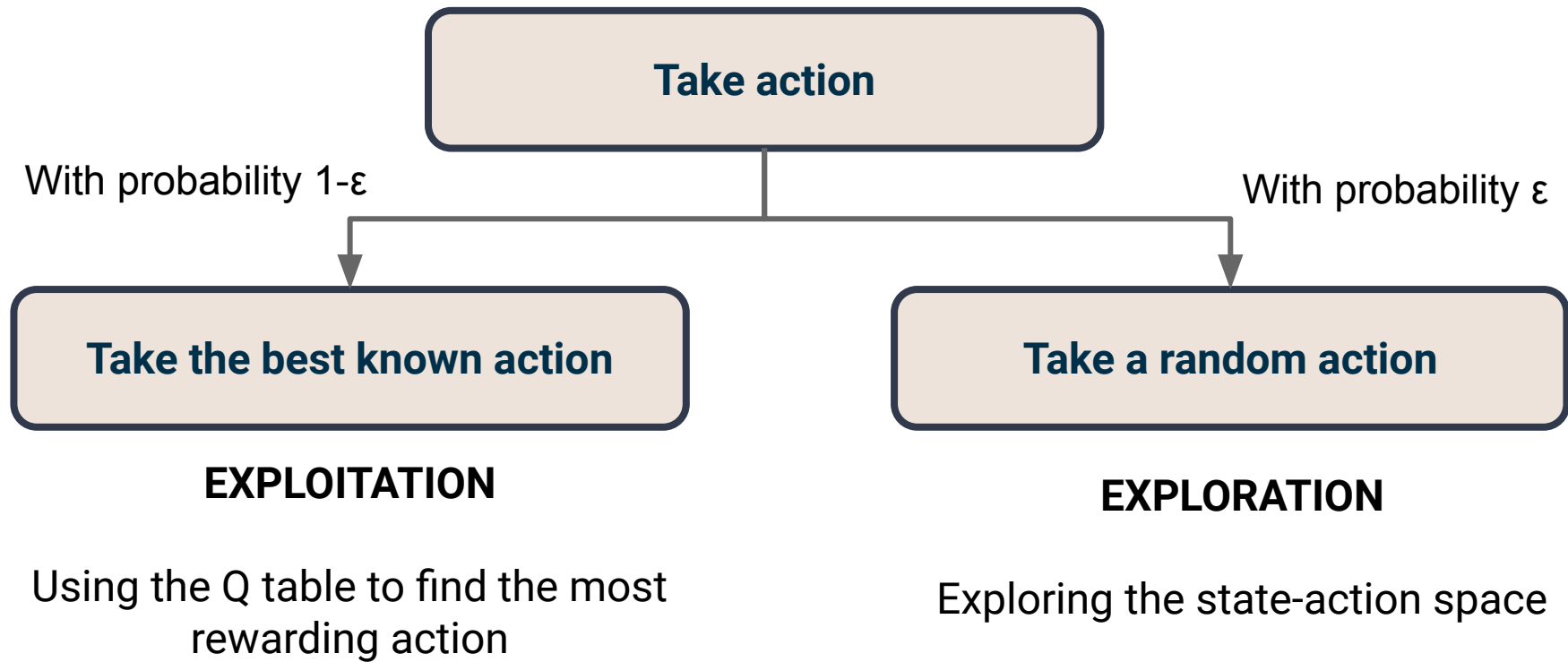## Compute reward

⬇

Update Q table

Game Board:

Current state (s): 0 0 0
0 1 0

Q Table: γ = 0.95

	0 0 0 1 0 0	0 0 0 0 1 0	0 0 0 0 0 1	1 0 0 0 0 0	0 1 0 0 0 0	0 0 1 0 0 0
⬆	0.2	0.3	1.0	-0.22	-0.3	0.0
⬇	-0.5	-0.4	-0.2	-0.04	-0.02	0.0
➡	0.21	0.4	-0.3	0.5	1.0	0.0
⬅	-0.6	-0.1	-0.1	-0.31	-0.01	0.0

With Q learning, agents base their decisions on a Q table

Take action

With probability 1-ε

With probability ε

Take the best known action

Take a random action

EXPLOITATION

EXPLORATION

Using the Q table to find the most rewarding action

Exploring the state-action space

Exploration vs. Exploitation – ε-greedy strategy

$$Q(S_t, A_t) \leftarrow Q(S_t, A_t) + \alpha[R_{t+1} + \gamma max_a Q(S_{t+1}, a) - Q(S_t, A_t)]$$

New Q-value estimation | **Former Q-value estimation** | **Learning Rate** | **Immediate Reward** | **Discounted Estimate optimal Q-value of next state** | **Former Q-value estimation**

TD Target

TD Error

- **TD Target (Temporal-Difference Target)**: estimate of the expected cumulative reward
- **TD Error (Temporal-Difference Error)**: quantifies how far off the current value estimate is from what is expected

The example of Q learning

Variants

Practical work

The notebook contains all the necessary instructions

Debrief

Debrief – G3

https://forms.gle/TiR9uaYTFF3iTX9e6

Debrief

What did we learn today?

What could we have done better?

What are we doing next time?