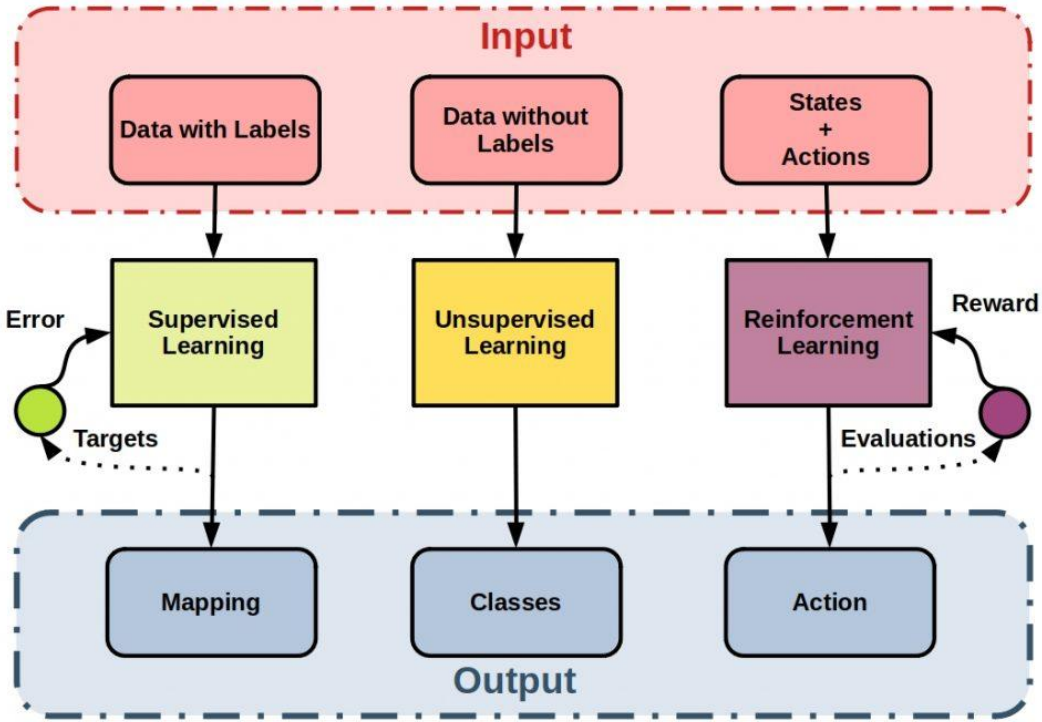
Reinforcement Learning

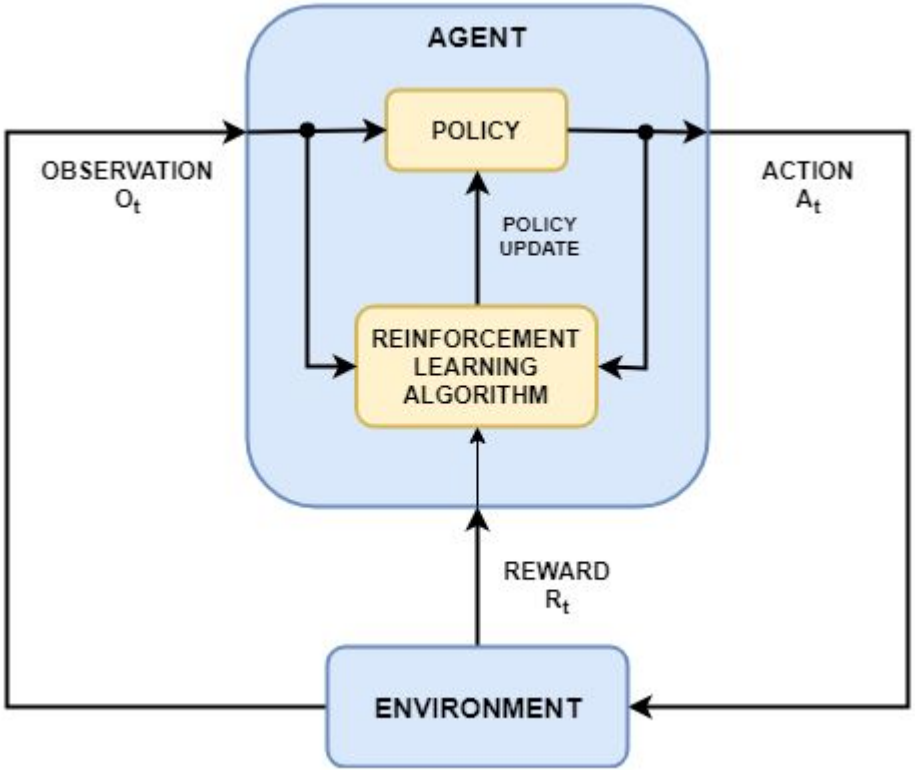
Cihan University – Erbil
Department of Computer Science
Assist. Lect. Soma Solaiman Zadeh

Machine Learning Paradigms



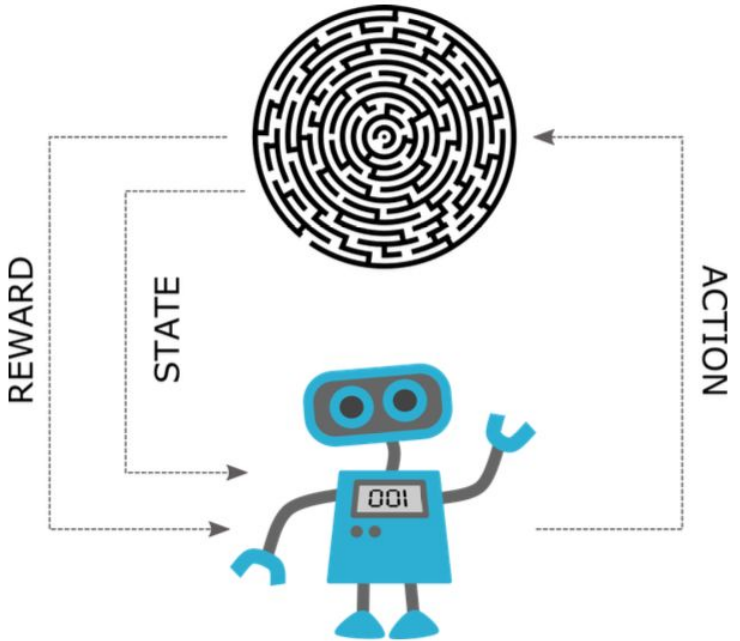
> In psychology, **reinforcement** is a term that refers to “anything that increases the likelihood that a response will occur” *(Cherry)*

Elements of Reinforcement Learning



Reinforcement Learning in a Simple Word

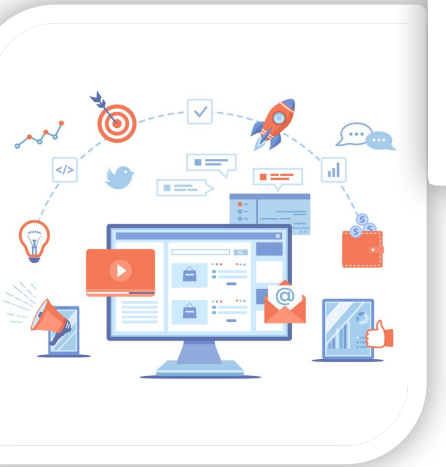
Reinforcement Learning is a subfield of machine learning that teaches an **agent** how to choose an **action** from its action space, within a particular **environment**, in order to **maximize rewards** over time.

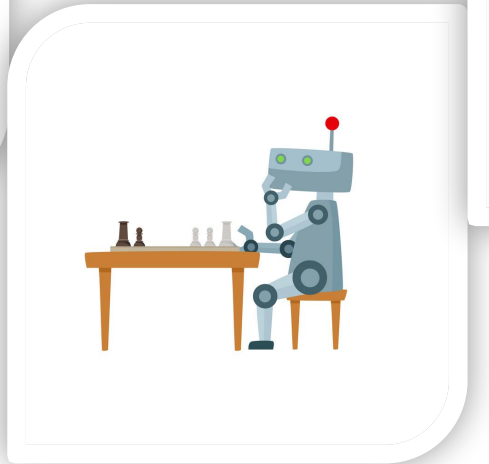


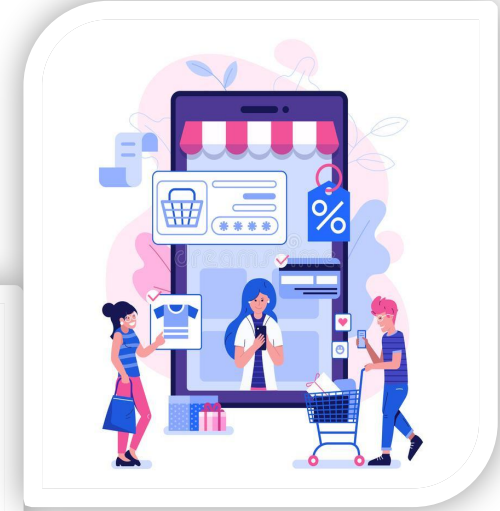
Reinforcement Learning Algorithms

- Markov Decision Process (MDP)
- Q-Learning
- Deep Reinforcement Learning

Reinforcement Learning Applications







RL in Self-driving Cars

- Speed limits at different streets, drivable zones, avoiding collisions, parking.
- **AWS DeepRacer** is an autonomous racing car that has been designed to test out **RL** in a physical track.

RL in Trading and Finance

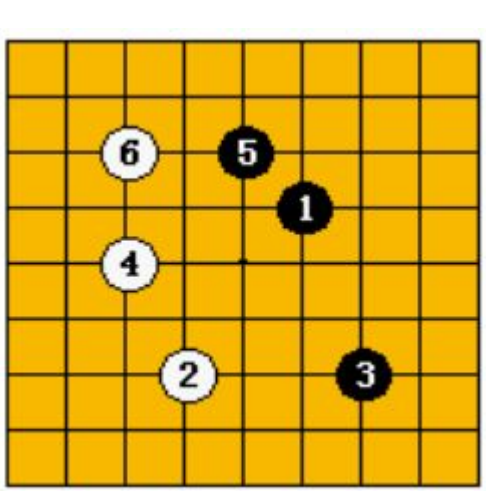
- An RL agent can determine the action to take at a particular stock price; hold, buy, or sell.
- The RL model is evaluated using market benchmark standards in order to ensure that it's performing optimally.

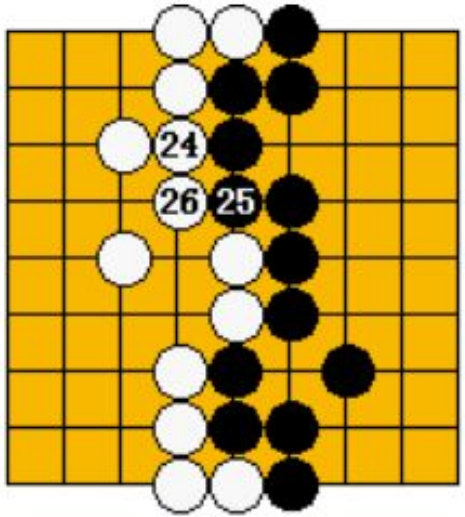
RL in Computer Games

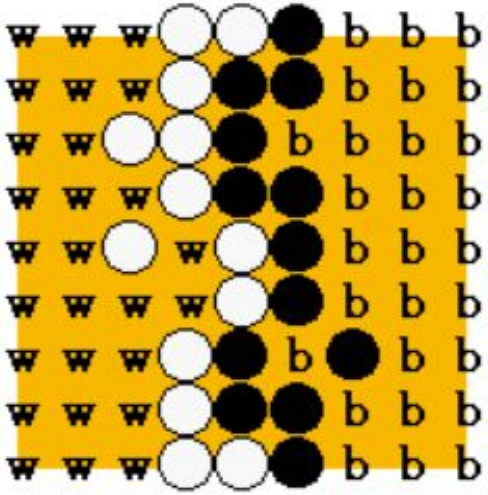
- Games like Atari, Chess, GO and sudoku have become testbeds of testing deep reinforcement learning algorithms.

AlphaGo Zero

Ancient Chinese Game of Go







Black □ 28 Points

White □ 27 Points

Winner: Black

DeepMind

AlphaGo

- 30 million moves of expert human play
- **Several months** for training

48 TPUs

AlphaGo Zero

- Never saw humans play
- 5 million games of self-play
- **3 days** for training and less computing power

4 TPUs

AlphaGo vs AlphaGo Zero

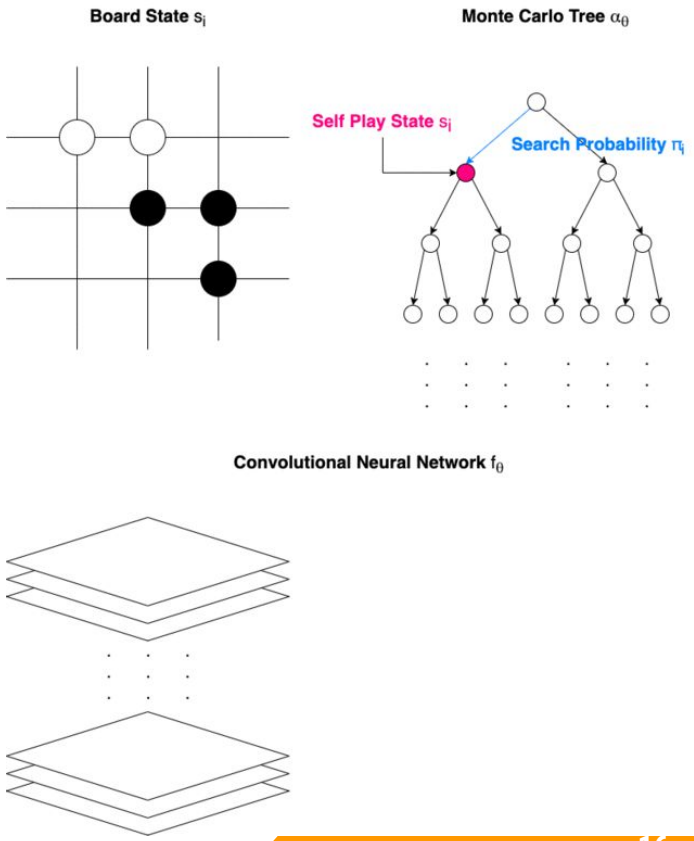
- **Original AlphaGo** defeated world champion **Lee Sedol** by 18 games to 0.
- **Original AlphaGo** defeated the world's leading Go player **Ke Jie** by 3 games to 0.
- **AlphaGo Zero** defeated the **original AlphaGo** in a 100-game match by 100 to 0.

AlphaGo and AlphaGo Zero

- Both original **AlphaGo** and **AlphaGo Zero** use deep reinforcement learning.
- Two basic things the program needs to learn:
  - **Policy:** the probability of making each of the possible moves in a given position.
  - **Value:** the probability of winning from any given position.

AlphaGo Zero Architecture

- It combines a **neural network** and **Monte Carlo Tree Search** in a policy iteration framework to achieve stable learning.



Conclusion

- Reinforcement Learning needs large datasets to make better decisions.
- Constant changes in the environment makes decision making difficult.
- Design of the reward and punishment structure of the model is the key for the success.

THANKS!

Any questions?