Introduction

Reinforcement
Learning

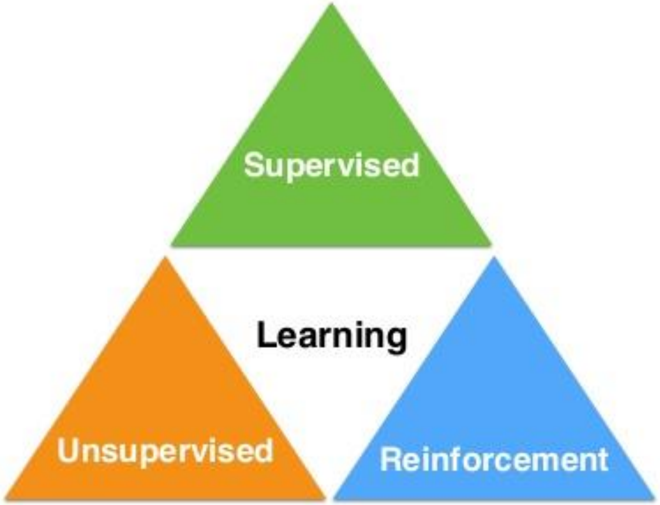
What is RL?

(Machine-based) **learning** how **agents** map **situations to actions** in an **environment** so as to maximize a numerical **reward signal.** (Sutton & Barto)

Machine Learning

Field of study that gives computers the ability to learn **without being explicitly programmed** (A. Samuel, 1959)

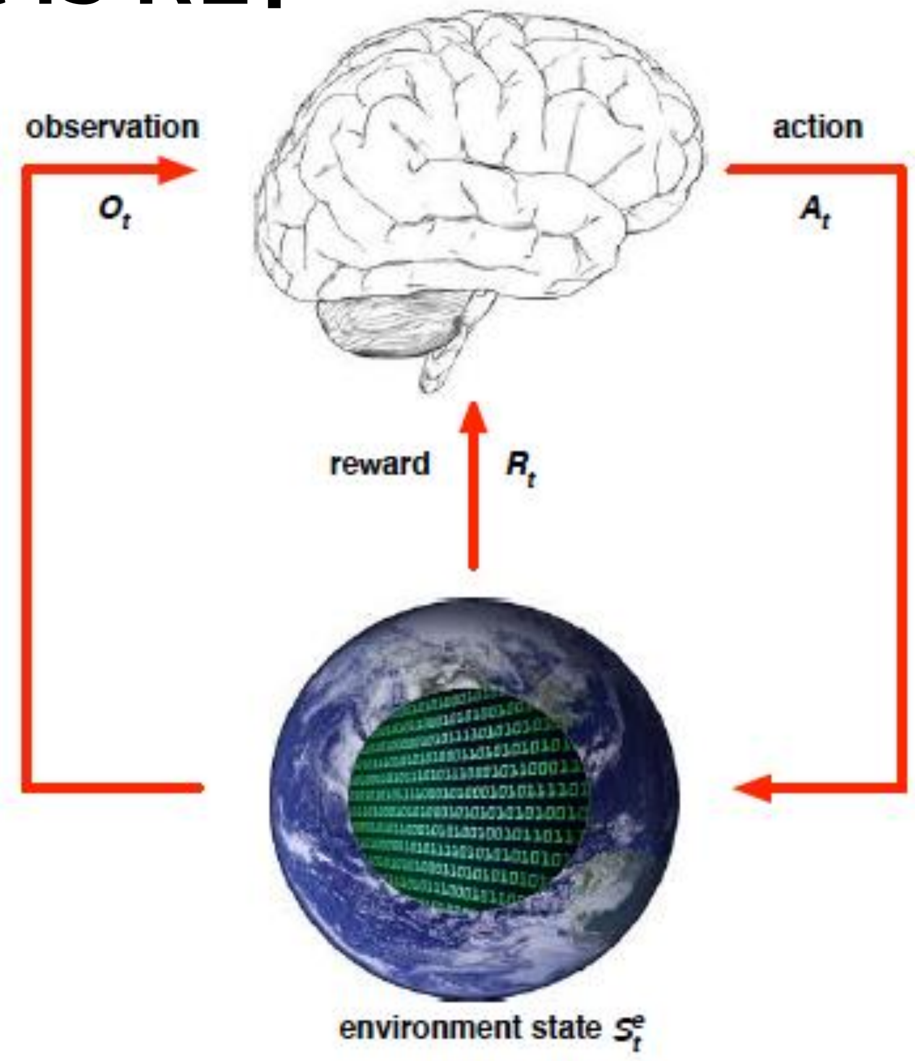
- Labeled data
- Direct feedback
- Predict outcome/future

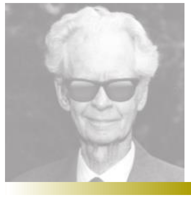


- No labels
- No feedback
- “Find hidden structure”

- Decision process
- Reward system
- Learn series of actions

What is RL?





RL Problems : Sequential Decision-making Problems

- Go player plans (anticipating possible replies & counter-replies)
- A gazelle struggles to its feet minutes after being born. Half an hour later it is running at 20 miles per hour.
- Robot vacuum cleaner needs to visit all the floor area.
- Multi-armed bandit problem.
- Grid World problem

All involve **interaction between an active decision-making agent and its environment**, within which the **agent seeks to achieve a goal despite uncertainty** about its environment

RL Problems : Sequential Decision-making Problems

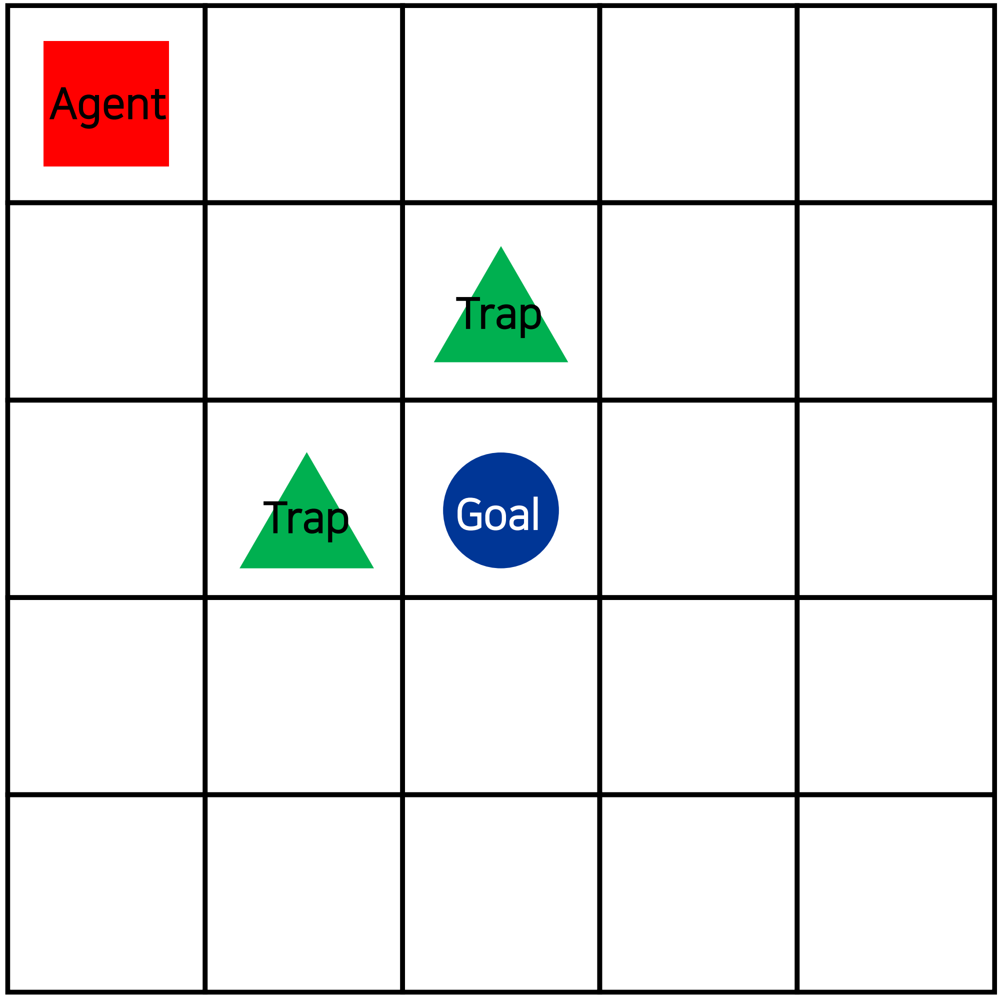
✓ **State** (e.g. Go position, robot's location & charge level of battery)

✓ **Action** (e.g. up/down movement, next Go position)

✓ **Reward** : the goal in RL problems (on time step basis)

✓ **Policy** : the learning agent's way of behaving at a given time.
Agent can maximize **reward** following the **optimal policy**.

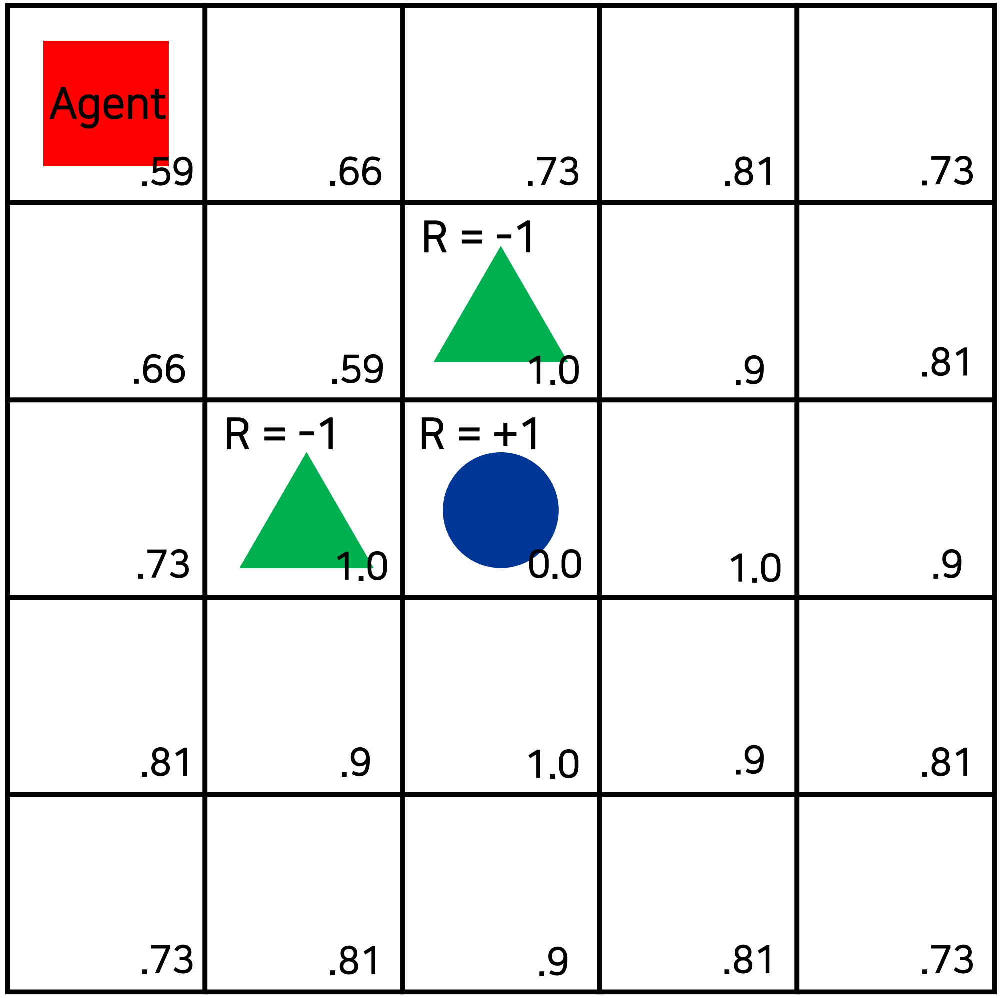
RL Example : when there are a handful of states



The 'most efficient' path?

✓ Several approaches

RL Example : when there are a handful of states



Agent .59	.66	.73	.81	.73
.66	.59	R = -1 1.0	.9	.81
.73	R = -1 1.0	R = +1 0.0	1.0	.9
.81	.9	1.0	.9	.81
.73	.81	.9	.81	.73

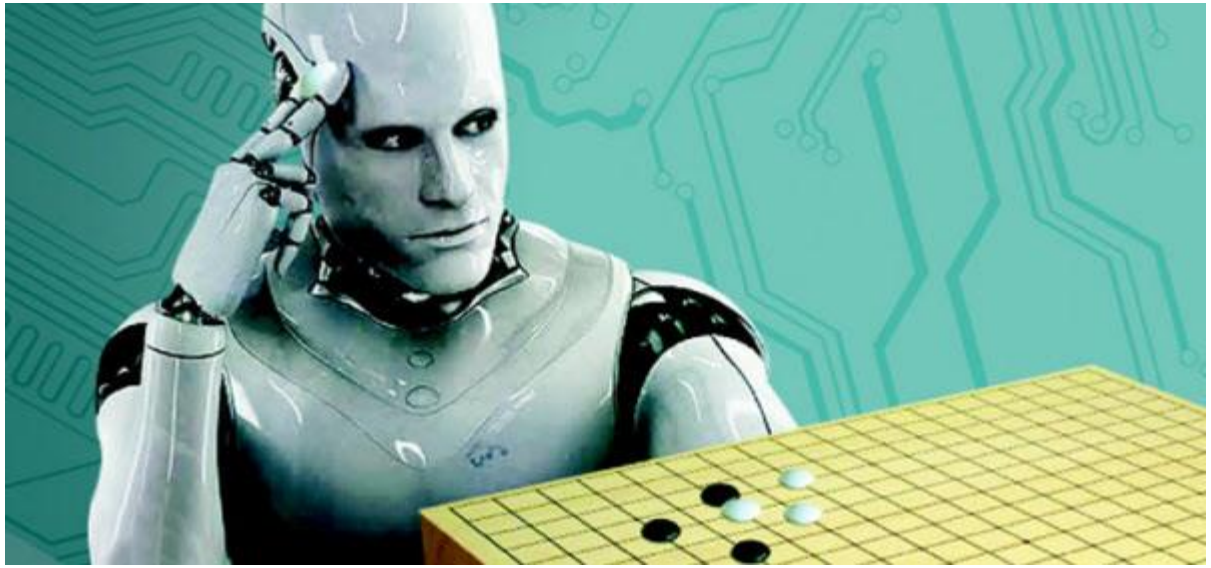
✓ (Action-) Value Function

= The " Guide Map "
from experience

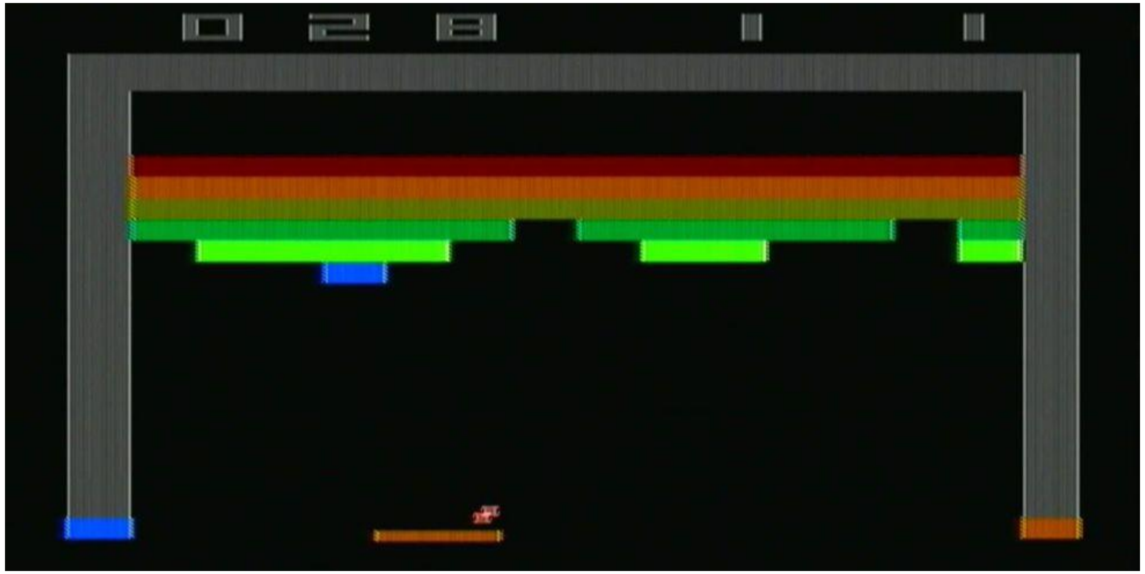
✓ Decision Making

= Reward + Value

RL Example : when there are a colossal number of states



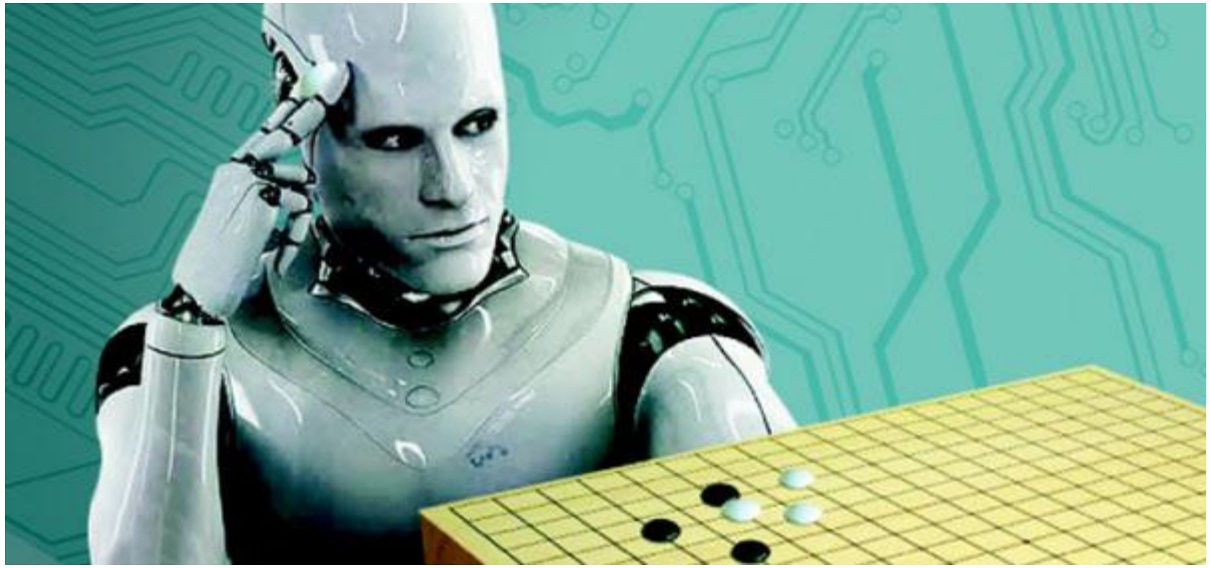
< 3^(19*19) ~ 2*10^172 number of states



Type "atari breakout" on Google Image Search

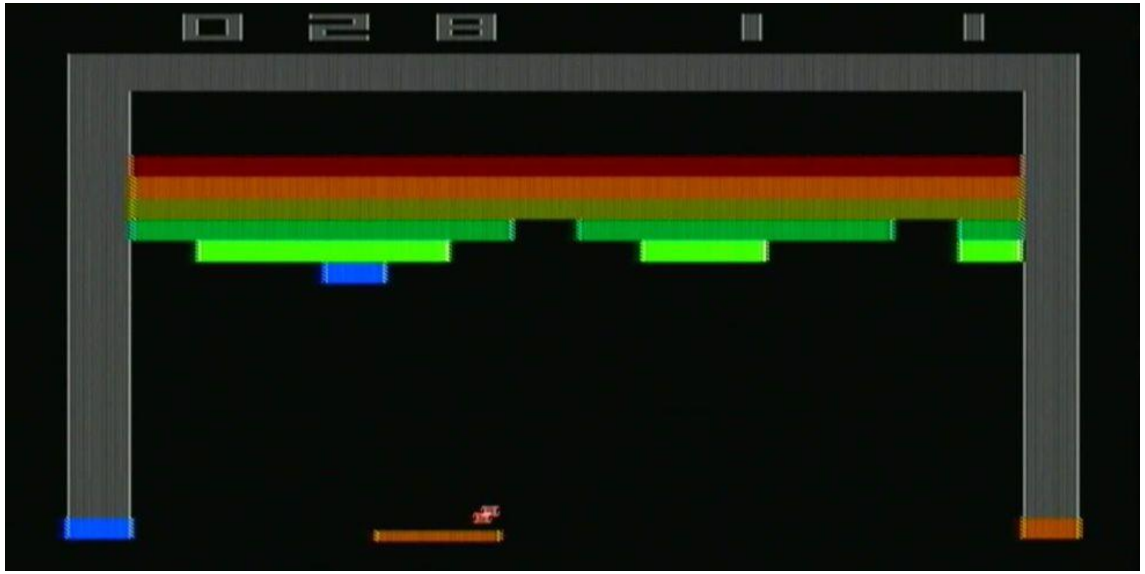
https://www.youtube.com/watch?v=V1eYniJ0Rnk

RL Example : when there are a colossal number of states

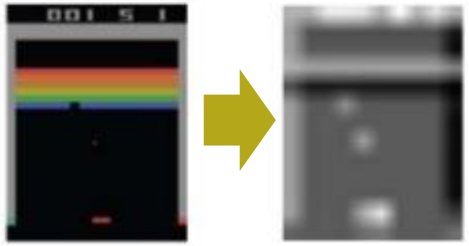


- ✓ Monte-Carlo Tree Search
- ✓ RL Policy Network
- ✓ Value Network

RL Example : when there are a colossal number of states



✓ Pre-processing (using **CNN**)



✓ Q-value prediction from **Deep Q-Network**

Q(s, a)

✓ Optimize **Deep Q-Network** using experience

Q(s, a) → Q*(s, a)