REINFORCEMENT LEARNING

"How an intelligent agent can make a good sequence of decisions where goodness is given by REWARDS"

RL is about how to map situations to action to maximize the rewards.

2 main characteristics

1. TRIAL AND ERROR SEARCH
2. DELAYED REWARDS

We have a LEARNING AGENT that must be able to:

- Sense the state of env.
- Take action to affect the state

Differences with ML: NO supervisor, NO training, feed delayed, time matters

RL is about LEARNING ONLINE, i.e. learning while interacting with a changing world.

REWARDS

A reward R_t is a scalar feed that indicates how well the learning agent is doing.

GOAL: Select actions to maximize the total future Reward

CHALLENGE: Exploration - Exploitation trade-off:
- to make better future decisions
- use past actions that obtain high rewards.

The agent must try a variety of actions and progressively favor those that appear to be best

IDEA

∀ step t the agent:
- execute action A_t
- receive obs O_t
- receive R_t

$t = t+1$

∀ step t the environment:
- Receive A_t
- Give O_t
- Give R_t

There is 1 for ENV S_e^t and 1 for AGENT S_a^t.

History $H_t = A_1, O_1, R_1 \dots A_t, O_t, R_t$

State $S_t = f(H_t)$ → info used to determine what happen next

An RL agent may include 1 or more:

- POLICY → defines the agent behaviour at a given time
  - → Deterministic / Stochastic
- VALUE FUNCTION → what is good in the long run.

$$V(s) = E[R_{t+1} + \gamma R_{t+2} + \dots \mid State = s]$$

- MODEL → what the environment will do next