

RL: Mapping situation to action, so as to maximize a reward.

- Is different than supervised and unsupervised learning.
- Is a third paradigm.

Learning Agent must:

- Have a goal
- sense the state it's in
- Be able to take action.
- explore and exploit;
 - exploit it has already experienced in order to obtain reward.
 - explore in order to make better action selections in the future.

Notation: at time step t :

- State $S_t \in S$
- Action $A_t \in A$ (action space)
- Reward $R_t \in R$

} formal framework of
Markov Decision processes

4 main subelements of an RL system.

- Policy
- Reward signal
- value function
- optional: model of the environment.

Policy: defines the agent's way of behaving at any given time.

Mapping perceived states of the environment to actions to be taken when in those states.

Reward signal: defines the goal of a RL problem.

At every time step, the RL agent receives a number called the reward.

An agent's goal is to maximize the total reward over the long term.

value function: what's good over the long run.
values must be estimated and re-estimated.
most important.

model of the environment:

something that mimics the behavior of the environment.
model free vs model-based methods.