A Quick Review on RL and MDP

Farnaz Adib Yaghmaie

Linkoping University, *Sweden*
farnaz.adib.yaghmaie@liu.se

April 6, 2021

Machine Learning

- Supervised Learning
- Unsupervised Learning
- **Reinforcement Learning**

Finding suitable actions to take in a given situation in order to maximize a reward [1]**.**

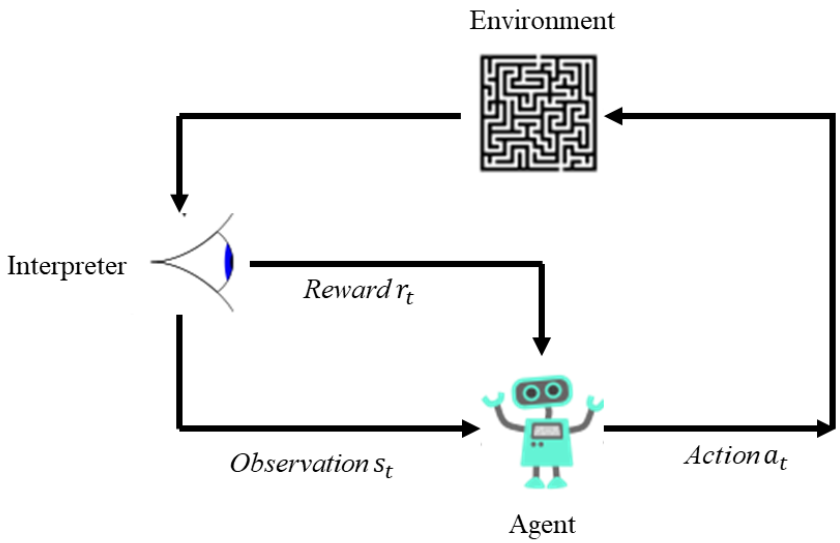
[1] Richard S Sutton & Andrew G Barto. *Reinforcement learning: An introduction*, volume 1. MIT press Cambridge, 1998.

How RL is different from other branches of ML?

- No supervisor; only a reward
- The action will effect subsequent data
- Dynamic data *vs.* Static data

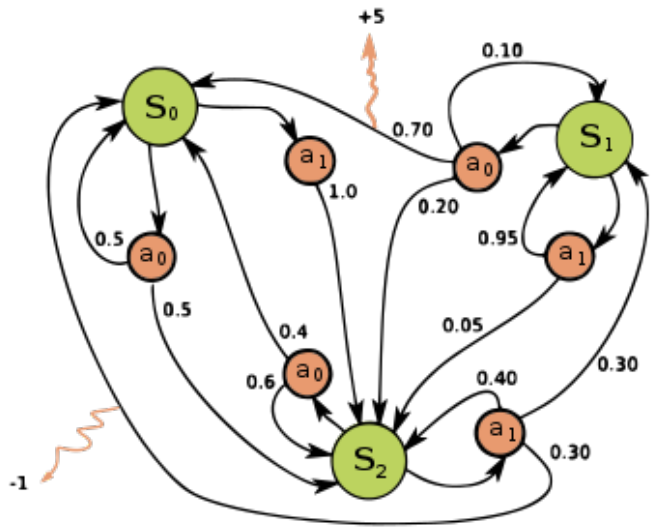
An RL framework

- Reward
- Environment
- Agent



A Markov Decision Process (MDP) is a tuple $< \mathcal{S}, \mathcal{A}, \mathcal{P}, \mathcal{R}, \gamma >$

- $\mathcal{S}$: The set of states.
- $\mathcal{A}$: The set of actions.
- $\mathcal{P}$: The set of transition probability.
- $\mathcal{R}$: The set of immediate rewards associated with the state-action pairs.
- $0 \leq \gamma \leq 1$: Discount factor.



Modified version of @ https://en.wikipedia.org/wiki/Markov_decision_process

- **States:** Describe internal status of MDP
- **Action:** Possible choices in each state of MDP
- **Transition probability:** The dynamics

$$[\mathcal{P}^a]_{ss'} = p[s_{t+1} = s' | s_t = s,\ a_t = a]$$

- **Reward:** $r_t = r(s,\ a)$

Total reward:	$R(T) = \sum_{t=1}^{T} \gamma^t r_t$
Average reward:	$R(T) = \lim_{T \to \infty} \frac{1}{T} \sum_{t=1}^{T} r_t$

- **Discount factor:** How to handle the future

$\gamma \to 0$:	Only care about the current reward not future rewards
$\gamma \to 1$:	Care about all rewards equally

RL goal

Generate actions to maximize the future rewards

- Policy: The agent's decision
  - Deterministic policy $a = \pi(s)$
  - stochastic policy $\pi(a|s) = P[a_t = a|s_t = s]$
- Value function: how good the agent does in a state

$$V(s) = \mathbf{E}\left[r_t + \gamma r_{t+1} + \gamma^2 r_{t+2} + ...|s_t = s\right]$$

- Model: The agent's interpretation of the environment

Not all components are necessary!

Email your questions to

farnaz.adib.yaghmaie@liu.se