Reinforcement Learning Fundamentals

Lecture 3: RL Framework

Dr Sandeep Manjanna
Assistant Professor, Plaksha University
sandeep.manjanna@plaksha.edu.in

Some material in this lecture is taken from

1. Prof. Ravindran's course: "Reinforcement Learning."
2. Dr Silver's course: "Reinforcement Learning."

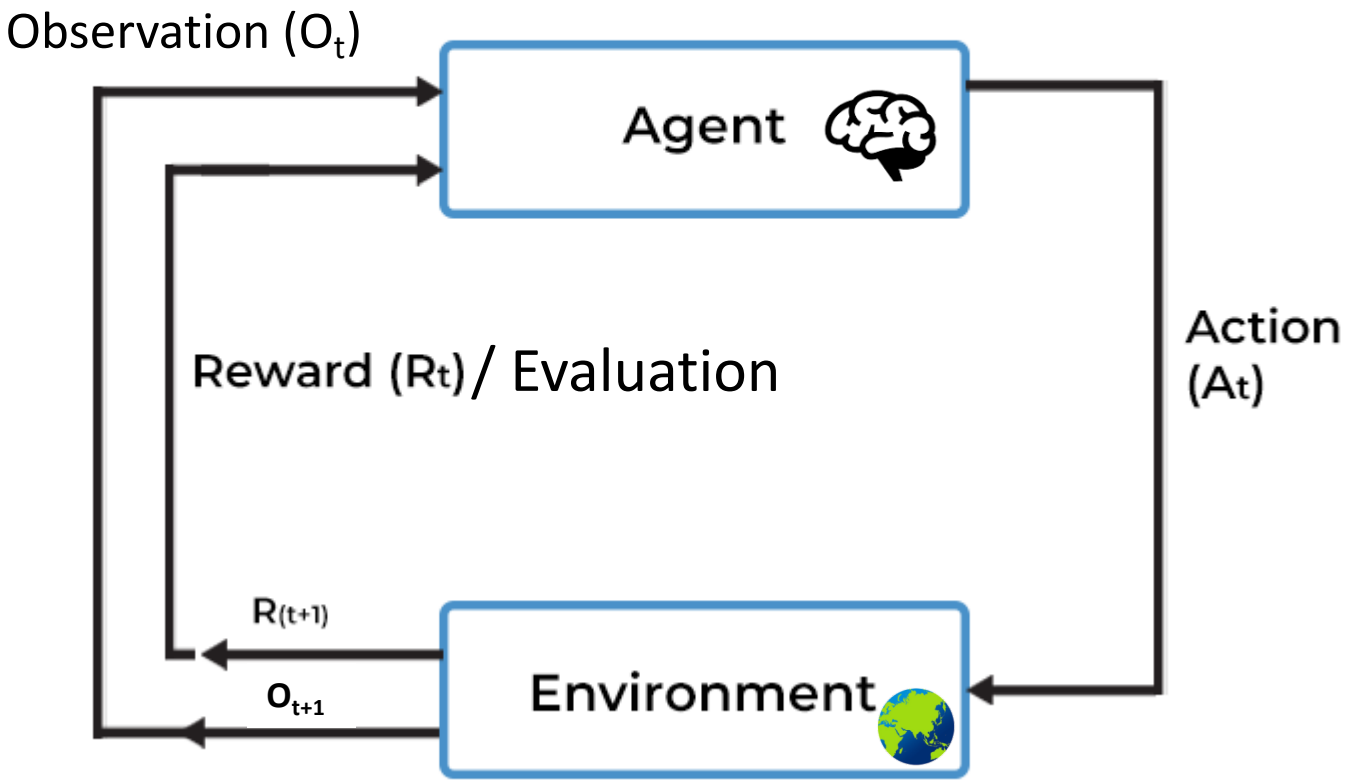
In today's class…

- RL Framework
- What are Rewards?
- What is a State?
- Special cases: Fully and Partially Observable environments
- Temporal Difference

RL Framework

Sequential Decision Making

- Goal: **Select actions or a sequence of actions to maximize the total future reward.**



At each step *t* the agent:

- Executes action A_t
- Receives observation O_t
- Receives scalar reward R_t

The environment:

- Receives action A_t
- Emits observation O_{t+1}
- Emits scalar reward R_{t+1}

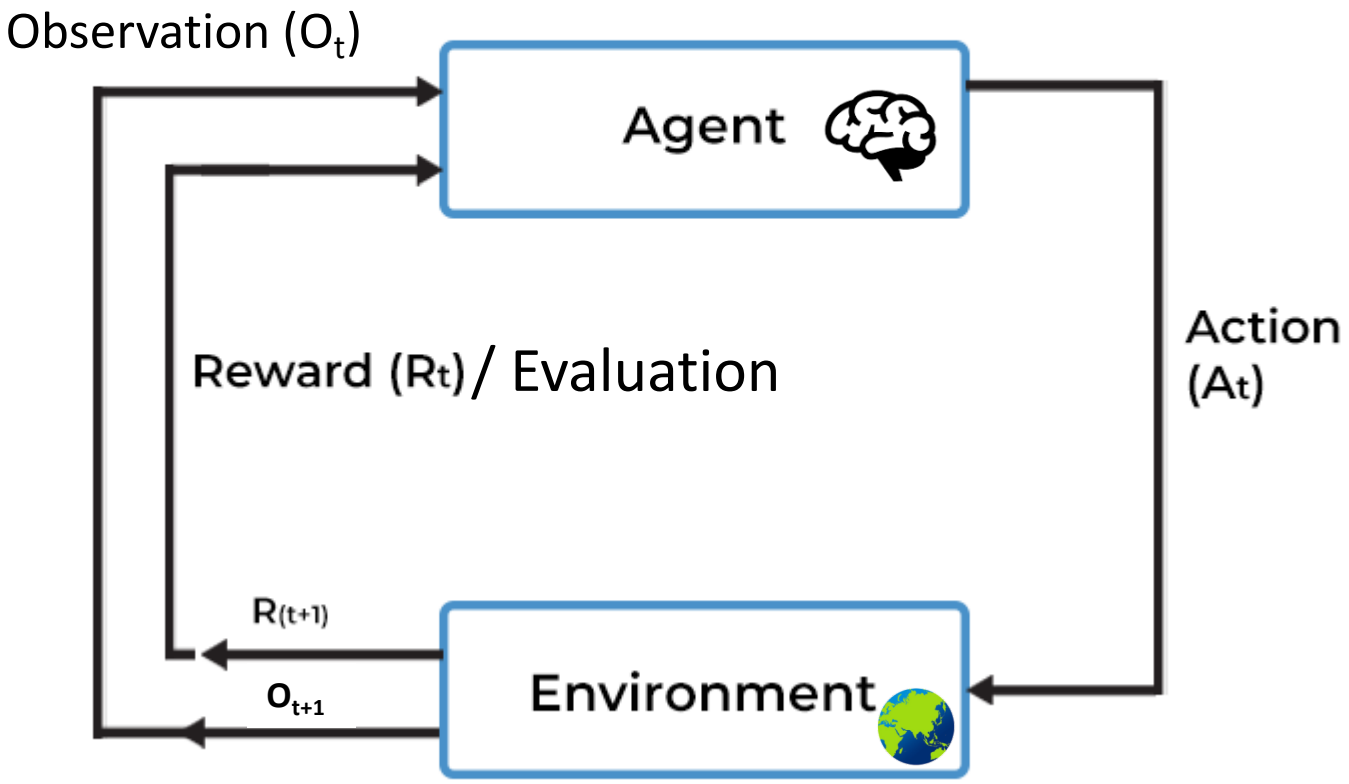
Really?

t increments at env. step

What is evaluation in supervised learning? How is it different in RL?

RL Framework

Sequential Decision Making



- Agent learns by interacting with the environment.
- Environment returns some rewards.

Really?

- Environment gives noisy delayed scalar evaluation.
- Environment also provides some observations.
- Environment can be stochastic.
- Goal is to maximize a measure of long-term performance. In this case it can be the reward.

What is State?

At time ***t=1***, the environment / the world returned some observation O_1 and reward R_1 and the agent took an action A_1. Then the world returned O_2 and R_2. Now the agent took the action A_2.

What is history at ***time = t*** ?

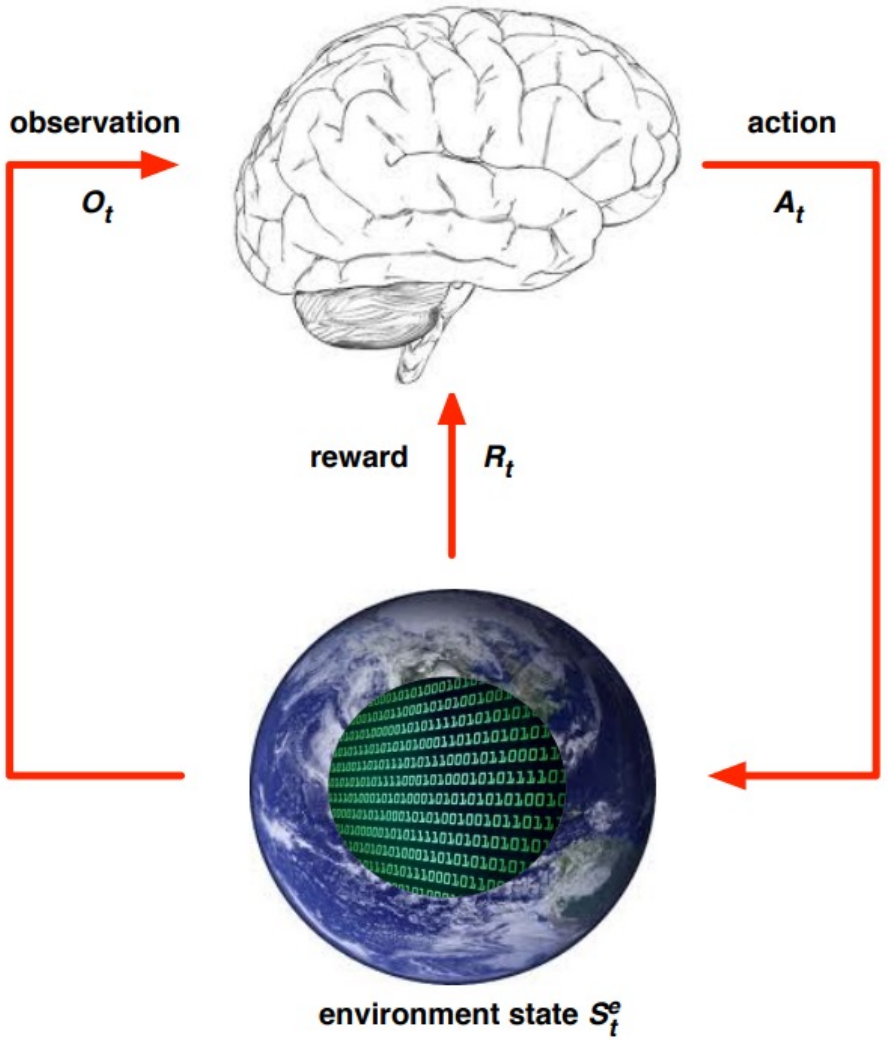
- The sequence of observations, actions, rewards from time = ***1*** to time = ***t*** is referred to as history.
- i.e., History is all the observable variables up to time ***t***.
- What happens next depends on the history.

State is the information used to determine what happens next. This is the **summary of the history**.

Formally, state if a function of history:

$$S_t = f(H_t)$$

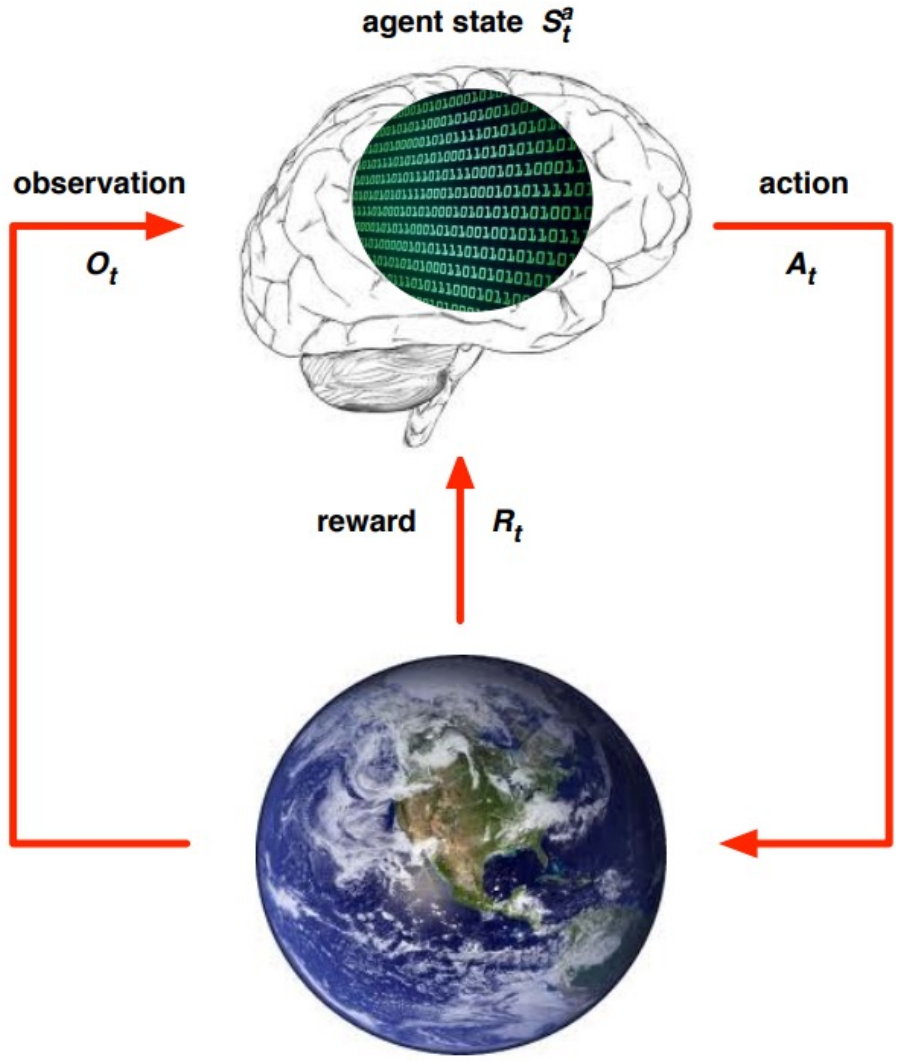
State: Environment State



- The environment state S_t^e is the environment's private representation
- i.e. whatever data the environment uses to pick the next observation/reward
- The environment state is not usually visible to the agent
- Even if S_t^e is visible, it may contain irrelevant information

Slide from David Silver's course: "Reinforcement Learning."

State: Agent State



- The agent state S_t^a is the agent's internal representation
- i.e. whatever information the agent uses to pick the next action
- i.e. it is the information used by reinforcement learning algorithms
- It can be any function of history:

$$S_t^a = f(H_t)$$

Slide from David Silver's course: "Reinforcement Learning."

State: Information State

An information state (a.k.a. Markov state) contains all useful information from the history.

What is Markov Property?

How does it map to Helicopter Example?

Slide from David Silver's course: "Reinforcement Learning."

State: Information State

An information state (a.k.a. Markov state) contains all useful information from the history.

Definition

A state S_t is Markov if and only if

$$\mathbb{P}[S_{t+1} \mid S_t] = \mathbb{P}[S_{t+1} \mid S_1, ..., S_t]$$

- "The future is independent of the past given the present"

$$H_{1:t} \rightarrow S_t \rightarrow H_{t+1:\infty}$$

- Once the state is known, the history may be thrown away
- i.e. The state is a sufficient statistic of the future
- The environment state S_t^e is Markov
- The history H_t is Markov

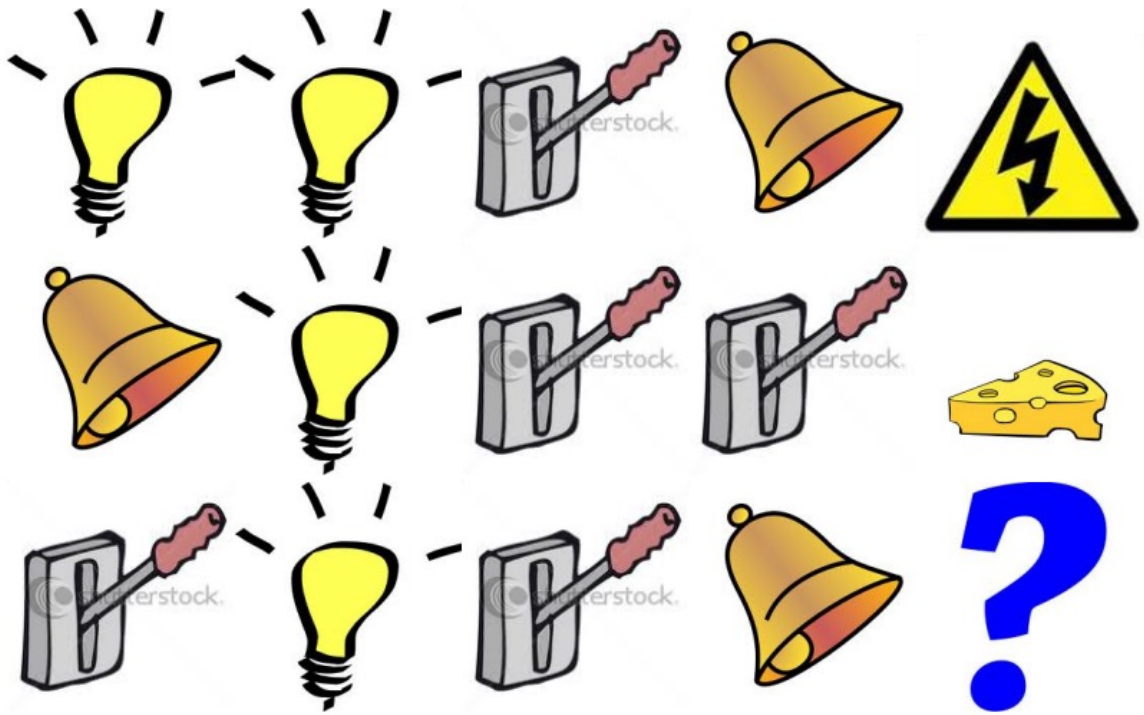
How does it map to Helicopter Example?

Slide from David Silver's course: "Reinforcement Learning."

How to pick a state?

Slide from David Silver's course: "Reinforcement Learning."

How to pick a state?



- What if agent state = last 3 items in sequence?
- What if agent state = counts for lights, bells and levers?
- What if agent state = complete sequence?

Slide from David Silver's course: "Reinforcement Learning."