Master 2 Computer Science
RL Course M2 AI

Basic Concepts in RL

Akka Zemmari

Introduction

Agent

Environment

State

Action

Reward

Episode

Introduction

Introduction

RL in a nutshell

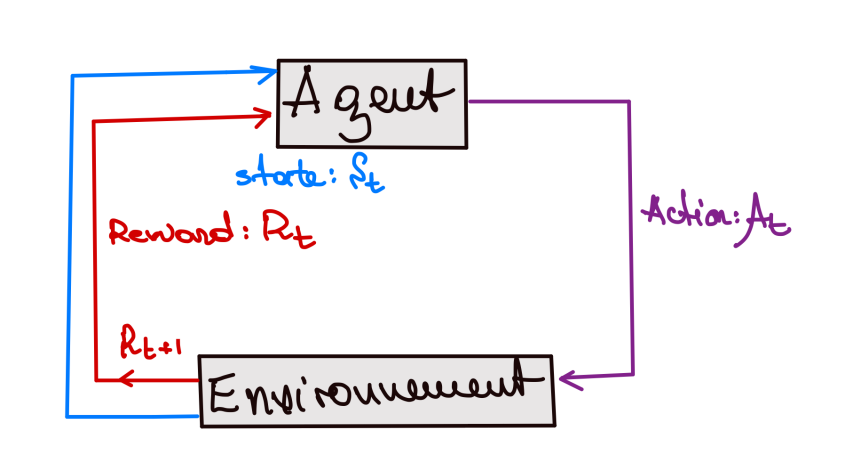

Figure 1: RL: An agent interacting with an environment

Introduction

Key Concepts:

- Agent
- Environment
- Action
- Reward
- Episode

Toy Example

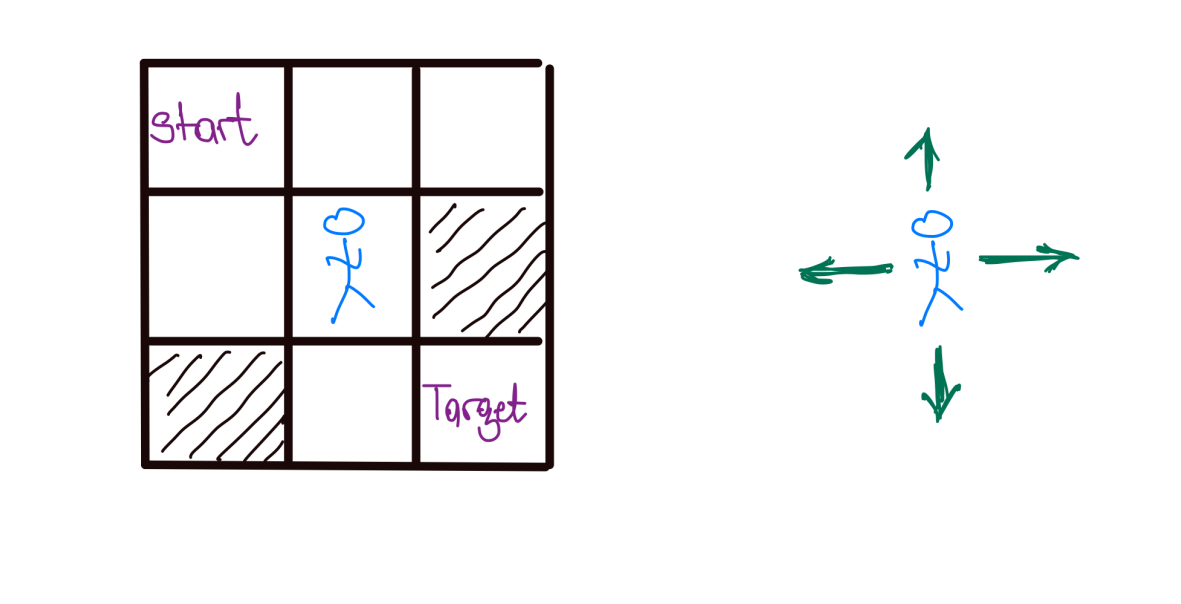

Figure 2: Toy Example: Grid World

- The agent navigates from a starting position to a goal while avoiding obstacles.

- Hands-on: see files `Grid.py` and `starter_grid.py`.

Agent

Agent

Agent: The learner or decision maker that interacts with the environment.

Components:

- Policy
- Value Function
- Model

- In the toy example, the agent is the robot navigating the grid world.

Environment

Environment

Environment: The external system with which the agent interacts.

Components:

- State
- Action
- Reward

State

State

State: A representation of the environment.

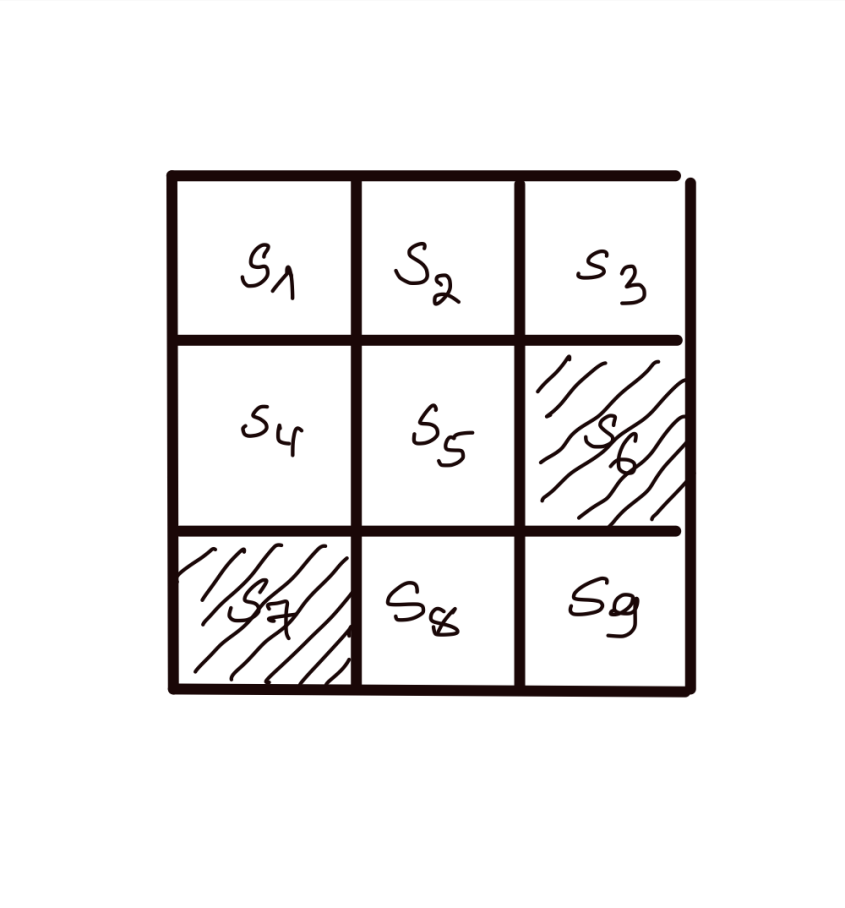

Figure 3: Toy Example: the state is the position of the robot in the Grid World.

Action

Action

Action: The set of possible moves the agent can make.

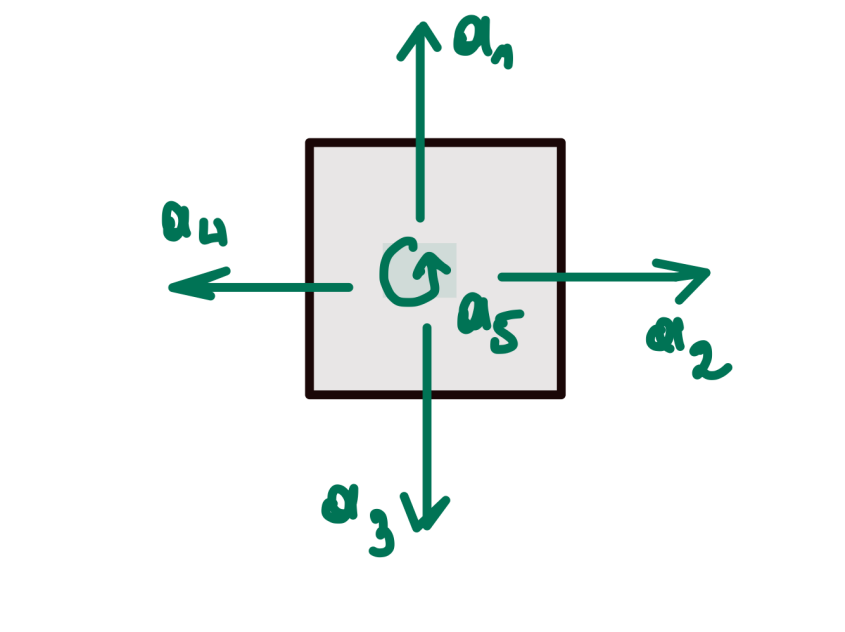

Figure 4: Toy Example: the agent can move up (a_1), right (a_2), down (a_3), left (a_4), or stay in its place (a_5).

Reward

Reward

Reward: A scalar feedback signal from the environment.

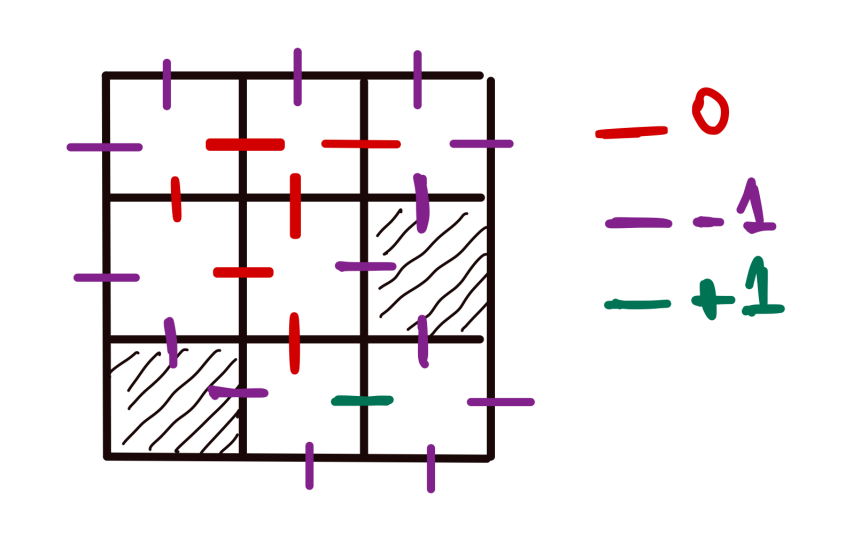

Figure 5: Toy Example: the agent receives a reward of $+1$ when reaching the goal, -1 when hitting an obstacle or get out of the boundary, and 0 otherwise.

Reward

Reward: Tabular representation (suitable for programming).

	a_1	a_2	a_3	a_4	a_5
s_1	-1	0	0	-1	0
s_2	-1	0	0	0	0
s_3	-1	-1	-1	0	0
s_4	0	0	-1	-1	0
s_5	0	-1	0	0	0
s_6	0	-1	+1	0	-1
s_7	0	0	-1	-1	-1
s_8	0	+1	-1	-1	0
s_9	-1	-1	-1	0	+1

Reward

Return: The sum of rewards over time steps.

A trajectory is a sequence of states, actions, and rewards:

$$s_1 \xrightarrow[a_1]{r_1} s_2 \xrightarrow[a_2]{r_2} s3 \xrightarrow[a_3]{r_3} \ldots \xrightarrow[a_{T-1}]{r_{T-1}} s_T. \tag{1}$$

The return is the sum of rewards:

$$return = r_1 + r_2 + \ldots + r_{T-1}. \tag{2}$$

Reward

Return: The sum of rewards over time steps.

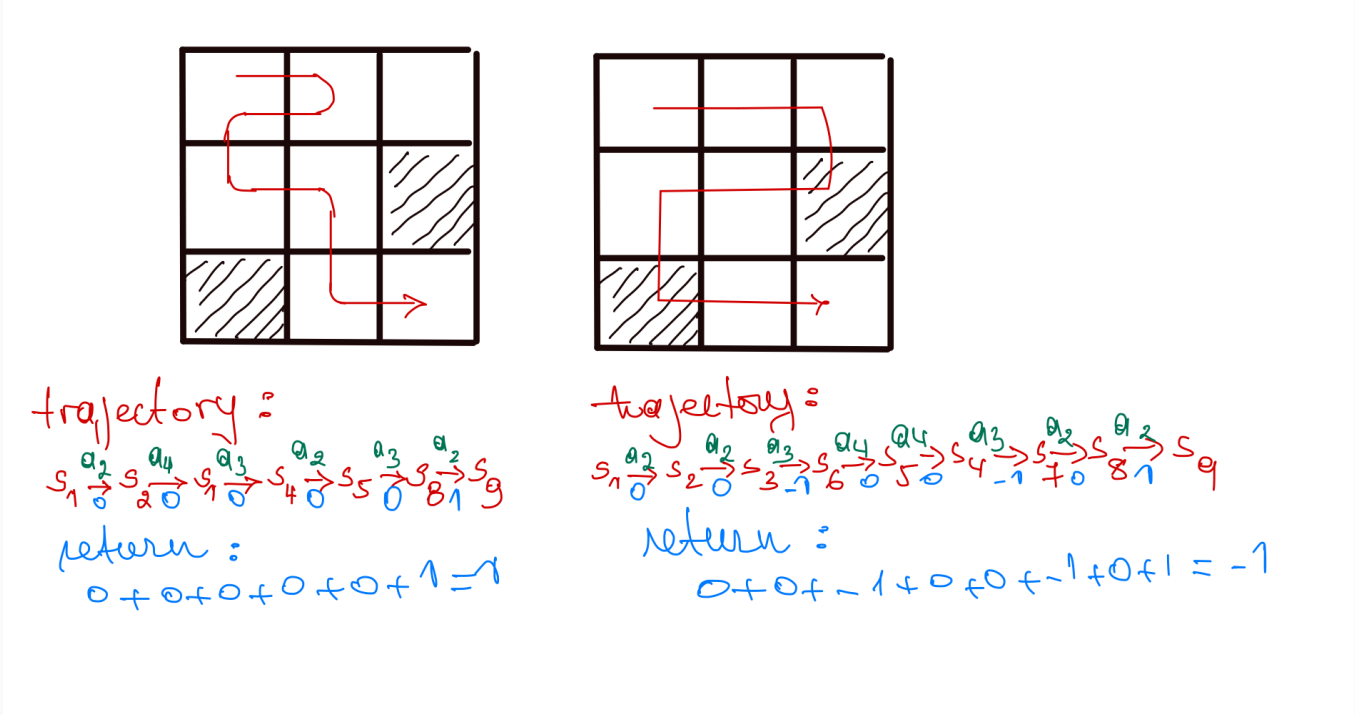

Figure 6: Toy Example: trajectories and returns.

Reward

Discounted Return: The sum of rewards over time steps with a discount factor γ.

Definition:

$$\textit{discounted return} = r_1 + \gamma r_2 + \gamma^2 r_3 + \ldots + \gamma^{T-1} r_{T-1}. \quad (3)$$

- $\gamma \in [0, 1]$ is the discount factor:

  - $\gamma = 0$: the agent is myopic (only cares about the immediate reward).
  - $\gamma = 1$: the agent is far-sighted (cares about all future rewards).

- Toy Example: see whiteboard.

Episode

Episode

Episode: A sequence of time steps where the agent interacts with the environment.

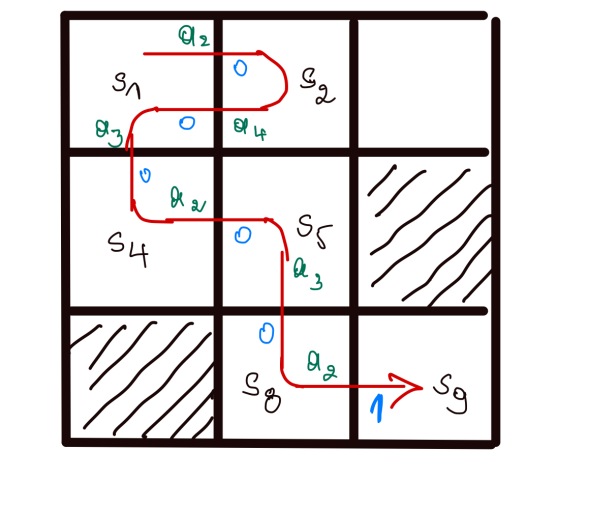

Figure 7: Toy Example: an episode in the grid world.

- An episode is usually assumed to terminate in a finite number of time steps. Tasks with episodes are called episodic tasks.