Introduction to Reinforcement Learning

Felipe José Bravo Márquez

December 30, 2022

Markov Decision Process

- A markov decision process is a tuple:

$$(S, A, \{P_{SA}\}, \gamma, R)$$

where

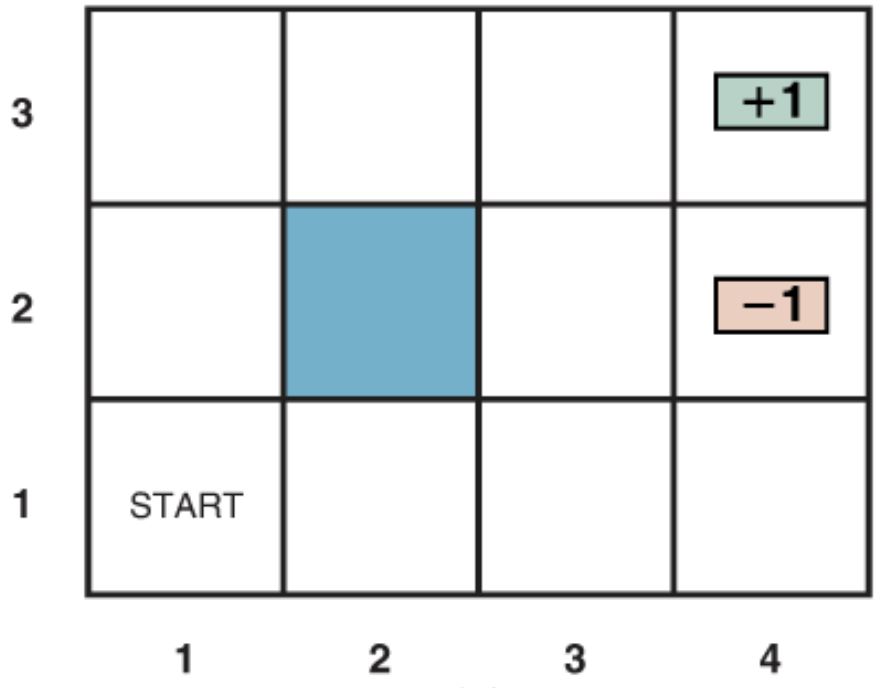
- S is a set os states.
- A is a set of actions.
- P_{SA} are the state transition probabilites:

$$\sum_{s'} P_{SA}(s') = 1, \quad P_{SA}(s') \geq 0$$

- $\gamma \in [0, 1)$ is a discount factor.
- R is a reward function.$R : S \rightarrow \mathcal{R}$.

Example

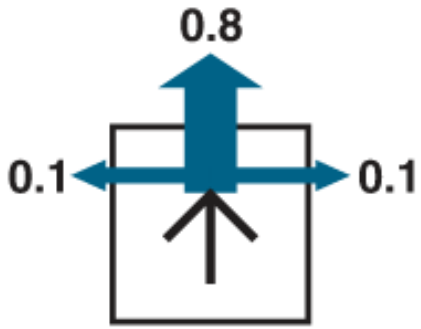
- Suppose that an agent is situated in the 4 $\times$ 3 environment shown in the Figure



- Beginning in the start state, it must choose an action at each time step.
- The interaction with the environment terminates when the agent reaches one of the goal states, marked +1 or –1.

Example

- The “intended” outcome occurs with probability 0.8, but with probability 0.2 the agent moves at right angles to the intended direction:



- A collision with a wall results in no movement.
- Transitions into the two terminal states have reward +1 and –1, respectively.
- All other transitions have a reward of –0.02 (to avoid the robot wasting time).

References I

Bickel, P. J., Hammel, E. A., and O'Connell, J. W. (1975).
Sex bias in graduate admissions: Data from berkeley: Measuring bias is harder than is usually assumed, and the evidence is sometimes contrary to expectation.
Science, 187(4175):398–404.

Hardt, M. and Recht, B. (2021).
Patterns, predictions, and actions: A story about machine learning.
arXiv preprint arXiv:2102.05242.