TD2

Artificial Intelligence - Introduction to Reinforcement Learning
ENSISA 2A

Ali El Hadi ISMAIL FAWAZ

November 16, 2025

Problem I: Monte-Carlo

Consider an undiscounted Markov Reward Process (MRP) with two states A and B. The MRP is unknown, meaning that the state transition matrix and the reward model are not visible for the agent.

However, the agent has these two sample episodes:

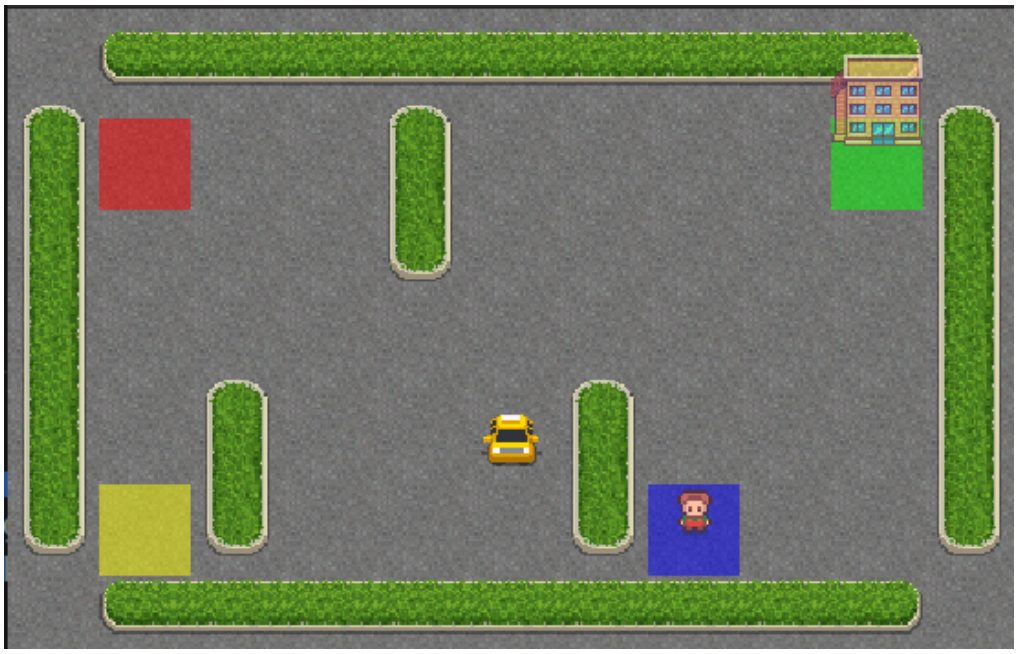
- $A + 3 \rightarrow A + 2 \rightarrow B - 4 \rightarrow A + 4 \rightarrow B - 3 \rightarrow TerminalState$
- $B - 2 \rightarrow A + 3 \rightarrow B - 3 \rightarrow TerminalState$

In the above episodes, $A + 3 \rightarrow A$ indicates a transition from state A to state A with a reward of $+3$ and so on.

Questions

1. Using first-visit Monte-Carlo evaluation, estimate the state-value function $V(a)$ and $V(B)$.
2. Same as question 1 but with every-visit Monte-Carlo evaluation.
3. Draw a diagram of the Markov Reward Process that best explains these two episodes. Show the transition probabilities and rewards on your diagram.
4. Recall the Bellman equation for an MRP.
5. Solve the Bellman equation to give exact state-value functions for states A and B.

Problem II: Self Driving Taxi with Q-Learning



Given the following environment made of a grid ($5x5$) parking lot, a passenger and a self-driving taxi. The goal is that the agent (the taxi) picks up the passenger and drop them into the desired location.

There exists four possible locations for the initial passenger location and the destination. These four locations are colored in Red, Blue, Yellow and Green in the figure above.

The possible actions that the agent can do are:

- Go West
- Go East
- Go North
- Go South
- Pickup

- Drop-off

The rewarding system is defined as follows:

- If the agent correctly drops the passenger into the corresponding destination, they get a reward of $+20$
- If the agent drops off or picks up at the wrong location they get a reward of -10
- For each time-step, the agent gets a reward of -1

At each episode, the agent starts at a random location, the passenger at one of the four possible locations (random) and the destination is also at one of the four possible locations (random).

Questions

1. Define the set of States.
2. Solve the Taxi problem by implementing the Q-Learning algorithm (hyper-parameters of your choosing).

PS: use code template