Civil and Environmental Engineering (CEE) 290:
Homework # 3

Instructor: Maria Laura Delle Monache
GSI: Dylan Kato

Handed out: 04/03/2023.
Due: 04/28/2023 at 9:59am (on bCourses)
Collaboration policy: You can work in study groups of two or three (no more than (3)) or on your own.

Each student must submit their own original paper in bCourses and all written work must be your own.

Submission policy: Submission must be done via bCourses.
Each paper must begin with the student's name and the name of the students they worked with (if any).

Example: Homework #3: firstname1, lastname1
Group: firstname2, lastname2 and firstname3, lastname3

Each uploaded paper and code must be named as follows:
S23-CE290-hw3-lastname-firstname.pdf /S23-CE290-hw3-lastname-firstname.py or S23-CE290-hw3-lastname-firstname.ipynb

---

IMPORTANT: We will not rerun the codes that you submit. So anything that needs to be considered for grading should be included in the PDF that you upload in bCourses, including figures and results.

Failure to follow these instructions will result in the paper not being graded.

Problem statement

The goal of this problem set is to learn how to construct a reinforcement learning problem and solve it.

Task 1: Markov Decision Process (35 points)

1. Think of two possible applications that would fit into the framework of reinforcement learning. Describe the applications and how would they fit a reinforcement learning problem. Be as precise as possible. For each one of the examples describe its state, actions and rewards. The examples need to be sufficiently diverse from each other and possibly not one of the examples seen in class. (20 points)

2. Pick one of the previous two applications, draw a simplified Markov Decision Process. Pick numerical values for the MDP.

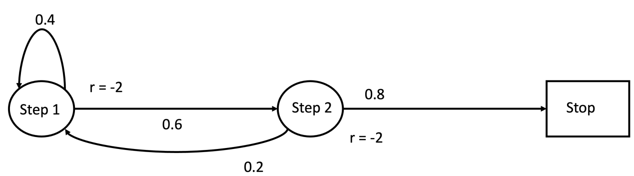
    - Define the state-value function (5 points)

- Consider the uniform random policy that takes all action from state s with equal probability, compute one iteration of the iterative policy evaluation and compute the new value function. (10 points)

Task 2: Reinforcement learning for the rebalancing of scooters sharing docking stations (40 points)
You are the manager for a scooter sharing company of two docking stations that can park 20 scooters each at most. Every day, people arrive to the docking stations to pick up scooters. If the scooter is available, the scooter is rented and you earn 4$. If there are no scooters available you lose business. You can move scooters from one station to another overnight at a cost of 1$ each. A maximum of 6 scooters can be moved overnight. Assume that the number of scooters requested and returned are Poisson Random Variables. And that at one location the expected number (average) of scooters requested is 4 and the expected number (average) of returned scooters is 4, while for the other location those numbers are respectively 5 and 3.

1. Formulate the problem as a Markov Decision Process. (10 points)
2. Find the optimal policy so that you have scooters always available whenever requested. Said otherwise, what is the optimal policy for rebalancing your scooters so that you maximize your income? (30 points)

Task 3: Monte Carlo methods (25 points)
Consider the following Markov Reward process:



The diagram shows the transition probabilities and the rewards.

1. Pick two random sample sequences. (or episodes) (5 points)
2. Estimate the state-value function for each state using
   - First-visit Monte Carlo evaluation. (5 points)
   - Every-visit Monte Carlo evaluation. (5 points)
3. Define and solve the Bellman equation for the Markov Reward process. (10 points)

Credit: Some exercises are derived and modified from [1], [2].

References

[1] David Silver. Lectures on reinforcement learning. URL: https://www.davidsilver.uk/teaching/, 2015.

[2] Richard S Sutton and Andrew G Barto. *Reinforcement learning: An introduction.* MIT press, 2018.