Reinforcement Learning
Exercise 4 - Solution

Jonathan Schnitzler - st166934
Eric Choquet - st160996

May 16, 2024

Task 1)

a) Advantages of Monte Carlo over dynamic programming

1. Does not require knowlegde of the environment $p(s'|s, a)$ and $r(s, a, s')$.

2. Simulate experience via simulator (or real world)

3. every episode vs every state

b) Example for Monte Carlo to learn the value function (over DP)
MC is preferable over DP when the distributions $p(s'|s, a)$ and $r(s, a, s')$ can not be sampled. For example, in blackjack it is not possible to sample them because they are up to chance.

Task 2)

Programming task
    The final result yields