CSCI 6963 - Reinforcement Learning
Homework 6

Christopher Okonkwo

The deadline is 11:59pm on Thursday, October 30.

Question 1

Why is the policy in the case of Mountain Car time-dependent? In other words, why does the value of a given state depend on the current time step?

Solution 1. *The policy in Mountain Car (MC) is time-dependent because of the finite horizon. MC has a finite time constraint with a maximum time step $T = 150$. The car must reach the goal within this time limit, or the episode fails. The value of a given state depends on the current time step because, in a finite-horizon problem, the number of future steps left to earn rewards reduces over time. That is to say that as time progresses or the closer you are to the horizon, the less reward you can collect, so $V_t(s)$ depends on t. While in infinite-horizon (Pendulum with discount $\gamma = 0.9$), there is no time dependency, every step with the same state always has the same optimal action, so the optimal policy $\pi^*(s)$ depends only on the state, not on time.*

Question 2

What is the trade-off of having more states in your MDP? Think about how closely the MDP approximates the simulator.

Solution 2. *There are fundamental accuracy and computation trade-offs when choosing the number of states in a discretized MDP. The more states there are, the finer the discretization. Having more states makes the MDP approximate to the real simulator more closely with lower bias, capturing minor differences in position/velocity more accurately, but increases computational and sample costs (higher variance and slower convergence).*