Homework 2

Kade E. Carlson

October 21, 2024

1 Abstract

The code reviewed for this homework uses Q-learning to solve an environment developed in OpenAI's Gym package. The environment is the Cliff Walking environment. The cliff walking environment is a 4x12 grid world where an agent starts in the bottom left corner and attempts to reach the goal in the bottom right corner. In between the start and goal location is a cliff and if the agent steps on anyone of these cliff squares it receives a negative reward and then starts again at the beginning location. The action space is up, down, left, or right movement.

The goal of the Q-learning algorithm here is to learn an optimal policy that reaches the goal without falling off the cliff. The code starts by initializing an epsilon-greedy policy where epsilon is a small value that determines whether a random action is taken or not. If no random action is taken, then it chooses the highest value from the Q-policy given an observation. Finally, do a temporal difference update using current reward, next state expected reward and a discount factor. This is done for t timesteps and n episodes.