CS 4782 Coding Assignment 5 Written Responses

Ryan Noonan rtn27, Chad Yu cky25

May 7, 2025

Note: For homework, you can work in teams of up to 2 people. Please include your teammates' NetIDs and names on the front page and form a group on Gradescope. (Please answer all questions within each paragraph.) Please show all the relevant steps in your solutions.

Question 1:

How does the behavior of the Q-Learning agent compare to the random policy visualized previously? What policy does the agent appear to have learned? Write 2-3 sentences below. (5 pts)

The behavior of the Q-learning agent is much more deliberate and logical than the visualized random policy. Not only does the Q-learning agent actually achieve the task, but the random policy takes so many unnecessary steps that does not contribute at all to making progress. The policy that the Q-learning agent seems to have learned is the shortest path to the passenger's location, then the shortest path to the target destination, which in this tabular case, is indeed the optimal policy.

Question 2:

How does the behavior of the DQN agent compare to the random policy visualized previously? What policy does the DQN agent appear to have learned? Write 2-3 sentences below. (5 pts)

The difference between the behavior of the DQN agent vs. the random policy is not as extreme as the Q-learning vs. the random taxi, which intuitively makes sense because there are only two actions in the Cartpole action space, but we can still tell that DQN balances the Cartpole slightly better. The policy that the DQN agent appears to have learned corrects for the left leaning pole that the random policy fails to account for, but appears to overcorrect for this as at some points, especially in the later episodes, the Cartpole is slightly leaning right, so the policy does not appear to be optimal.

Question 3:

Compare your results for using DQN and Reinforce. Discuss the performance, convergence, and behavior of the two algorithms. Consider factors such as the stability of learning, the quality of the learned policies, and the behavior of the final policy. Write 3-5 sentences below. (5 pts)

Visually, the performance of the learned policy is much better in Reinforce vs. DQN, as the behavior of the final policy as displayed in the GIF for Reinforce looks much more stable than the DQN policy's behavior. Not only does the Cartpole rock back and forth with a lot less amplitude, but we can also see that the absolute horizontal movement is a lot less as well. Also, looking at the graphs of the rewards per episode in DQN vs. Reinforce, we can tell that the stability of learning and quality of learned policy is much higher for Reinforce, as the rewards per episodes are steadily increasing the whole time for Reinforce, while the DQN graph is steadily increasing at the beginning and then takes on a really jagged shape afterwards, demonstrating instability.