Meadow Monticello

Dr. Forouraghi
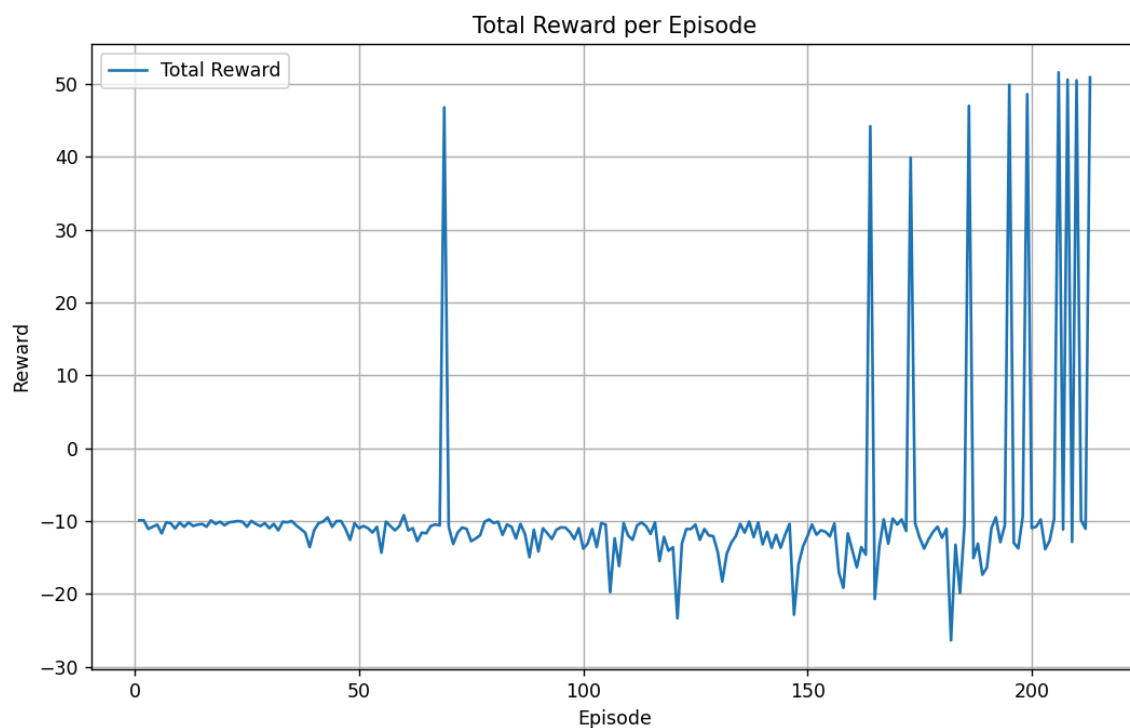
CSC 362

9 December 2025

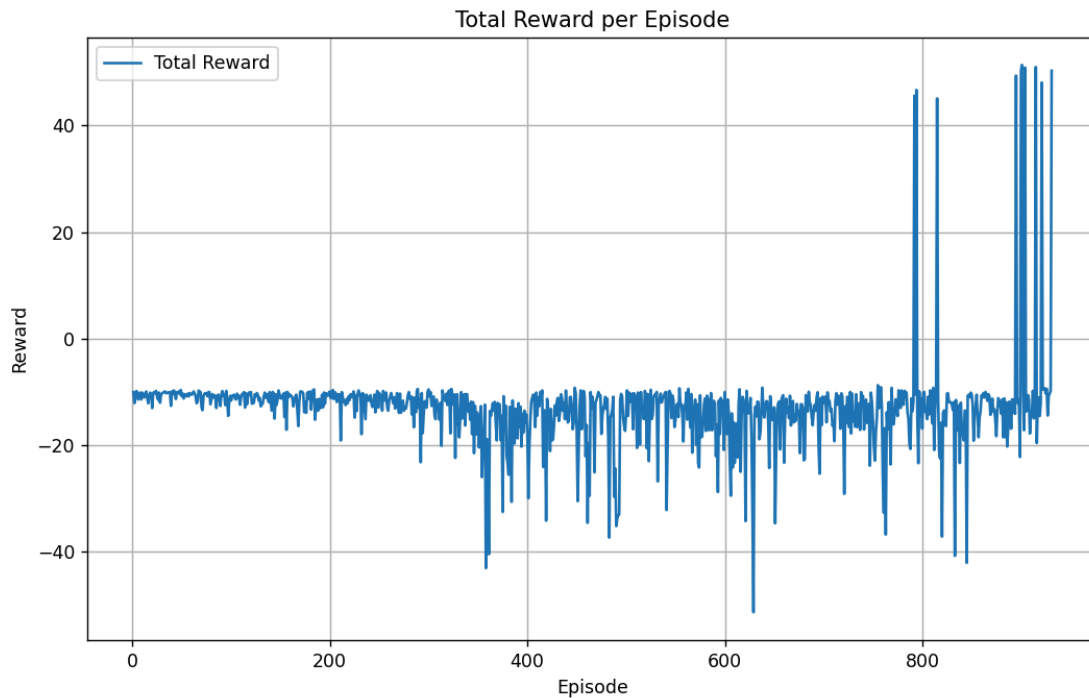<div align="center">Assignment 5</div>

Case: 1A

Epsilon Decay: 1250 (high)

The epsilon decay rate was set to 1250, which is high. This means that exploitation is rapidly favored. The agent completed 10/10 goals with about 200 episodes. The successes towards the end of the run-time (right side) are very close together. When the agent finds a successful path, it repeats and exploits it with very minimal exploration. If there were a bigger grid and there was a more significant size difference between the longest and shortest paths for the agent, the high epsilon rate's downsides would be more apparent.
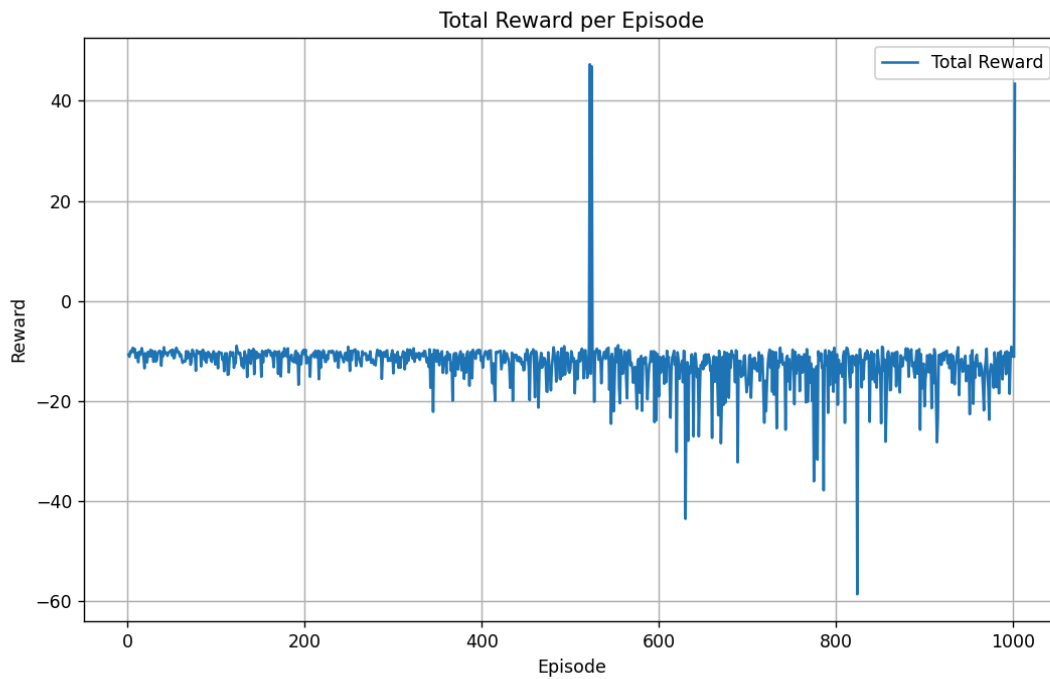
1B

Total Reward per Episode

Epsilon Decay: 2500 (medium)

This epsilon decay rate was balanced at 2500. The agent balances exploration and exploitation. The agent achieves the 10/10 goals in under 1000 episodes. As I stated earlier, if the grid space was larger, we'd more clearly see the optimality of this epsilon decay rate. We can see the goals were still achieved and there was more exploration at the beginning to find an optimal path to then eventually exploit.
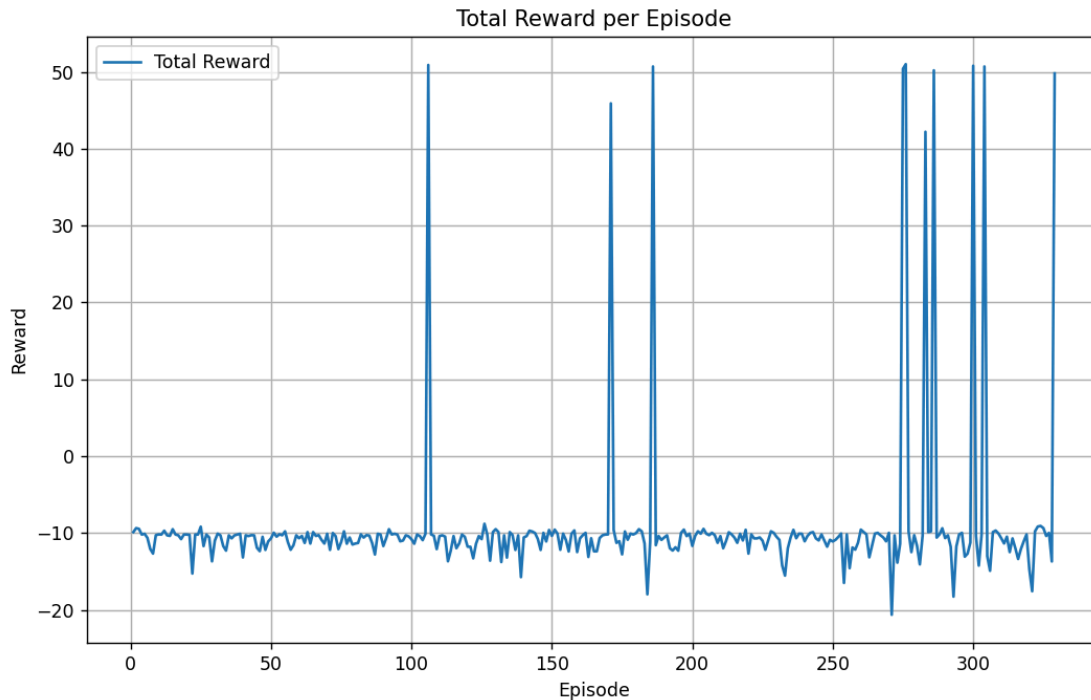
1C

Epsilon Decay: 5000 (low)

The epsilon decay rate is low here at 5000. The agent spends more time exploring than exploiting and takes its time. While exploring is helpful in finding an optimal path, too much time dedicated to exploration prevents significant progress. There were only 2 goals reached before it quit at 1000 episodes.
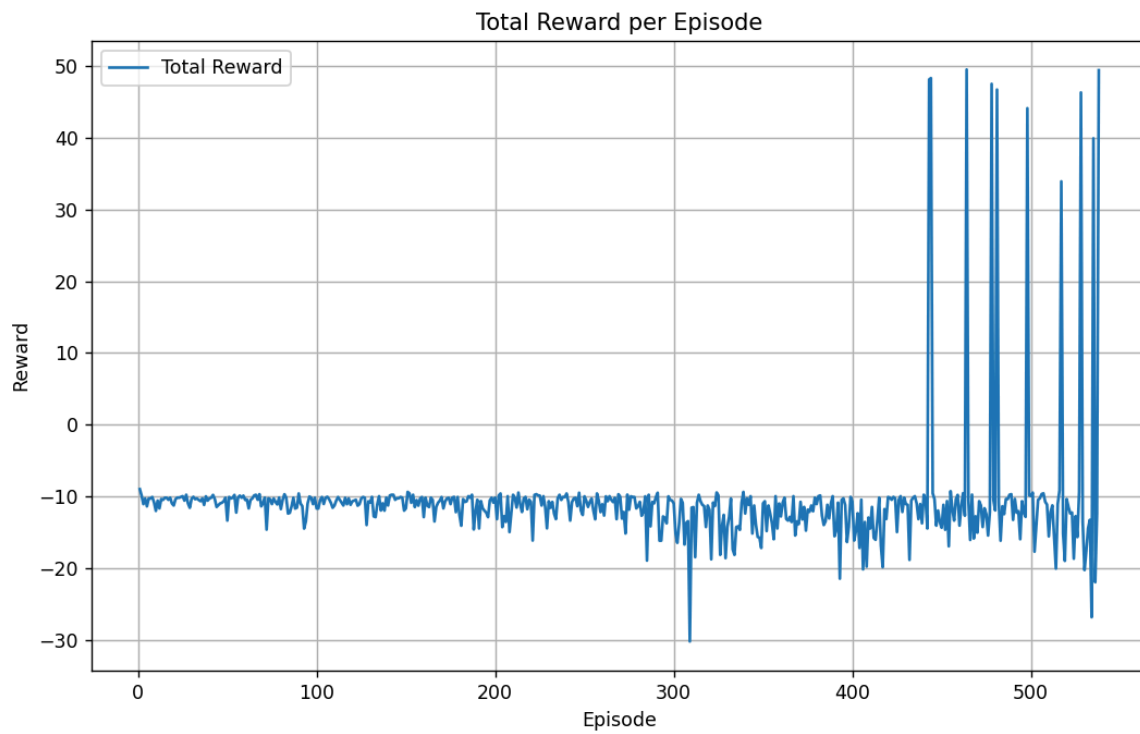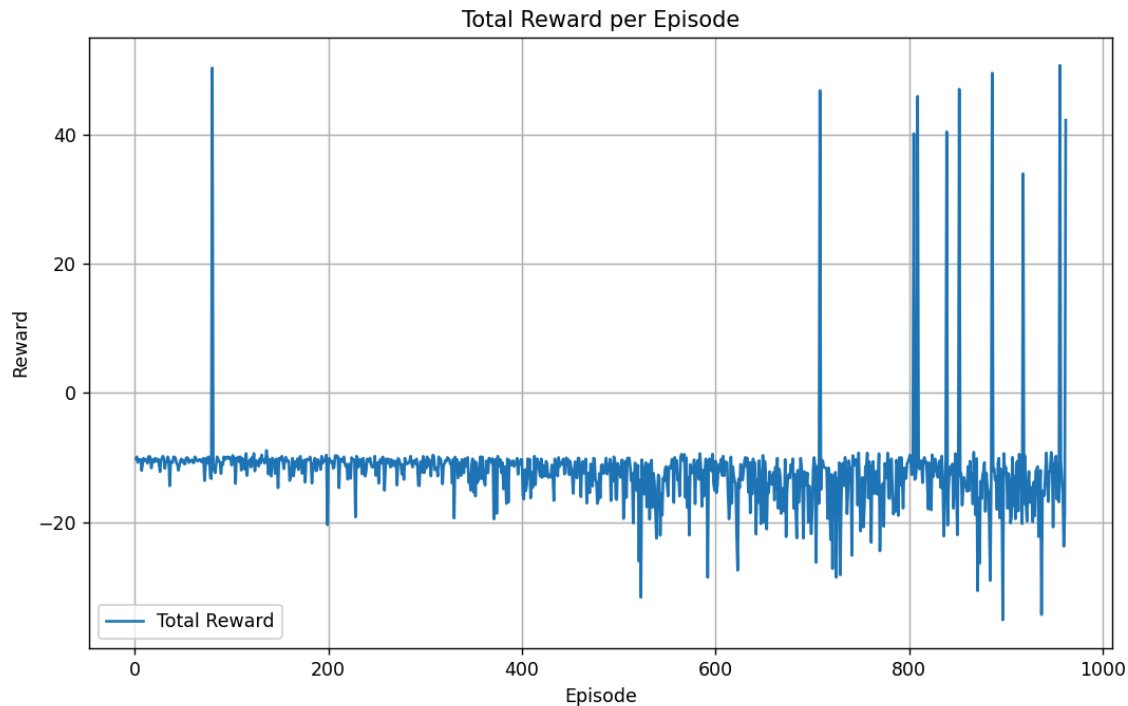
PART 2:

Case: 2A

Gamma = 0.5

In the chart above we can see the results of the DQN when the gamma value or discount value is set to 0.5. Gamma is low here and it means that the function is short-sighted. It was anticipated that it would take longer to find the optimal path since it prioritized instant rewards over long term success. Surprisingly, the goal was reached 10/10 in under 350 episodes. On the right side of the graph, we can see that the agent began to rapidly and repeatedly achieve the goal. This makes sense because at this point it knows a proper path and can repeat it until all 10 goals have been achieved, even if the path is not optimal. I hypothesize that if the agent was navigating a larger grid, the drawbacks of a short-sighted agent would become more apparent. Since this model was a relatively small grid, there was not a significant difference in distance between the longest and shortest possible paths.

Gamma = 0.3. For extra fun and experimentation, I also ran with Gamma even lower at 0.3. Here we can see that about 550 episodes were needed for goal completion.
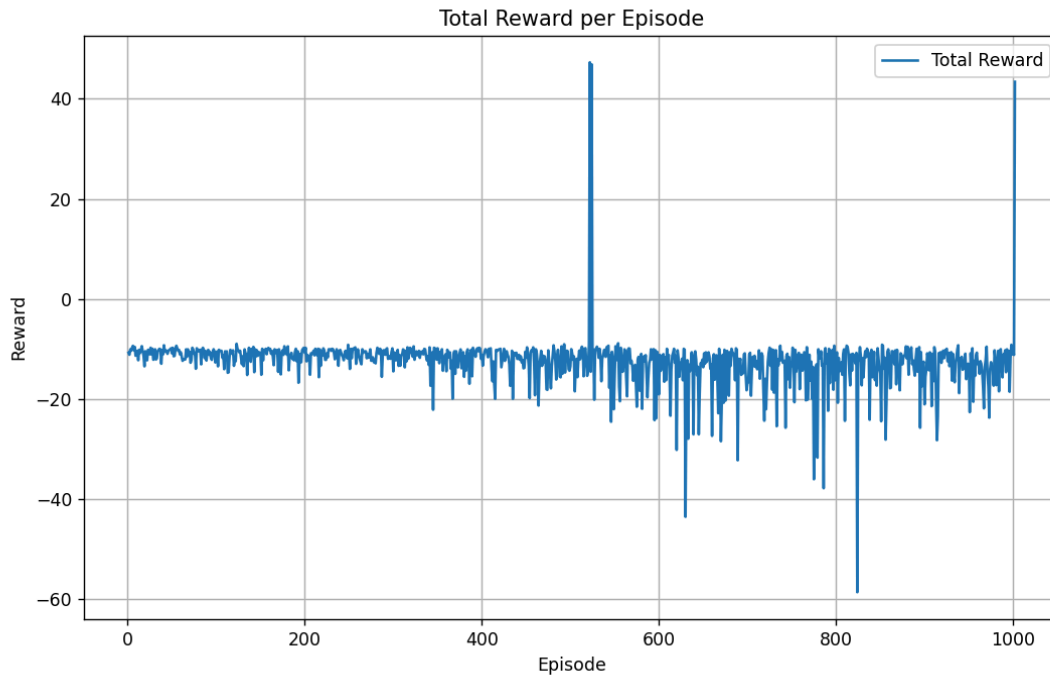
2b

Gamma = 0.8

In the above graphic we can see Gamma or the discount factor was balanced. We can see that the agent was rewarded in one of the early episodes and not again towards the later episodes. Compared to low-gamma, there are also more episodes between times where the goal was reached. This could be due to the agent balancing short and far-sightedness. It will not settle for a very long path that the agent finds, but it will also not exhaust all trials locating the shortest possible path. It is fairly rewarded throughout the trial and was able to be successful 10/10 in under 1000 episodes.

2c

Gamma = 0.99

The Gamma or discount factor is high and the agent prioritizes finding the shortest possible path. It took longer to reach a goal state then the other gamma trials (~500 episodes) and continued its search slowly. Long term success was valued higher and ultimately slowed down the agent in achieving the goals in a timely manner. I force timed-out all trials at 1000 episodes.