CS 541: Artificial Intelligence, Winter 2022
Programming Assignment #3

The goal of this assignment is to have Robby the robot learn to correctly pick up cans and avoid walls in his grid world using Q-learning. Robby the Robot lives in a 10 x 10 grid, surrounded by walls and some grid squares contain soda cans. Robby has five possible actions - move North, move South, move East, move West and Pick-up-can.

Robby gets rewards in for the following :

- +10 for each can he picks up
- -5 if he crashes into a wall
- -1 if he picks up a can in empty square

According to the Q-learning method, the agent (robot) needs to choose its action which has the highest q value. Initially, all actions have a q-value of zero.The agent can either explore the environment to determine q values for particular actions or it can exploit the environment by using the information already available to it. In each episode, in 200 steps, the robot is going to explore or exploit the environment as the learning process. Rewards are calculated every 100 episodes.

Epsilon is set to 0.1 to begin, and decreases by .05 every 50 episodes until it reaches 0. After this point, epsilon remained zero until all episodes were complete. After training, the test was run again with the q-matrix obtained during training, and epsilon set at 0.1 (unchanging). The Test-average and Test-standard-deviation values in the results correspond to this test run. Once the model is trained, we can plot the graph of Sum of the rewards vs Number of episodes.

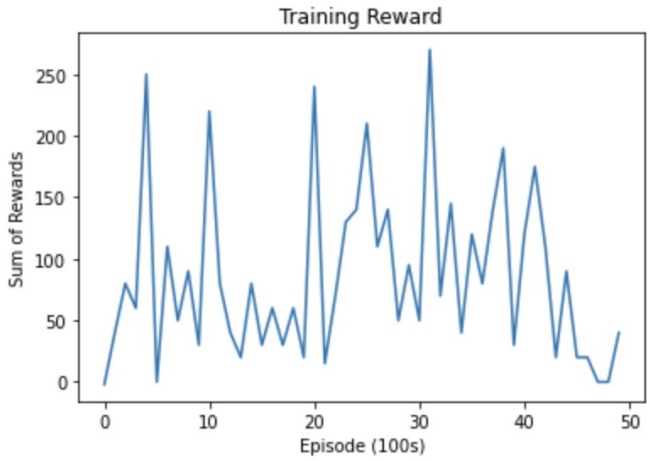
Results:

Part1:

For , N= 5000, M= 200, η = 0.2, γ = 0.9
The standard deviation we achieve here is 84.22

Here, we have achieved the average value from the test rewards to be fairly close to the actual reward value. Hence, it can be determined that the agent has performed well.

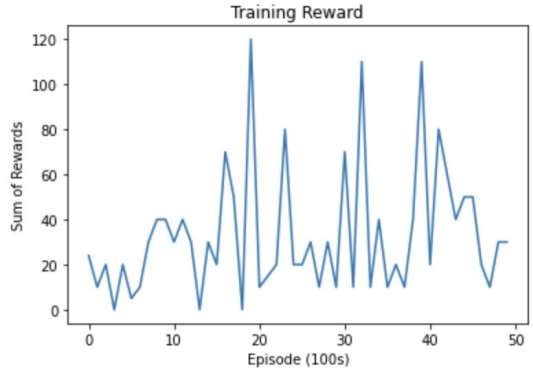
```
The Test-average is:  169.96
The Test-standard-deviation is:  84.22896413942178
```



Part 2: Experiment with learning rate:

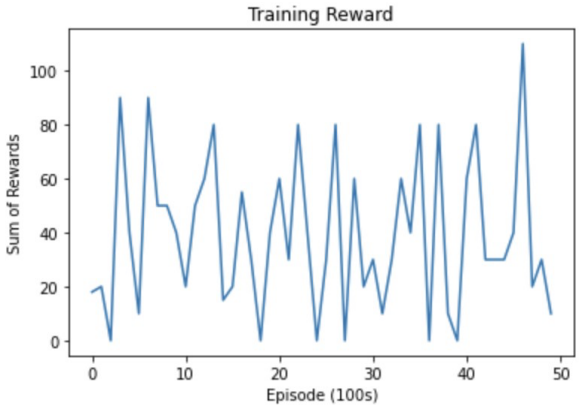
1. Learning rate = 0.8

```
The Test-average is:  75.16
The Test-standard-deviation is:  36.57067677798703
```



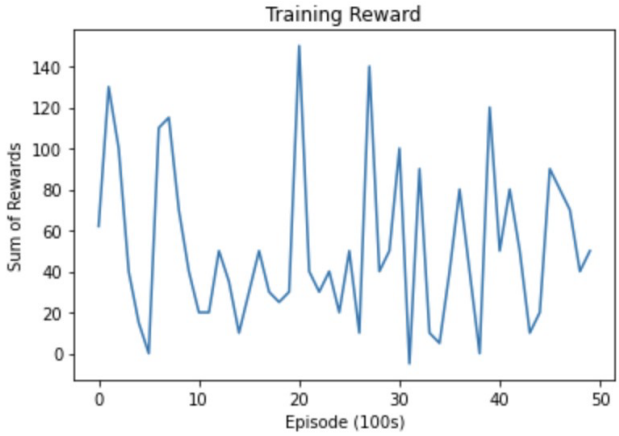
2. Learning rate = 0.6

```
The Test-average is:  48.26
The Test-standard-deviation is:  40.85770918688418
```



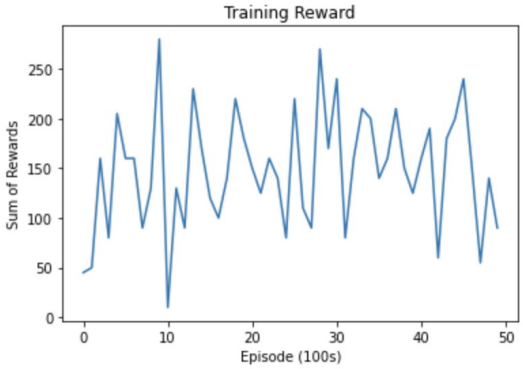
3. Learning rate = 0.4

```
The Test-average is:  92.38
The Test-standard-deviation is:  49.92389808498531
```



4. Learning rate = 0.2

```
The Test-average is:  190.12
The Test-standard-deviation is:  65.22411823857797
```

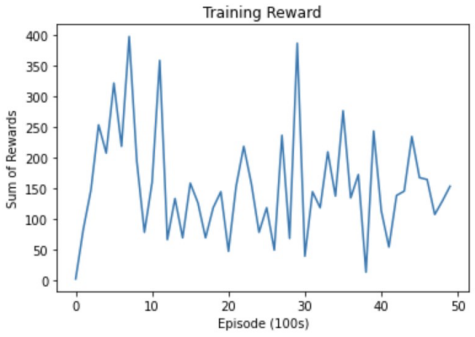


As the learning rate decreases, we notice that the testing average and standard deviation values decrease. This occurs because lower the learning rate, the less likely the agent is inclined to learn from the environment. This means that it is less likely to abandon information that it has already gathered.

Part 3: Experiment with epsilon

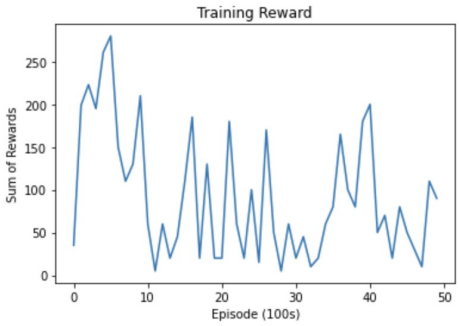
1. Epsilon = 0.9

```
The Test-average is:  181.84
The Test-standard-deviation is:  86.540478390173
```



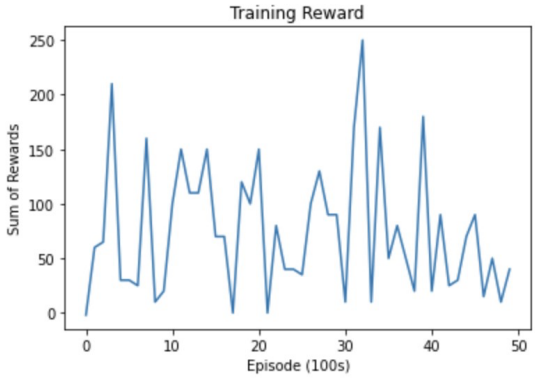
2. Epsilon = 0.65

```
The Test-average is:  166.2
The Test-standard-deviation is:  77.79023074911142
```



3. Epsilon = 0.1

```
The Test-average is:  147.6
The Test-standard-deviation is:  61.53470565461413
```



As the epsilon decreases, we notice that the testing average and standard deviation values decrease. This occurs because lower the epsilon value, the more likely the agent performs the non greedy action from the current state.