

**CS 541: Artificial Intelligence, Winter 2023**  
**Programming Assignment 3 – Report**  
**Submitted by: Chitradevi Maruthavanan**  
**PSU ID: 950828319**

The goal is to use Q-learning to teach Robby the robot to accurately pick up cans and avoid walls in his grid. Robby lives in a 10 x 10 grid surrounded by walls and with soda cans in certain grid squares. Robby can take one of the below five actions:

- move up
- move down
- move left
- move right
- pick-up-can

Robby gets the below rewards for his actions:

- +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 approach, the agent (Robby the robot) must choose the action with the highest q value. All actions start with a q-value of zero. The agent can do one of the two steps below:

- Explore the environment to determine q values for actions
- Exploit the environment by utilizing the current information

As part of the learning process, the robot will explore or exploit the surroundings in 200 steps for each episode. Every 100 episodes its rewards are calculated.

Epsilon begins at 0.1 and decreases by .05 every 50 epochs until it reaches zero. Epsilon will remain zero after this until all episodes are completed. The test was repeated after training with the q-matrix obtained during training and epsilon set to 0.1. This test run is represented by the Test-average and Test-standard-deviation values in the results. We can plot the graph of Sum of Rewards vs. Number of Episodes once the model has been trained.

Results:

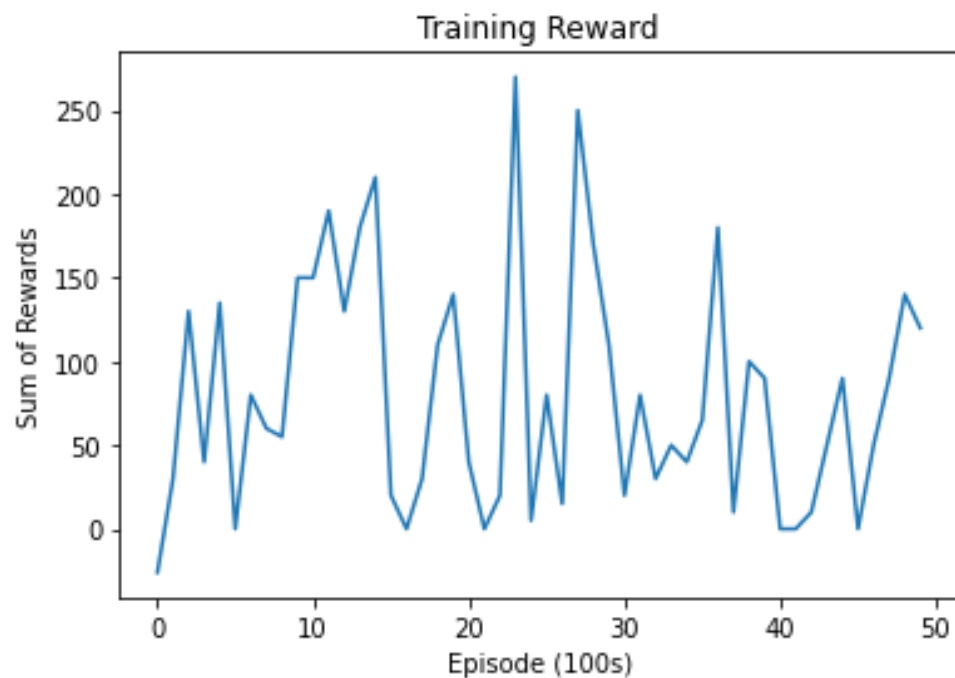
**Part 1:** For,  $N = 5000$ ,  $M = 200$ ,  $\eta = 0.2$ ,  $\gamma = 0.9$

The standard deviation is 84.22.

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

**The Test-average is: 172.92**

**The Test-standard-deviation is: 93.84473133852534**

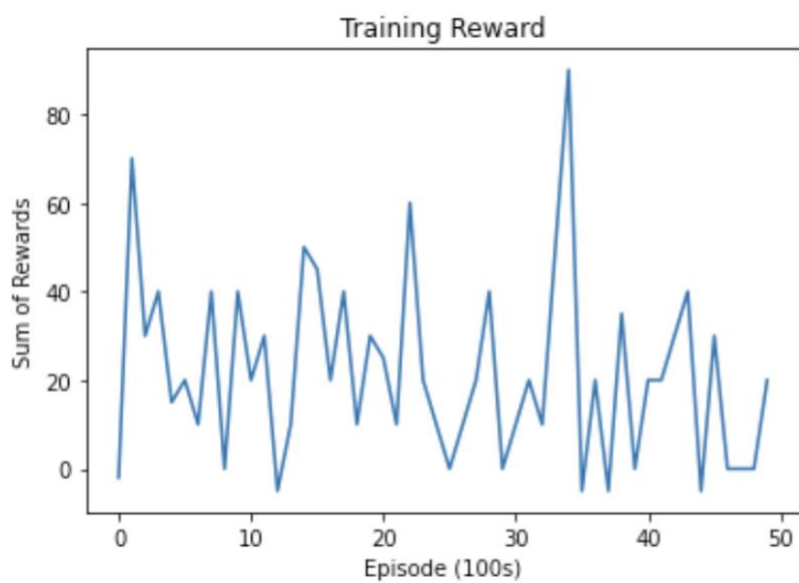


**Part 2:** Experiment with learning rate:

1. Learning rate = 0.9

**The Test-average is: 64.94**

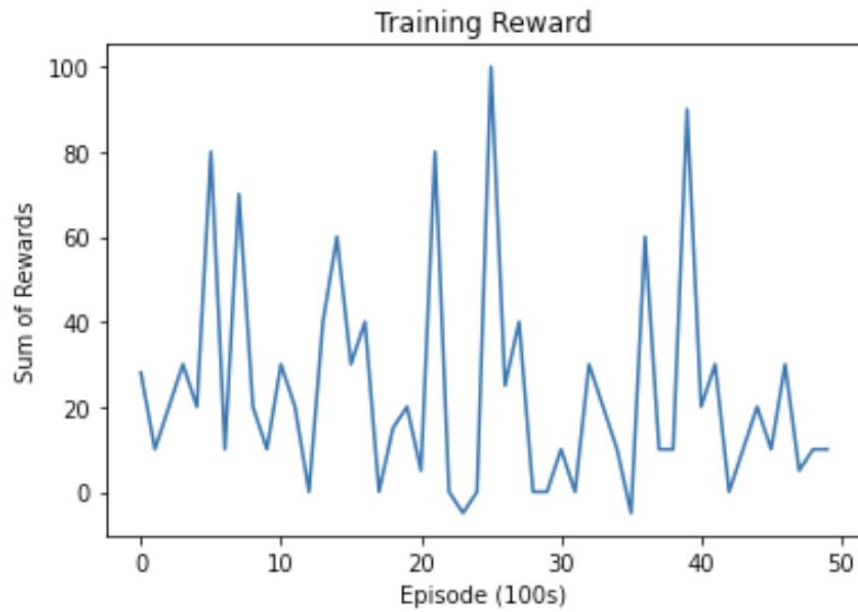
**The Test-standard-deviation is: 38.366344626508265**



2. Learning rate = 0.7

**The Test-average is: 50.46**

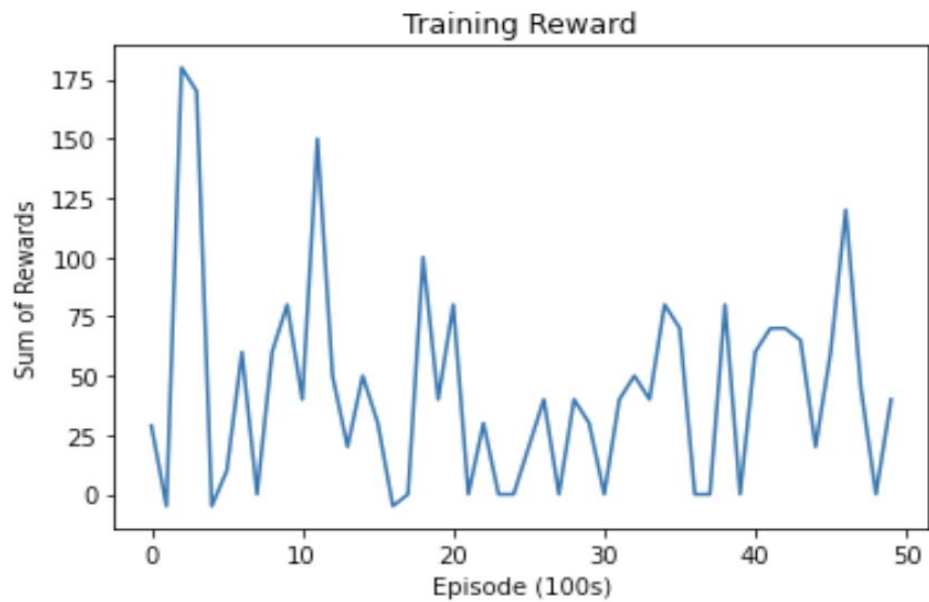
**The Test-standard-deviation is: 29.430739032514968**



3. Learning rate = 0.5

**The Test-average is: 128.16**

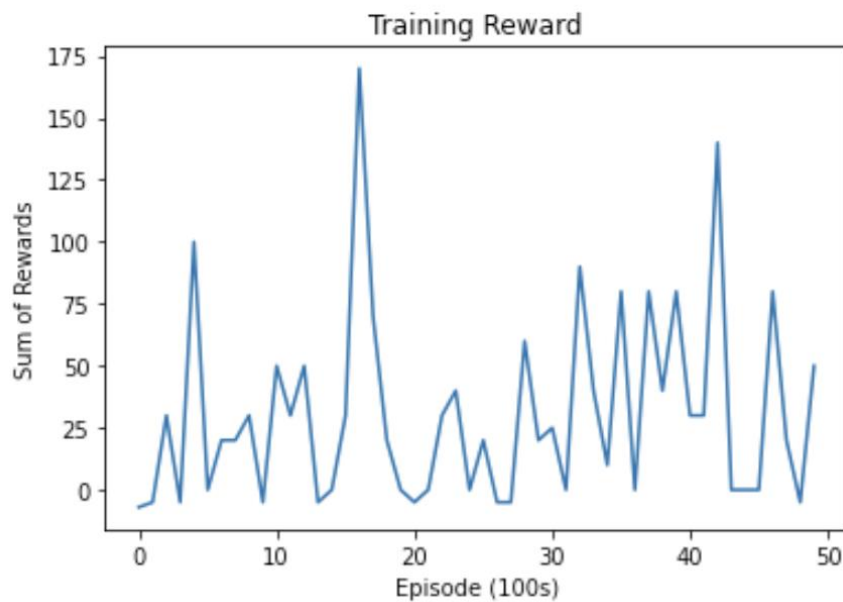
**The Test-standard-deviation is: 72.9677627449273**



- Learning rate = 0.2

**The Test-average is: 86.94**

**The Test-standard-deviation is: 46.16509937171153**



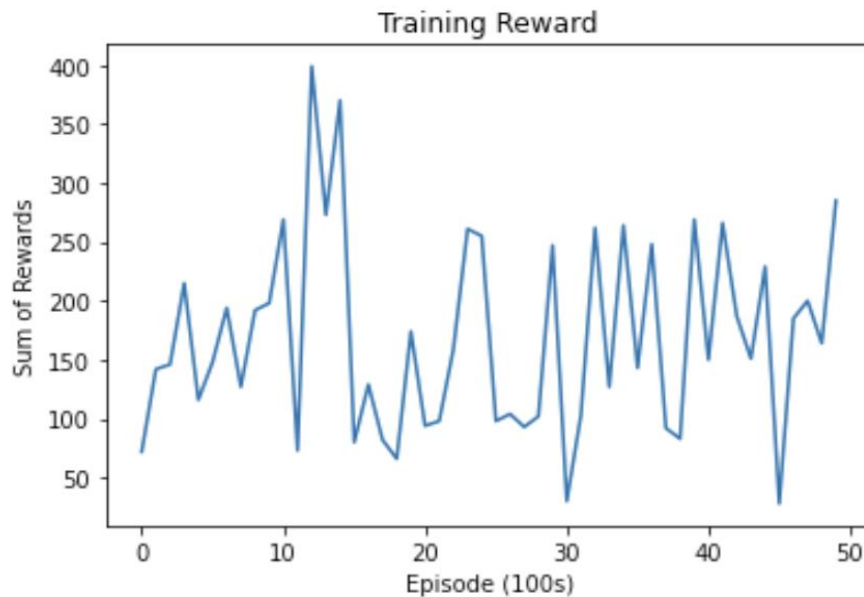
As the learning rate decreases, we notice that the testing average and standard deviation values decrease. This occurs because lower the learning rate, the less likely the agent is inclined to learn from the environment. This means that it is less likely to abandon information that it has already gathered.

### Part 3: Experiment with epsilon

- Epsilon = 0.8

**The Test-average is: 169.76**

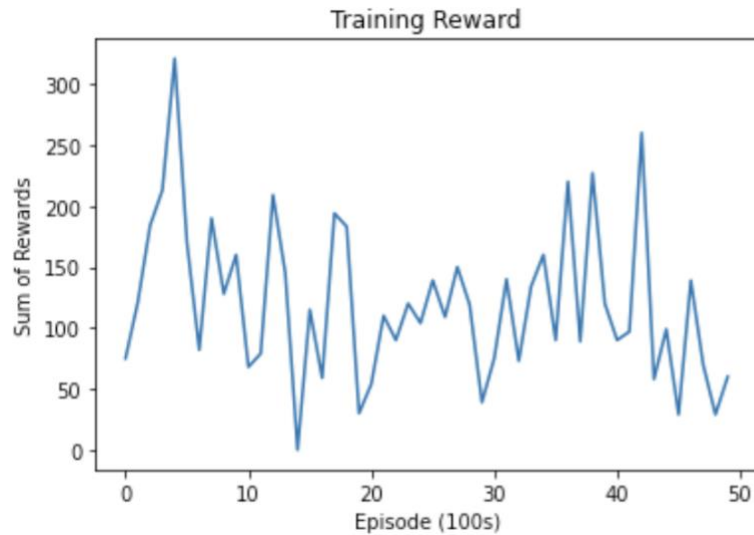
**The Test-standard-deviation is: 68.29569825398961**



2. Epsilon = 0.67

**The Test-average is: 158.08**

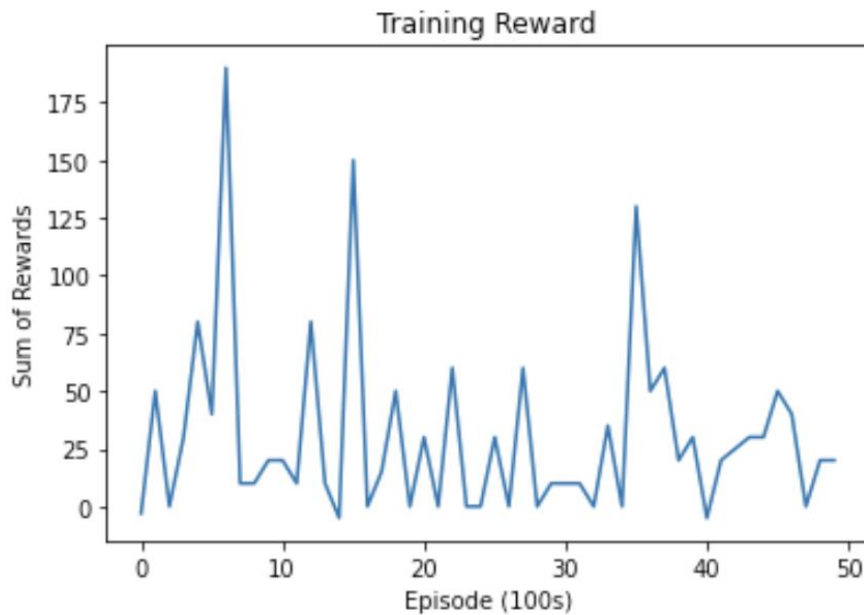
**The Test-standard-deviation is: 67.22316267478048**



3. Epsilon = 0.1

**The Test-average is: 117.5**

**The Test-standard-deviation is: 63.241521170825735**



There is a decrease in the testing average and standard deviation when the epsilon decreases. This occurs because the lower the epsilon number, the more likely the agent will perform a non-greedy action from its current condition.

#### Part 4: Experiment with negative reward for each action

Negative point reward for each action such as:

CAN = 9.5

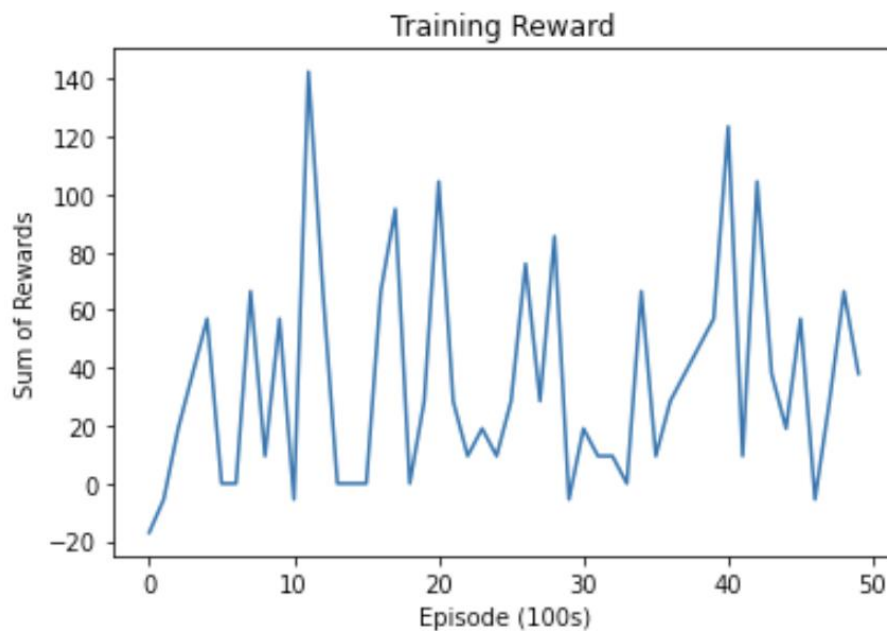
WALL = -5.5

EMPTY = -1.5

The test- average and test standard deviation is reduced when we decrease the reward.

The Test-average is: 61.95

The Test-standard-deviation is: 46.24145867076427



#### Part 5: Own experiment

Tried changing step values drastically to see how it affects the results. Decreasing the number of steps at a high rate reduces the average value of the tests by a high factor. Standard deviation decreases proportionate to the decrease in average.

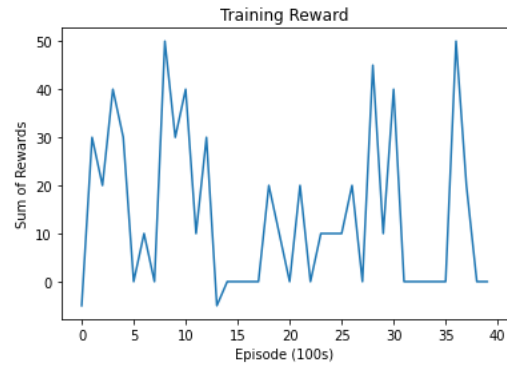
Sample 1:

N=4000

M=90

The Test-average is: 47.775

The Test-standard-deviation is: 39.32968821386714



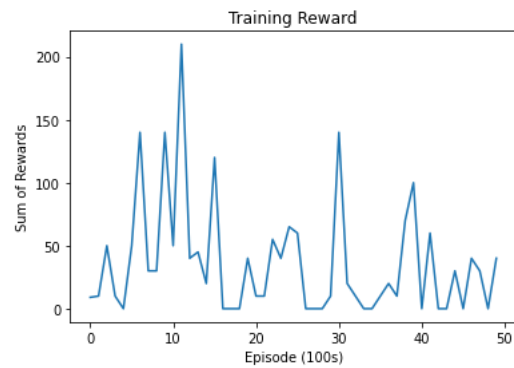
Sample 2:

N = 5000

M = 90

The Test-average is: 73.26

The Test-standard-deviation is: 44.06214248081906



Sample 3:

N=3000

M=70

The Test-average is: 17.433333333333334

The Test-standard-deviation is: 23.904927432551546



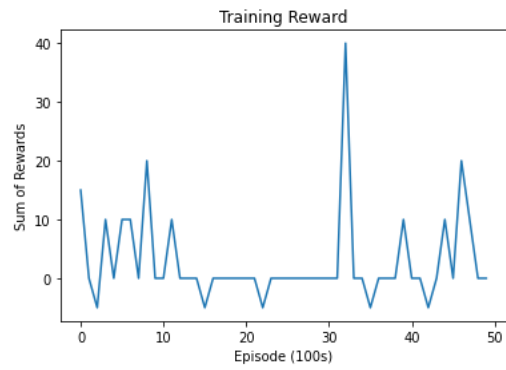
Sample 4:

N = 5000

M = 20

The Test-average is: 3.18

The Test-standard-deviation is: 8.252732880688674



Sample 5:

N=5000

M=5

The Test-average is: 0.6

The Test-standard-deviation is: 3.104834939252005

