Reinforcement Learning

Short description:

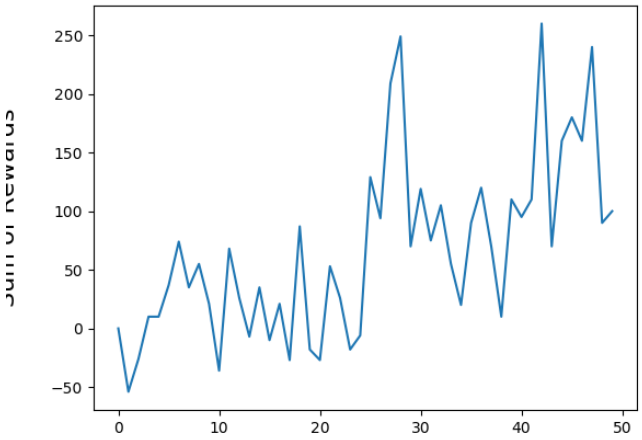
In this experiment I trained Robby the robot to pick up cans from 10x10 grid consisting walls at the border. I have used Q Learning algorithm. (Reinforcement Learning technique for that)

If Robby picks up the can he gets $10 reward, loses $1 if he tries to pick up a can from an empty slot. Otherwise there is no reward.

In all graphs, y axis represents cumulative records per episodes and x axis represents no of episodes, which is scaled down by 100. Original episode values can be obtained by multiplying them with 100.

Experiment-1

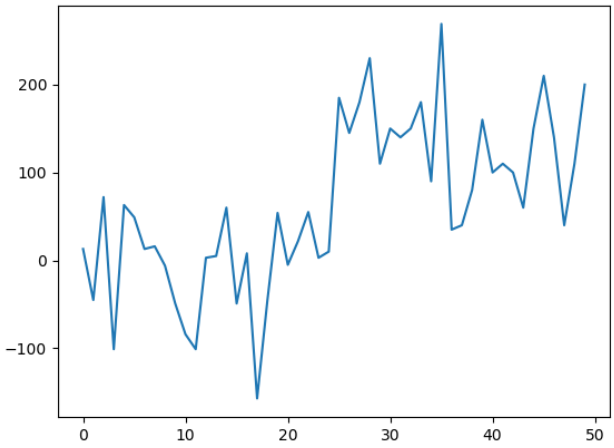
In this experiment, I am observing the cumulative reward per episode. In this experiment, learning rate is 0.2, gamma is 0.9, no of episodes are 5000 and there are 200 steps per episode. It can be observed that, cumulative reward per episode increases gradually and it increases more at around 25 that is 2500th episodes. This is because according to my code epsilon goes below 0.5 at 2500th epoch then Robby goes into exploration phase and thus we can see rewards increasing there.



Testing average: 105.6
Testing Standard Deviation: 54.374

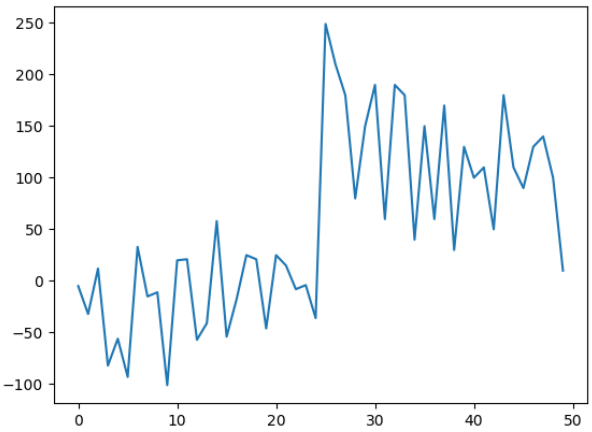
Experiment 2:

1) **Learning rate = 0.5**



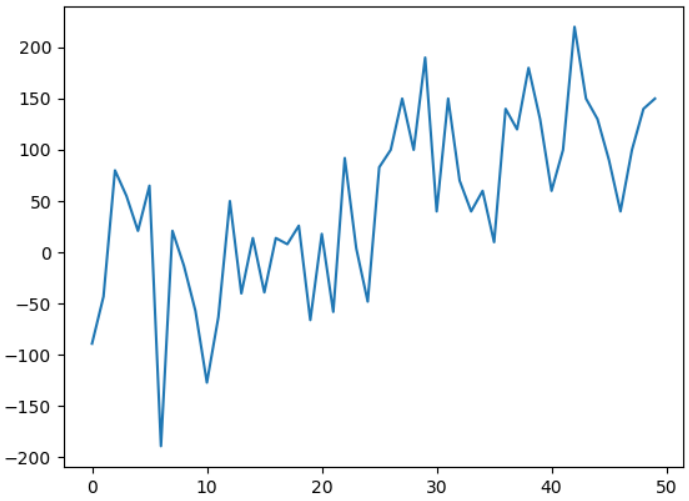
Testing Average: 110.2
Testing Standard Deviation: 59.2786

2) **Learning rate 0.9**



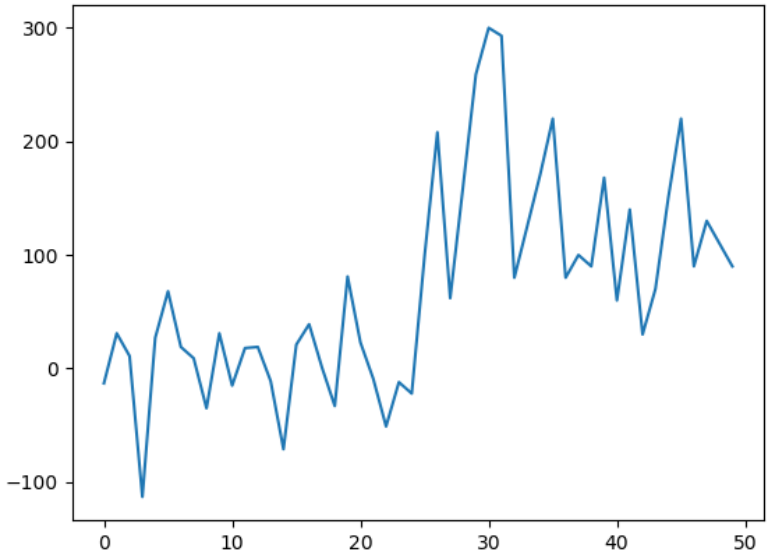
Testing Average: 110.8
Testing Standard deviation: 43.8105

3) Learning rate 0.75



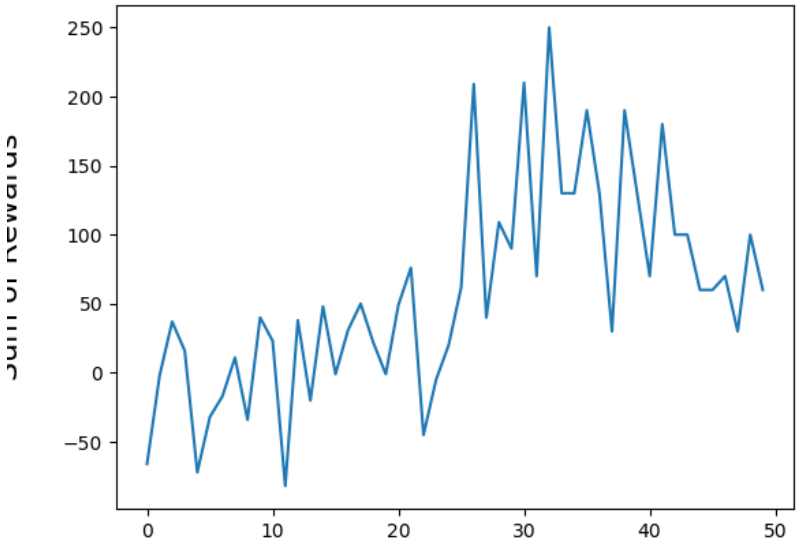
Testing Average: 102.0
Testing Standard deviation: 48.5798

4) Learning rate 0.1:



Testing average: 123.2
Testing standard deviation: 54.311

5) Learning rate 0.25:

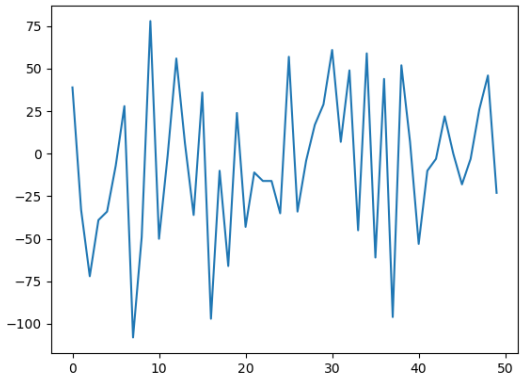


Testing average: 84.04
Testing standard deviation: 77.333

Discussion:
As I increased the learning rate testing average increases. This is because when I increase the learning rate, Robby learns much better at each step. And thus, while testing we can see that Robby gets better rewards. Moreover, as a result of "better learning", standard deviation is also decreased!

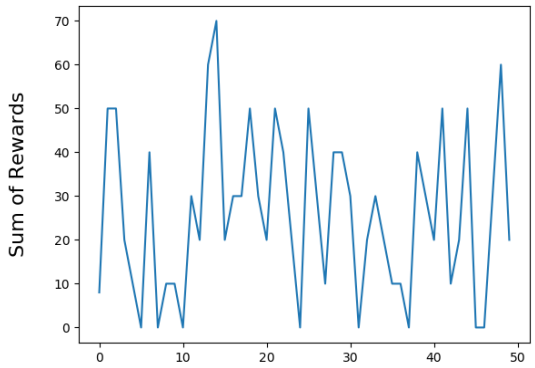
Experiment – 3

1. Epsilon = 0.75



Testing Average = -56.3
Testing Standard deviation = 425.22

2. Epsilon = 0.25



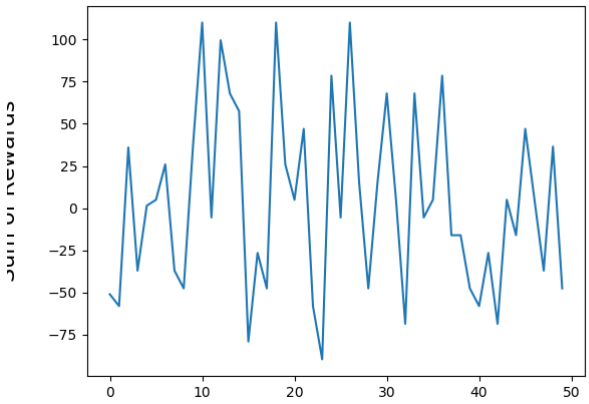
Testing Average = 21.6
Testing Standard deviation = 18.36

Discussion:

Basic concept behind decreasing epsilon is at a certain point it goes below 0.5 so that epsilon becomes less than (1 - epsilon) becomes true. Now, when this condition becomes true, according to epsilon greedy algorithm, it selects action which maximizes the reward. That is, it goes into exploitation phase from exploration phase. And therefore, decreasing epsilon is important so that, at certain point, Robby should switch into exploitation phase and start exploiting the knowledge he has gained so far.

From graphs above, we can observe that graph is neither increasing nor decreasing. This is because in first case it has only exploration phase. And thus testing average is very poor. Because, always, it selects the random action. In second graph Robby is always in the exploitation phase but, without any prior knowledge. Thus, here too, we get poor results. However, because it doesn't learn much, standard deviation is very low. Thus, we should have a fair trade off between these 2 phases by manipulating with epsilon.

Experiment – 4



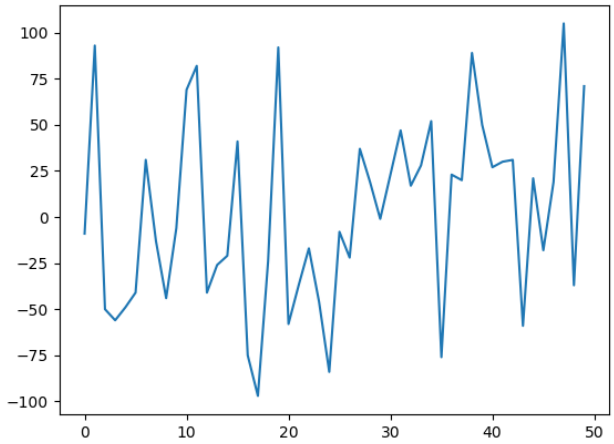
Testing Average = 21.8
Testing Standard deviation = 54.397

Discussion:

In this experiment we ask Robby to find cans with minimum steps. And therefore, compared to the first experiment, we get higher rewards at fewer no of episodes. However, while exploring the steps we lose rewards and thus, we don't get reward. Moreover, we learn better ad thus we get better standard deviation.

Experiment – 5

Gamma = 0.1



Testing Average: 24.7

Testing Standard Deviation: 405.7245

Discussion:

In this experiment I manipulated value of gamma. Value of gamma in this experiment is 0.1. Low gamma means that we don't consider what is already there in the Q Matrix. Rather we focus on the rewards received in the current state and action. And thus, Robby failed to learn much and we don't quite get increasing graph. And the result that Robby does not learn much standard deviation is too high and average is comparatively low.