Ajay Babu Gorantla

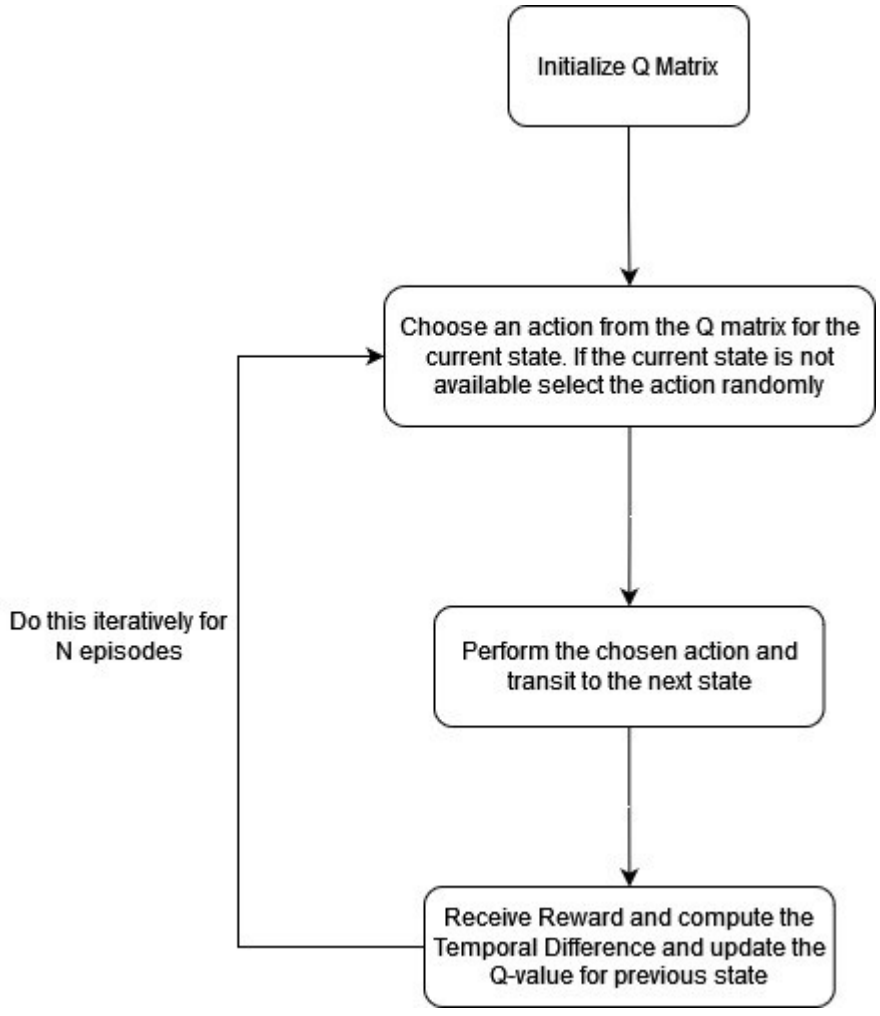
Artificial Intelligence

Programming 3

Q Learning on Robby the Robot

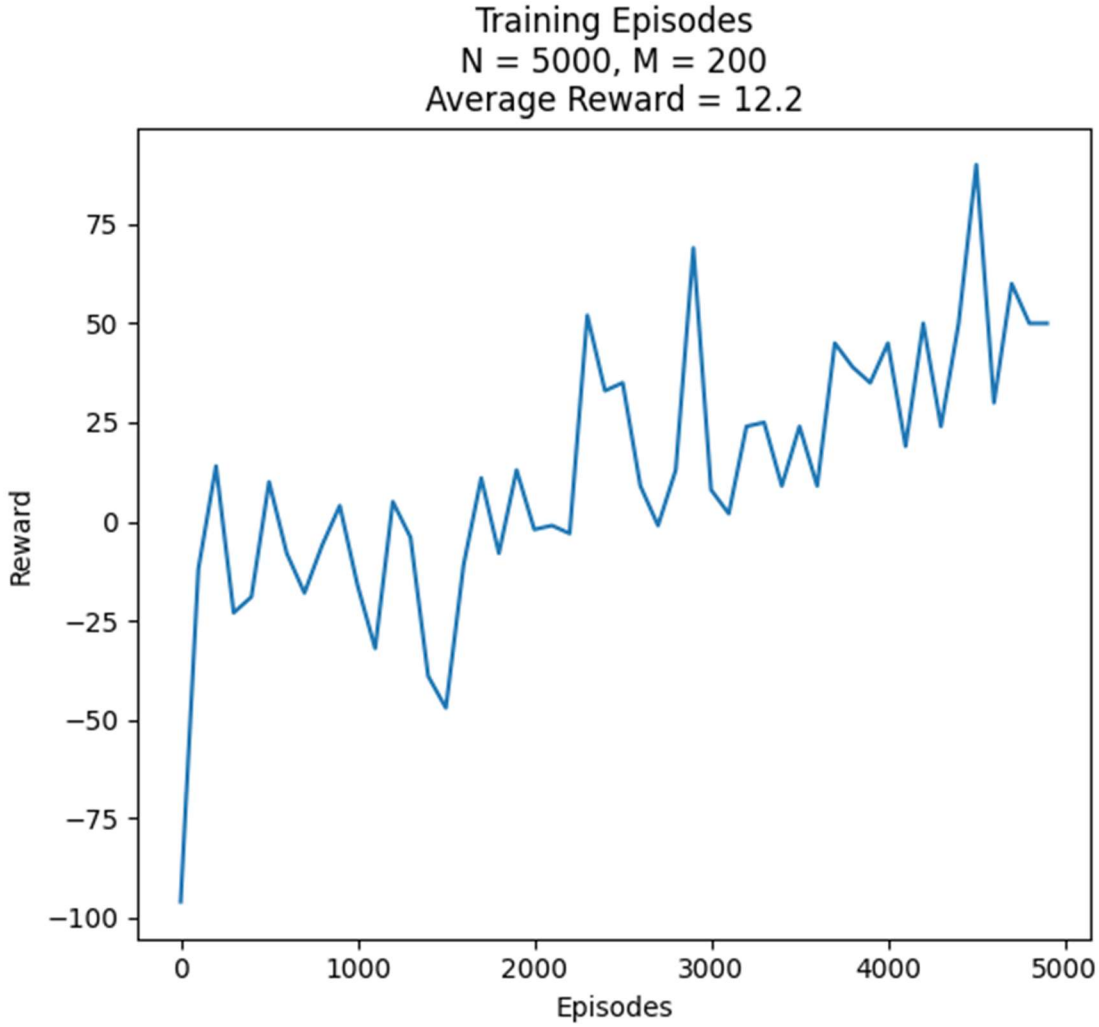
The objective of this work is to train a robot and make it learn without any prior knowledge of the environment. We use a reward mechanism where the robot is incentivized for performing correct actions and penalized for performing undesirable actions. The main objective is to maximize the total reward.

The whole program follows the following control flow graph



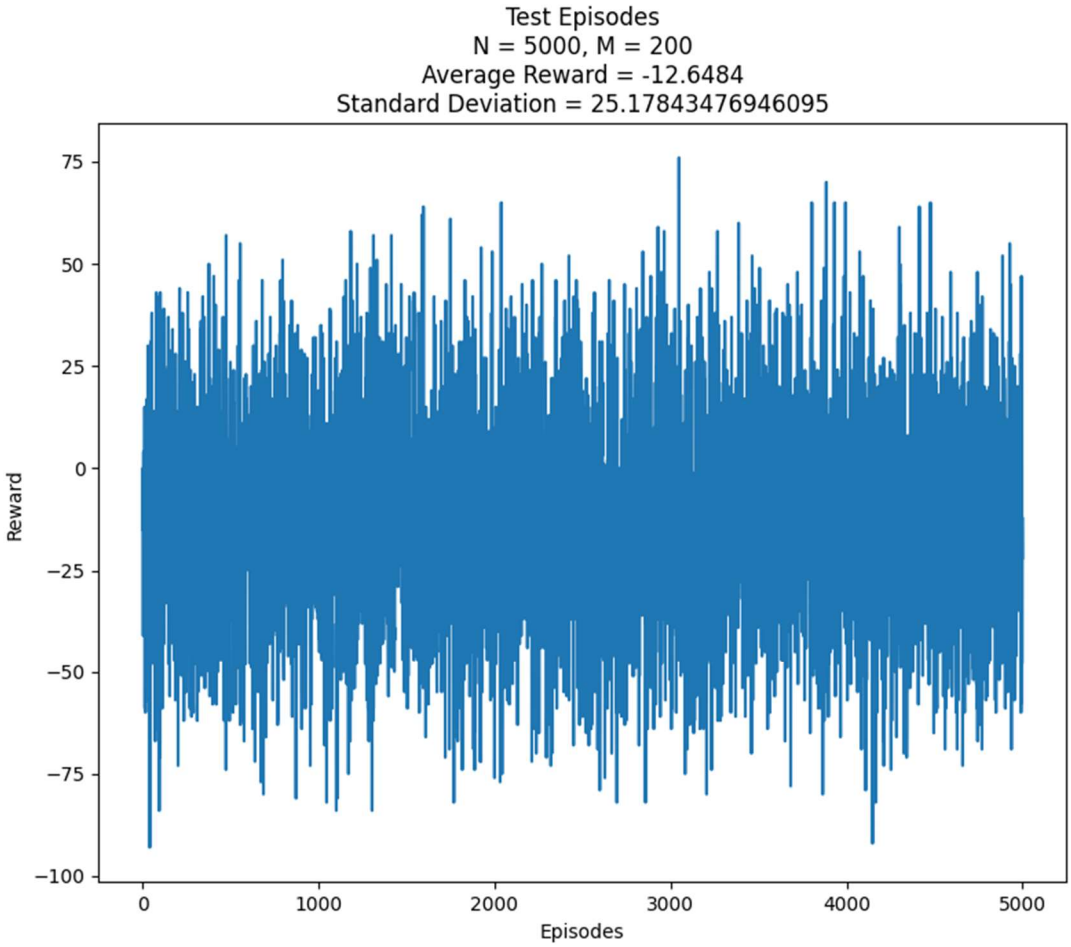
After the training for N episodes is complete, we use the Q-Matrix obtained from the training to test the robot on a new set of environments for N episodes, where each episode will have a new environment

As mentioned in the assignment, I have considered few hyperparameters and plotted the rewards over N Episodes. The below graph depicts the rewards observed at every 100^{th} episode.



Average Reward for Training is: 12.2

The below graph depicts the reward outcomes of testing over N episodes using the Q-matrix obtained from training.



The above graph has been plotted for every episode unlike the training one where the reward was plotted for every 100th episode.

The Average Reward for Testing is: -12.6484

Standard Deviation: 25.18

I have experimented for different values of hyperparameters like N, M, epsilon value, learning rate and gamma value.

One observation was there should be enough number of episodes on which the robot should be trained for, else there are chances that the robot might not explore all the possible states. In that case, when testing is done, the robot might encounter a state which it has not explored and will not know what action to perform and will get stuck.

So, it is important to train the robot for many number of episodes (also important to balance the tradeoff between exploration and exploitation).