Assignment 5 - Artificial Intelligence

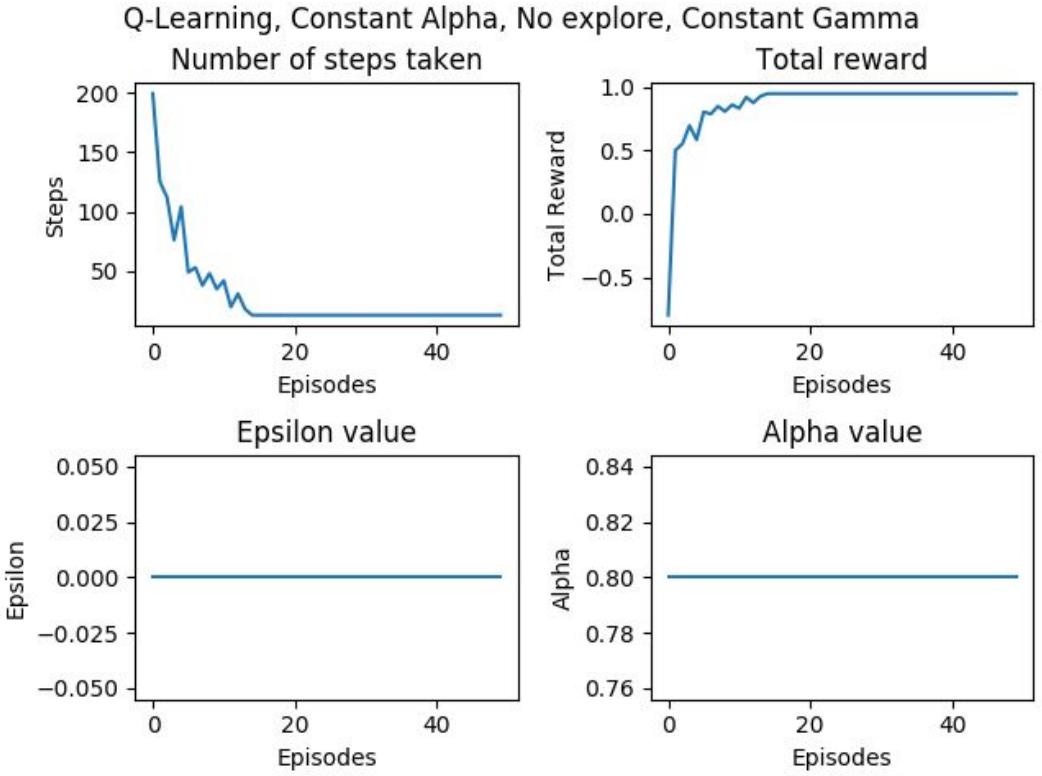
Name - Viresh Gupta

Roll Number - 2016118

Environment Used - Maze-sample-5x5-v0

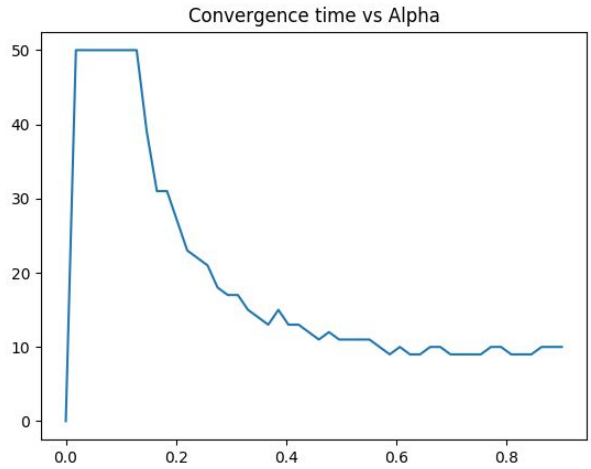
Part 3:

a) Simple learning curve for a Q-Learning Agent implementation



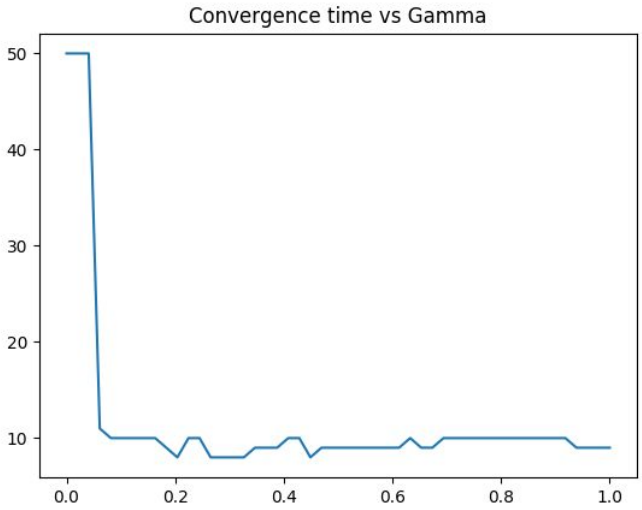
b) Impact of various hyperparameters

Effect of Changing Alpha (versus Convergence time):



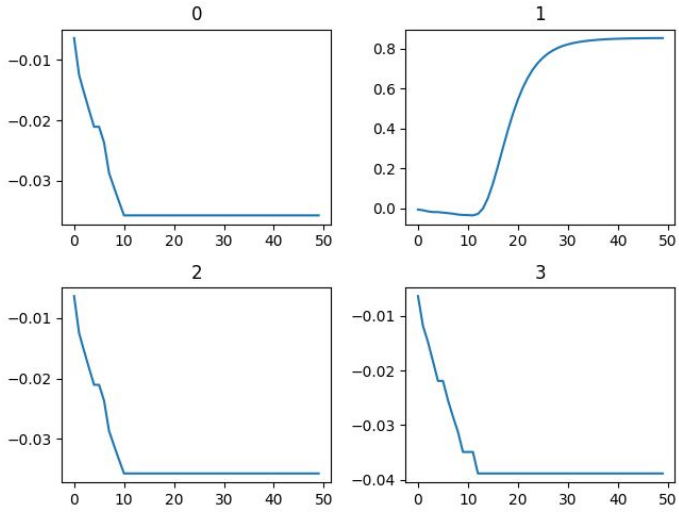
Best alpha = 0.8

Effect of Changing Gamma:



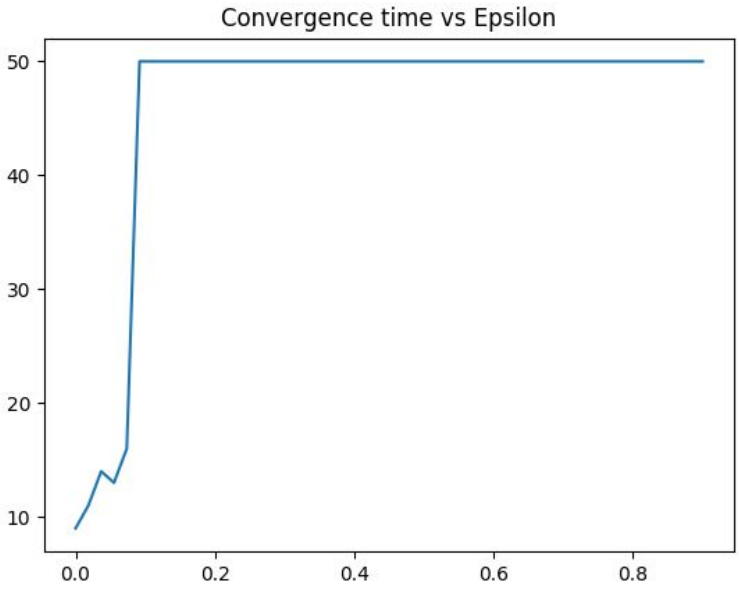
Best gama = 0.99

c) Study probability change of a given state
For Initial State (0,0), with time, Q-table values tend to a particular move.



The probabilities are directly proportional to Q-table values, hence above is equivalent to plotting probabilities

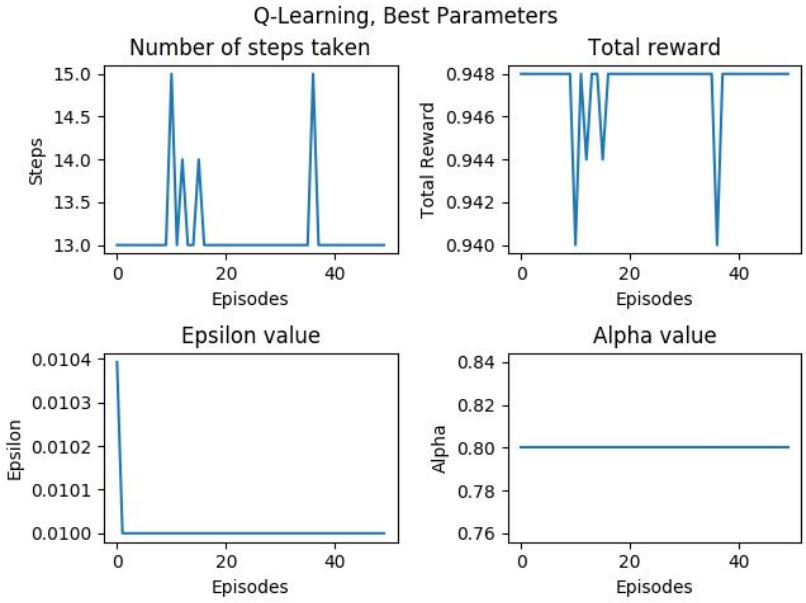
d) Implement epsilon greedy



Best Epsilon - 0.03
On increasing epsilon values, the agent's takes too long to converge (so we either need a decay factor, or we need a very small explore rate)

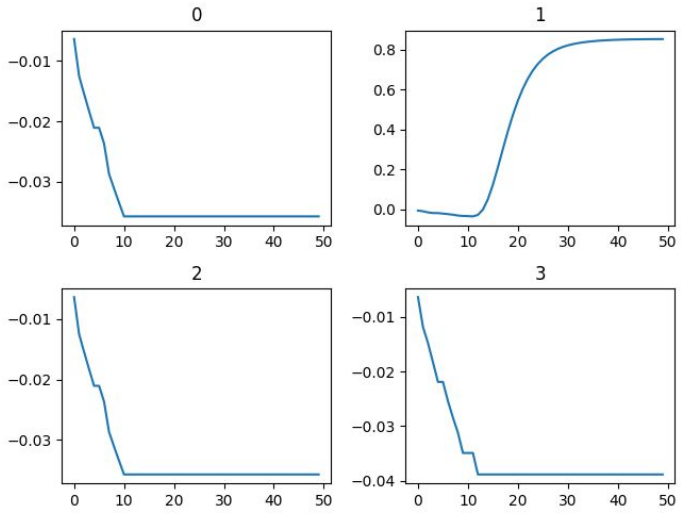
e) Use best Parameter Agent for comparing state probability
Alpha = 0.8
Gamma = 0.99

Epsilon = 0.03

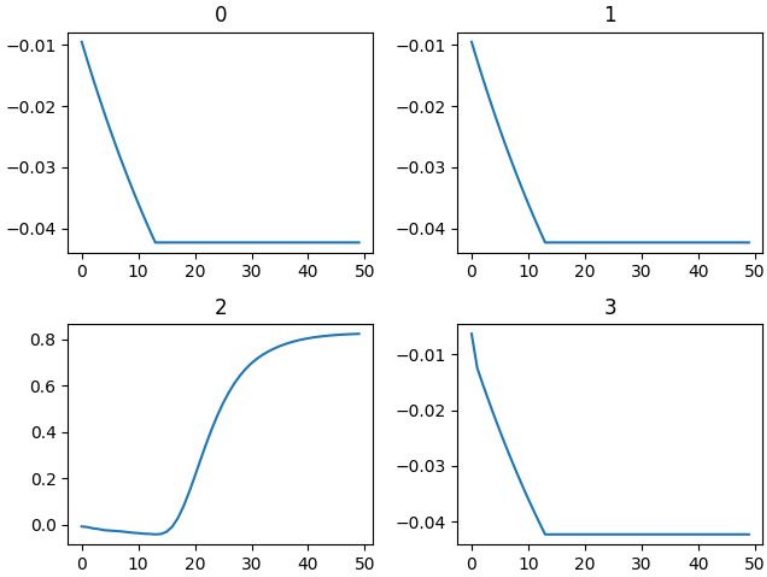


Two intermediate states :

0,0 -



0,2 -



Question 4:

The trained agent doesn't work for Another maze environment, because it has learnt values specific to the training environment. This is because the states are a function of location. This can be averted by either - Retraining on the new environment or by Making states such that they are independent of location (and are rather things like near two walls, near three walls etc etc). Also we can add history information to avoid this.

Bonus:

Figure for SARSA algorithm: