AI, Assignment 4
Katari Saketh
2015045

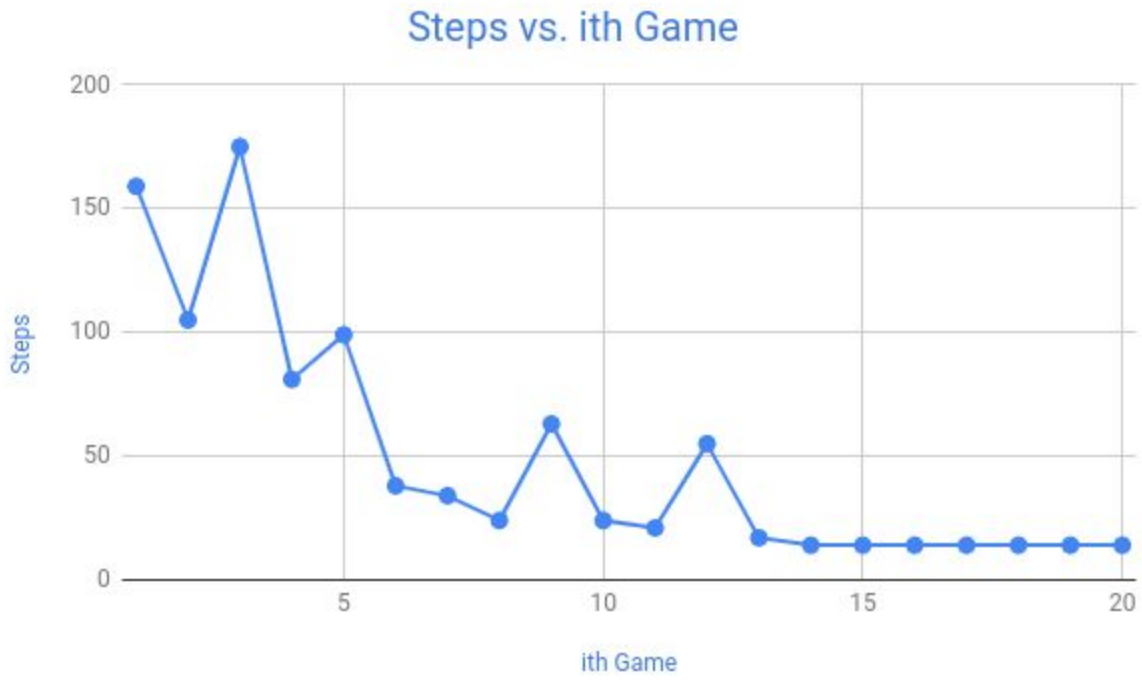
Default Parameters:

- Gamma: 0.3
- Alpha: 0
- Learning Rate: 0.3
- Games: 1

3.a. Performance of Q-Learning:

Manhattan distance is calculated at each step (1 to 159) with default parameters. As we progress towards the goal state, the manhattan distance got converged to 0, as shown below (raw data: link) -

The same Q-table is used for 20 games, and the no. of steps taken in each game got reduced till 14th game, after which, it remained constant, as shown below -

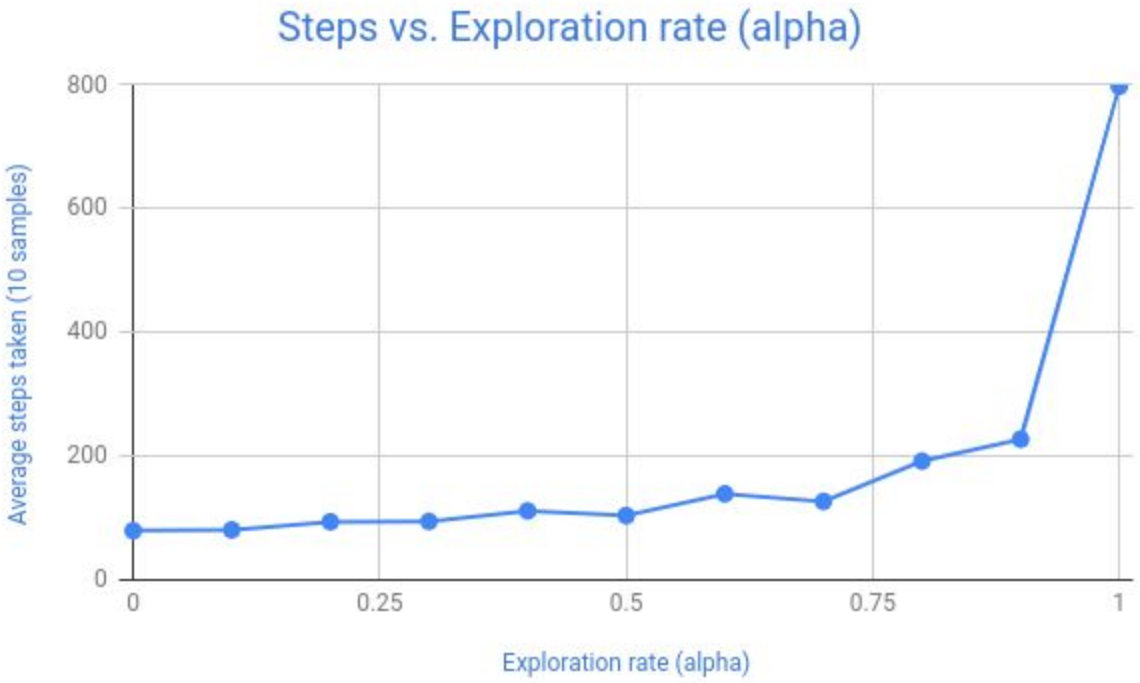


3.b. Hyper-parameters:

- Gamma a.k.a. discount factor: There's no change in the behaviour of Q-learning in reaching the goal state of 5x5 maze i.e it takes same 159 steps with any value of gamma. *Reason:* $new_q = reward + (\gamma * max_q)$. In our 5x5 maze, reward is always negative, except for the goal state; max_q is always $max(0, negative_q)$, which is 0, except for the goal state. As max_q is always 0, there's no effect of γ ($0 * \gamma$ *is always* 0).
- Learning Rate: It has no impact on our 5x5 maze. Reason: On top of $\gamma's$ reasoning, $reward + (\gamma * max_q)$ is always negative due to negative rewards, except for goal state. Learning rate is always positive. $learning\ rate * negative\ value = 0$, irrespective of the magnitude of learning rate.

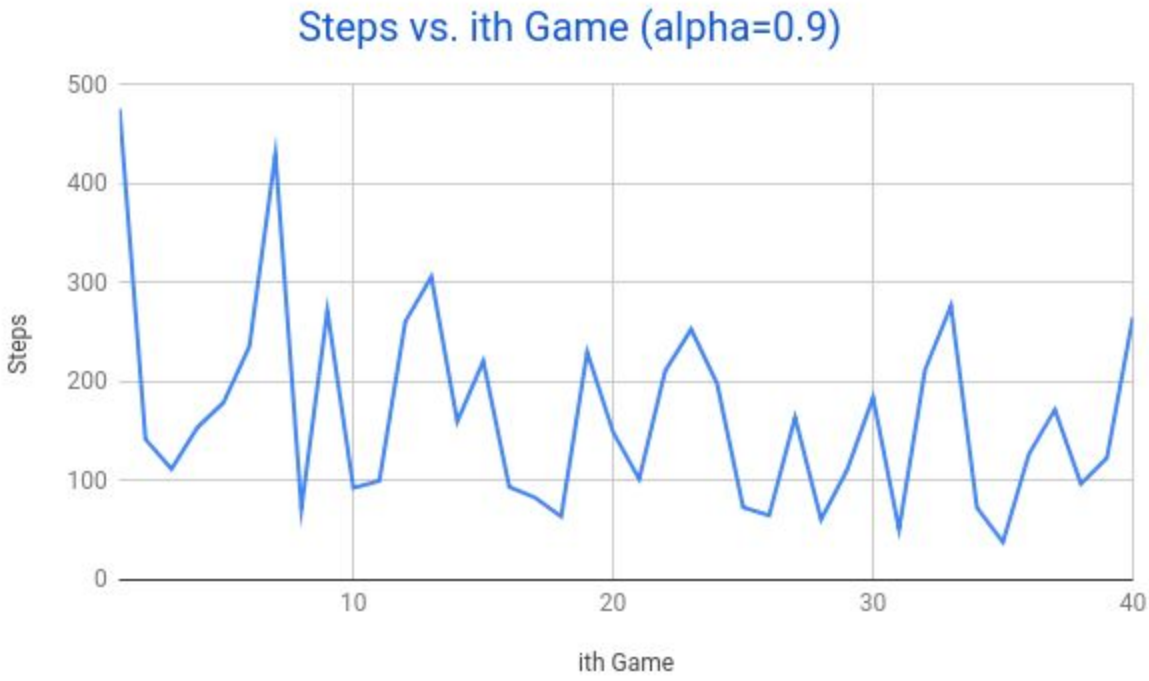
3.d. Random Exploration (α):

The exploration rate, α, is varied from 0 to 1 at an interval of 0.1 (0, 0.1, 0.2, 0.3…0.8, 0.9, 1), taking the average of 10 games (continuous i.e same q-table) on a single exploration rate. The ones with less exploration rate took fewer steps to reach goal state, as shown below (raw data: link) -



With alpha=0.1, it took almost 16 games to converge to 14 steps, to reach goal state, as shown below -

With alpha=0.9, it didn't converge for even 40 games, as shown below -



Lesser the alpha, better the model.

3.e Best params, 2 states:
alpha=0. State values of a particular state can be analysed by printing the data on the terminal (during demo).

4. After reaching a threshold say, 14 steps to reach the goal state, it cannot be improved further i.e saturation point or the shortest path