CPS841 Assignment 3 Report

First Visit = 10 as G_{10}= 10 when gamma = 1.

Every Visit Average = 5.5

$G_n = Qn + 1/n (R_t + Qn)$

$G_0 = 0$

$G_1 = 10+1/1 * (10-0)$

$G_2 = 9.5+1/2 * (9-10)$

$G_3 = 9+1/3 * (8-9.5)$

$G_4 = 8.5+1/4 * (7-9)$

$G_5 = 8+1/5 * (6-8.5)$

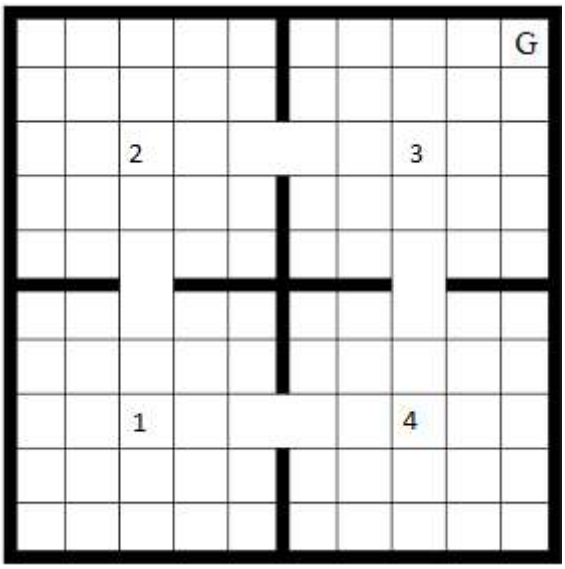
$G_6 = 7.5+1/6 * (5-8)$

$G_7 = 7+1/7 * (4-7.5)$

$G_8 = 6.5+1/8 * (3-7)$

$G_9 = 6+1/9 * (2-6.5)$

$G_{10} = 5.5$



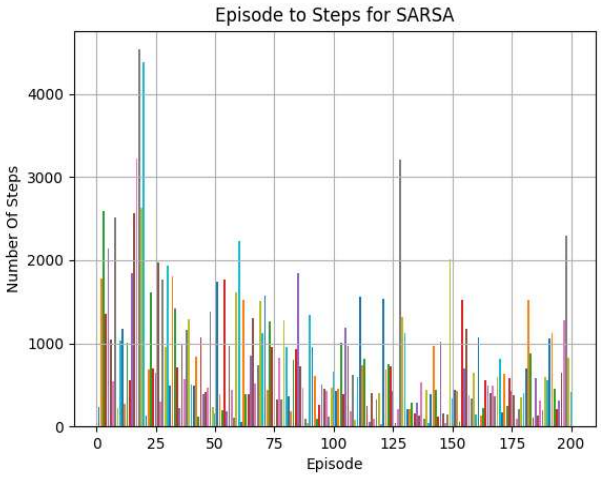
Grid Order as pictured

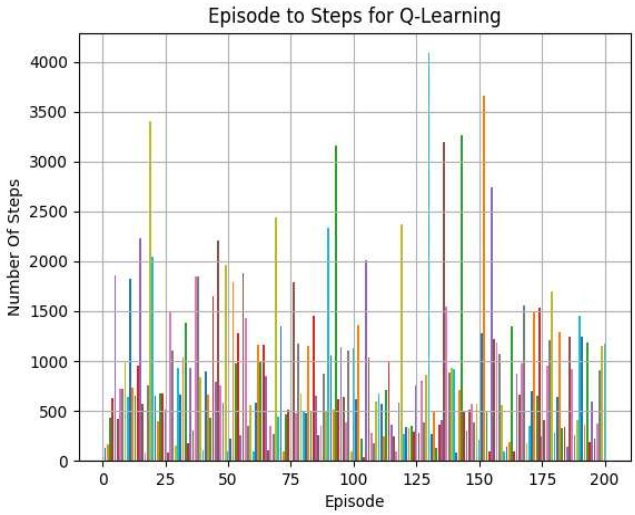
We found that Monte Carlo would, once it found a correct path, have it's values converge quickly and find a path that worked best for it. There were some scenarios that would cause infinite loops, such as when it got stuck in a corner and due to a maximum number for steps would have it's episode terminated. Now that whole corners values would be incredibly negative and any future episodes that involved that area could be forced into actions that prevented from leaving.

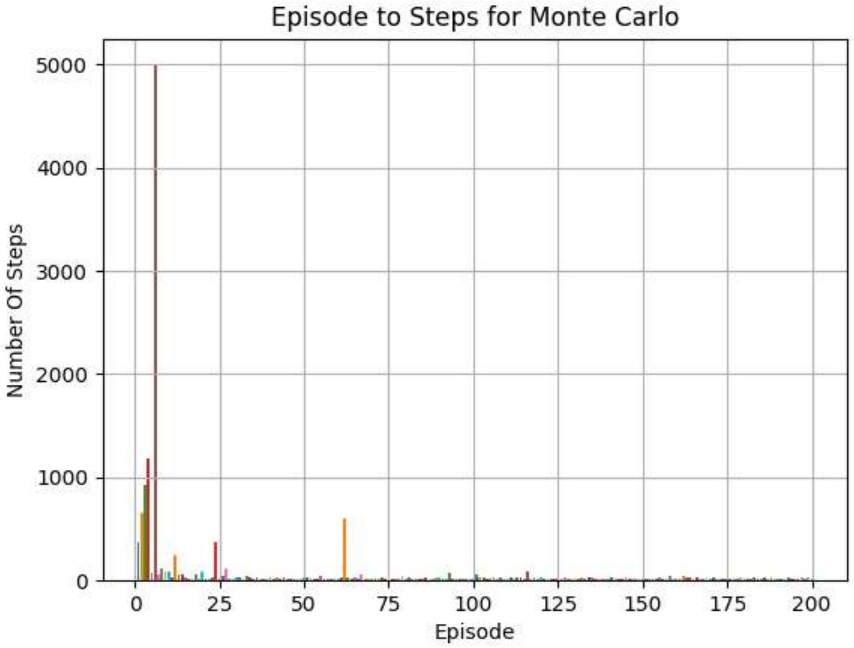
SARSA and Q-learning are both very good at getting to the goal state and given enough steps (3000 - 4000) they have almost a 100% success rate of getting to the Goal State. The issue is that these algorithms have issues creating an optimal policy that lowers the number of steps. Instead they seem to always take larger number of steps, and changing the alpha and epsilon can help increase the likelihood of reaching the Goal State.

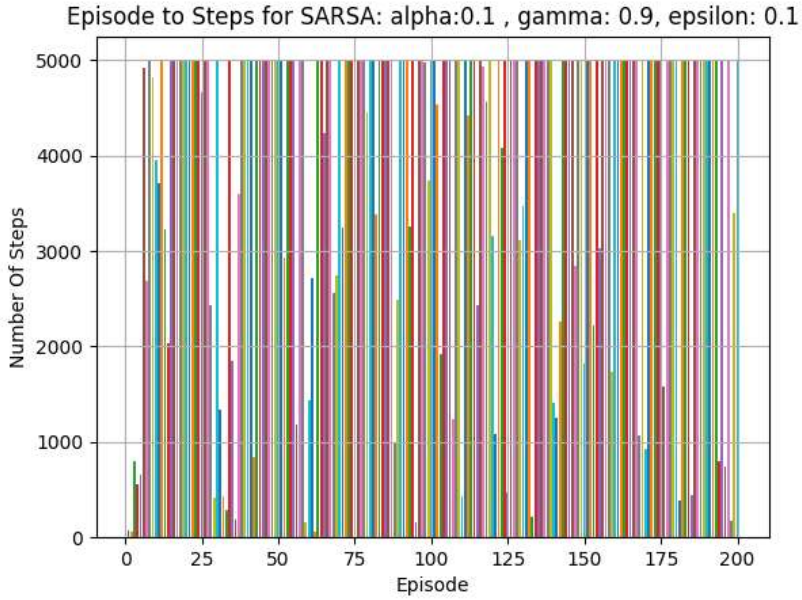
Please find below a large amount of graphs we have plotted over many runs of our algorithms that help agree with the above observations.
Ones with “Drop” in the title are ones where the epsilon was lowered to 0 over the run

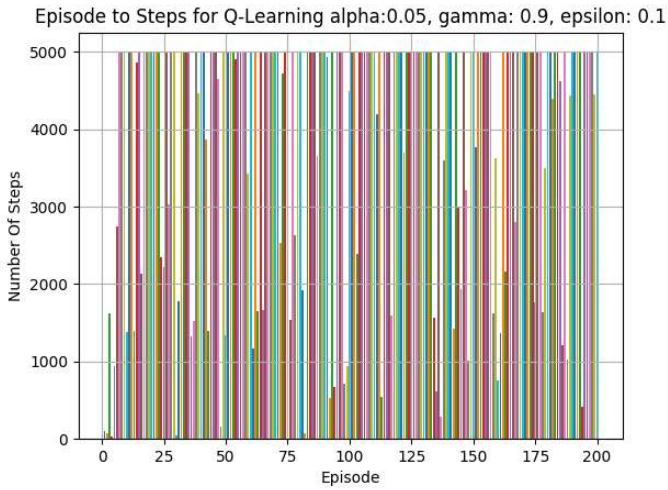
Agent starting {X:2,Y:2,Grid:1}

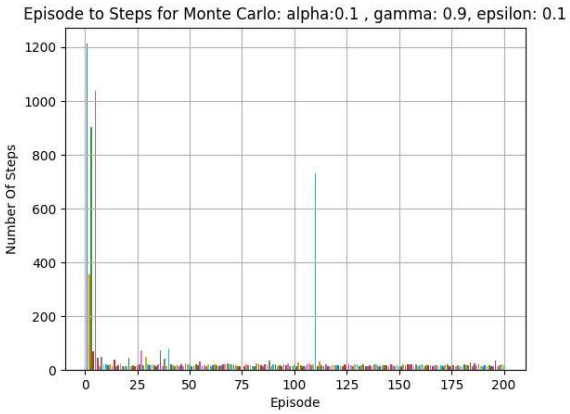


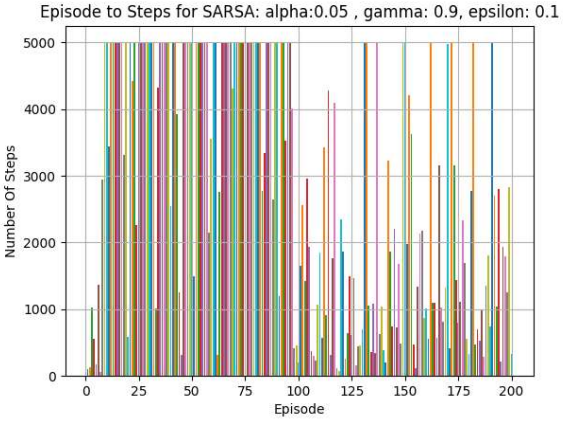


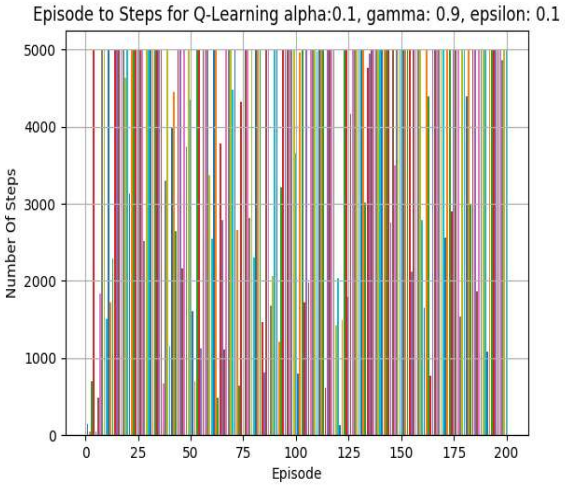


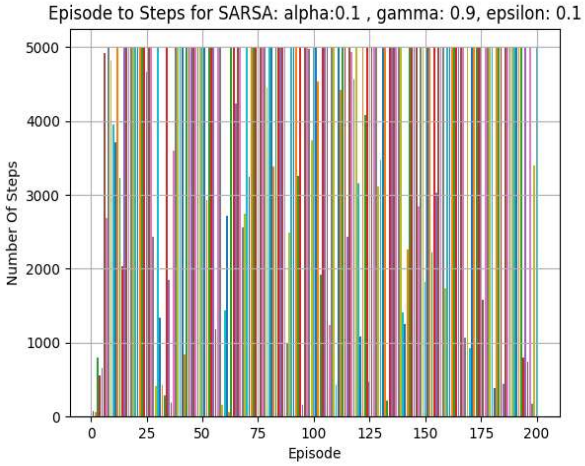

Episode to Steps for SARSA: alpha:0.1 , gamma: 0.9, epsilon: 0.1
Number Of Steps
Episode
0
1000
2000
3000
4000
5000
0
25
50
75
100
125
150
175
200

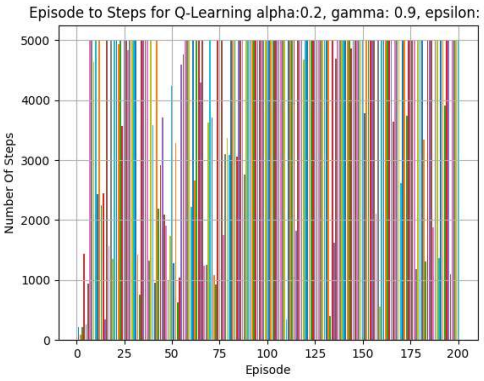

Episode to Steps for Q-Learning alpha:0.05, gamma: 0.9, epsilon: 0.1
Number Of Steps
Episode
0
1000
2000
3000
4000
5000
0
25
50
75
100
125
150
175
200

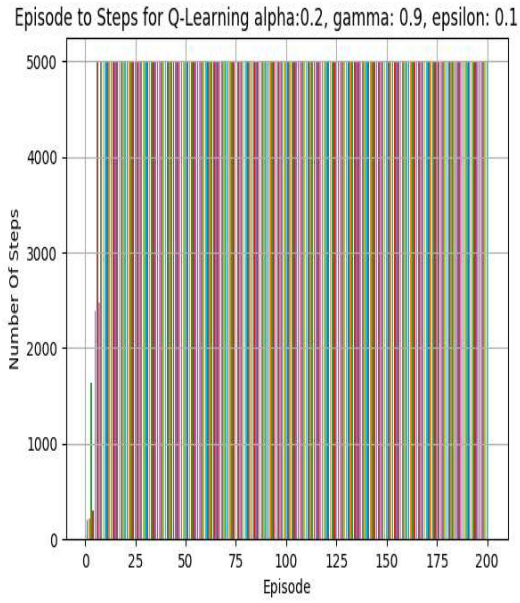

Episode to Steps for Monte Carlo: alpha:0.1 , gamma: 0.9, epsilon: 0.1
Number Of Steps
Episode
0
200
400
600
800
1000
1200
0
25
50
75
100
125
150
175
200

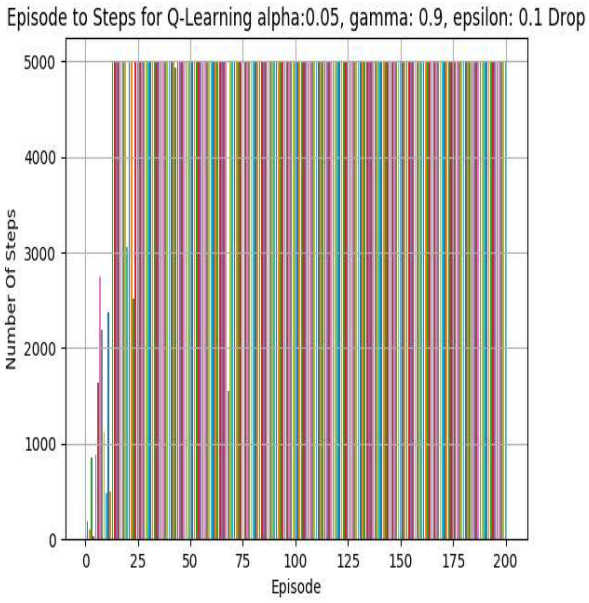

Episode to Steps for SARSA: alpha:0.05 , gamma: 0.9, epsilon: 0.1
Number Of Steps
Episode
0
1000
2000
3000
4000
5000
0
25
50
75
100
125
150
175
200

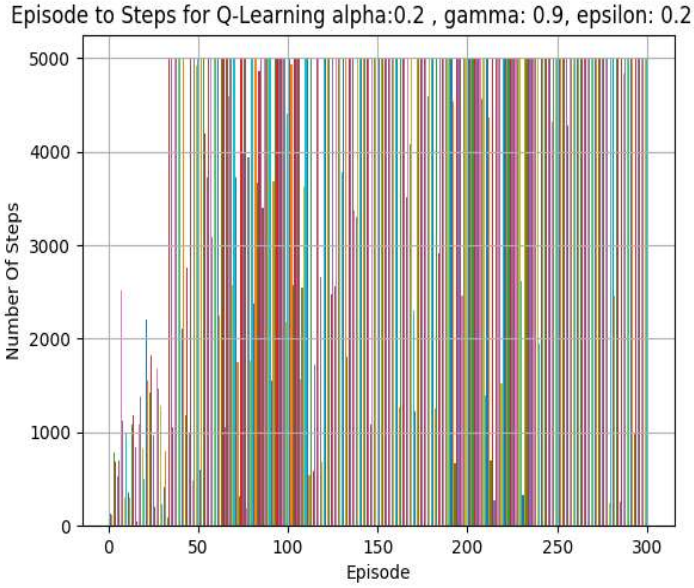

Episode to Steps for Q-Learning alpha:0.1, gamma: 0.9, epsilon: 0.1
Number Of Steps
Episode
0
1000
2000
3000
4000
5000
0
25
50
75
100
125
150
175
200

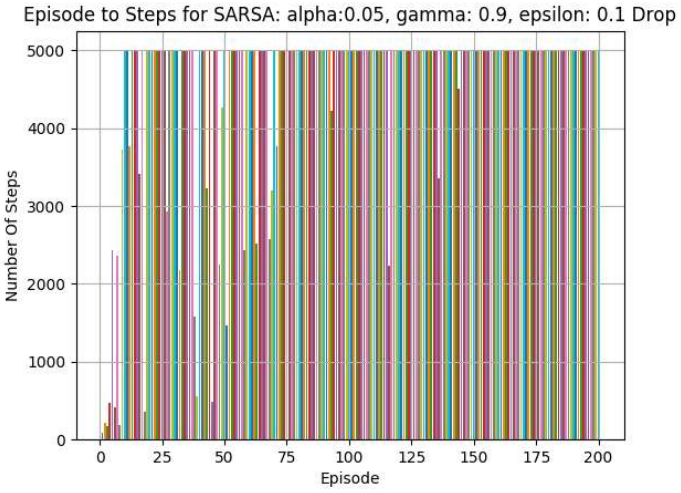

Episode to Steps for SARSA: alpha:0.1 , gamma: 0.9, epsilon: 0.1
Number Of Steps
Episode
0
1000
2000
3000
4000
5000
0
25
50
75
100
125
150
175
200

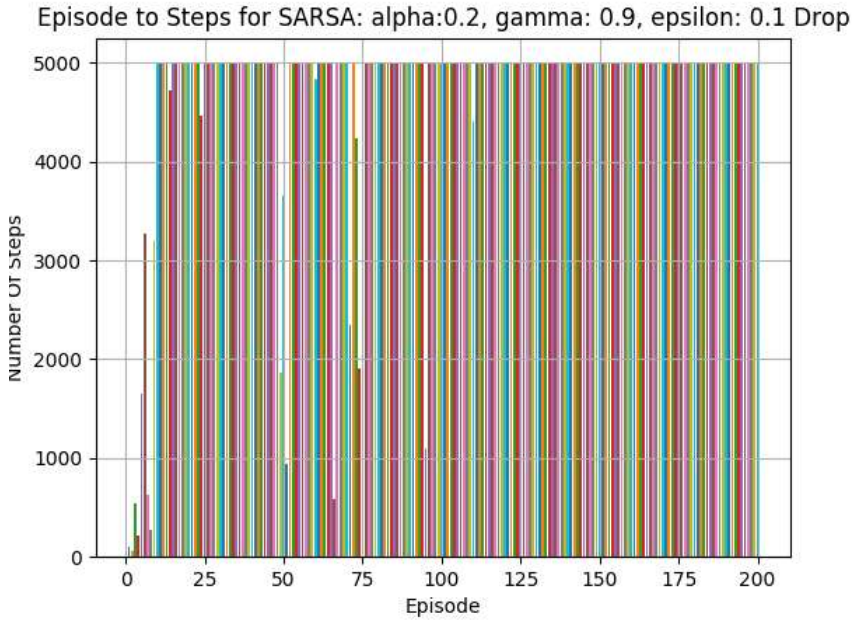

Episode to Steps for Q-Learning alpha:0.2, gamma: 0.9, epsilon: 0.1
Number Of Steps
Episode
0
1000
2000
3000
4000
5000
0
25
50
75
100
125
150
175
200

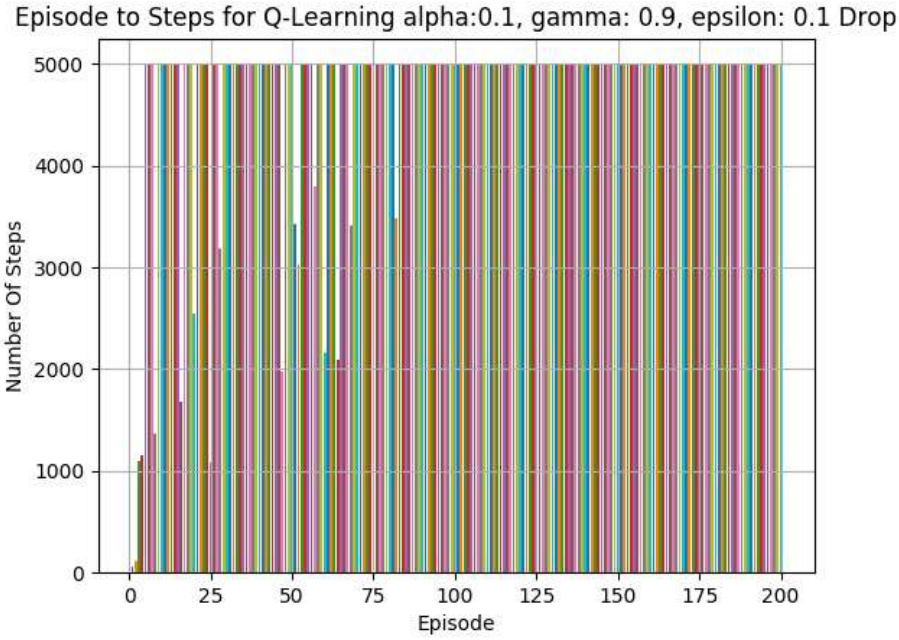

Episode to Steps for Q-Learning alpha:0.2, gamma: 0.9, epsilon: 0.1 Drop
Number Of Steps
Episode
0
1000
2000
3000
4000
5000
0
25
50
75
100
125
150
175
200


Episode to Steps for Q-Learning alpha:0.05, gamma: 0.9, epsilon: 0.1 Drop
Number Of Steps
Episode
0
1000
2000
3000
4000
5000
0
25
50
75
100
125
150
175
200


Episode to Steps for Q-Learning alpha:0.2 , gamma: 0.9, epsilon: 0.2 Drop
Number Of Steps
Episode
0
1000
2000
3000
4000
5000
0
50
100
150
200
250
300


Episode to Steps for SARSA: alpha:0.05, gamma: 0.9, epsilon: 0.1 Drop
Number Of Steps
Episode
0
1000
2000
3000
4000
5000
0
25
50
75
100
125
150
175
200


Episode to Steps for SARSA: alpha:0.2, gamma: 0.9, epsilon: 0.1 Drop
Number Of Steps
Episode
0
1000
2000
3000
4000
5000
0
25
50
75
100
125
150
175
200


Episode to Steps for Q-Learning alpha:0.1, gamma: 0.9, epsilon: 0.1 Drop
Number Of Steps
Episode
0
1000
2000
3000
4000
5000
0
25
50
75
100
125
150
175
200