CS747 : FOUNDATIONS OF INTELLIGENT AND LEARNING AGENTS.

On Policy TD control on Windy Gridworld.

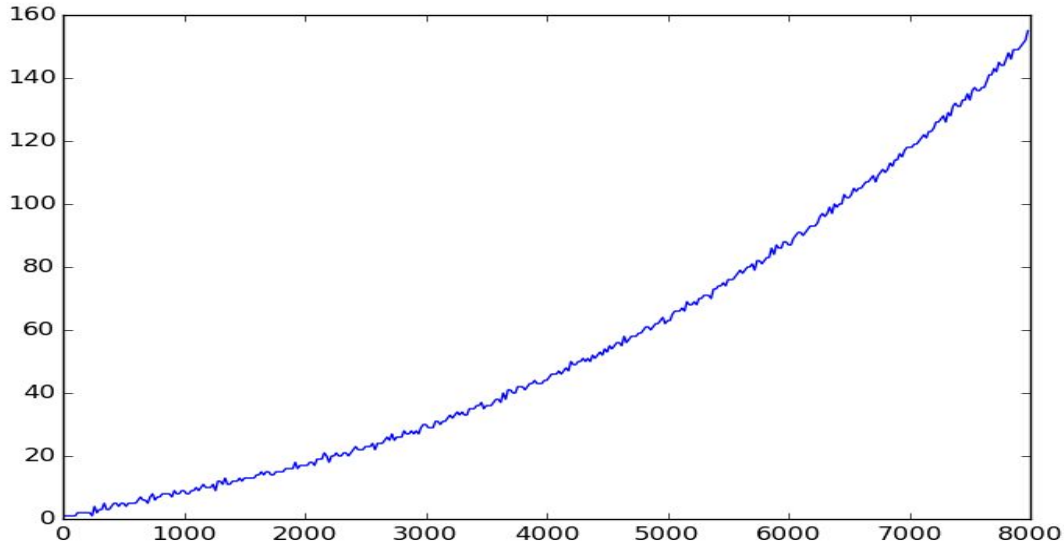
Observations

Simple Windy Gridworld

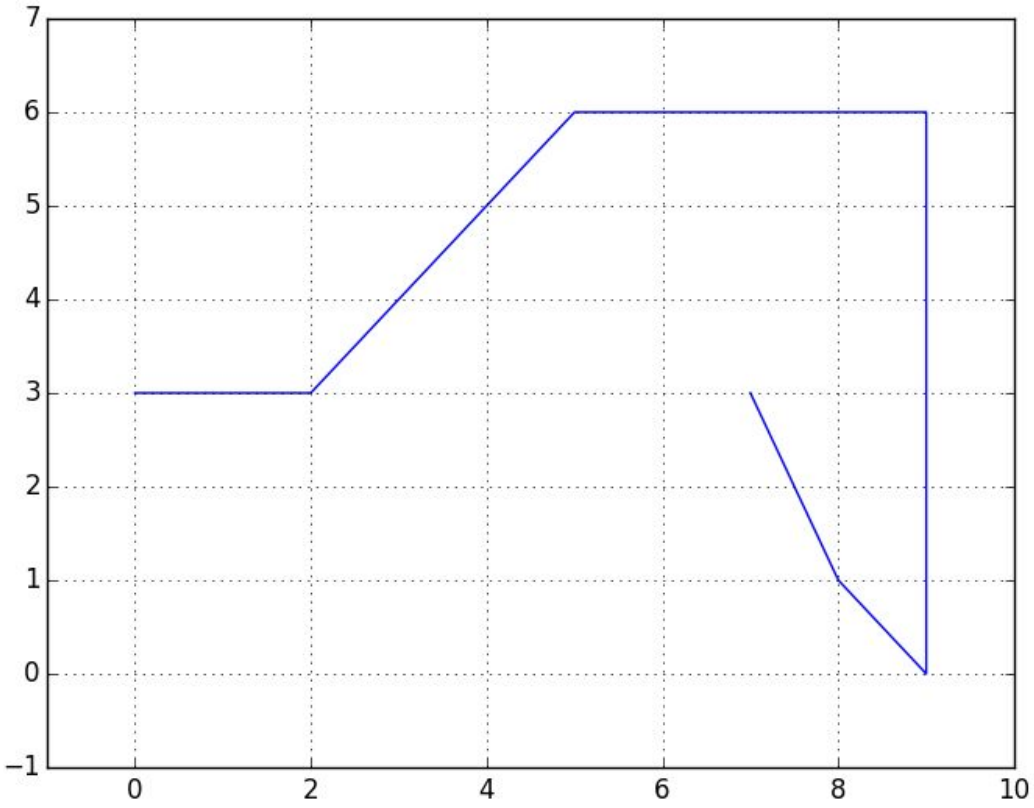
Sarsa On-Policy TD control converges fast but not as fast as it has been demonstrated in book. I have plotted the graph above using scale of 20 timesteps starting from 1.

Since the number of actions are less, it will take longer time to reach the goal state as compared to King's moves. But exploration is happening faster since available actions are less hence it will converge faster as you can see in the graph below.

Average episode length upon convergence with 8000 timesteps is 17.
Once the Q-values has been converged, greedy policy plot is shown below.



Windy GridWorld Simple timesteps(X axis) vs episodes(Y axis).


7
6
5
4
3
2
1
0
−1
0
2
4
6
8
10

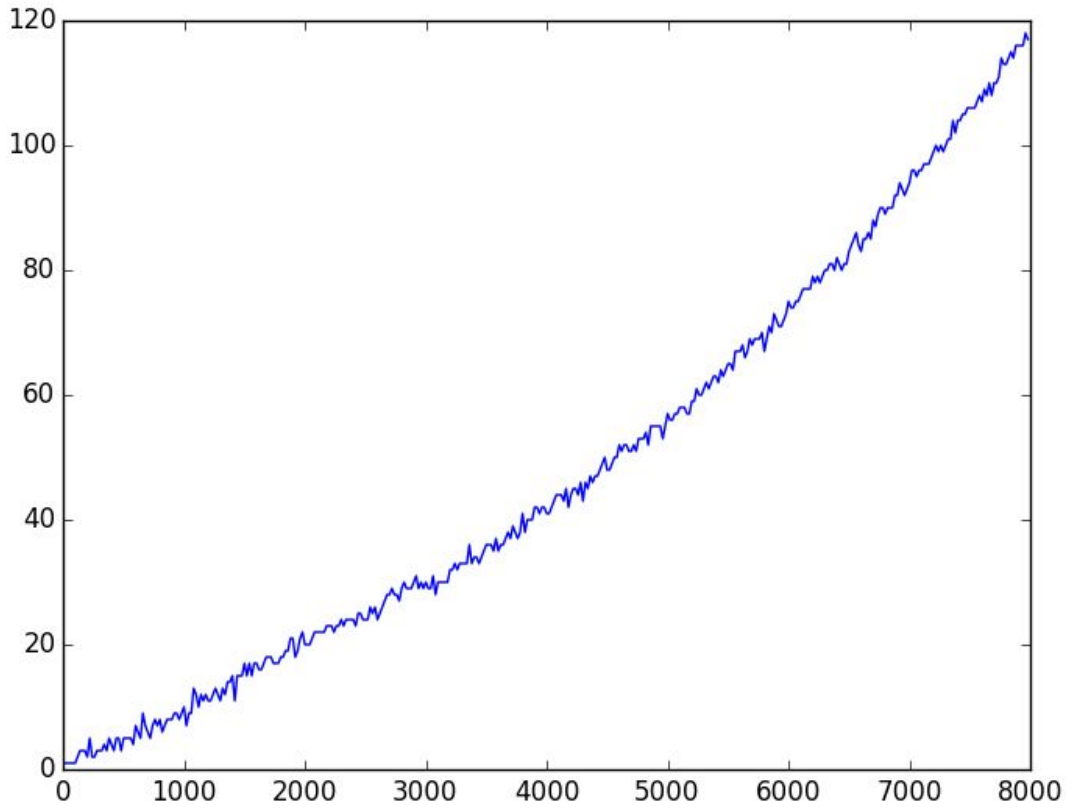
Windy Gridworld with King's Moves.

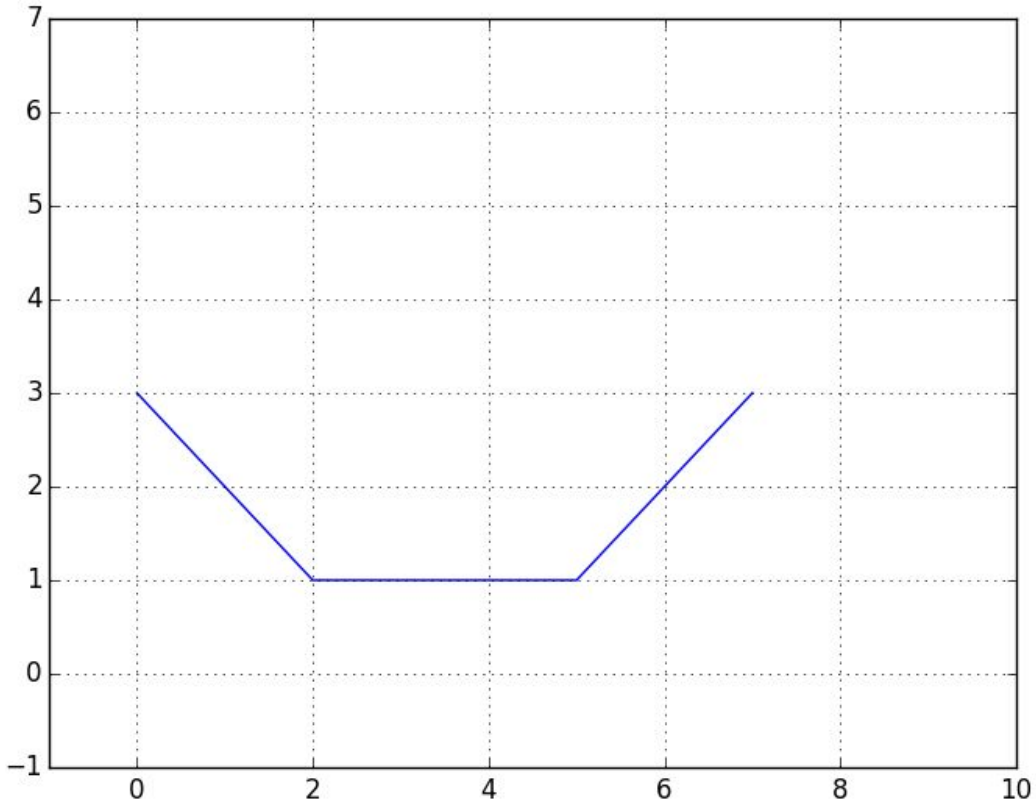
It takes less number of steps as compared to simple windy gridworld. Since the agent have extra moves now. It can move diagonally. But availability of extra moves reduces convergence rate which is clearly visible in the timestep vs episode graph below

Though adding a neutral move does not help much except in some special situations. Overall i did not observe significant change in episode length due to neutral move.

Average episode length upon convergence with 8000 timesteps is approximately 7.

Once the Q-values has been converged, greedy policy plot is shown above.

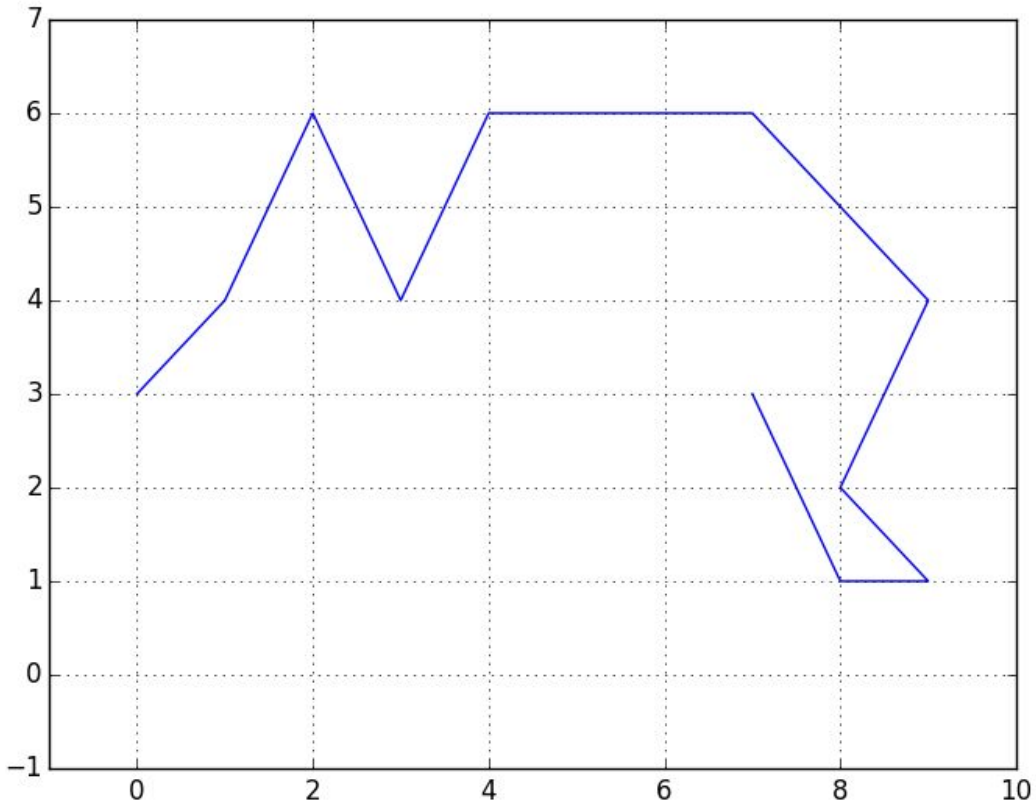


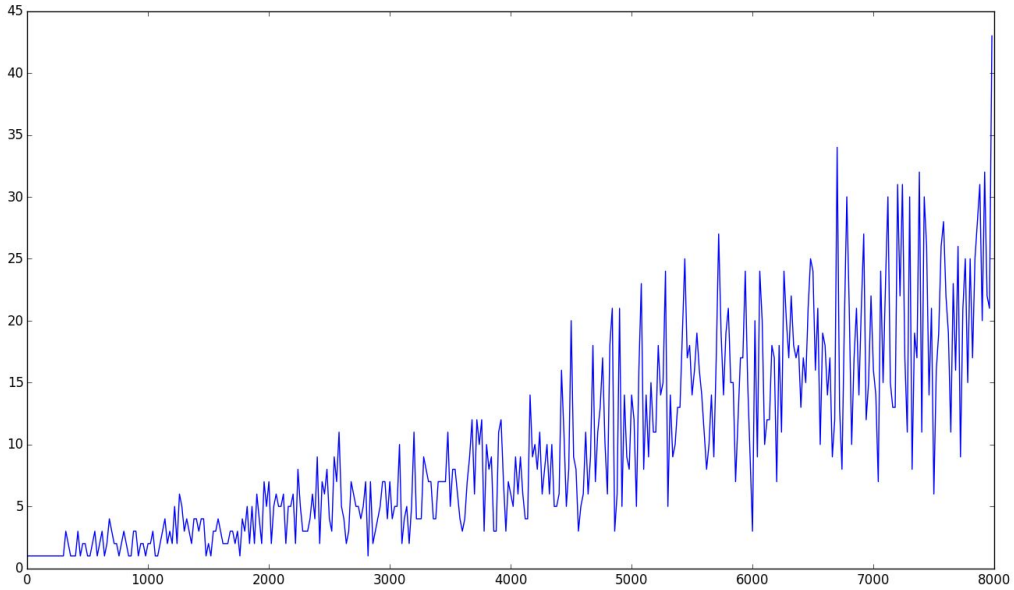

7
6
5
4
3
2
1
0
−1
0
2
4
6
8
10

Windy Gridworld with King's Moves and Stochastic wind.

Stochastic wind causes larger episode length. Empirically speaking It is performing better than simple agent which can make only 4 moves. But agent is able to still learn the behavior. Performance is unpredictable and any comment on a different configuration can be made empirically. Though there is certain relationship among number of actions and performance. We need more timesteps to learn random behavior.

Once the Q-values has been converged, greedy policy plot is shown above.




0
5
10
15
20
25
30
35
40
45
0
1000
2000
3000
4000
5000
6000
7000
8000