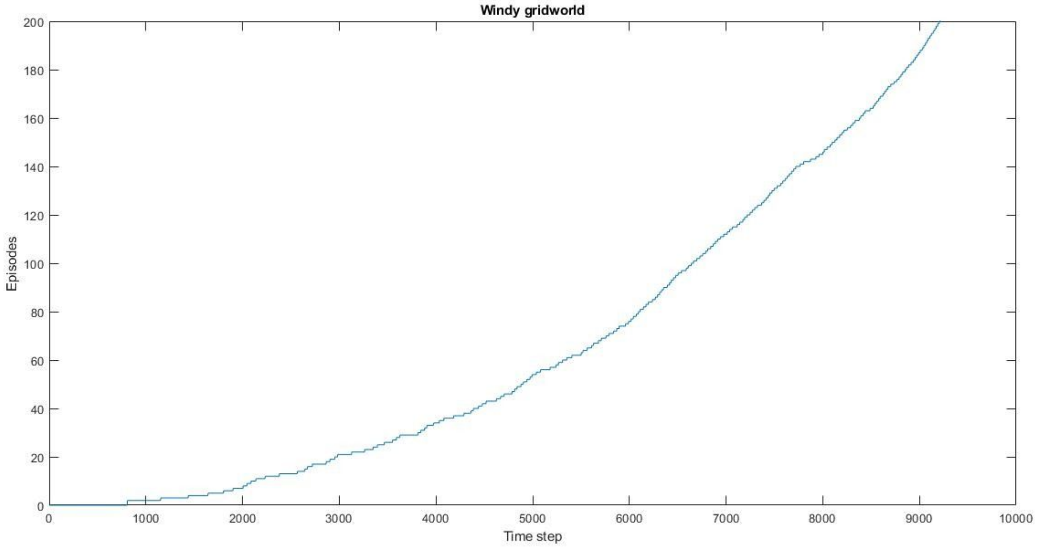
Reinforcement Learning - Homework 5
Vishal I B, 917809310

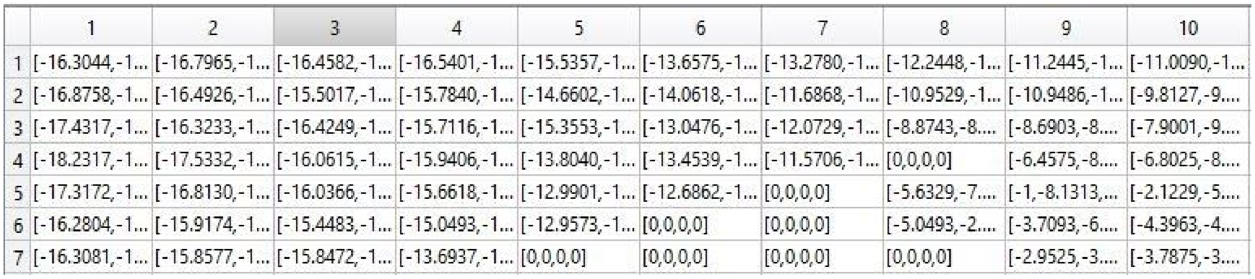
SARSA

Sarsa is an on-policy TD control algorithm that uses quintuple of events (St, At, Rt+1, St+1, At+1).
The Q(s, a) values are initialized to zero for all values of s, a. Looping for each episode, we initialize the state, and for each step of the episode, we take action and observe the reward and the next state. The next action is chosen by being epsilon-greedy with respect to Q. Q is updated. The state and action are also updated. This continues until the state reaches the final state. Q is initialized as a cell array of vectors for each state, action pair.
The wind in each state is defined in a matrix.
Every time the next state leaves the boundary it is set to the current state.

Graph:



Q:



	1	2	3	4	5	6	7	8	9	10
1	[-16.3044,-1...	[-16.7965,-1...	[-16.4582,-1...	[-16.5401,-1...	[-15.5357,-1...	[-13.6575,-1...	[-13.2780,-1...	[-12.2448,-1...	[-11.2445,-1...	[-11.0090,-1...
2	[-16.8758,-1...	[-16.4926,-1...	[-15.5017,-1...	[-15.7840,-1...	[-14.6602,-1...	[-14.0618,-1...	[-11.6868,-1...	[-10.9529,-1...	[-10.9486,-1...	[-9.8127,-9....
3	[-17.4317,-1...	[-16.3233,-1...	[-16.4249,-1...	[-15.7116,-1...	[-15.3553,-1...	[-13.0476,-1...	[-12.0729,-1...	[-8.8743,-8....	[-8.6903,-8....	[-7.9001,-9....
4	[-18.2317,-1...	[-17.5332,-1...	[-16.0615,-1...	[-15.9406,-1...	[-13.8040,-1...	[-13.4539,-1...	[-11.5706,-1...	[0,0,0,0]	[-6.4575,-8....	[-6.8025,-8....
5	[-17.3172,-1...	[-16.8130,-1...	[-16.0366,-1...	[-15.6618,-1...	[-12.9901,-1...	[-12.6862,-1...	[0,0,0,0]	[-5.6329,-7....	[-1,-8.1313,...	[-2.1229,-5....
6	[-16.2804,-1...	[-15.9174,-1...	[-15.4483,-1...	[-15.0493,-1...	[-12.9573,-1...	[0,0,0,0]	[0,0,0,0]	[-5.0493,-2....	[-3.7093,-6....	[-4.3963,-4....
7	[-16.3081,-1...	[-15.8577,-1...	[-15.8472,-1...	[-13.6937,-1...	[0,0,0,0]	[0,0,0,0]	[0,0,0,0]	[0,0,0,0]	[-2.9525,-3....	[-3.7875,-3....