# Name – Suhit Sinha
# Roll No: 153050005
# M. Tech - 1
# Assignment 4 Report

SARSA algorithm in Windy Grid world with 4 actions [up, right, down, left] has been performed as part of this assignment. Then the same task has been performed using 8 actions including the diagonal ones. At the end wind has been considered as stochastic varying by 1 from the mean values given for each column.
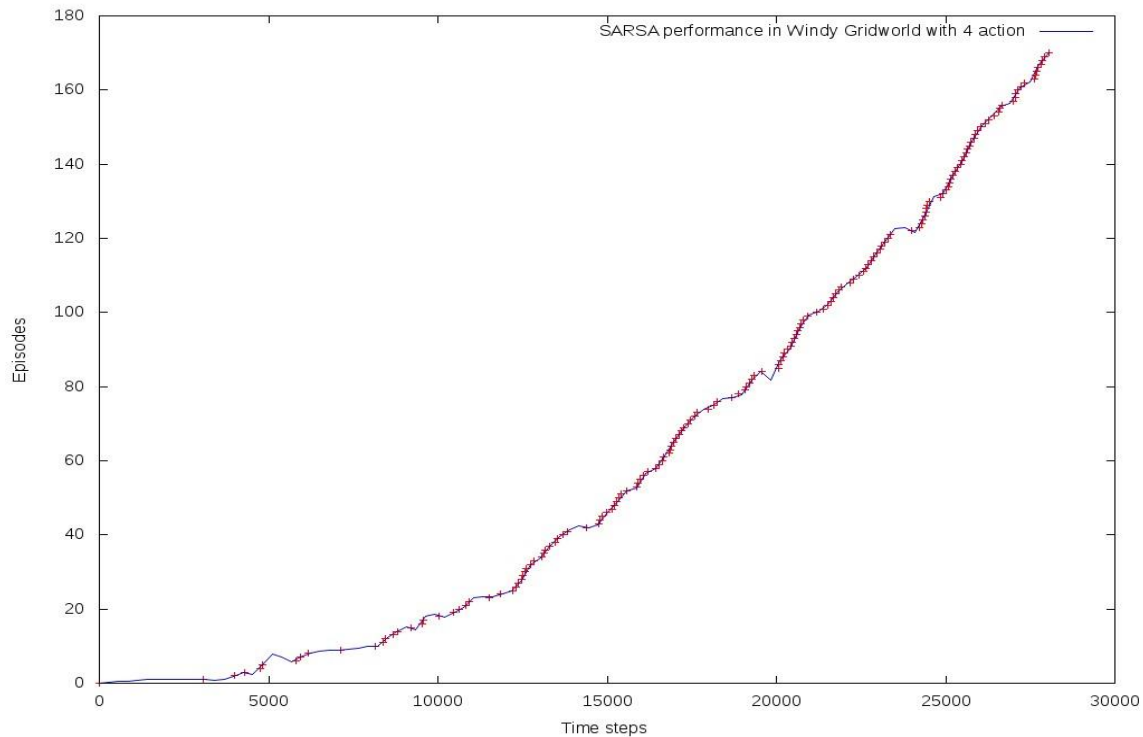
The experiment has been started using epsilon value as 0.1 and alpha value 0.1 in the above 3 experiments.
Below are the results when run with the above given epsilon and alpha values on 170 episodes:

**Windy Gridworld with 4 actions:**

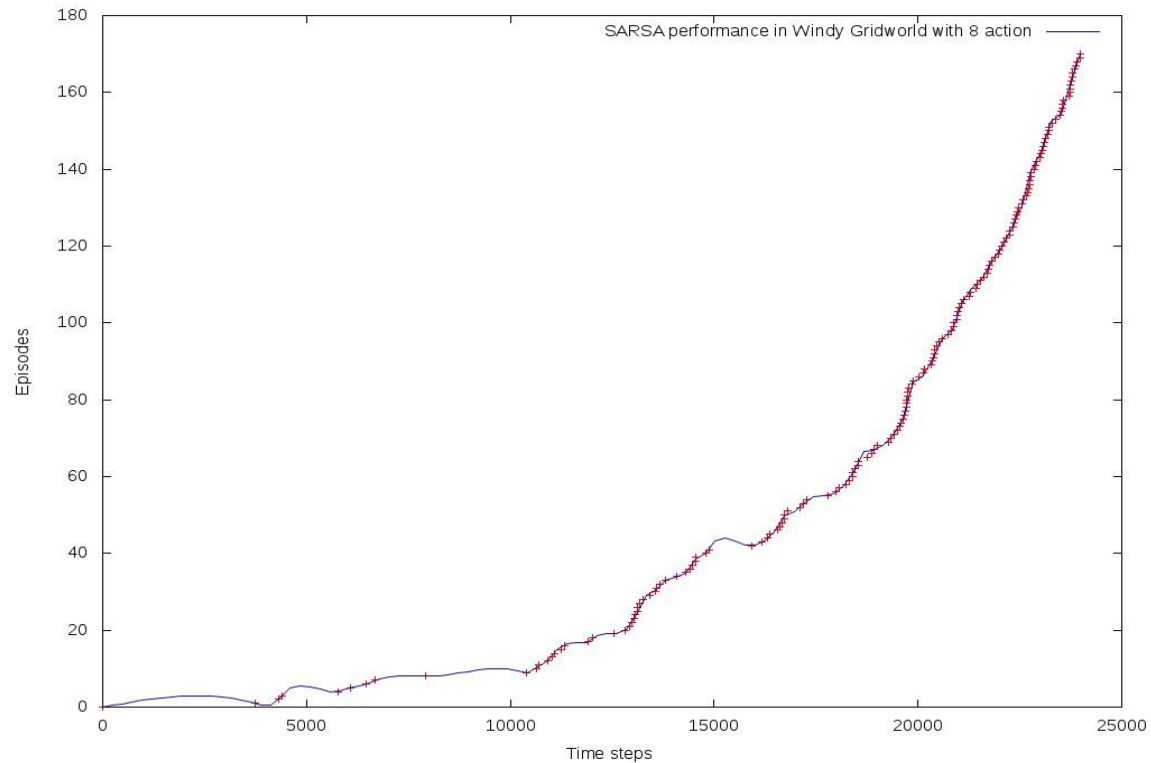Average time step per episode in 4 action SARSA on Windy Grid world = 166.629412
Minimum steps required is: 18

**Windy Gridworld with 8 actions:**

Average time step per episode in 8 action SARSA on Windy Grid world = 150.505882
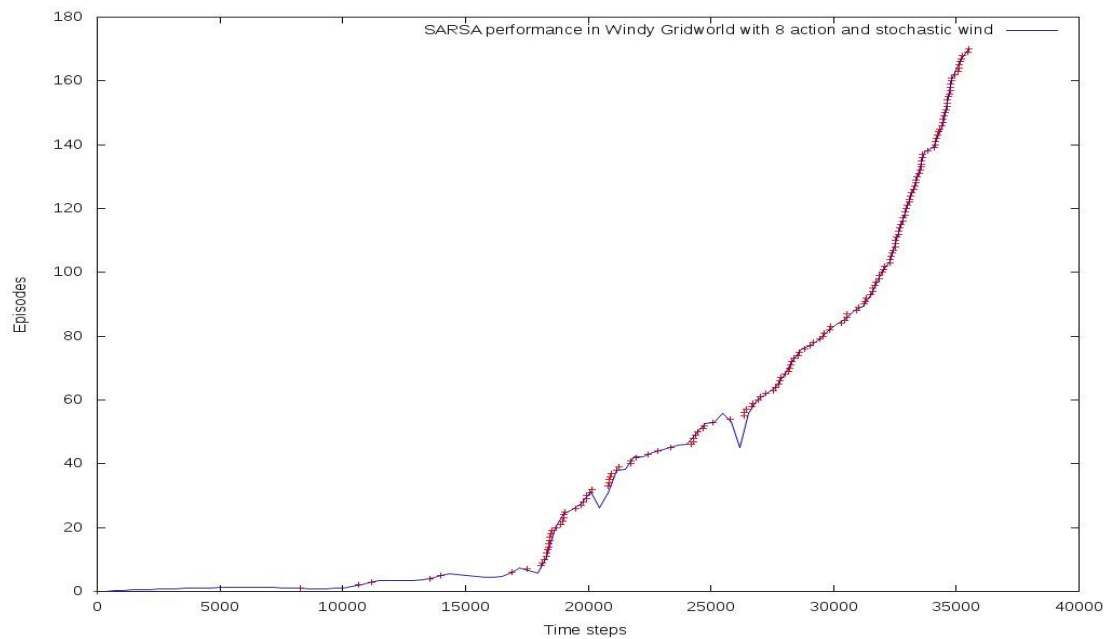Minimum steps required is: 7



**Windy Gridworld with 8 actions and stochastic wind varying by 1:**

Average time step per episode in 8 action SARSA with stochastic wind on Windy Grid world = 206.041176

Minimum steps required is: 7

We can see that for 170 episodes, with epsilon = 0.1 and alpha = 0.1 8 action SARSA gives slightly better performance than the other 2 variants. Also 4 actions SARSA with epsilon greedy approach performs better than 8 actions SARSA with stochastic wind. But the minimum value for 4 action grid should be 15. Hence with the values of this epsilon and alpha, convergence takes more time.
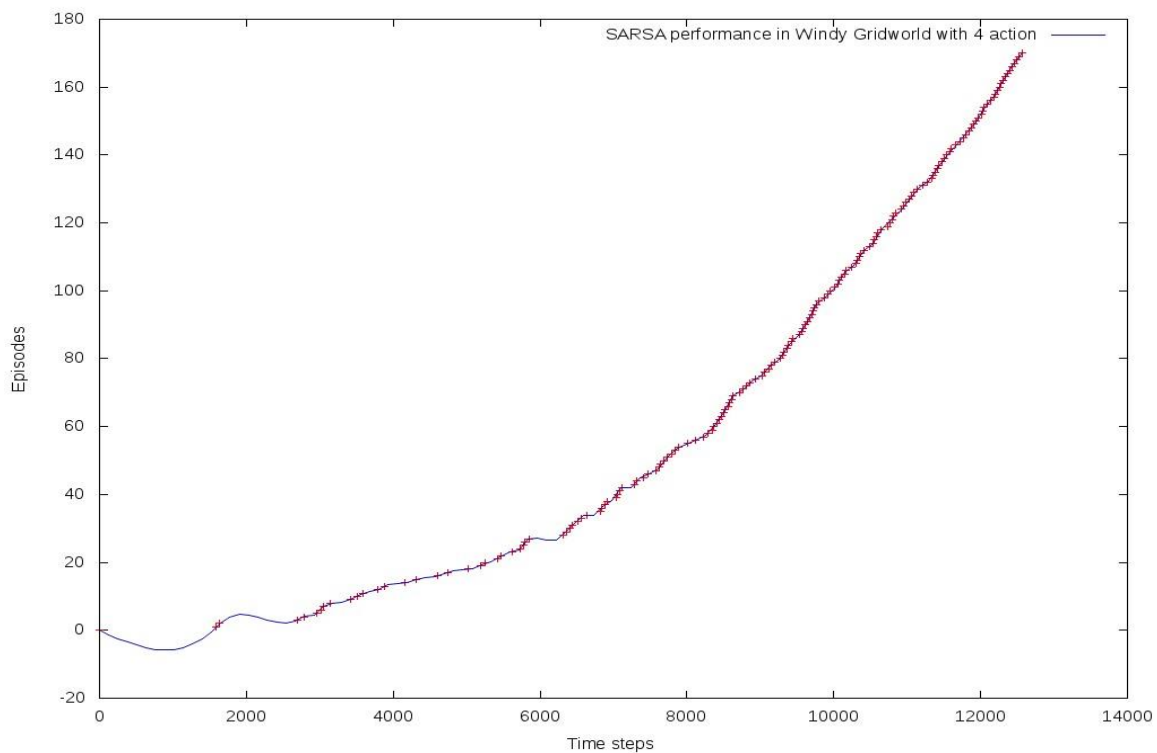
Hence we try to increase the value of alpha and epsilon and aloha one by one. After several experiment, alpha = 0.35 is giving the fast convergence.

**Results with epsilon = 0.1 and alpha = 0.35 is given below when running for 170 episodes:**

**Windy Grid world with 4 actions:**

Average time step per episode in 4 action SARSA on Windy Grid world = 73.900000
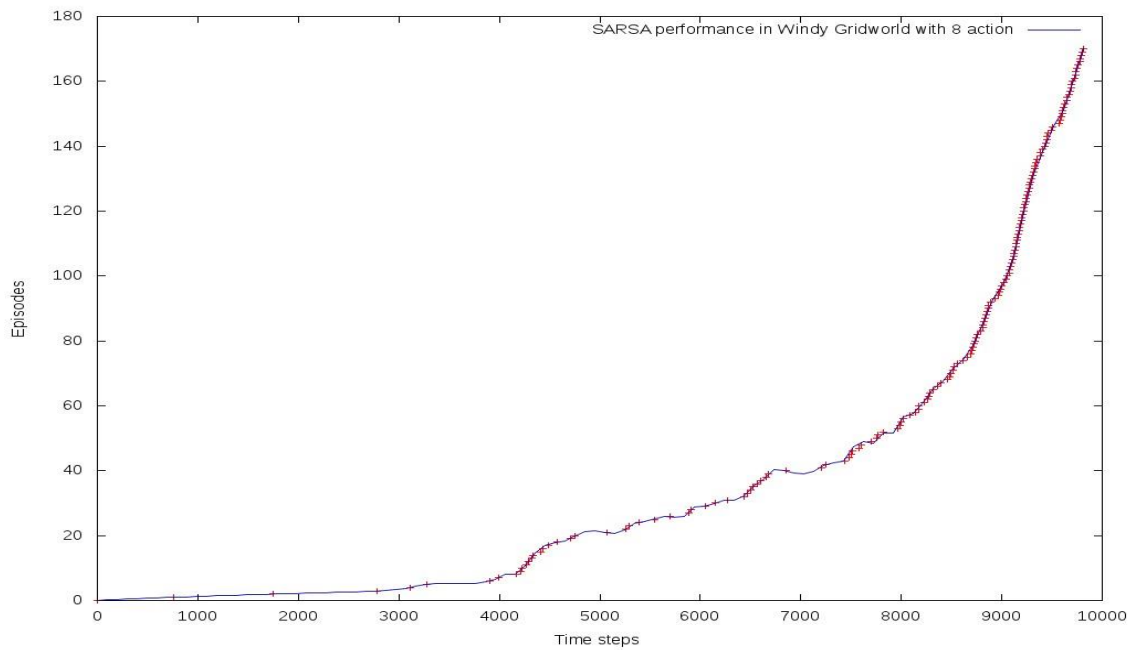Minimum steps required is: 15

**Windy Grid world with 8 actions:**

Average time step per episode in 8 action SARSA on Windy Grid world = 56.176471
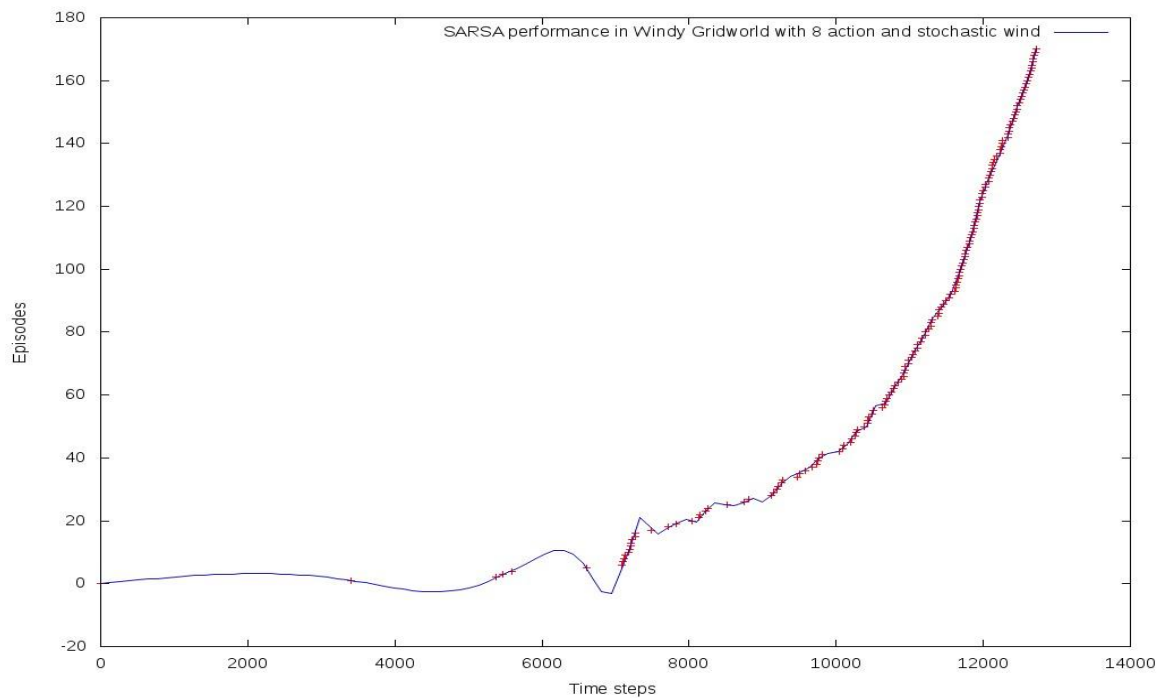Minimum steps required is: 7



**Windy Grid world with 8 actions and stochastic wind varying by 1:**

Average time step per episode in 8 action SARSA with stochastic wind on Windy Grid world = 73.152941
Minimum steps required is: 7

Here we can see that increasing alpha value to 0.35 helped to converge the values within 170 samples for 4 action variants. At this point, the 4 action grid and the 8 action grid is giving nearly same performance on 170 episodes simulation.
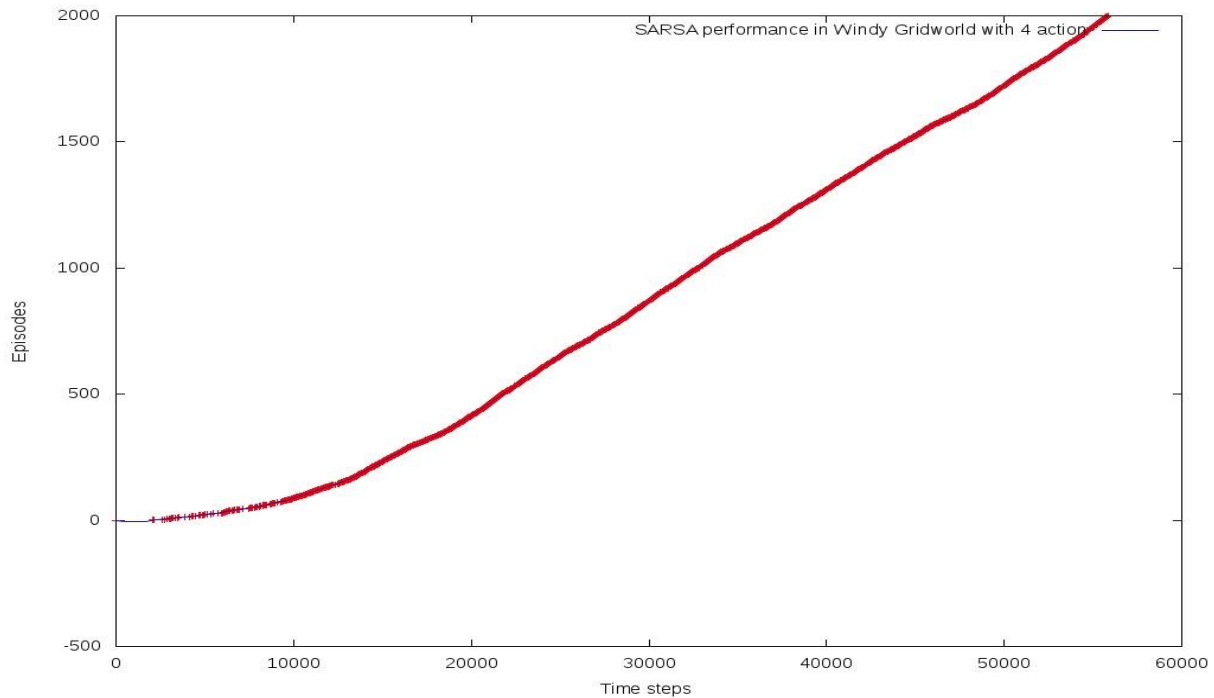
**Now we can test those variants for some large number of episodes [2000 episodes]. The results are given below with alpha = 0.35 and epsilon = 0.1**
Results are given below:

**Windy Grid world with 4 actions:**

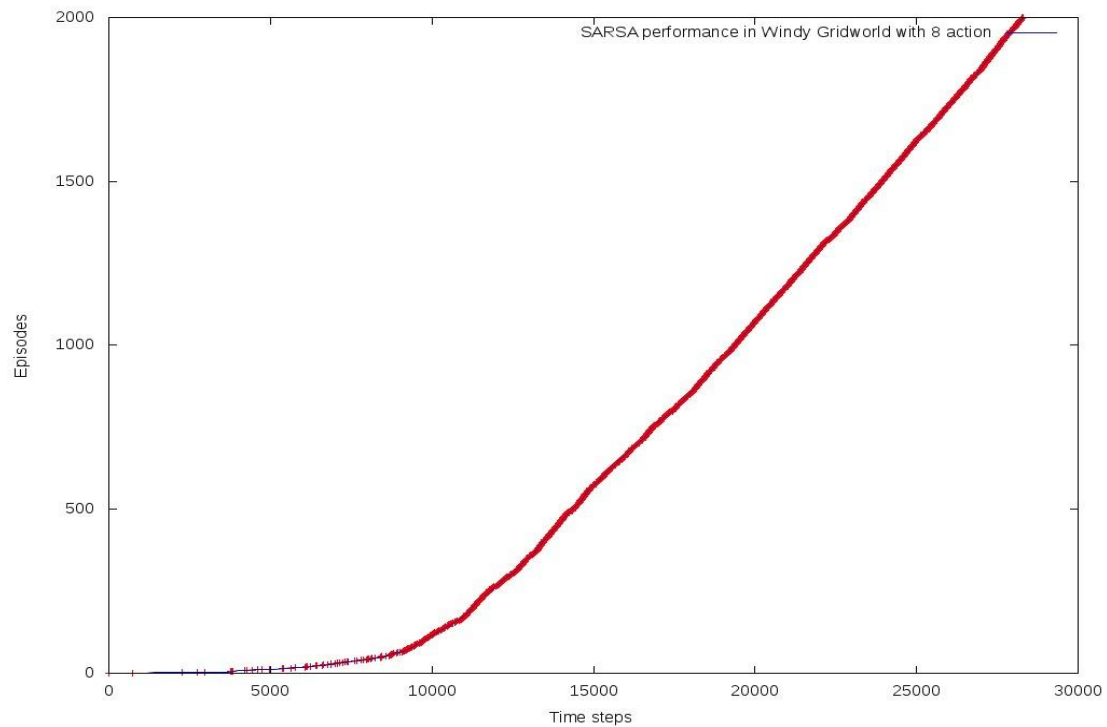Average time step per episode in 4 action SARSA on Windy Grid world = 27.926500
Minimum steps required is: 15

**Windy Grid world with 8 actions:**

Average time step per episode in 8 action SARSA on Windy Grid world = 14.143000
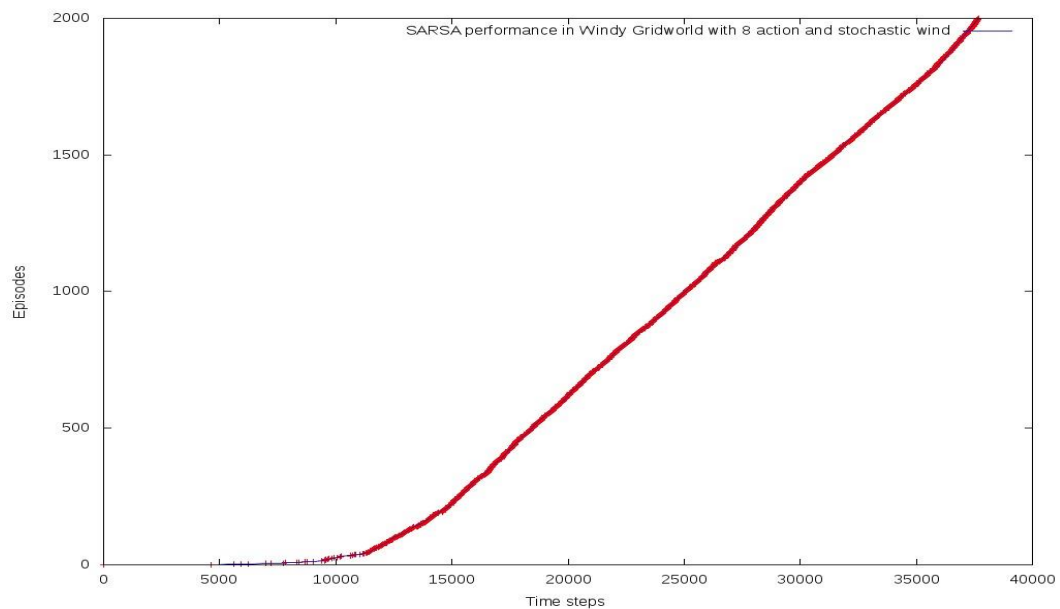Minimum steps required is : 7



**Windy Grid world with 8 actions and stochastic wind varying by 1:**

Average time step per episode in 8 action SARSA with stochastic wind on Windy Grid world = 18.837500
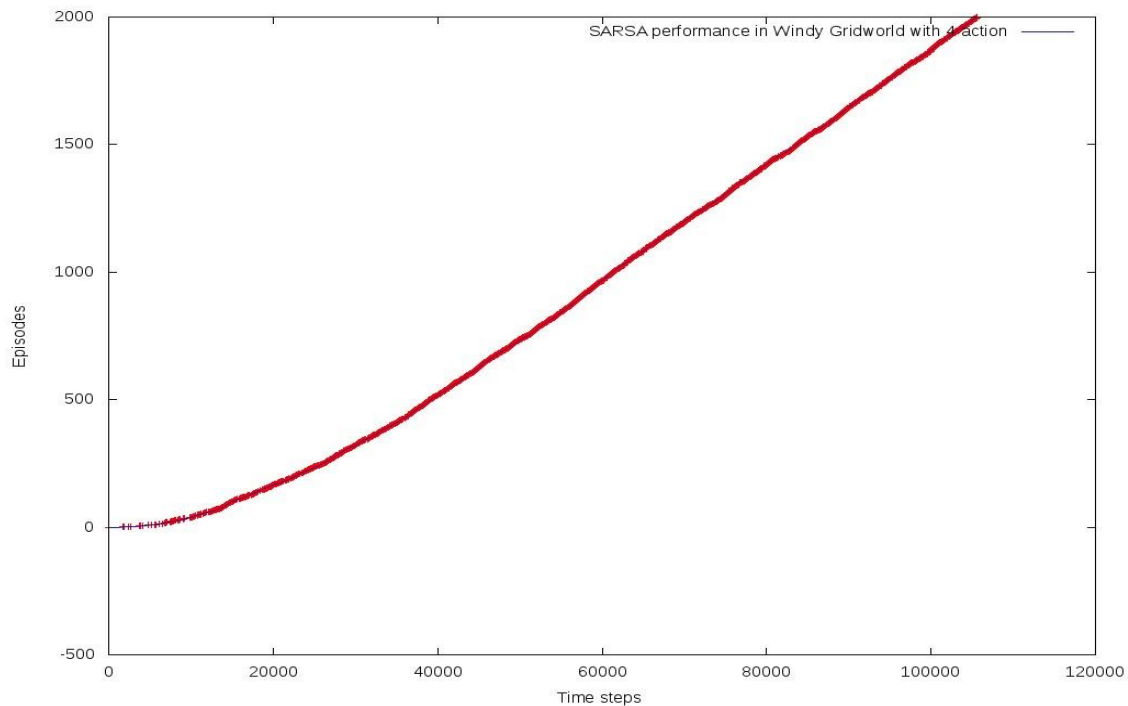Minimum steps required is: 7

Here we can see that, with the increasing number of episodes, 8 action grid with stochastic wind is performing better than 4 action grid. Also the best performance is giving by 8 action non stochastic wind grid.

Now we try to increase epsilon to see the effect of the same. We can test those variants for some large number of episodes [5000 episodes]. The results are given below with alpha = 0.35 and epsilon = 0.35 Results are given below:

**Windy Grid world with 4 actions:**

Average time step per episode in 4 action SARSA on Windy Grid world = 52.800500
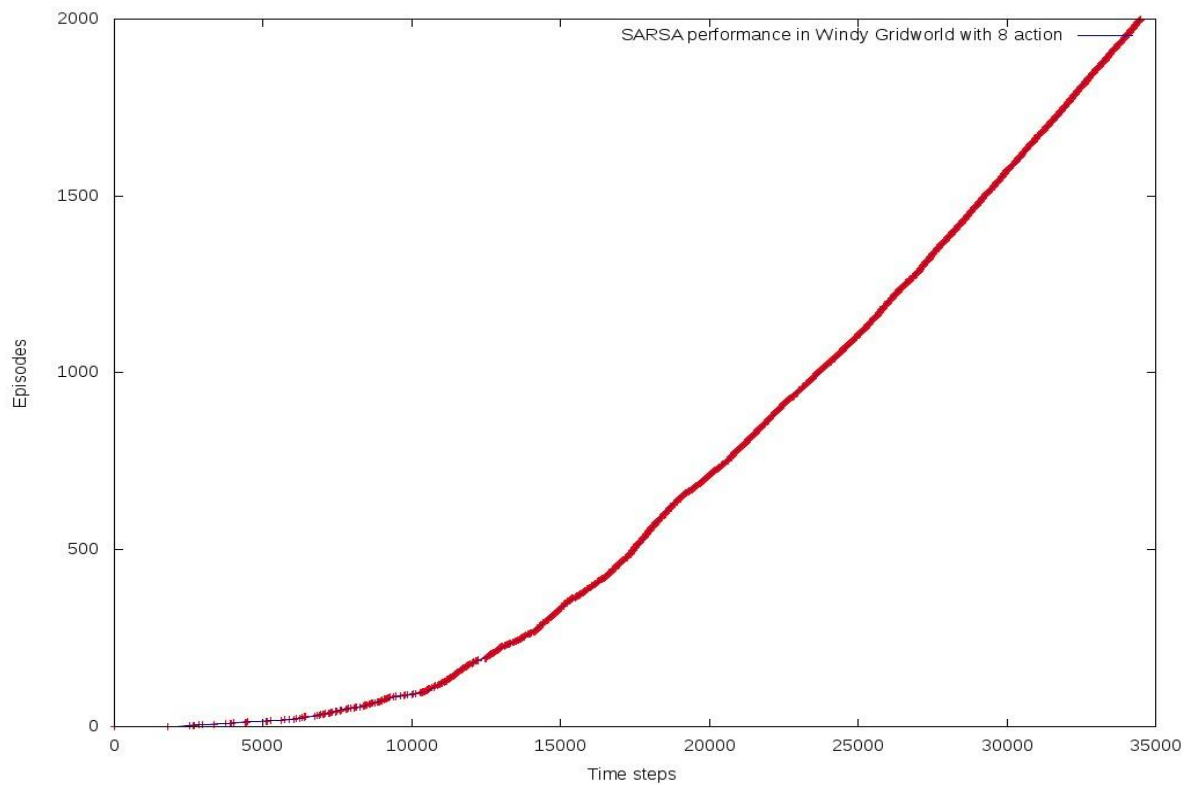Minimum steps required is: 15

**Windy Grid world with 8 actions:**

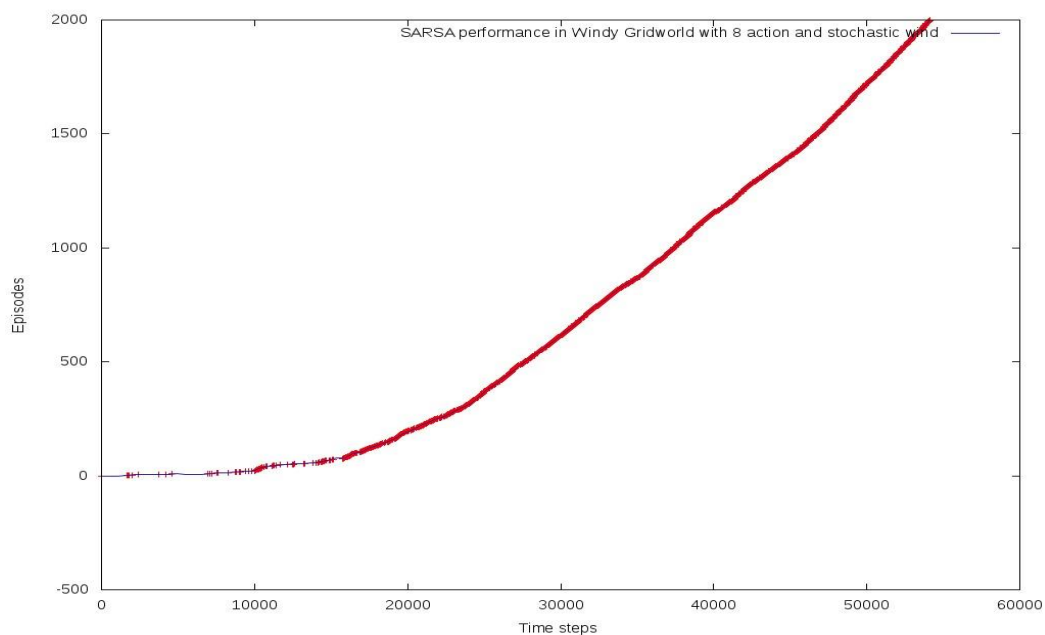Average time step per episode in 8 action SARSA on Windy Grid world = 17.251000
Minimum steps required is: 7



**Windy Grid world with 8 actions and stochastic wind varying by 1:**

Average time step per episode in 8 action SARSA with stochastic wind on Windy Grid world = 27.079500
Minimum steps required is: 7

Here we can see that increasing the value of epsilon is having adverse effect of the average value and also on the convergence of the optimal value.

## Key observations:
1. Increasing the value of Alpha increase the chance of quick convergence to the optimal value path for the grid.
2. For low number of episodes, the average time step per episode should be higher.
3. Increasing the number of episodes help to perform the agent well because after finding an optimal path with highest reward, it continues to that path with probability 1 – epsilon, and explore for new path with probability epsilon. As epsilon is large compare to 1 – epsilon hence the chances are high that it will take the optimal path after finding one.
4. By experiment alpha value of 0.35 is giving good results compare to other values.
5. Increasing the value of epsilon helps the agent to explore more at the beginning. But in case of large episodes, after finding an optimal path, i.e. the path with highest reward, it searches for other paths with epsilon probability. It degrades the performance in long run.
6. Keeping nine steps will also not create a difference in large episodic runs, because the straight line distance is 7 which can be the minimum one, hence adding one additional idle action would not improve the performance.