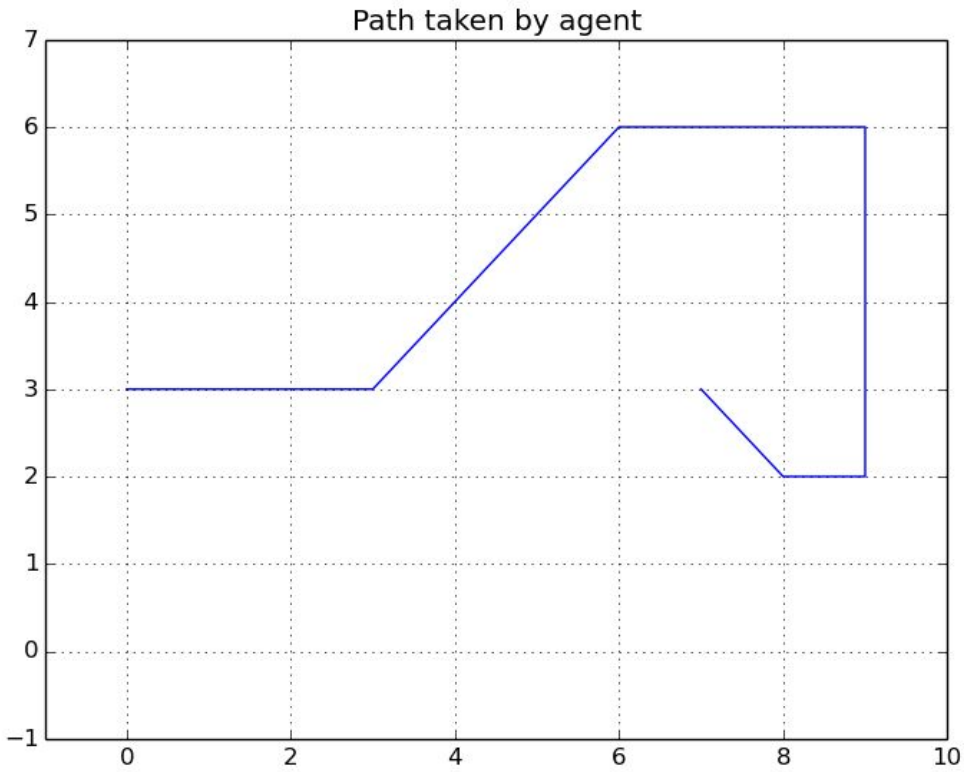
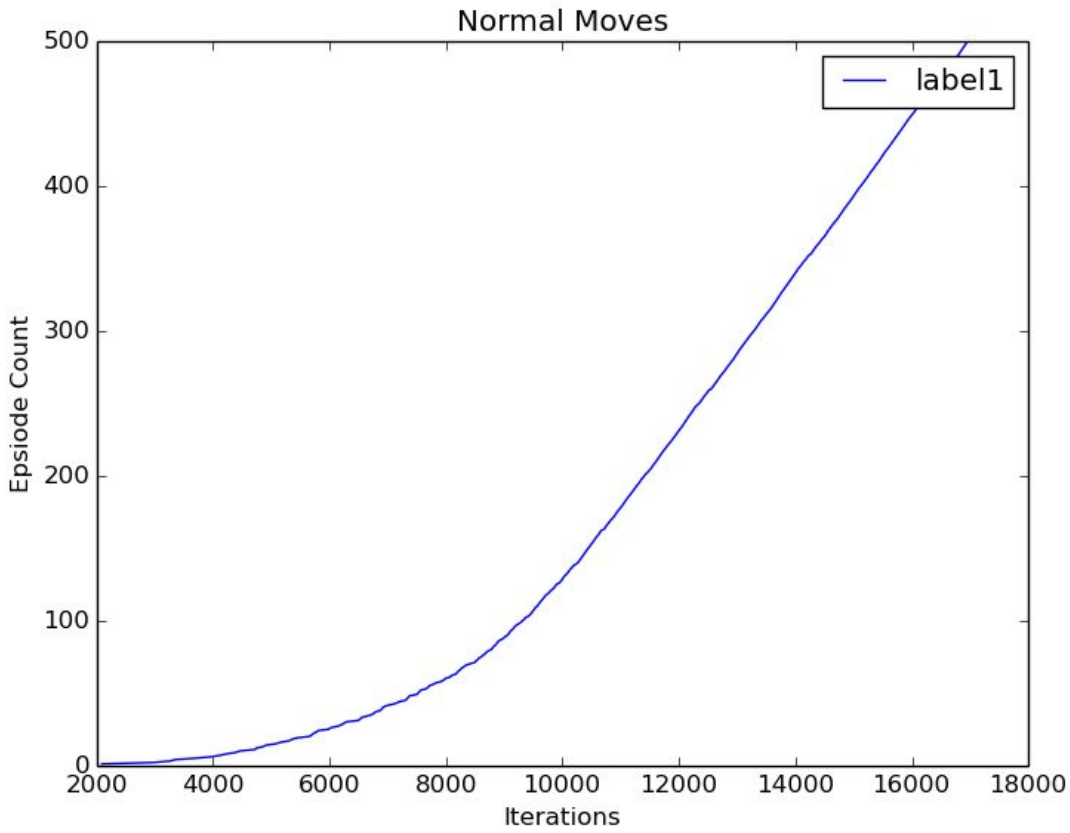
Observation :

Baseline Plot : 4 standard moves are allowed .

Compared to king's moves , this model took more iterations to converge to optimal policy. Average steps to reach goal state from start state is 19(when reward was set 0 on reaching goal state).

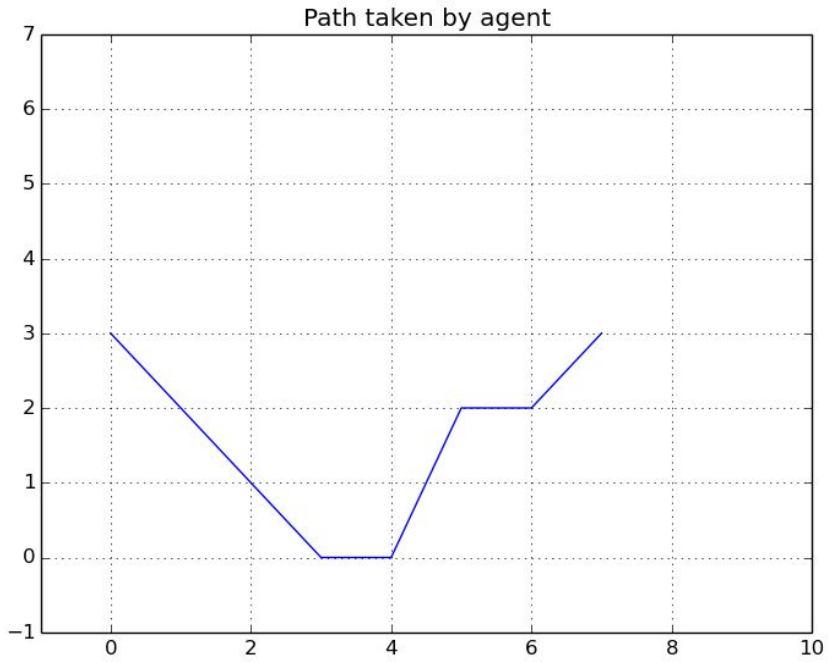
It is also observed , change in reward to reach goal state(giving some reward >0 on reaching goal state) resulted in good performance of the model . Average steps to reach goal state reduced to 15(on following optimal policy) when reward was set to 1000 on reaching goal state.

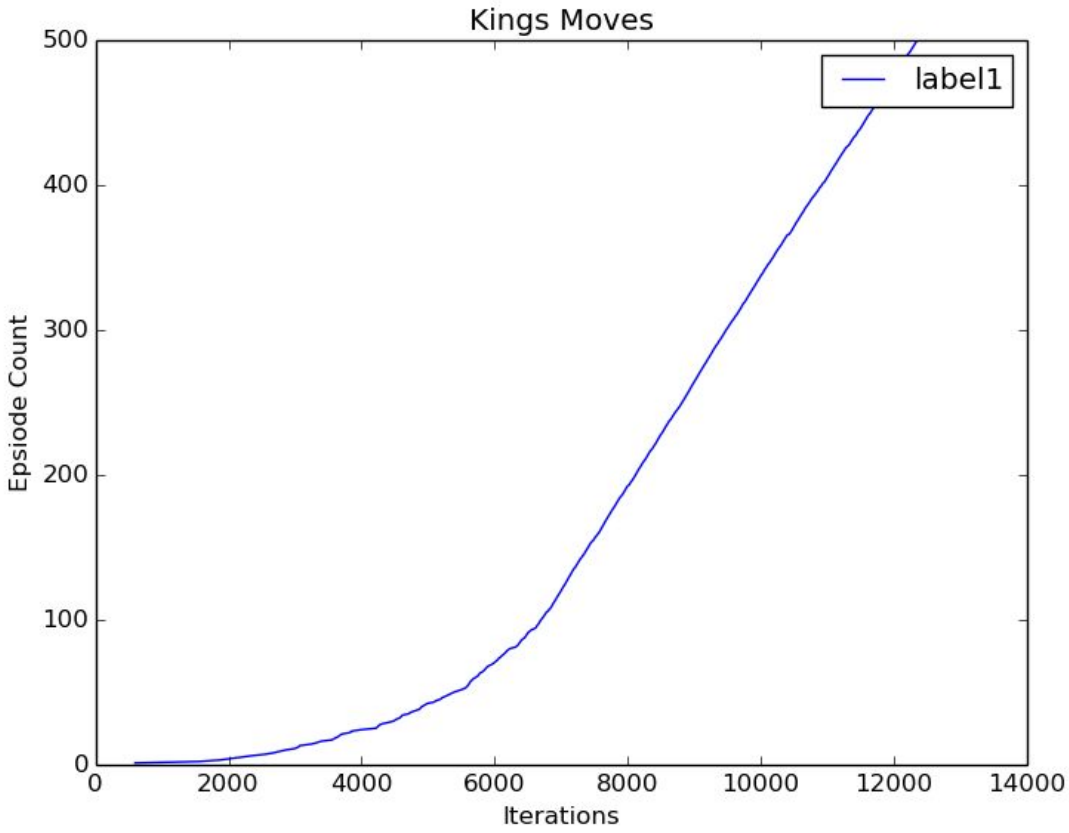




King's Moves: 8 moves are allowed

This model converged to optimal policy at less iterations . Average iterations per episode is 11 when reward on reaching goal state is set to 0 and -1 otherwise . But when the reward on reaching on goal state is set to 10 and -1 otherwise, decreased in average iterations per episode to 7 on following optimal policy and average to 9 on following epsilon greedy (based on qvalues)





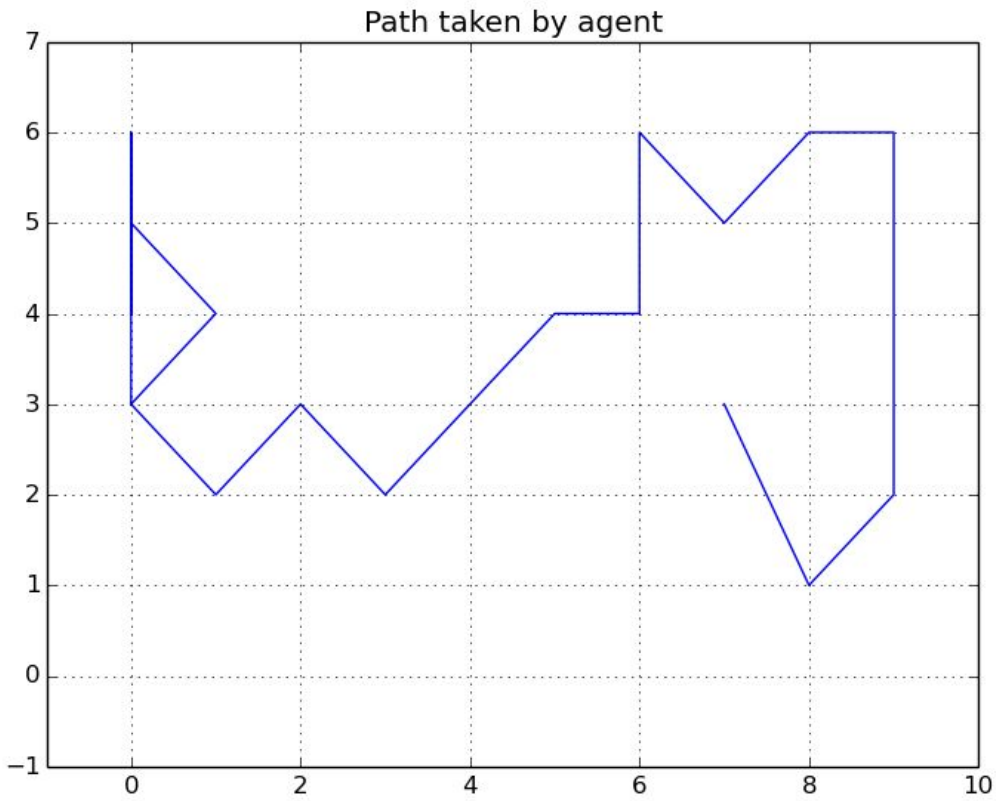
Stochastic wind :

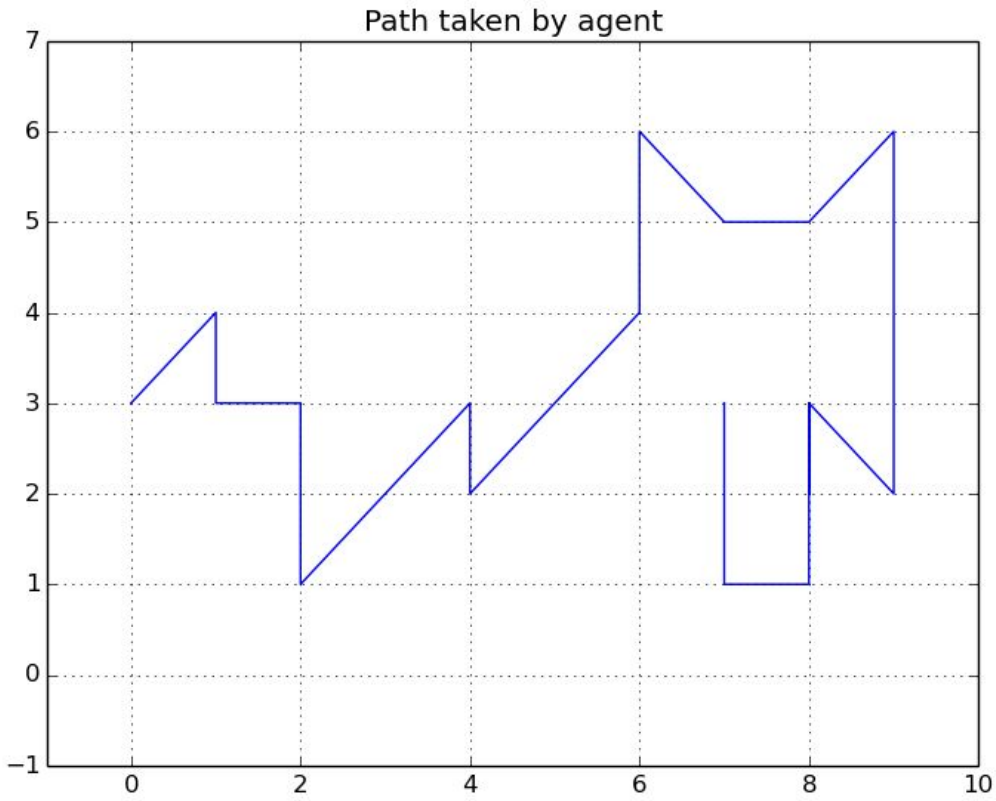
Effect of stochastic wind , sometimes varying by 1 from the mean values given for each column. That is, a third of the time you move exactly according to these values, as in the previous exercise, but also a third of the time you move one cell above that, and another third of the time you move one cell below that.

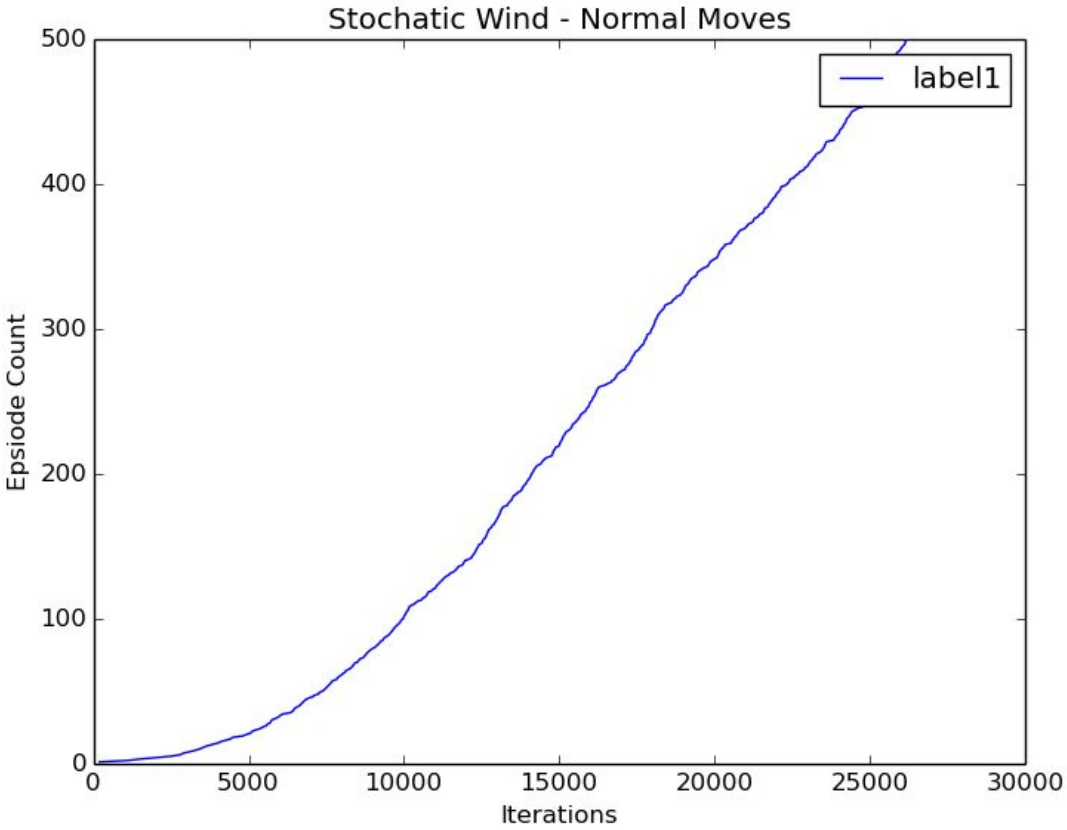
Stochastic wind and 4 Moves are allowed :
Running the model in stochastic environment , even after 50,000 iterations and 400+ episodes , agent didn't learn the optimal policy. Average iterations per episode is 26

It is also observed on setting to different random seed , agent converged to different policies

Two paths of different random seed:



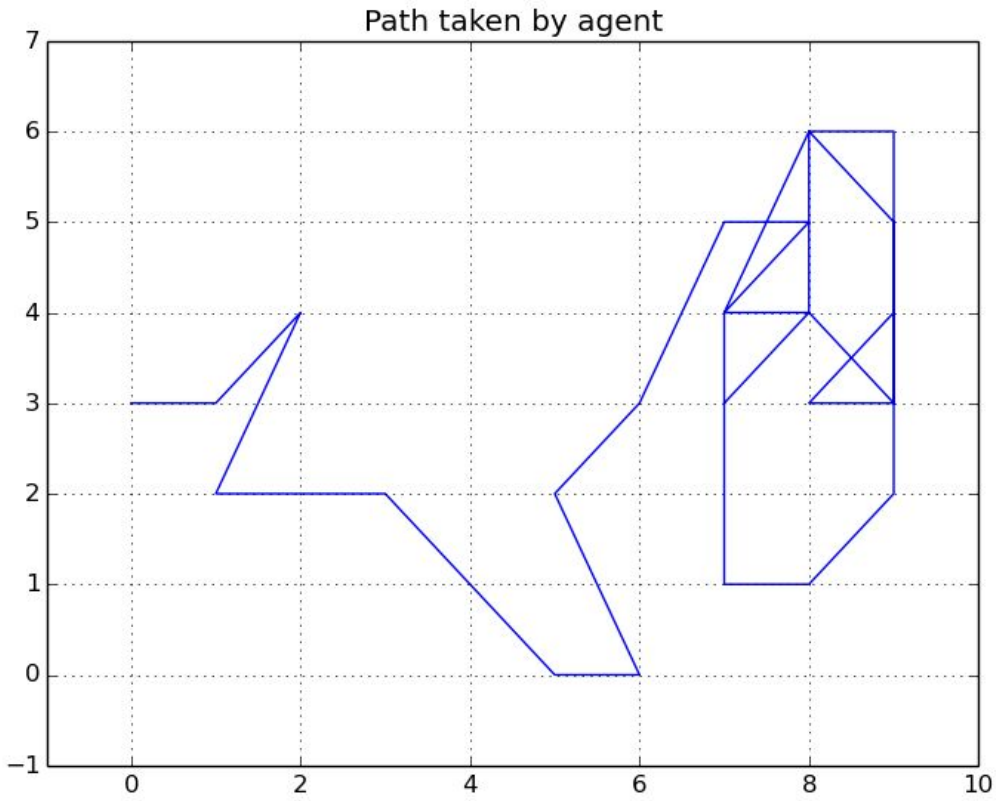

Path taken by agent
7
6
5
4
3
2
1
0
−1
0
2
4
6
8
10

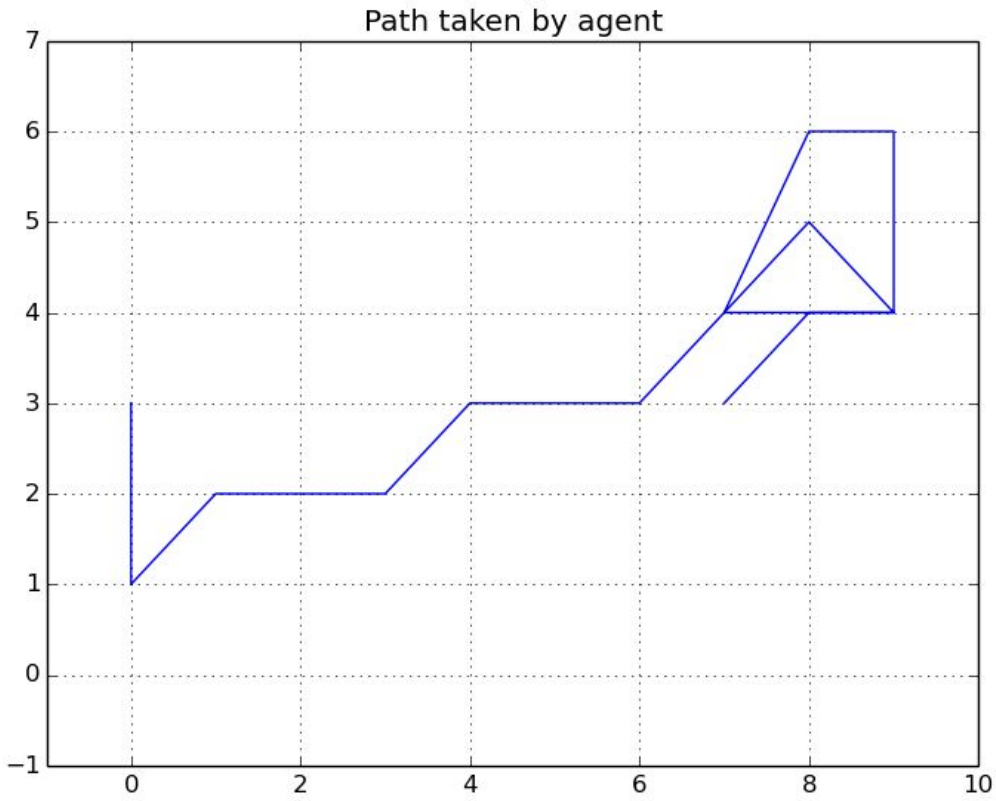

Stochatic Wind - Normal Moves
label1
Epsiode Count
Iterations
500
400
300
200
100
0
0
5000
10000
15000
20000
25000
30000

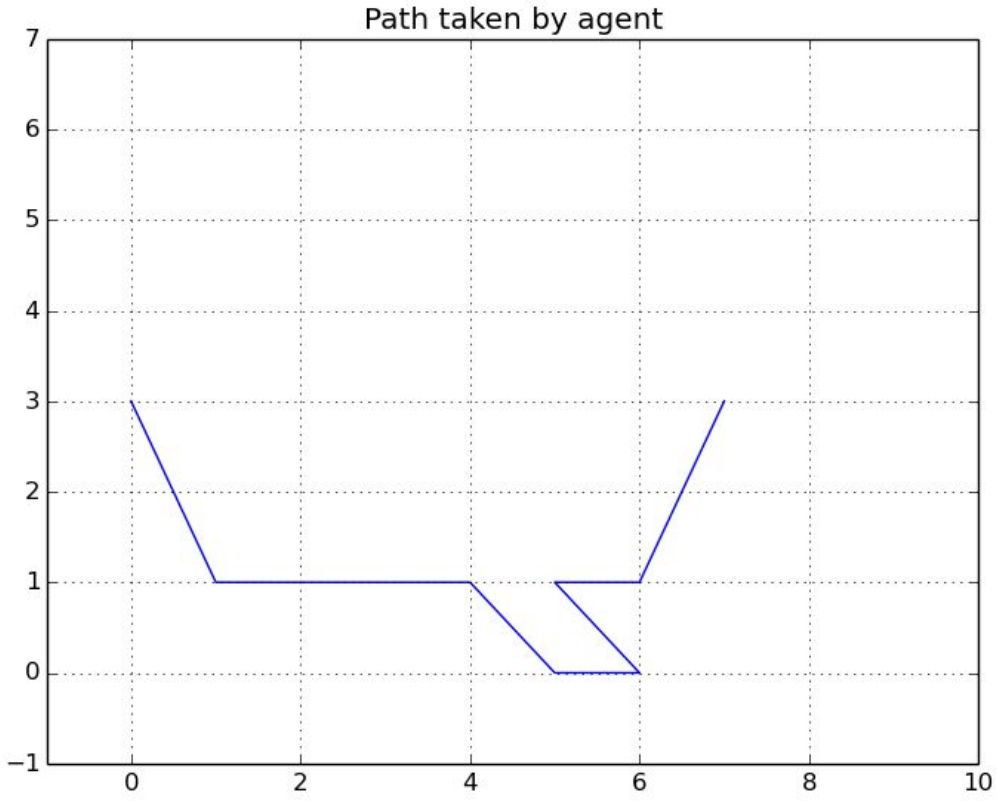
Stochastic wind and King's Moves

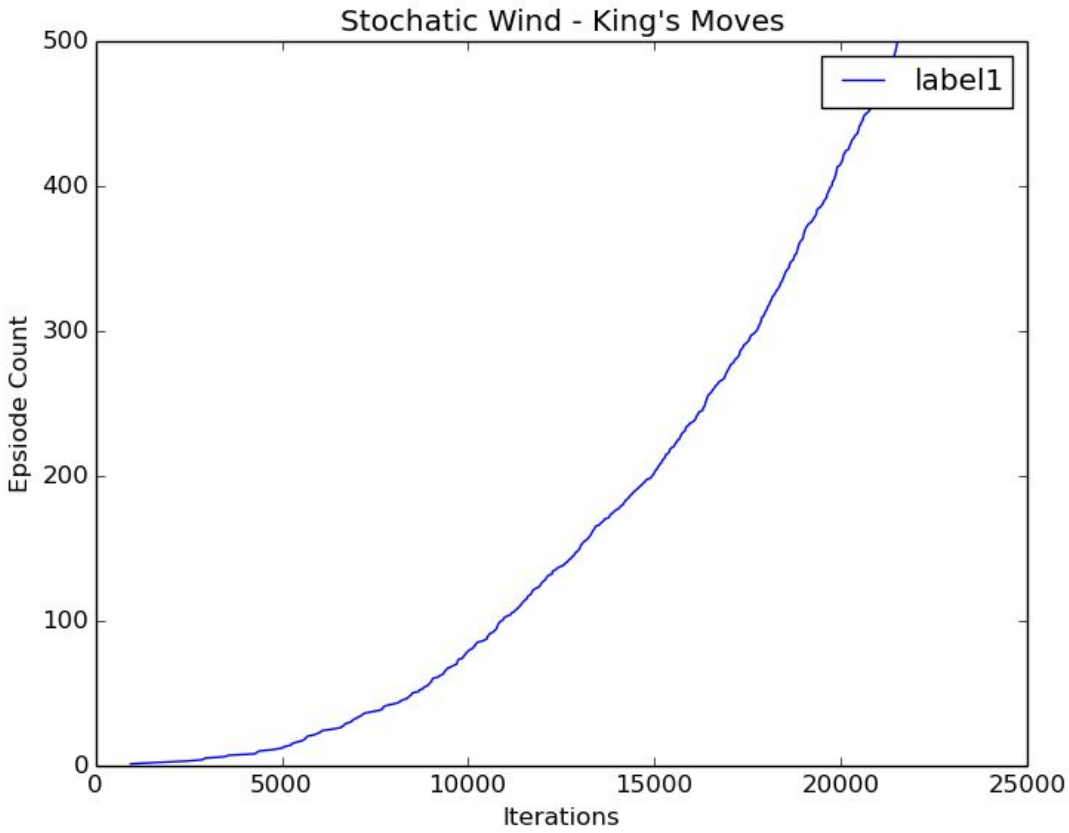
Even in stochastic environment , king's moves model learnt the optimal policy in less than 30,000 iterations but average iterations per episode is 16 .

In king's moves , for some random seed agent took more than 60 steps to reach goal state. Below are the three sample paths taken by agent to reach goal state on different random seeds.




Path taken by agent
7
6
5
4
3
2
1
0
−1
0
2
4
6
8
10


Path taken by agent
7
6
5
4
3
2
1
0
−1
0
2
4
6
8
10



Common observation :

1) Reward also plays role in the speed of convergence to optimal policy
2) As the number of iterations tends to infinity , model will converge to optimal policy even in stochastic environment .
3) For some random seeds, policy didnt converge to optimal policy .