CS 747: Programming Assignment 4

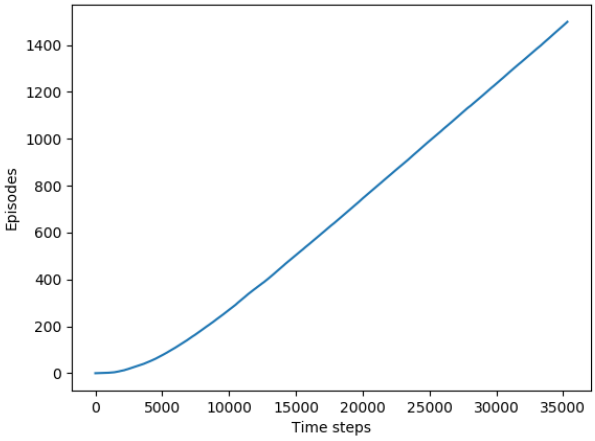
16D070041 - Saarthak Kapse

In all three task I am using learning rate i.e alpha = 0.5 and exploration rate i.e epsilon = 0.1 same as in book.

In all the task start state is set to [3,0] and goal state [3,7].
And there is no horizontal wind and vertical wind is [0,0,0,1,1,1,2,2,1,0].

Task2:
In this task the objective is to train a agent having 4 actions ('up', 'down', 'left', 'right') and a grid of 7 rows and 10 columns to reach goal state [3,7] starting at [3,0].

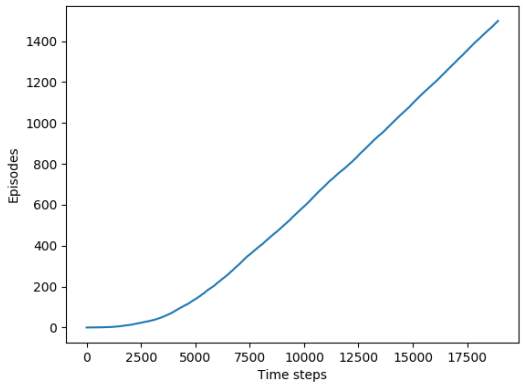
Plot:-



During initial episodes as agent is naive so it should take more time steps to reach goal state and after some episodes it should start learning and start reaching goal state in less time step.
We can observe from graph that at initial episodes slope is very less and it increases with episodes.
Slope = # of episodes / Time steps.
So as at initial episodes slope tells us Time steps for each episode is so high as slope is low. And as slope increases time to do one episodes decreases as slope, basically it is learning. After some 300 episodes slope becomes constant and it shows the learning has now converge and agent has learnt best possible policy using sarsa(0) algorithm

Task3:
In this task the objective is to train a agent having 8 actions ('up', 'down', 'left', 'right', 'up right', 'up left', 'down right', 'down left') and a grid of 7 rows and 10 columns to reach goal state [3,7] starting at [3,0].

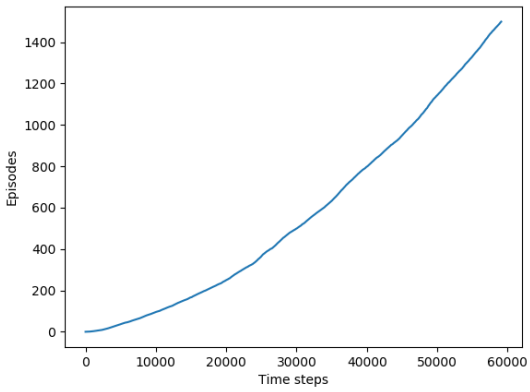
Plot:-



Same explaination as last task. But the main difference here is as action can take 8 actions and 4 actions are combination of two actions so it should take less time to reach goal state.
And it can be observed by the plot that total time steps to do 1500 episodes is 19000 time step by king move and 35000 time steps by normal actions. So king additional actions let it reach goal state faster.
Here too it converges in somewhat 300-350 episodes

Task4:
In this task the objective is to train a agent having 8 actions ('up', 'down', 'left', 'right', 'up right', 'up left', 'down right', 'down left') and a grid of 7 rows and 10 columns to reach goal state [3,7] starting at [3,0] but with stochashtic wind i.e if their is a wind in a columnns then its wind speed can be wind_speed+1 or wind_speed or wind_speed-1 all with same probability.

Plot:-



Here as environment is stocastic so it is obvious to take more time to converge as well as more time to reach goal step after converging as environment is tricky for the agent. Here it takes around 1200 episodes to converge and total time step is also very large as compared to other two plots, which shows it takes more time to complete episodes as compared to other two tasks

Comparison:

I am taking average of time steps of last 10 episodes for each plot and it will be the approximately the Fastest route time to complete the task.

-	Task1	Task2	Task3
Time steps to complete 1500 episodes	35000	19000	60000
Fastest Route Time	17.36	8.54	21.96