REPORT

CS 747 Assignment 4

Samarjeet Sahoo
150100017
CSE B.Tech

Settings:
Parameters used:

rows = 7
cols = 10

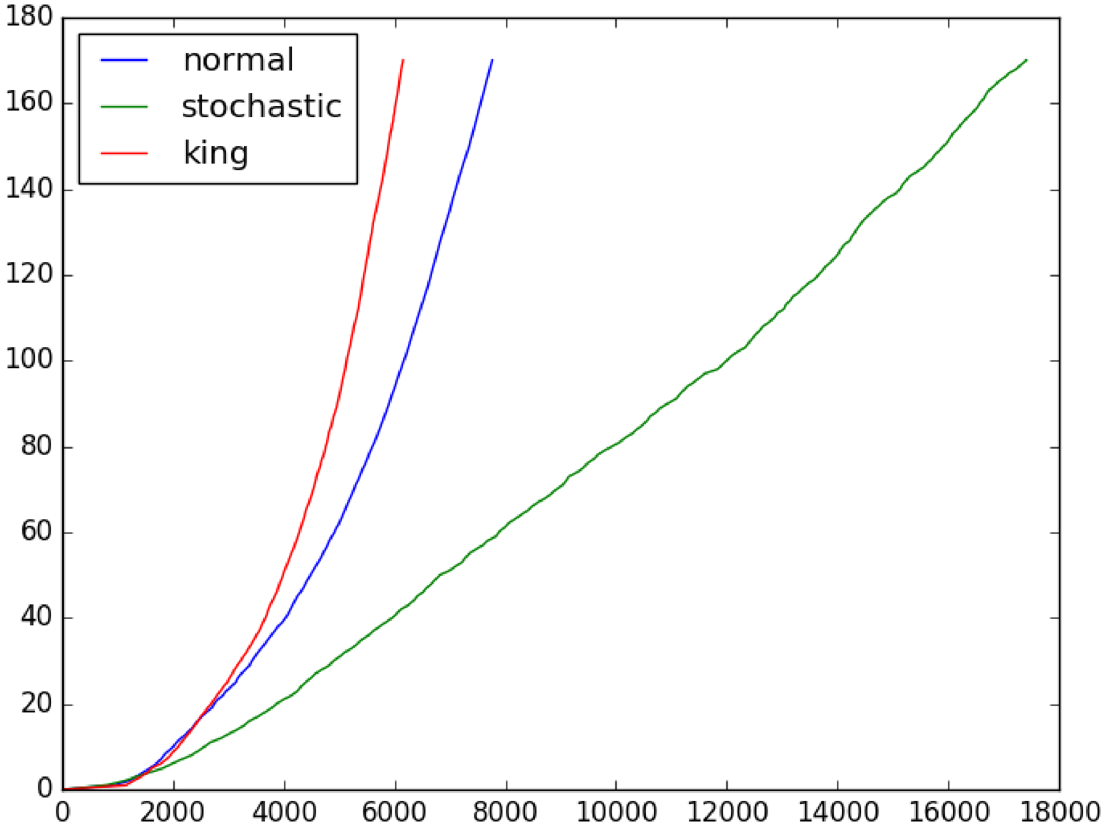
start = [3,0]
goal = [3,7]

alpha   = 0.5
epsilon = 0.1
episodes = 170

Handling boundary cases:
I first add the effects of the wind and the action and then apply the bounds. For example, if the agent is at the bottom most row and takes down, and the wind is upwards 2, then it moves upwards 1. This can also be easily observed in get_next_state_reward(state,action) function in the code

Observations:

Here is the graph with the number of episodes on the y axis and the average number of cumulative time steps in the x-axis.



The following observations can be made from the graph:

1. The stochastic case takes more number of steps (compared to the other 2 cases, both cumulative and per episode) to reach the goal even after 170 episodes, as can be seen by comparing its slope with the others. This is expected, because when the wind is stochastic, one can't really learn a fixed optimal policy to reach the goal because the variation in the wind can affect the new state reached after taking an action. Hence, the optimal policy takes more number of steps (As we can see, the slope is lesser than the others, that means it takes more number of steps per episode)

2. The king's moves cases takes least number of steps in its optimal policy as its slope is the highest. This is also expected, because the king's move has more actions compared to the normal case, and hence we can learn a more "detailed"/better path to reach our goal and are not constrained to only 4 actions. Since the 4 actions of the normal case are contained in the king's moves, we can expect king's case to do at least as good as the normal case anyway. Also, the 4 extra actions allow us to move a step (up/down) and a step (left/right) simultaneously and hence we can potentially replace 2 actions of the normal case by a single action. Hence, we can effectively take lesser steps to reach a destination. Indeed, it ends up performing better because of more specificity, granularity and coverage in the actions and hence better(shorter) path is learnt.

3. We can also see that the the cumulative number of steps is initially higher in the king's case compared to the normal case. This is because with more actions to explore, it takes more time to learn the exact optimal policy. But, once learnt, the optimal policy of the king's case takes lesser number of steps compared to the normal case for reasons specified in 2nd observation.