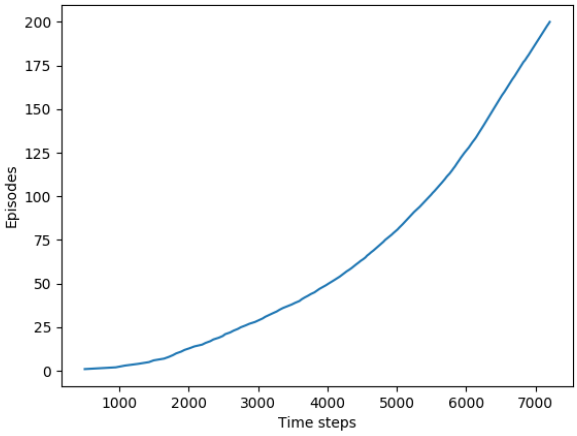
CS 747: Programming Assignment 4

Vinayak K (150050098)

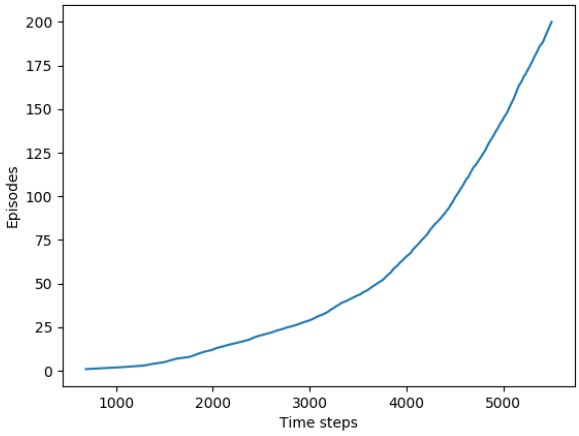
10 Nov 2018

Plots

Experiments were run on a windy grid with configuration (size, start, goal and wind speeds) similar to the example provided in Sutton and Barto. Following are the episodes vs timesteps plot for 4 experiments with different combinations of wind and moves. In each experiment, learning rate $\alpha = 0.5$, discout factor $\gamma = 1.0$ and exploration rate $\epsilon = 0.01$.

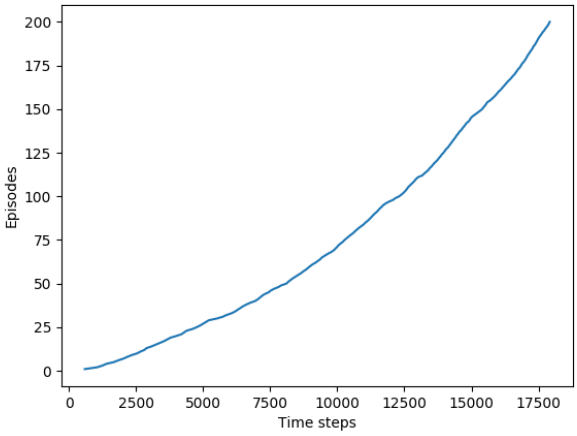


(a) Deterministic wind, Without King's move

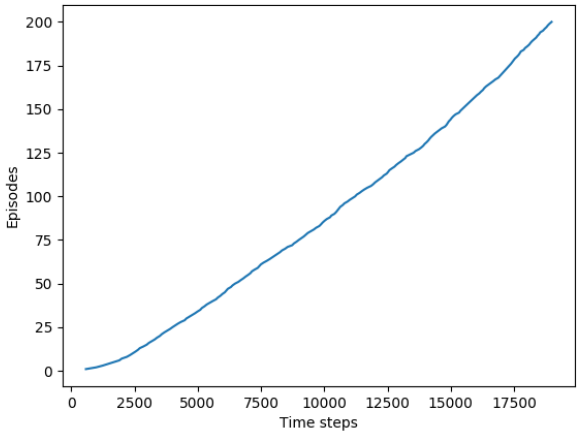


(b) Deterministic wind, With King's move

Figure 1: 1



(a) Stochastic wind, Without King's move



(b) Stochastic wind, With King's move

Figure 2: 2

Observations

* In the case with deterministic wind without king's move, the best path from start to goal takes 15 time steps and the agent has learned that path by approximately 100 episodes.
* Clearly if King's move is allowed there is a better path from start to goal and the algorithm found that path (which is 7 steps long).
* **SARSA** algorithm is not ideal for stochastic environments. It does not define how the policy should be derived from action-values. So we it can only learn some approximation to stochastic policies. This is clear from plots (2a) and (2b). It took much more number of steps to finish 200 episodes compared to deterministic wind case.