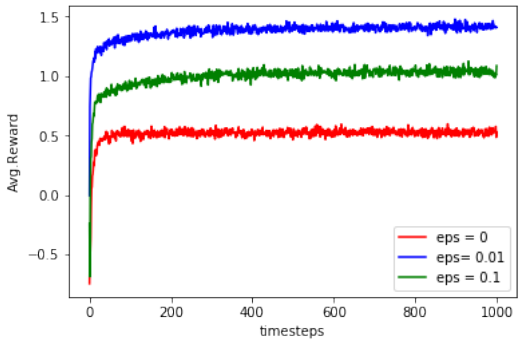
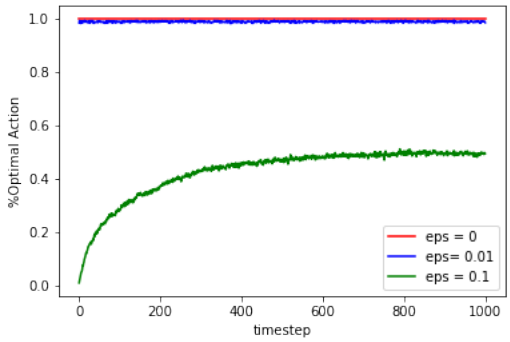
RL-Programming Assignment

Rishabh Samra

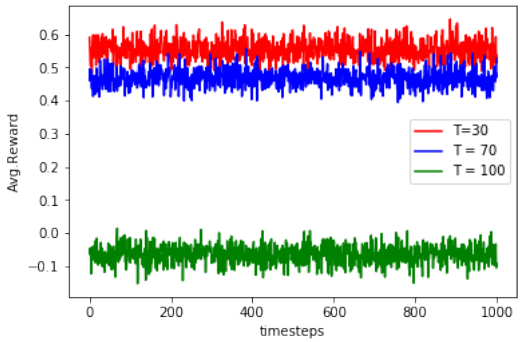
March 2019

Q1. Epsilon Greedy

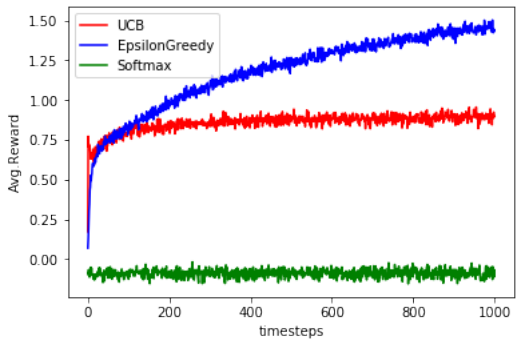




Q2 Softmax action selection using Gibbs Distribution.

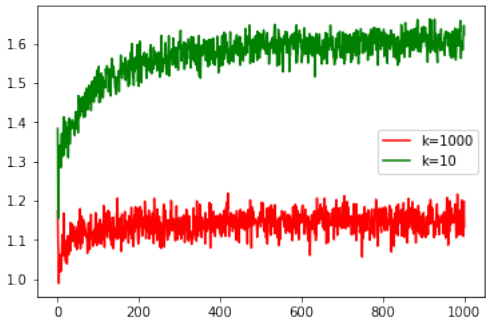


Q3. Comparision between UCB, Epsilon Greedy and Softmax.



In case of epsilon greedy we get maximum average saturation reward but it takes longer time to converge.Also since in epsilon greedy and softmax there is asymptotic convergence and in UCB we select optimal arm initially only thus the convergence rate varies.

Q4 What happens as the number of arms grows? Run experiments on a 1000 arm bandit setup and compare.



We get average rewards of 10 arm bandit setup to be more as compared to 1000 arm testbed.