CS6700: Reinforcement Learning
Programming Assignment #1

Bikash Gogoi (CS14B039)

February 21, 2018

1 ϵ-greedy

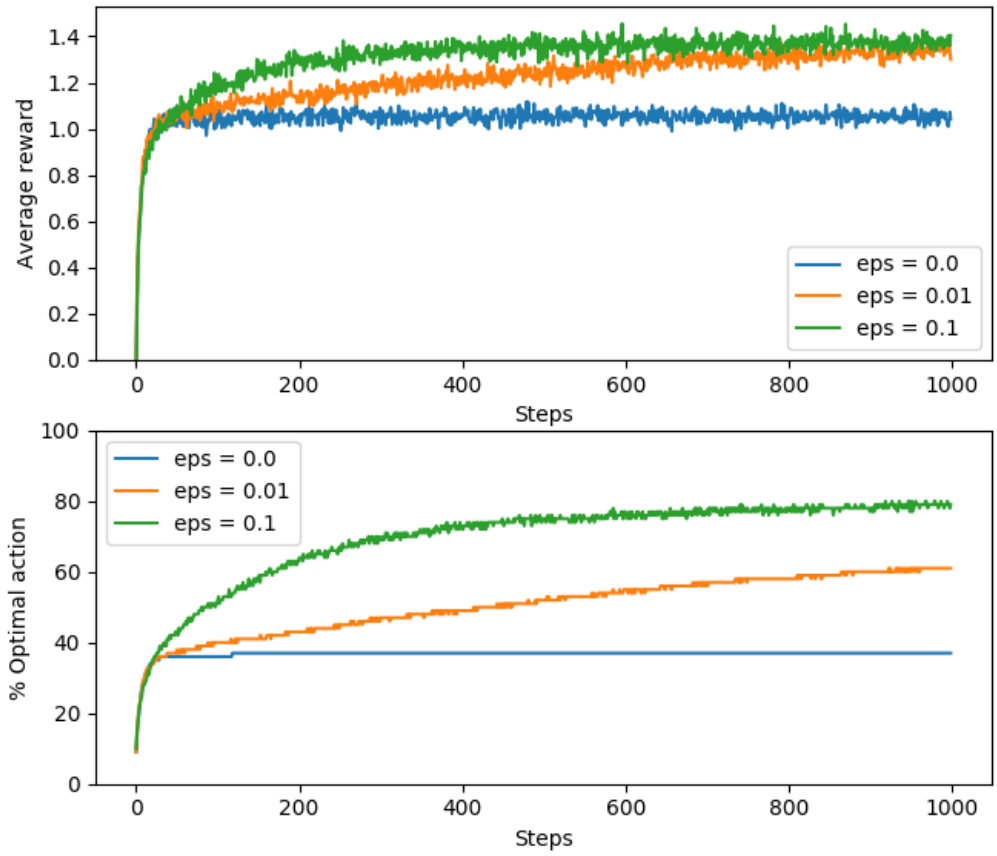

Figure 1: ϵ-greedy

2 Softmax

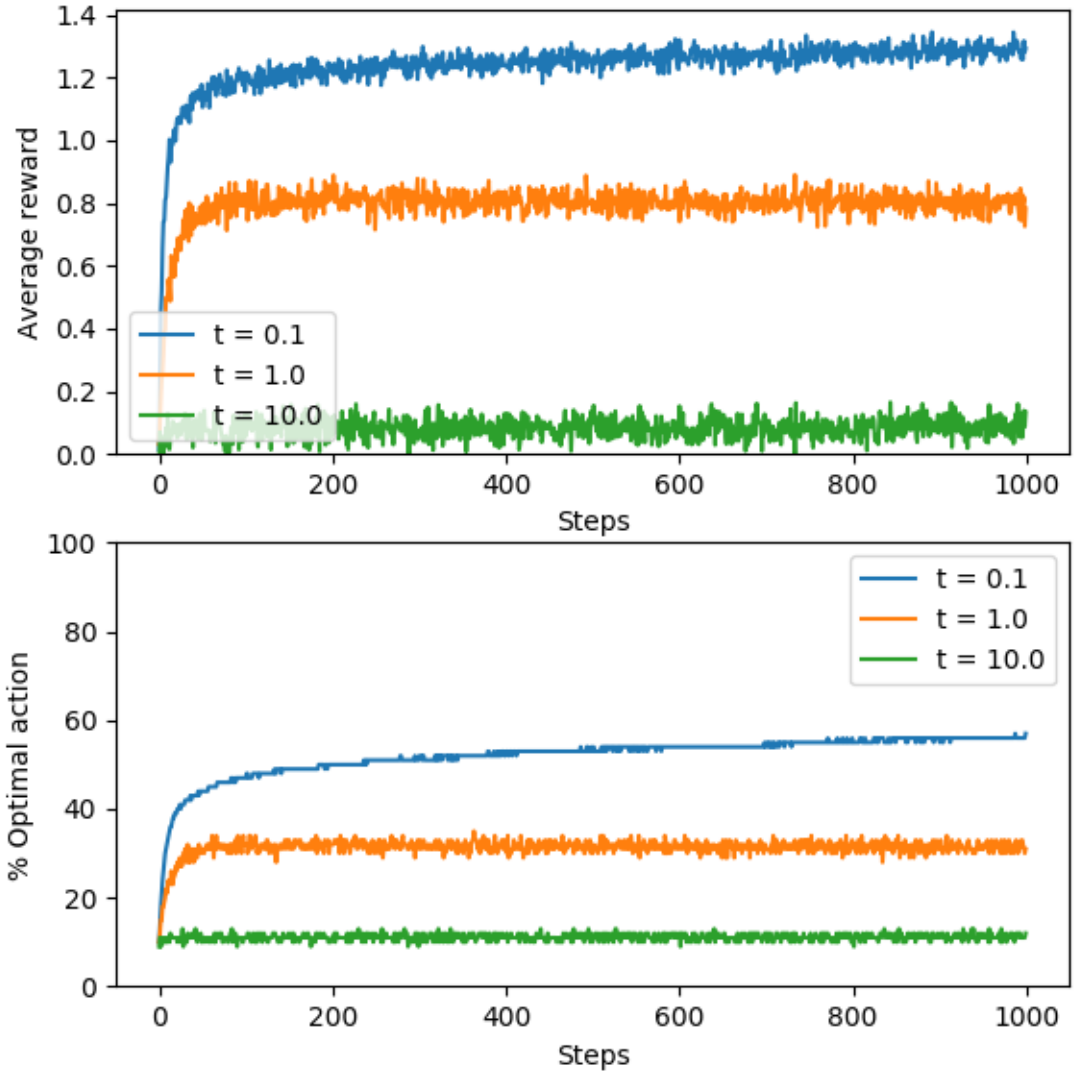

Figure 2: softmax

Inference: As we increase temperature, algorithm tries to pick actions more randomly as a result average reward decreases and also the percentage of optimal action decreases. As we increase temperature, the difference in probability of selecting two arm decreases, so algorithm explores more and less number of times optimal action is picked.

3 UCB-1

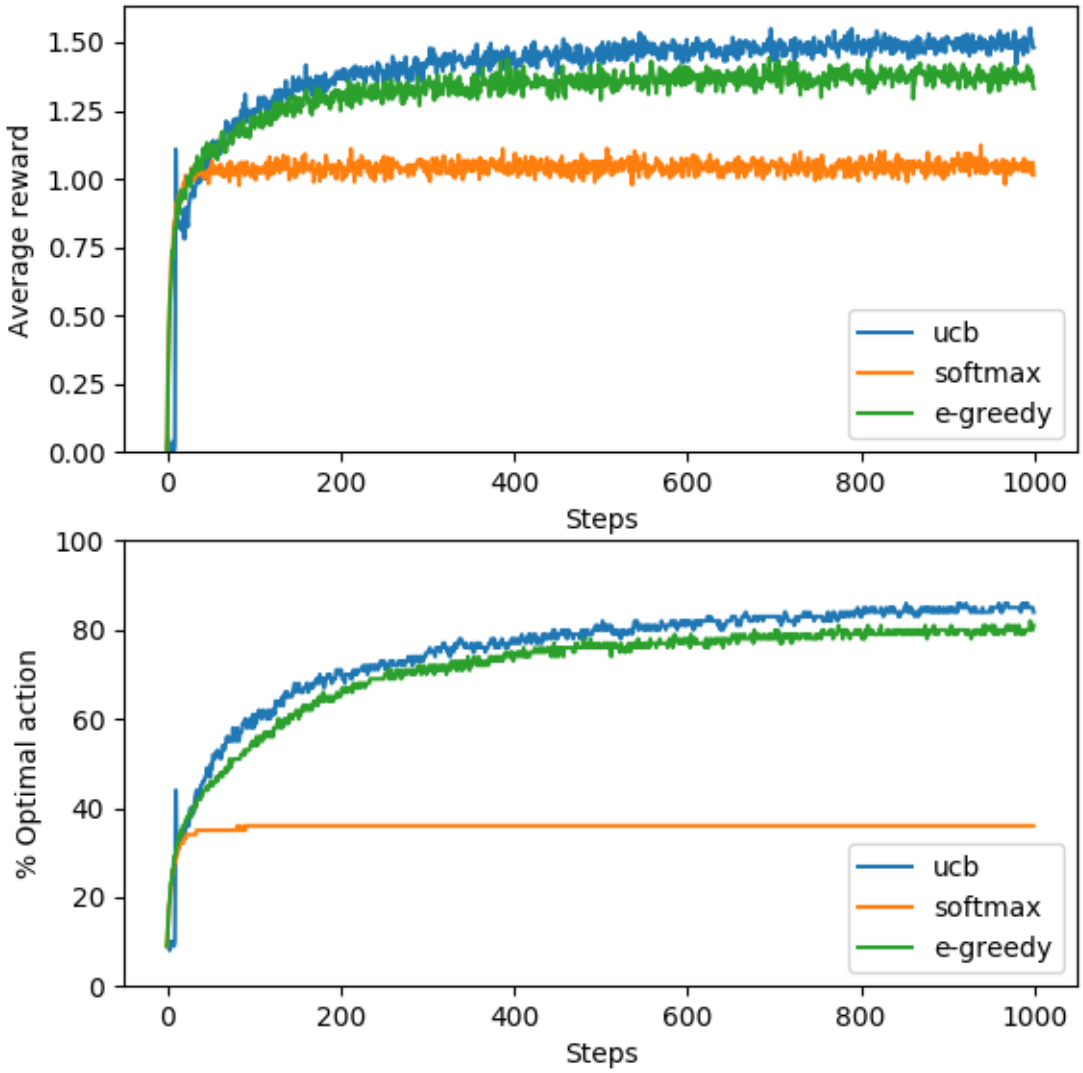

Figure 3: ucb-1

Inference: UCB-1 algorithm performs better than both ϵ-greedy and softmax. UCB keeps a balance between exploration and exploitation. It explores actions based on how close it's value to maximal value and how uncertain it's value is.

4 Median Elimination

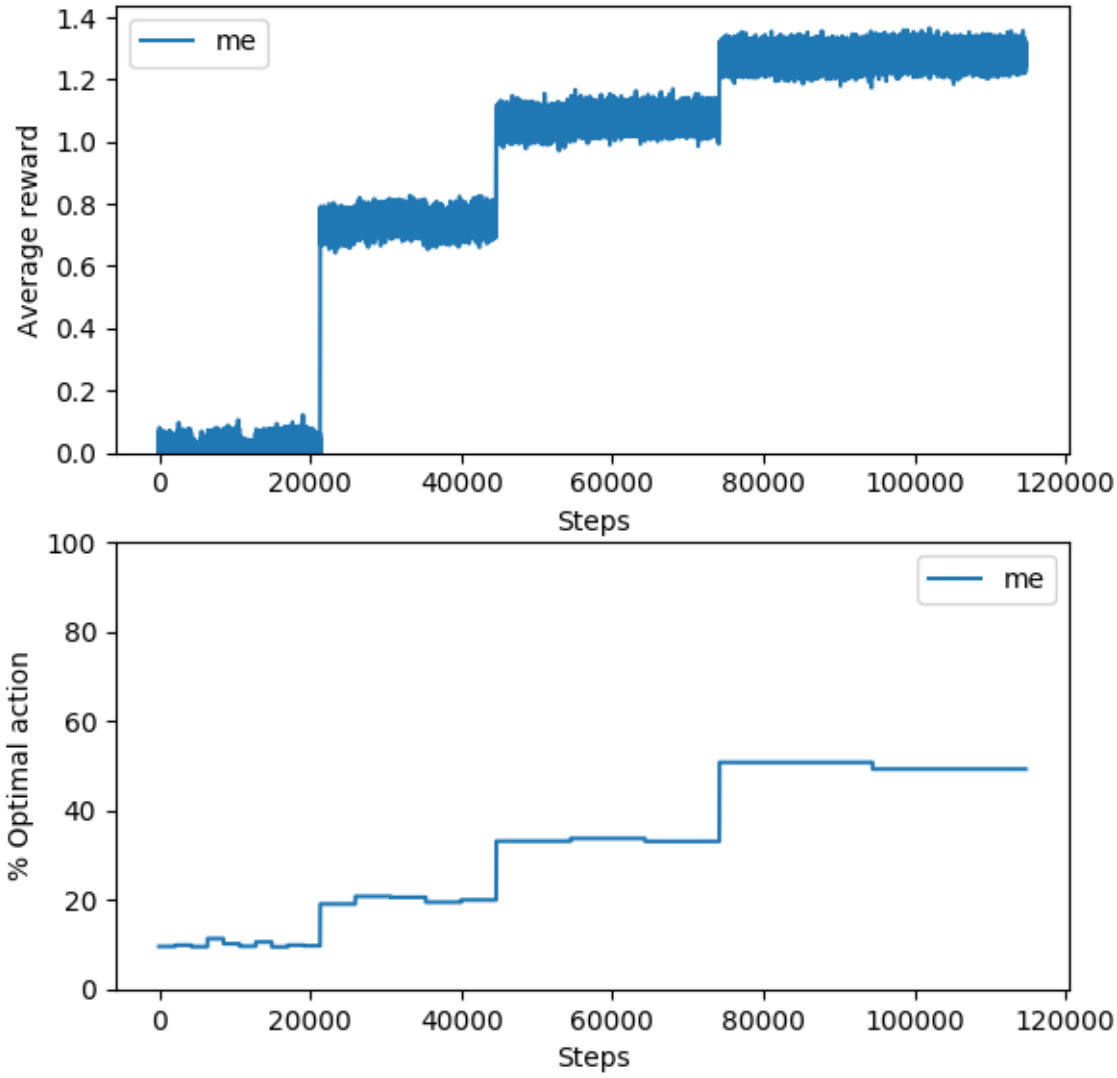

Figure 4: median elimination

Inference: Since median elimination takes lots of steps to find ϵ-optimal action, it is difficult to compare with other algorithms. Median elimination performs poor for less number of steps, since it explores a lot.

5 1000 Arm Bandit

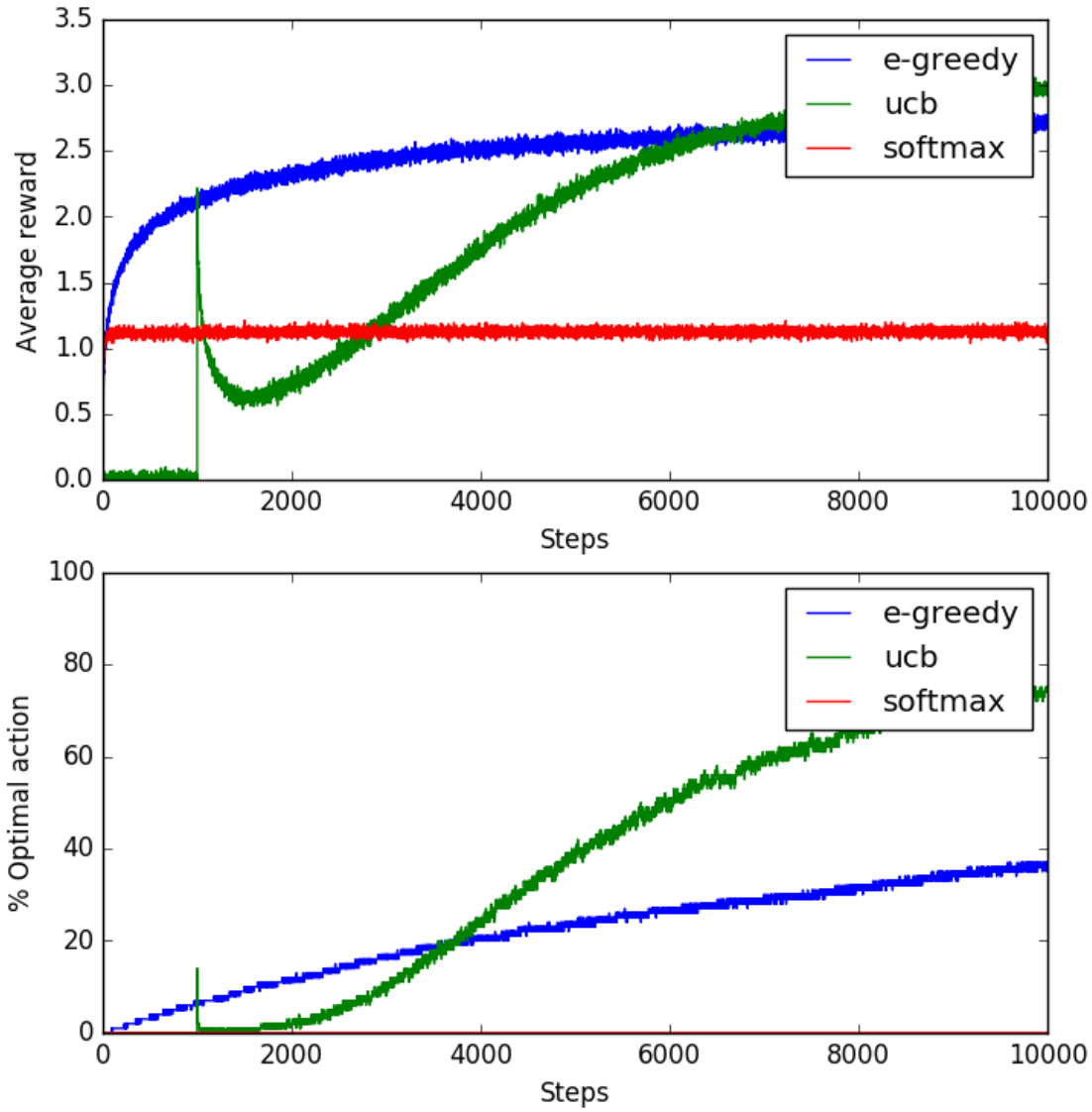

Figure 5: 1000 Arm Bandit

Inference: As the number of arms increases, steps required to converge to optimal action increases. Softmax performs worst. Initially UCB-1 performance is poor than ϵ-greedy as it pulls each actions once at the beginning. At longer time UCB-1 performs better than other algorithm.