

ME537: Homework 2

Kevin Kemper

For all algorithms and trials: $\epsilon = 0.1$, $\alpha = 0.001$, $\gamma = 0.1$ and $T = 100$

1 Part 1 - n-Arm bandit

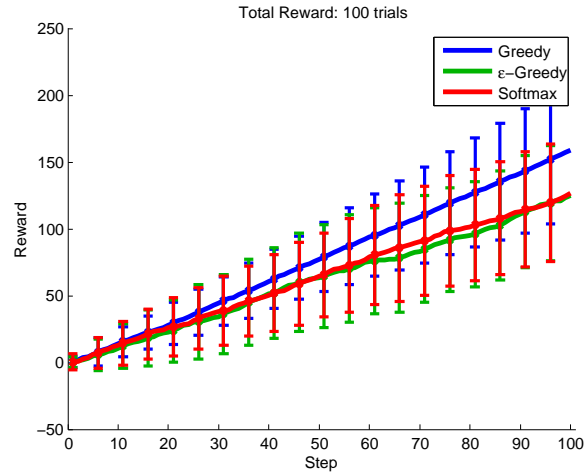


Figure 1: Results for three different action-value functions. Greedy ends up performing better but the variance is very high for all cases.

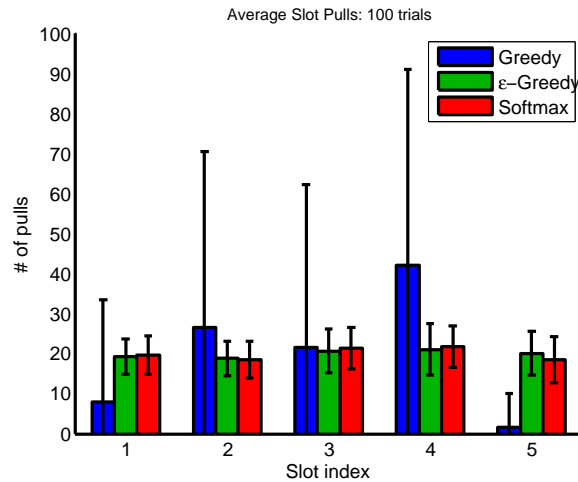


Figure 2: Number of slot pulls for the three different action-value functions. Greedy tended to choose slot 4 while the others had even distribution.

2 Part 2 - Gridworld

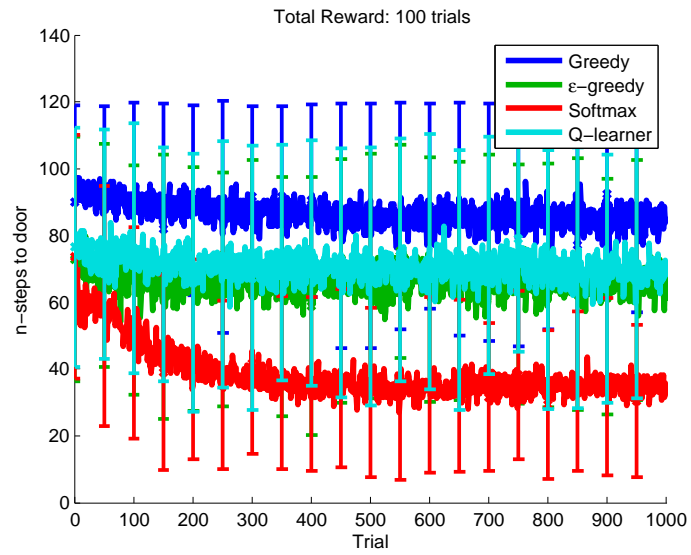


Figure 3: Average number of steps to find the goal. Softmax outperformed all the others with ϵ -greedy doing the next best.