EECS 498 HW1

Casper Guo

September 2023

1 Bandits

1. $(1 - 0.5) + \frac{\epsilon}{2} = 0.75$
2. •
3. The $\epsilon = 0.01$ method will perform the best in the long run in terms of both cumulative reward and probability of selecting the best action. Both ϵ-greedy methods are guarenteed to eventually identify the local actions but then the $\epsilon = 0.1$ will select it 91% of the time whereas the $\epsilon = 0.01$ method will select it 99.1% of the time. Thus the later method will eventually achieve higher average reward.

2 MDP

3 DP

4 Monte Carlo