Artificial Intelligence

Lab Assignment 2

Markov Decision Process
BANDIT

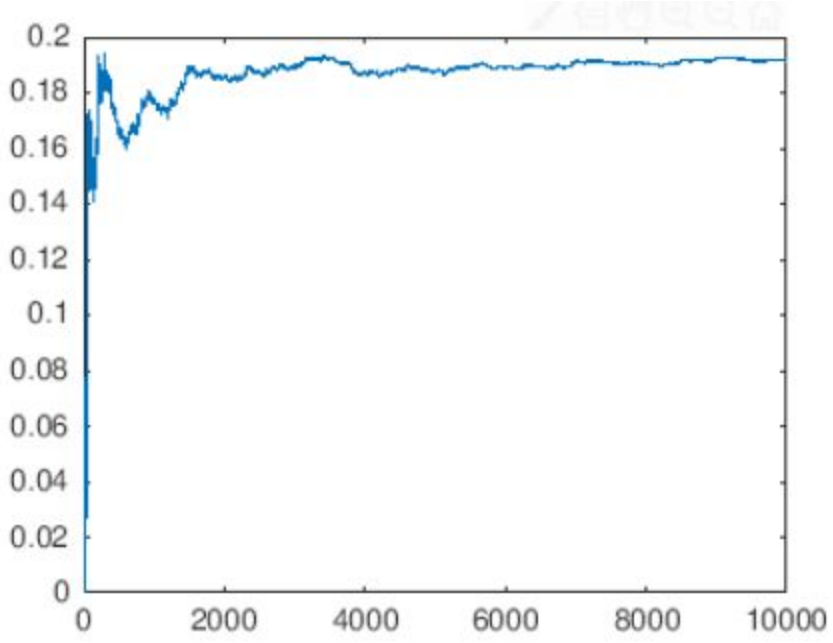
Group 02
Radheyshyam Jangid 201652020
Kulshreshtha goyal 201652009
Vishal Katiyar 201652030

Problem 1

Consider a binary bandit with two rewards 1-success, 0-failure. The bandit returns 1 or 0 for the action that you select, i.e. 1 or 2. The rewards are stochastic (but stationary). Use the epsilon-greedy algorithm discussed in class and decide upon the action to take for maximizing the expected reward.

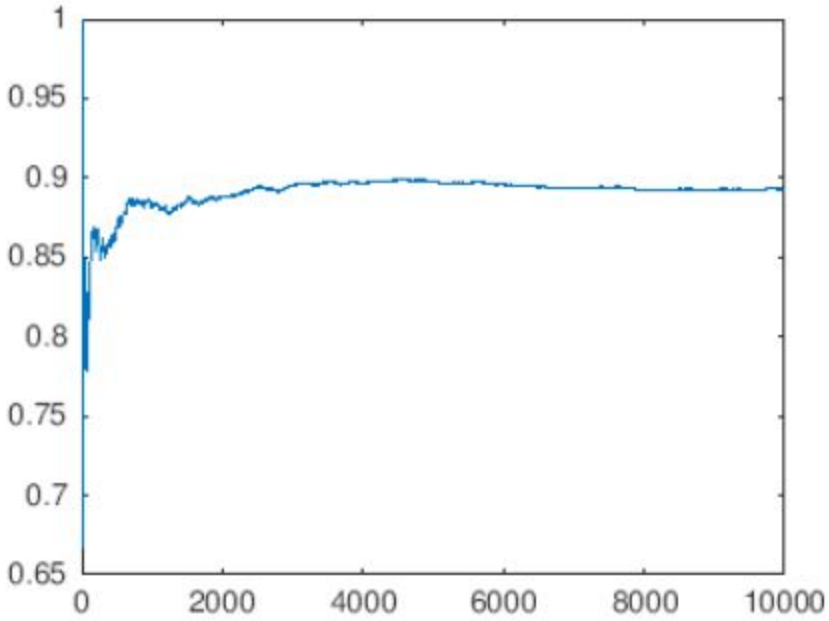
Observation

For binaryBanditA() the maximum reward is converging between 0.1 and 0.2 over 1000 iterations and reward values for the two actions of bandit A are shown in g. below. From the values and the graphs we can conclude that action 2 produces more reward as compared to action 1 in long term.



Binary BanditA(iteration vs reward)

For binaryBanditB() the maximum reward is converging between 0.8 and 0.9 over 1000 iterations and reward values for the two actions of bandit B are shown in g. below. From the values and the graphs, we can conclude that action 2 produces more reward as compared to action 1 in long term.



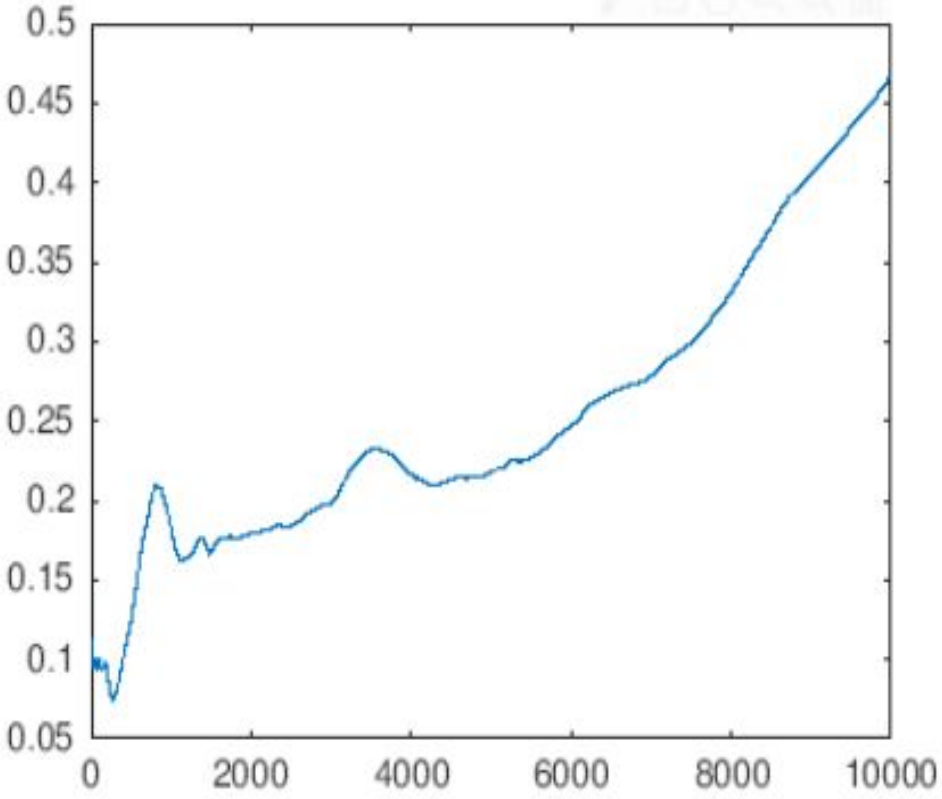
BinaryBanditB (Iteration vs reward)

Problem 2

Develop a 10-armed bandit in which all ten mean-rewards start out equal and then take independent random walks (by adding a normally distributed increment with mean zero and standard deviation 0.01 to all mean-rewards on each time step). function [value] = bandit nonstat(action).

Observation

A standard greedy-epsilon method is used to determine which action produces more reward but only for stationary rewards. And on increasing the value of epsilon randomness increases which increases the probability of exploration. Therefore, it is not converging using this method and we need a modified algorithm to make it converge.



Non-stationary bandit (iteration vs rewards) using a standard epsilon-greedy algorithm