PSG College of Technology

Department of Applied Mathematics and Computational Sciences

VIII Semester DS

Reinforcement Learning Lab

Assignment 2

Implement the following three bandit algorithms:

1. Incremental Uniform. This algorithm repeatedly loops through the arms pulling each arm once each time through the loop. The number of pulls for any two arms will never differ by more than 1. The average rewards for each arm are tracked and for the simple regret objective, the arm with the best average reward is returned as the best arm.

2. (**ε)** epsilon-Greedy. This is the -Greedy algorithm from the course notes, where 0 < **ε** < 1 is a parameter to the algorithm. The arm that currently looks best is selected with probability and otherwise a random arm is selected from among the other arms. Note that if there are k arms, then the (1/k)-Greedy algorithm will behave very much like Incremental Uniform, it is just a randomized version of that approach.

3. UCB. This is the UCB algorithm from the notes. In the case of simple regret, when asked to return the best arm, UCB returns the arm that accumulated the largest average reward (it ignores the exploration term of the UCB rule for picking the best arm).

Refer : https://rl-lab.com/multi-armed-bandits/