Assignment II: Multi-Arm Bandit for Article Recommendation

This assignment is designed to give you practice with bandit algorithms. The assignment includes two modules: simulator, which implements the simulation environment, and bandit scripts, which include the implementation of classic bandit algorithms. We will apply bandit algorithms on an article recommendation scenario. The environment consists of articles and users. At each time step, the algorithm selects an article for each user and makes a recommendation to it, then receives a reward from the user for the recommended article.

Simulator: The reward of different articles is assumed to be independent across articles and users. Each article has an index and its reward for a user is a gaussian distribution with a mean randomly selected from [1,5] and a variance randomly selected from [0,1]. Note that we are in a stationary setting, i.e. the parameters of the reward distribution of articles remain fixed over time.

MAB: Implement the following bandit algorithms:

- Greedy
- ε-greedy
- UCB

You may choose the techniques for learning action values from the lectures. The average reward at each step is the average reward over all users.

The algorithms and simulator consider the following hyper parameters: number of articles, number of users, number of iterations, ε (in ε-greedy), c (in UCB), and Q_0 (initial reward).

What to include in your report:

(default hyperparameters: ε=0.1, c=2, Q_0=1 for all action)

- Clearly state your design choices including the techniques you use for learning action values
- Average reward vs. number of steps (0 to 1K, step size 250)
- Parameter study plot of ε={0.1, 0.2, 0.5}, c={1,2}, and Q_0={0,2,5} for all action, for all implemented algorithms (report average reward over 1K first steps for each parameter setting)
- Average reward over 1K first steps vs. number of articles (20, 200, 400, 600, 800, 1K)
- Average reward over 1K first steps vs. number of users (20, 200, 400, 600, 800, 1K)
- Any interesting and insightful plot of your choice
- Your insights and analyses over the above plots.