Zmh6339
AI & Machine Learning, CS-UH 3260 Ziad Hassan
Professor Keith Ross January 31, 2024

Assignment 1

Part 1: Continuous Bandit Algorithm

Part 2: Theory

a) *Proof.*

For the algorithm to consider taking the *one* greedy action, two scenarios must be taken into consideration.

The first scenario is that the algorithm decides to take the greedy action, which occurs with probability $1 \cdot (1-\epsilon)$, where ϵ is the probability of taking a random action, and 1 is the probability of taking the one greedy action.

The second scenario is that the algorithm decides to take a random action, which occurs with probability ϵ.
Additionaly, the algorithm chooses a random action with an equal probability for each action; so, the probability of choosing the greedy action is $\frac{\epsilon}{k}$, where k is the number of actions.

Given that the greedy action can be chosen during exploration *or* exploitation, the above probabilities must be added together.
Therefore, the probability of the algorithm taking the greedy action is $(1-\epsilon) + \frac{\epsilon}{k}$.

■

b) i) *Proof.*

To determine the probability that the greedy action was chosen for the first time at time T, we need to consider that it was not chosen at any time before T, and that it was chosen at time T.
Thus, the following equation should be quantified:

$$P(\text{greedy at } T) = P(\text{not greedy before } T) \cdot P(\text{greedy at } T)$$

Therefore, the probability that the greedy action was chosen for the first time at time T is:

$$\begin{aligned} P(\text{greedy at } T) &= P(\text{not greedy before } T) \cdot P(\text{greedy at } T) \\ &= \left(1 - 1 + \epsilon - \frac{\epsilon}{k}\right)^{T-1} \cdot \left(1 - \epsilon + \frac{\epsilon}{k}\right) \\ &= \left(\epsilon - \frac{\epsilon}{k}\right)^{T-1} \cdot \left(1 - \epsilon + \frac{\epsilon}{k}\right) \end{aligned}$$

■

ii) *Proof.*

To get the expected number of steps, $\mathbb{E}(T)$, until the the greedy action is chosen for the first time is a sum over all possible time steps, each weighted by its probability of being the first time the greedy action is chosen.
Thus, the following equation should be quantified:

$$\mathbb{E}(T) = \sum_{t=1}^{\infty} t \cdot P(\text{greedy at } t)$$

Following, the equation is

$$\begin{aligned}\mathbb{E}(T) &= \sum_{t=1}^{\infty} t \cdot P(\text{greedy at } t) \\ &= \sum_{t=1}^{\infty} t \cdot \left(\epsilon - \frac{\epsilon}{k}\right)^{t-1} \cdot \left(1 - \epsilon + \frac{\epsilon}{k}\right)\end{aligned}$$

It can be observed that the above equation is a geometric series, which can be simplified to the following:

$$\begin{aligned}\mathbb{E}(T) &= \sum_{t=1}^{\infty} t \cdot \left(\epsilon - \frac{\epsilon}{k}\right)^{t-1} \cdot \left(1 - \epsilon + \frac{\epsilon}{k}\right) \\ &= \frac{1}{\left(\epsilon - \frac{\epsilon}{k}\right)^{t-1} \cdot \left(1 - \epsilon + \frac{\epsilon}{k}\right)}\end{aligned}$$

The above simplifciation is valid due to the definiton of the expected value of a geometric series.

■