Autumn Semester 2020 **March 19, 2021**

IE 613: Online Machine Learning

Solutions: Assignment 1

Solution 4

Given: Let K denotes the number of arms. A policy π is function which determines the probability distribution of the action that should be played next in a given sequence of past actions and rewards. We have a deterministic policy π and environment ν. Output is a single action for a deterministic policy. Given a deterministic policy π it will be shown that, there exists a sequence $X \in [0,1]^{K\times T}$ for T number of rounds. We define the expected regret as :

$$\mathcal{R}_n(\pi,\nu) = \max_{i\in[K]} \sum_{t=1}^{T} X_{ti} - \mathbb{E}\left[\sum_{t=1}^{T} X_{tI_t}\right]$$

To show: For any deterministic policy π there exists a $\nu \in [0,1]^{K\times T}$ such that

$$\mathcal{R}_T(\pi,\nu) \geq T\left(1-\frac{1}{K}\right)$$

Proof:. For a given deterministic policy π we will construct the bandit sequence as follows:

The policy being deterministic in nature, the adversary may construct a reward sequence $X^* \in [0,1]^{K\times T}$ as,

$$X_{ti} = \begin{cases} 0, & \text{if } i = I_t \\ 1, & \text{if } i \neq I_t \end{cases} \tag{1}$$

i.e. giving 0 reward to the arm selected by the policy at time t. Hence, total aggregated reward by this policy is 0.

It will be possible to define such a sequence of rewards because the policy π is deterministic in nature. According to the defined X^* we will have,

$$\mathbb{E}\left[\sum_{t=1}^{T} X_{tI_t}\right] = 0 \tag{2}$$

$$\Rightarrow \mathcal{R}_n(\pi,\nu) = \max_{i\in[K]} \sum_{t=1}^{T} X_{ti}$$

Now for such a sequence X^* we want to determine the arm which was pulled by the least number of times by the policy π and hence will gather maximum reward over the time horizon for the defined sequence X^*. When this arm will be pulled

it will earn a reward of 0 and other arms will have reward as 1. As this arm was pulled for the lowest number of times, the reward of other arms got lesser reward cumulatively. We know the total number of pulls will sum upto T. Since the least pulled arm can be pulled for atmost $\frac{T}{K}$ times because other arms are pulled more times than this and the total number of pull will sum upto T. So

$$\max_{i\in[K]} \sum_{t=1}^{T} X_{ti} \geq \frac{1}{K} \sum_{i=1}^{K} \sum_{t=1}^{T} X_{ti} = \frac{T(K-1)}{K}$$
$$\Rightarrow \max_{i\in[K]} \sum_{t=1}^{T} X_{ti} \geq T - \frac{T}{K} \tag{3}$$

Combining the results of Eq.(2) and (3) we get,

$$\mathcal{R}_n(\pi, \nu) \geq T - \frac{T}{K}$$

□

Solution 5

Given: The regret is defined by,

$$\mathcal{R}_T^{track}(\pi, \nu) = \mathbb{E}\left[\sum_{t=1}^{T} \max_{i\in[K]} x_{ti} - \sum_{t=1}^{T} x_{tI_t}\right]$$

where I_t is the arm chosen out of K number of arms by the policy π and x_{tI_t} is the reward observed in the round t. So, here I_t in round t now can be a random variable. The environment is $\nu \in [0,1]^{K\times T}$. At the first sight this definition seems like the right thing because it measures what we actually care about. Unfortunately, however, it gives the adversary too much power.

To show: For any policy π (randomized or not) there exists a $\nu \in [0,1]^{K\times T}$ such that

$$\mathcal{R}_T^{track}(\pi, \nu) \geq T\left(1 - \frac{1}{K}\right)$$

Proof:. The regret is given by,

$$\begin{aligned}\mathcal{R}_T^{track}(\pi, \nu) &= \mathbb{E}\left[\sum_{t=1}^{T} \max_{i\in[k]} x_{ti} - \sum_{t=1}^{T} x_{tI_t}\right] \\ &= \mathbb{E}\left[\sum_{t=1}^{T} \max_{i\in[K]} x_{ti}\right] - \mathbb{E}\left[\sum_{t=1}^{T} x_{tI_t}\right]\end{aligned}$$

Since in a round t, $\max_{i\in[k]} x_{ti}$ is a deterministic value

$$\Rightarrow \mathcal{R}_T^{track}(\pi,\nu) = \sum_{t=1}^{T} \max_{i\in[K]} x_{ti} - \mathbb{E}\left[\sum_{t=1}^{T} x_{tI_t}\right] \tag{1}$$

Now for a given policy π we will construct the bandit sequence as follows:

Let us consider that the arm is chosen as,

$$I_t^* = \arg\min_{i\in[k]} \mathbb{P}\left\{I_t = i\right\}$$

i.e in each round t, I_t^* returns that index value of arm $i \in [K]$ which has the minimum chances to get selected by the policy π.

The reward x_{ti} for each round t is given as,

$$x_{ti} = \begin{cases} 1, & \text{if } i = I_t^* \\ 0, & \text{otherwise} \end{cases} \tag{2}$$

In this above setting Eq. (2),

$$\mathbb{E}\left[x_{tI_t}\right] = \mathbb{P}\left\{I_t = I_t^*\right\}$$

I_t^* denotes that arm which has minimum chances to be selected and all other arms are chosen more than I_t^*.

$$\mathbb{E}\left[x_{tI_t}\right] \leq \frac{1}{K} \tag{3}$$

Therefore,

$$\begin{aligned} \mathbb{E}\left[\sum_{t=1}^{T} x_{tI_t}\right] &= \sum_{t=1}^{T} \mathbb{E}\left[x_{tI_t}\right] \\ &\leq \frac{T}{K} \quad \text{[By Eq3]} \end{aligned} \tag{4}$$

In each round x_{ti} assigns 1 to that arm which is selected for minimum number and $\max_{i\in[K]} x_{ti} = 1 \forall i \in [K]$. Summing these values over T rounds we will get

$$\sum_{t=1}^{T} \max_{i\in[K]} x_{ti} = T \tag{5}$$

Putting the values of Eq.(3) and Eq.(5) in Eq.(1),

$$\begin{aligned} \mathcal{R}_T^{track}(\pi,\nu) &\leq T - \frac{T}{K} \\ &= T(1 - \frac{1}{K}) \end{aligned} \qquad \square$$

Solution 6

Given: Let $p \in P_k$ be a probability vector where k denotes the number of arms and p_i is the i^{th} component of probability vector p. Suppose $\hat{X} : [k] \times \mathbb{R} \to \mathbb{R}$ is a function such that for all $x \in \mathbb{R}^k$, if A,

$$\mathbb{E}\left[\hat{X}(A, x_A)\right] = \sum_{i=1}^{k} p_i \hat{X}(i, x_i) = x_1$$

To show: There exists an $a \in \mathbb{R}^k$ such that $\sum_{j=1}^{k} a_j p_j = 0$ and $\hat{X}(i, x_i) = a_i + \frac{\mathbb{I}\{i=1\}x_1}{p_1}$

Proof:. It is given that,

$$\begin{aligned} \mathbb{E}\left[\hat{X}(A, x_A)\right] &= \sum_{i=1}^{k} p_i \hat{X}(i, x_i) \\ &= x_1 \end{aligned} \tag{1}$$

Let us choose $a \in \mathbb{R}^k$ such that $\hat{X}(i, x_i) = a_i + \frac{\mathbb{I}\{i=1\}x_1}{p_1}$ where,

$$\{i = 1\} = \begin{cases} 1, & \text{if } i = 1 \\ 0, & \text{otherwise} \end{cases} \tag{2}$$

Hence,

$$\begin{aligned} \mathbb{E}\left[\hat{X}(A, x_A)\right] &= \sum_{i=1}^{k} \left(a_i + \frac{\mathbb{I}\{i=1\}x_1}{p_1}\right) p_i \\ &= \sum_{i=1}^{k} a_i p_i + \sum_{i=1}^{k} \frac{\mathbb{I}\{i=1\}x_1}{p_1} p_i \end{aligned}$$

By Eq.(2) we get

$$\begin{aligned} &= \sum_{i=1}^{k} a_i p_i + \frac{x_1}{p_1} p_1 \\ &= \sum_{i=1}^{k} a_i p_i + x_1 \end{aligned} \tag{3}$$

Since $\mathbb{E}\left[\hat{X}(A, x_A)\right] = x_1$, so equating with Eq. (3) we get, $\sum_{i=1}^{k} a_i p_i = 0$. Therefore, there exists an $a \in \mathbb{R}^k$ as $\sum_{j=1}^{k} a_j p_j = 0$ and $\hat{X}(i, x_i) = a_i + \frac{\mathbb{I}\{i=1\}x_1}{p_1}$ such that $\mathbb{E}\left[\hat{X}(A, x_A)\right] = x_1$. □