IE613: Online Machine Learning **Jan-Apr 2020**

Assignment 1: February 7

Instructions: *You are free to code in Python/Matlab/C/R. Discussion among the class participants is highly encouraged. But please make sure that you understand the algorithms and write your own code. If you share any code with any other student then you will be penalized and can be given 0 mark for that question.*

Submit the code and report by 11:59PM, 16th February on Moodle. Late submission will not be evaluated and given 0 mark. **This assignment has 100 Points.**

Question 1 (Full information setting, 20 Points) *Consider the problem of prediction with expert advice with $d = 10$. Assume that the losses assigned to each expert are generated according to independent Bernoulli distributions. The adversary/environment generates loss for experts 1 to 8 according to $Ber(0.5)$ in each round. For the 9th expert, loss is generated according to $Ber(0.5 - \Delta)$ in each round. The losses for the 10th expert are generated according to different Bernoulli random variable in each round– for the first $T/2$ rounds, they are generated according to $Ber(0.5 + \Delta)$ and the remaining $T/2$ rounds they are generated according to Bernoulli random variable $Ber(0.5 - 2\Delta)$. $\Delta = 0.1$ and $T = 10^5$. Generate (pseudo) regret values for different learning rates (η) for Weighted Majority algorithm. The averages should be taken over at least 20 sample paths (more is better). Display 95% confidence intervals for each plot. Vary c in the interval [0.1 2.1] in steps of size 0.2 to get different learning rates. Implement Weighted Majority algorithm with $\eta = c\sqrt{2\log(d)/T}$.*

Question 2 (Bandit setting, 30 Points) *Consider the problem of multi-armed bandit with $K = 10$ arms. Assume that the losses are generated as in Question 1. For each of the following algorithms generate (pseudo) regret for different learning rates (η) for each of the following algorithms. The averages should be taken over atleast 50 sample paths (more is better). Display 95% confidence intervals for each plot. Vary c in the interval [0.1 2.1] in steps of size 0.2 to get different learning rates.*

- *EXP3. Set $\eta = c\sqrt{2\log(K)/KT}$.*
- *EXP3.P. Set $\eta = c\sqrt{2\log(K)/KT}$, $\beta = \eta$, $\gamma = K\eta$.*
- *EXP-IX. Set $\eta = c\sqrt{2\log(K)/KT}$, $\gamma = \eta/2$.*

Question 3 (5 Points) *In Question 2, which one of EXP3, EXP3.P and EXP3-IX performs better and why?*

Question 4 (10 Points) *Show that for any deterministic policy π there exists an environment ν such that $R_T(\pi, \nu) \geq T(1 - 1/K)$ for T rounds and K arms.*

Question 5 (15 Points) *Suppose we had defined the regret by*

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

where I_t is the arm chosen by the policy π and x_{tI_t} is the reward observed in the round t. At first sight this definition seems like the right thing because it measures what you actually care about. Unfortunately,

however, it gives the adversary too much power. Show that for any policy π (randomized or not) there exists a $\nu \in [0,1]^{K\times T}$ such that

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

Question 6 (15 Points) *Let $p \in \mathcal{P}_k$ be a probability vector and suppose $\hat{X} : [k]\times\mathbb{R}\to\mathbb{R}$ is a function such that for all $x\in\mathbb{R}^k$, if $A\sim p$,*

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

Show there exists an $a\in\mathbb{R}^k$ such that such that $\sum_{j=1}^k a_j p_j = 0$ and $\hat{X}(i,x) = a_i + \frac{\mathbb{I}\{i=1\}x_1}{p_1}$.

Question 7 (5 Points) *Suppose we have a two-armed stochastic Bernoulli bandit with $\mu_1 = 0.5$ and $\mu_2 = 0.55$. Test your implementation of EXP3 from the Question 2. What happens when $T = 10^5$ and the sequence of rewards is $x_{t1} = \mathbb{I}\{t \leq T/4\}$ and $x_{t2} = \mathbb{I}\{t > T/4\}$?*

Submission Format and Evaluation: Your should submit a report along with your code. Please zip all your files and upload via Moodle. The zipped folder should named as YourRegistrationNo.zip e.g.154290002.zip. The report should contain two figures: first figure should have one plot corresponding to algorithm in Q.1 and the other should have 3 plots one corresponding to each algorithm in Q.2. For each figure, write a brief summary of your observations. We may also call you to a face-to-face session to explain your code.

Note: Please calculate (pseudo) regret for each algorithm in Q.2 for a given set of parameters as follows:

Let μ_t^i denote the mean of arm i in round t. Suppose an adversary generates sequence of loss vectors $\{l_t\}_{t=1}^T$ and an algorithm generates sequence of pulls $\{I_t\}_{t=1}^T$, the (pseudo) regret for this sample path is

$$\sum_{t=1}^{T}\mathbb{E}[l_t(I_t)] - \min_i \sum_{t=1}^{T}\mathbb{E}[l_t(i)] \tag{1.1}$$

$$= \sum_{t=1}^{T}\mu_t^{I_t} - \min_i \sum_{t=1}^{T}\mu_t^i \tag{1.2}$$

Note that in this calculation we only considered the mean values of losses, not the actual losses suffered. It is Okay if this value turns out to be negative. There is no expectation over random choices of I_ts here. Now generate 20 such sample paths and take their average.