IE613: Online Machine Learning **Jan-Apr 2019**

Assignment 2: February 13

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, 23th February on Moodle. Late submission will not be evaluated and given 0 mark. **This assignment has 100 Points.**

Question 1 (30 points) *Consider a K-armed bandit problem where each arm is a Bernoulli random variable. Fix a bandit instance for each $K = 10, 20, 30$ as follows:*

$$\mu_i = \begin{cases} \frac{1}{2} \text{ if } i = 1 \\ \frac{1}{2} - \frac{i}{70} \text{ if } i = 2, 3, \cdots, K \end{cases}$$

where μ_i denotes the mean reward of i^{th} arm. Generate plots for cumulative regret vs number of samples (T) for following algorithms.

1. *ϵ_t-greedy with $\epsilon_t = \frac{1}{t}$*
2. *UCB with $\alpha = 1.5$*
3. *Thompson Sampling*
4. *KL-UCB with $c = 0$*

The averages should be taken over at-least 20 sample paths (more is better). Display 95% confidence intervals for each plot for $T = 25000$.

Question 2 (5 points) *Suppose that X is $\sigma-$subgaussian and X_1 and X_2 are independent and σ_1 and σ_2-subgaussian respectively, then:*

1. *$\mathbb{E}[X]=0$ and $\mathbb{V}ar[X] \leq \sigma^2$.*
2. *cX is $|c|\sigma-$subgaussian for all $c \in \mathbb{R}$.*
3. *$X_1 + X_2$ is $\sqrt{\sigma_1^2 + \sigma_2^2}-$subgaussian.*

Question 3 (5 points) *Suppose that X is zero-mean and $X \in [a, b]$ almost surely for constants $a < b$.*

1. *Show that X is $(b-a)/2-$subgaussian.*
2. *Using Cramer-Chernoff method shows that if $X_1, X_2, \ldots, X_n$ are independent and $X_t \in [a_t, b_t]$ almost surely with $a_t < b_t$ for all t, then prove*

$$\mathbb{P}\left(\sum_{t=1}^{n}(X_t - \mathbb{E}[X_t]) \geq \epsilon\right) \leq \exp\left(-\frac{2\epsilon^2}{\sum_{t=1}^{n}(b_t - a_t)^2}\right)$$

Question 4 (10 points) *Show that*

$$R_T \leq \min\left\{T\Delta, \Delta + \frac{4}{\Delta}\left(1 + \max\left\{0, \log\left(\frac{T\Delta^2}{4}\right)\right\}\right)\right\}$$

implies the regret of an optimally tuned Explore-then-Commit (ETC) algorithm for subgaussian $2-$*armed bandits with means* $\mu_1, \mu_2 \in \mathbb{R}$ *and* $\Delta = |\mu_1 - \mu_2|$*, satisfies* $R_T \leq \Delta + C\sqrt{T}$ *where* $C > 0$ *is a universal constant.*

Question 5 (10 points) *Fix* $\delta \in (0, 1)$*. Modify the ETC algorithm to depend on* δ *and prove a bound on the pseudo-regret* $R_T = T\mu^\star - \sum_{t=1}^{T} u_{A_t}$ *of ETC algorithm that holds with probability* $1 - \delta$ *where* A_t *is the arm chosen in the round* t*.*

Hint: *Choose 'm' appropriately in the regret upper bound of ETC algorithm which is proved in the class.*

Question 6 (10 points) *Fix* $\delta \in (0, 1)$*. Prove a bound on the random regret* $R_T = T\mu^\star - \sum_{t=1}^{T} X_t$ *of ETC algorithm that holds with probability* $1 - \delta$*. Compare this to the bound derived for the pseudo-regret in the question 5. What can you conclude?*

Question 7 (15 points) *Assume the rewards are* $1-$*subgaussian and there are* $k \geq 2$ *arms. The* $\epsilon-$*greedy algorithm depends on a sequence of parameters* $\epsilon_1, \epsilon_2, \ldots$*. First it chooses each arm once and subsequently chooses* $A_t = \arg\max_i \hat{\mu}_i(t-1)$ *with probability* $1 - \epsilon_t$ *and otherwise chooses an arm uniformly at random.*

1. *Prove that if* $\epsilon_t = \epsilon > 0$*, then* $\lim_{T\to\infty} \frac{R_T}{T} = \frac{\epsilon}{k}\sum_{i=1}^{k} \Delta_i$*.*
2. *Let* $\Delta_{min} = \min\{\Delta_i : \Delta_i > 0\}$ *where* $\Delta_i = \mu^\star - \mu_i$*, and* $\epsilon_t = \min\left\{1, \frac{Ck}{t\Delta_{min}^2}\right\}$ *where* $C > 0$ *is a sufficiently large universal constant. Prove that there exists a universal* $C' > 0$ *such that*

$$R_T \leq C' \sum_{i=1}^{k}\left(\Delta_i + \frac{\Delta_i}{\Delta_{min}^2}\log\max\left\{e, \frac{T\Delta_{min}^2}{k}\right\}\right).$$

Question 8 (15 points) *Fix a* $1-$*subgaussian* $k-$*armed bandit environment and a horizon* T*. Consider the version of UCB that works in phases of exponentially increasing length of* $1, 2, 4, \ldots$*. In each phase, the algorithm uses the action that would have been chosen by UCB at the beginning of the phase.*

1. *State and prove a bound on the regret for this version of UCB.*
2. *How would the result change if the* l^{th} *phase had a length of* $\lceil \alpha^l \rceil$ *with* $\alpha > 1$*?*

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 one figure with four plots corresponding to each algorithm in Q.1. Write a brief summary of your observations. We may also call you to a face-to-face session to explain your code.