IE613: Online Machine Learning **Jan-Apr 2018**

Assignment 3: April 18

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, 18^{th} April on Moodle. Late submission will not be evaluated and given 0 mark.

Question 1 *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 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 $c = 0$*
5. *UCB-V with $\beta = 1.2$ and $c = 1$*

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 (Pure exploration, best arm selection) *Consider a K-armed bandit problem where each arm is a Bernoulli random variable. We would like to identify the best arm with probability at-least $1 - \delta$, i.e., $\Pr\{\hat{I} = I^*\} \geq 1 - \delta$, where $\hat{I}$ is the estimated best arm and I^* is the best arm.*

Consider a K-armed bandit problem where each arm is a Bernoulli random variable. Fix a bandit instance for each $K = 5, 10, 15, 20, 25$ as follows:

$$\mu_i = \begin{cases} \frac{3}{4} \text{ if } i = 1 \\ \frac{3}{4} - \frac{i}{40} \text{ if } i = 2, 3, \cdots, K \end{cases}$$

where μ_i denotes the mean of i^{th} arm. Run the following algorithms for different values of K. Set confidence parameter, $\delta = 0.1$.

- *KL-LUCB [1] with parameters $\epsilon = 0, \alpha = 2, \kappa_1 = 4e + 4, \beta(t, \delta) = \log(\kappa_1 K t^\alpha/\delta) + \log\log(\kappa_1 K t^\alpha/\delta)$.*
- *Median Elimination [2] with parameters $\epsilon = 0.01$.*

Run both algorithms at-least 10 times (more is better!) and average the sample complexities for each problem instance. Similarly mistake probability should be calculated as the fraction of times non best arm is returned till stop signal is received. In one figure, plot average sample complexity v/s number of arms and average mistake probability v/s number of arms with 95% confidence.

References

[1] E. Kaufmann and S. Kalyanakrishnan. *Information Complexity in Bandit Subset Selection*. COLT 2013.

[2] Even-Dar E, Mannor S, Mansour Y. *Action elimination and stopping conditions for the multi-armed bandit and reinforcement learning problems*. Journal of machine learning research. 2006

Submission Format and Evaluation: You 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 four figures: one figure for each K which should have five plots corresponding to each algorithm in Q.1 and one figure for 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.