CS6046 Problem Set 4

Instructor: Dr. Kota Srinivas Reddy

Jul - Nov 2023, Deadline : 16/10/2023

1. (10 marks) Consider a 1-subgaussian k-armed bandit environment and a horizon n. Consider the version of UCB that works in phases of exponentially increasing length $1, 2, 4 \ldots$. That is, in each phase, the algorithm plays the action that would have been chosen by UCB at the beginning of the phase for an exponential number of times.

   Input k and δ
   Choose each arm once
   for $l = 1, 2, \ldots$ **do**
      Compute $A_l = \arg\max_i UCB_i(t-1, \delta)$
      Choose arm A_l exactly 2^l times
   end for

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

2. (20 marks) In this exercise you will investigate the empirical behavior of UCB on a two-armed Gaussian bandit with means $\mu_1 = 0$ and $\mu_2 = -\Delta$. The horizon is set to $n = 1000$, and the sub-optimality gap Δ is varied between 0 and 1 as follows: $\Delta \in \{0, 0.1, 0.2, ..., 1\}$. Plot the expected regret of UCB relative to ETC for a variety of choices of commitment time m. Repeat the experiment 100 times for each value of Δ, and take the average value to get the average regret. Explain your results.

3. (10 marks) Let $\mathbb{V}_t[U] \triangleq \mathbb{E}_t[(U - \mathbb{E}_t[U])^2]$. Find $\mathbb{V}_{t-1}[\hat{y}_{t,i}]$.

4. (Practice) Show that $1 - \frac{1}{x} \leq \ln x \leq x - 1 \quad \forall x > 0$.