CS6046: Multi-Armed Bandits
Assignment 3
Course Instructor : Arun Rajkumar.
Release Date : May-18, 2021
Submission Date: On or before 5 PM on June-06,2021

SCORING: There are 2 question in this assignment which contributes 10 points each towards your final grade.

WHAT SHOULD YOU SUBMIT? You should submit a zip file titled 'Solutions_ roll-number.zip' Your assignment will NOT be graded if it does not contain all of the following:

- A PDF file which includes explanations regarding each of the solution as required in the question. Title this file as 'Report.pdf'
- Source code for all the programs that you write for the assignment clearly named.

CODE LIBRARY: You are expected to code all algorithms from scratch. You cannot use standard inbuilt libraries for the algorithms. You are free to use inbuilt libraries for plots. You can code using either Python or Matlab or C.

GUIDELINES: Keep the below points in mind before submission.

- Plagiarism of any kind is unacceptable. These include copying text or code from any online sources. These will lead to disciplinary actions according to institute guidelines.
- Any graph that you plot is unacceptable for grading unless it labels the x-axis and y-axis clearly.
- Don't be vague in your explanations. The clearer your answer is, the more chance it will be scored higher.

LATE SUBMISSION POLICY You are expected to submit your assignment on or before the deadline to avoid any penalty. Late submission incurs a penalty equal to the number of days your submission is late by.

Q1:Gaussian Explore Then Commit : Investigate the empirical behaviour of the Explore then Commit algorithm when the loss of the two arms are standard Gaussian random variables with mean $\mu_1 = 0$ and $\mu_2 = \Delta$.

- Write a piece of code for a function that takes T and δ as input and returns the number of rounds T_0 needed for exploration which minimizes the regret.
- For $\Delta = 0.1$ and $T = 1000$, plot the regret as a function of T_0. Plot error bars for the regret for each value of T_0 over $10,000$ runs.
- Explain your insights from the plots and the theoretical bound you obtained in the first part of the question.
- Instead of standard Gaussian arms, use Gaussians with same mean as above but variance σ^2 for both arms. Try different choices of σ^2 and explain your insights for the above problem.

Q2: Thompson vs UCB
Consider the same setting of two armed Standard Gaussian Bandits as in Question 1.

- How will you adapt the UCB and Thompson sampling algorithms to work for this setting?
- Fix $T = 10000$ and $\Delta = 0.1$ and compare the regret bound of both the algorithms. What are your observations.
- Change Δ in steps of 0.1 from 0.1 till 1. How do these algorithms perform? What are your observations?