Assignment 1

EE675: Introduction to Reinforcement Learning

January 20, 2024

Instructions

- Kindly name your submission files as ‘`RollNo_Name_A1_PartA/B.ipynb`‘, based on the part you are submitting. Marks will be deducted for all submissions that do not follow the naming guidelines.
- You are required to work out your answers and submit only the iPython Notebook. The code should be well commented and easy to understand as there are marks for this.
- You may use the notebook given along with the assignment as a template. You are free to use parts of the given base code but may also choose to write the whole thing on your own.
- Submissions are to be made through HelloIITK portal. Submissions made through mail will not be graded.
- Answers to the theory questions, if any, should be included in the notebook itself. While using special symbols use the $\LaTeX$ mode
- Make sure your plots are clear and have title, legends and clear lines, etc.
- Plagiarism of any form will not be tolerated. If your solutions are found to match with other students or from other uncited sources, there will be heavy penalties and the incident will be reported to the disciplinary authorities.
- In case you have any doubts, feel free to reach out to TAs for help.

Part-A (Deadline - 28th Jan 2024)

(Chernoff Bound) `[10 Marks]` Suppose $X_1, X_2, \cdots, X_n$ are i.i.d. copies of a $\mathcal{N}(0, \sigma^2)$ r.v. Then for $X = \frac{1}{n}\sum_{i=1}^{n} X_i$ we know that

$$\mathbb{P}[X \geq \varepsilon] \leq \exp\left(\frac{-n\varepsilon^2}{2\sigma^2}\right)$$

.

Write a Python code to run Monte Carlo simulations that verify the inequality. Specifically, for a given ε and σ, generate n samples from the zero mean Gaussian distribution $x_1, x_2, \ldots, x_n$ and check whether the sample average is more than ε. Repeat this experiment 500 times and observe in how many experiments out of those 500 experiments, the sample average is more than ε. This will gives us an empirical estimate of $P[X \geq \varepsilon]$.

Take $\sigma = 0.1,\ \varepsilon = 0.05$, and plot the empirical estimate as a function of $n \in \{100, 200, \dots, 1000\}$. In the same plot, include the Chernoff upper bound as a function of n.

Part-B (Deadline - 4th Feb 2024)

(**Multi arm bandits | Explore-then-Commit and UCB**) `[20 Marks]` Consider a two-armed Bernoulli bandit scenario with true means given by $\mu_1 = \frac{1}{2}, \mu_2 = \frac{1}{2} + \Delta$, for some $\Delta < \frac{1}{2}$. Let the time horizon be $T = 10000$.

1. Take $\Delta = \frac{1}{4}$ and run the Monte Carlo simulations to estimate the expected regret of the ETC algorithm which explores each arm $m = T^{2/3}(\log T)^{1/3}$ times before committing. Specifically, you run the ETC algorithm to compute the sample regret

   $$\mu_2 \cdot T - \sum_{t=1}^{T} R_t,$$

   where R_t is the reward obtained in time step t.

   Repeat this experiment 500 times and estimate the expected regret by taking the average of the sample regrets you obtained in all those 500 experiments. `[5 Marks]`

2. Repeat the above for various values of $\Delta \in \{0.05, 0.1, 0.2, 0.3, 0.4, 0.45\}$ and plot the estimated regret as a function of Δ and verify whether it satisfies the regret upper bound we derived in class. `[5 Marks]`

3. Repeat the experiment with the UCB algorithm and plot the comparison with ETC. `[10 Marks]`

4. **(Bonus)** In the ETC algorithm, assume that we know Δ, and choose a better m as function of Δ and repeat the experiments and compare with UCB. What did you observe? Hint: Check how many samples of exploration are required to make $\varepsilon < \frac{\Delta}{2}$ with a high probability of $1 - \frac{1}{T}$. `[5 Marks]`