Assignment 1

August 28, 2025

Adventures in Online Advertising

You have just joined the data science team at a cutting-edge AdTech company. Millions of users stream through the platform every day, and behind the scenes, sophisticated algorithms are deciding which ads they should see.

Your mission is clear but challenging: **design algorithms that learn which ads maximize user engagement**.

Each time you display an ad, the system produces a noisy **engagement reward** — a single numerical score reflecting how the user responded. This reward could depend on many factors (clicks, dwell time, conversions, etc.), but you only observe the final number. Each ad has its own hidden average reward, and your task is to discover the best ones through intelligent exploration and exploitation.

Connection to Multi-Armed Bandits: This problem is a direct instance of the multi-armed bandit framework: each ad is an arm, showing an ad is pulling an arm, and the engagement reward is the payoff. The challenge is balancing *exploration* and *exploitation*.

Your Mission

You will investigate and compare several strategies for ad selection, using a simulation of the online advertising environment:

- Simulate **a platform with 10 ads for 1000 steps**, each generating engagement rewards from a Gaussian distribution with an unknown mean.
- Repeat each experiment for **2000 independent runs** to ensure reliable averages.
- Plot the **average engagement reward** and the **percentage of times the best ad was selected** for the following strategies:
  (a) ϵ-greedy with $\epsilon = 0.1$
  (b) ϵ-greedy with $\epsilon = 0.2$

(c) Decaying ϵ-greedy with $\epsilon = \frac{1}{n+1}$ for number of iterations $\in [50n, 50(n+1))$, $n \geq 0$ and $n \in \mathbb{N} \cup \{0\}$.

(d) UCB (Upper Confidence Bound) with parameter $c = 1$

(e) UCB with parameter $c = 10$

Starter Code

We will provide you with Python starter code that sets up the environment, agents and plotting framework.

- The environment already simulates the ads and their engagement rewards.
- Plotting functions are implemented.
- Sections of the algorithms are left as `# TODO` for you to complete.

Your task is to carefully study the code and **fill in the missing parts**. This ensures you both understand the logic and practice implementing the strategies yourself.

Deliverables

You must submit **a single structured report (PDF)** containing the following:

- **Code snippets:** Your completed implementations of the `# TODO` parts (ϵ-greedy, decaying ϵ-greedy, and UCB).
- **Plots:** For each of the 5 strategies listed above, include:
  (a) The average engagement reward vs. time.
  (b) The percentage of times the optimal ad was chosen vs. time.
- **Discussions:** A clear comparison of the strategies that have been used.

Important Instructions

- This is an **individual assignment**. Collaboration is not allowed.
- Deadline: 7$^{\text{th}}$ **September 2025, 11:59 PM**.

Good luck, and may your ads win the clicks!