CS 688: Homework 5

Due: Friday May 3 by 10:00 PM

Notes:

- Homework is due **by 10:00 PM on the due date.** Remember that you may not use more than 2 late days on any one homework, and you only have a budget of 5 in total.
- Please keep in mind the collaboration policy as specified in the course syllabus. If you discuss questions with others you **must** write their names on your submission, and if you use any outside resources you **must** reference them. **Do not look at each others' writeups, including code.**
- There are 2 problems on 2 pages in this homework.
- Remember that you will be graded on the quality of your report and writing. We may or may not even look at your code. So please write carefully and clearly.
- For your code, please create a single zip file labeled `yourlastname_hw5.zip` and submit it to the "HW5 Code" assignment on Gradescope. You should provide a readme explaining how to run your code and replicate the results in your report.
- All graphs should have clearly labeled axes.

Problems:

1. (25 points) Consider the following scenario. There are two coins, with probabilities p_1 and p_2 respectively of coming up heads. At each time I pick one coin to toss; the probability that I pick the first coin is π_1 and the probability that I pick the second is $1-\pi_1$. Let $\theta = (p_1, p_2, \pi_1)$.

   (a) (5 points) Suppose you observe, for T time steps, both the coin that I flip (say $z_t \in \{1, 2\}$ at time step t) and an outcome variable x_t which is 1 if the coin comes up heads and 0 otherwise. Write down the log likelihood function $\log \mathcal{L}(\theta)$ and then use the maximum likelihood approach to find estimates $\hat{p}_1, \hat{p}_2, \hat{\pi}_1$ for the unknown parameters.

   (b) (5 points) Now, suppose you only observe the outcome of the flip (x_t) and not which coin was flipped. Thus, z_t is a hidden variable. Show that the log likelihood function is given by:

   $$\sum_{t=1}^{T} \log \left(p_1^{x_t}(1-p_1)^{(1-x_t)}\pi_1 + p_2^{x_t}(1-p_2)^{(1-x_t)}(1-\pi_1) \right)$$

   (c) (5 pts) Using the above, describe how the E-step of the EM algorithm would proceed in this case by deriving $\Pr(z_t = 1 | x_t)$

(d) (10 pts) Finally, using the above, describe the M-step of the EM algorithm, to update p_1, p_2, π_1 based on all the observations over the entire T periods.

2. (25 points) Implement three different algorithms for multi-armed Bernoulli bandits: the greedy algorithm (**not ϵ-greedy**), UCB1, and Thompson Sampling. You should write a function for each of the three, which takes as input (1) a vector containing the probabilities with which each arm gives a payoff of 1, and (2) the number of iterations to run for. Test your algorithms on two different settings: (1) an eleven-armed bandit with payoff probabilities $0, 0.1, 0.2, \ldots, 1.0$, and (2) a five-armed bandit with payoff probabilities $0.3, 0.5, 0.7, 0.83, 0.85$.

   Use these test cases to analyze the empirical properties of these algorithms, at least in terms of regret over time and probability of choosing the best action over time. Minimally, you should present a graph of average regret vs. time for each of the algorithms, and some kind of visualization of which action they are choosing over time. Be sure to label your axes and caption your figures appropriately. I would suggest trying the algorithms with at least $10^3, 10^4, 10^5$ time steps, and also re-running them from the start many times and averaging (so, for example, run the experiment with 10^3 time steps several hundred times). What do your results tell you about the properties of these three different algorithms?

   On this question in particular, I urge you to experiment and also present and write up any other interesting insights you may have (for example, playing with different probability vectors or different ways of analyzing the data).