Homework 2 (due 10/20)

DS-563 / CD-543 @ Boston University

Fall 2021

Before you start...

Collaboration policy: You may verbally collaborate on required homework problems, however, you must write your solutions independently. If you choose to collaborate on a problem, you are allowed to discuss it with at most 4 other students currently enrolled in the class.

The header of each assignment you submit must include the field "Collaborators:" with the names of the students with whom you have had discussions concerning your solutions. A failure to list collaborators may result in credit deduction.

You may use external resources such as textbooks, lecture notes, and videos to supplement your general understanding of the course topics. You may use references such as books and online resources for well known facts. However, you must always cite the source.

You may **not** look up answers to a homework assignment in the published literature or on the web. You may **not** share written work with anyone else.

Submitting: Solutions should be submitted via Gradescope (entry code: BPDKV8). You are allowed to submit your solutions both in handwriting or typed. If you decide to hand-write your solutions, make sure they are as readable as possible. If you decide to submit a typed version, we suggest using LaTeX.

Grading: Whenever we ask for an algorithm (or bound), you may receive partial credit if the algorithm is not sufficiently efficient (or the bound is not sufficiently tight).

Questions (up to 10 points for each)

1. The handout with useful probabilistic inequalities on the webpage was updated with information about the birthday paradox, which was also covered in a discussion section.

   - Read the section titled "Collisions (the Birthday Paradox)."
   - Go to `https://www.xe.com/currencytables/` or any similar webpage. Select an arbitrary pair of currencies and check their exchange rate yesterday. Look at the two least significant digits. (Example: If the exchange rate is `3.523432`, then these digits are `32`.) Then look at the exchange rate two days ago, three days ago, and so on, always restricting your attention to the two least significant digits. How many days do you have to look back to see two days on which the exchange rate has~~a pair of exchange rates with~~ identical two least significant digits.

Note (10/20): The statement of the problem has been updated to avoid ambiguity. See Question "Problem 1" on Piazza for more information.

- Repeat this experiment for four more pairs of currencies. Tell us what the results of your experiments are. Do you feel these results match theory, assuming that the two least significant digits are distributed uniformly?

2. Consider an experiment in which an event $\mathcal{E}$ occurs with probability p. In class, we said that it suffices to run the experiment $\Omega(1/p)$ times to see $\mathcal{E}$ occur at least once with constant probability. Formalize this argument.

   Hint: What is the probability that you repeat the experiment k times and it does not occur? You may find the following inequality useful: $1 - x \leq e^{-x}$ for all $x \in \mathbb{R}$.

3. In class, we said that deterministic streaming algorithms (i.e., streaming algorithms that use no randomness) are adversarially robust. Why is this the case?

4. You have a weighted graph with all edges of weights in $\{1, \ldots, \Delta\}$. How much space is required to compute the exact minimum spanning tree in

   (a) the insertion–only stream model,

   (b) the insertion/deletion stream model?

   Explain why.

5. Same as above, can the space be improved if you are computing a $(1+\epsilon)$-multiplicative approximation for some $\epsilon \in (0, 1)$? By how much and why?

6. In class we focused on proving that a randomly selected linear transformation preserves Euclidean distances between all points with non-zero probability. What if want to select such a transformation with probability at least $1 - \frac{1}{n^k}$ for some positive integer k, where n is the number of points? How many dimensions do we have to use? Explain why. State exact constants if possible. Asymptotic bounds may achieve partial credit.

   You can follow the analysis from the lecture or the slightly improved analysis in the notes on the webpage.

7. In class, we had to select a point uniformly at random from the k-dimensional sphere. But how can this be done in practice? Suppose that you know how to draw samples from $\mathcal{N}(0, 1)$, i.e., the Gaussian distribution. First, select a vector v with each coordinate drawn independently from $\mathcal{N}(0, 1)$. Then normalize it, i.e., divide it by $\|v\|$. Why does this work?

   Hint: Look at the probability density function of v.

8. How much time (approximately) did you spend on this homework? Was is too easy or too hard?

   Note: You will not be evaluated based on your answer to this question.