Khoury College of Computer Sciences | Fall 2023
Northeastern University | 27 October 2023
CS 7870: Seminar in Theoretical Computer Science | Rajmohan Rajaraman

Problem Set 3 (due Tuesday, November 21)

Problem 1 (3 points) Random projections

Given a set X of n points in n-dimensional space, consider the problem of find two points x and y in X that minimize $|x - y|$.

(a) Give an $O(n^3)$-time algorithm for solving the problem.

(b) Using random projections, give an $O(n^2 \log n)$ time algorithm that, with probability at least $1 - 1/n$, finds two points x and y, whose distance is at most 1% more than the distance between two closest points.

Problem 2 (3 points) Dimensionality reduction via JL Transform

Suppose you have an algorithm for finding a TSP tour for n points in $\Re^d$ whose cost is within $(1 + \varepsilon)$ of the optimal in time $T(n, d, \varepsilon)$. Show how to use the Johnson-Lindenstrauss transform to derive an algorithm that finds a TSP tour for n points in $\Re^d$ whose cost is within $(1 + \varepsilon)$ of the optimal in time $T(n, O(\log n/\varepsilon^2), O(\varepsilon))$ with probability at least $1 - 1/n$. (Note that this could be significantly more efficient if $d \gg \Theta(\log n)$.

Problem 3 (1 point) Pseudoinverse and SVD

Suppose X is a square matrix, but not necessarily invertible. Consider the singular value decomposition of X given by $U^T DV$. If u_i denotes the ith column vector of U and v_i the ith column vector of V, we can also write this as $X = \sum_i \sigma_i u_i v_i^T$. Define Y as the matrix $\sum_i \frac{1}{\sigma_i} v_i u_i^T$. Prove that $Y X v = v$ for all v in the span of the right singular vectors of X. The matrix Y is referred to as a pseudoinverse of X and can play the role of X^{-1} in many applications.

Problem 4 (3 points) Heavy hitters using compressed sensing

Suppose we are trying to maintain frequencies of the heavy-hitters in a long stream of elements, where the elements are drawn from the set $[n]$. Each update in the stream is an increment or decrement (by one) of the frequency of one of the elements. Assume that the initial frequency of all elements is 0.

A simple way to maintain the frequencies is to maintain an array of size n, where we keep a count for each element. This requires space $\Theta(n)$ (assuming each count requires $O(1)$ space).

(a) Consider the case where we know that only $k \ll n$ elements will ever show up in the stream. Show that you can accomplish the task using space $O(k)$ and time $O(k)$ per update. Show how to improve this to space $O(k)$ and expected time $O(1)$.

(b) Now consider the case where we know that there are only $k \ll n$ heavy-hitters, but there is noise in the system so that other counts may also be non-zero. In other words, the frequency vector f

is almost k-sparse. From compressed sensing, we know that there exists an $m \times n$ matrix A for $m = O(k \log(n/k))$ such that the k most significant entries of f can be recovered from Af. Show how to maintain Af (which incurs space $O(m)$) without explicitly maintaining f. You may assume that you have access to A at every update.

Problem 5 (10 points) Project Proposal

Please submit a proposal for the course project. The proposal must be approximately 2 pages long for an individual project and 3-4 pages long for a pair project. Use 1-inch margins on all sides, and 11-pt font. The proposal should include the following:

- Tentative title and author(s)
- High-level description of the problem or topic: this could include formal problem definitions and/or models being considered.
- Motivation for the problem or topic: describe why the problem or topic is of interest.
- A brief summary of related work on the problem or topic: this could include a summary of results of say two of the papers in the relevant domain.
- Proposed work (could be a survey, theoretical exploration, experimental work): if you have a specific open problem in mind, highlight it; otherwise, you can list potential directions for work; if you are conducting a survey, list the papers you are planning to review and present, and discuss why you think these are important.
- References (not counted in the page length)

Prepare a project proposal presentation (5-7 minutes long + 3-5 minutes for questions) to be presented on Nov 14 or Nov 21.