Math 126 Problem 10: Applications of the Johnson-Lindenstrauss Lemma

Spring 2027

Submission Instructions

Please upload your work to Gradescope by **11:59pm on Monday, March 1, 2027 (Pacific time).**

- You are encouraged to discuss ideas and work with your classmates. However, you *must acknowledge* your collaborators at the top of each solution on which you collaborated with others and you *must write* your solutions independently.
- Your solutions to theory questions must be written legibly, or typeset in LaTeX or markdown.
- I recommend that you write your solutions in a Python notebook using Google Colab.
- You should submit your solutions as a *single PDF* via the assignment on Gradescope.

Grading: The point of the problem set is for *you* to learn. To this end, I hope to disincentivize the use of LLMs by *not* grading your work for correctness. Instead, you will grade your own work by comparing it to my solutions. This self-grade is due the Friday *after* the problem set is due, also on Gradescope.
Original notebook: Google Colab.

Applications of the Johnson-Lindenstrauss Lemma

Consider vectors $\mathbf{x}_1, \ldots, \mathbf{x}_n \in \mathbb{R}^d$. Let $k = O\left(\frac{\log n}{\epsilon^2}\right)$. The JL lemma tells us that, for a special random matrix $\mathbf{\Pi} \in \mathbb{R}^{k \times d}$,

$$(1-\epsilon)\|\mathbf{x}_i - \mathbf{x}_j\|_2^2 \leq \|\mathbf{\Pi}\mathbf{x}_i - \mathbf{\Pi}\mathbf{x}_j\|_2^2 \leq (1+\epsilon)\|\mathbf{x}_i - \mathbf{x}_j\|_2^2$$

with probability 9/10.

Part A: Inner Product

Show that we can use the JL lemma as stated above to show that inner-products are also preserved. In particular, under the same conditions as above,

$$|\langle \mathbf{x}_i, \mathbf{x}_j \rangle - \langle \mathbf{\Pi}\mathbf{x}_i, \mathbf{\Pi}\mathbf{x}_j \rangle| \leq \frac{1}{2}\epsilon(\|\mathbf{x}_i\|_2^2 + \|\mathbf{x}_j\|_2^2)$$

with probability 9/10.

Part B: Join Size Estimation

One powerful application of sketching is in database applications. For example, a common goal is to estimate the *inner join size* of two tables without performing an actual inner join (which is expensive, as it requires enumerating the keys of the tables).

Consider two sets of unique keys $X = \{x_1, \ldots, x_m\}$ and $Y = \{y_1, \ldots, y_n\}$ which are subsets of $1, 2, \ldots, U$. Our goal is to estimate $|X \cap Y|$ based on small space compressions of X and Y.

Using your result from above, describe a method based on inner product estimation that constructs independent sketches of X and Y of size $k = O\left(\frac{\log n}{\epsilon^2}\right)$ and from these sketches can return an estimate Z for $|X \cap Y|$ satisfying

$$|Z - |X \cap Y|| \leq \epsilon(|X| + |Y|)$$

with probability 9/10.