CS698C: Sketching and Sampling for Big Data Analysis

Subspace Embedding using Normal Distribution: Discussion and Exercises

Note: All norms are ℓ_2 norms $\|\cdot\|_2$ unless explicitly specified.

1 Subspace Embeddings using Normal Distributions

We assume that A is some arbitrary but fixed n by d orthonormal matrix of reals, and S is a $k \times n$ matrix of iid $N(0, \frac{1}{k})$ distributed random variables.

This is for Q2. A k-sparse n-dimensional vector x is an n-dimensional vector such that there are at most k non-zero coordinates, that is, coordinates $r_1, \ldots, r_k$ such that $x_{r_1}, x_{r_2}, \ldots, x_{r_k} \neq 0$, and all the other coordinates of x in $\{1, \cdots n\} \setminus \{r_1, \ldots, r_k\}$ have zero values.

The following inequality is useful. For $n \geq k \geq 1$ and integral,

$$\binom{n}{k} \leq \left(\frac{ne}{k}\right)^k \quad .$$

The following is a proof for this statement.

Proof. Stirling's approximation is a very well-known approximation of the factorial function. For any integer $n \geq 1$,

$$n! = \sqrt{2\pi n}\left(\frac{n}{e}\right)^n \left(1 + \frac{1}{12n} + o\left(\frac{1}{n}\right)\right)$$

This implies a reasonable approximation of $n! \geq \sqrt{2\pi n}\left(\frac{n}{e}\right)^n$. Applying this to $k!$,

$$\binom{n}{k} = \frac{n(n-1)\cdots(n-k+1)}{k!} \leq \frac{n^k}{\sqrt{2\pi k}(k/e)^k} \leq \left(\frac{ne}{k}\right)^k \quad .$$

□

This implies that $\log \binom{n}{k} \leq \log\left(\frac{ne}{k}\right)^k = k + k\log\frac{n}{k}$.

Give complete arguments for the statements below.

1. Show that if $k = \frac{O(1)}{\epsilon^2}(d + \log(1/\delta))$, then,

$$\mathrm{P}\left[\text{for all } x \in \mathbb{R}^d,\ \|SAx\|_2^2 \in (1 \pm \epsilon)\|Ax\|_2^2\right] \geq 1 - \delta \quad .$$

2. Suppose B is an arbitrary n by t matrix of reals. Assume $t > ed$. Let S be a k by n matrix of iid $N(0, \frac{1}{k})$ random variable entries. We say that S approximately preserves the norms of all images of d-sparse vectors under B to within factors of $1 \pm \epsilon$ if: ,

$$\text{for all } d\text{-sparse vectors } x \in \mathbb{R}^t,\ \|SBx\|^2 \in (1 \pm \epsilon)\|Bx\|^2$$

   with probability $e^{-\Theta(d \log \frac{t}{d})}$.

   Show that if $k \geq \frac{1}{\epsilon^2}\left(\Theta(d \log \frac{t}{d})\right)$, then S satisfies the above property.

3. Complete the following alternate argument for column space embedding which is slightly weaker than the one covered in the class.

   Let A be an $n \times d$ orthonormal matrix and S be a $k \times n$ random matrix consisting of iid $N(0, \frac{1}{k}$ entries. Let y be an arbitrary unit vector in the column space of A. Let M be a γ net for the unit sphere in the column space of A. Find the range of value of γ for the following calculation to go through. Then, y is written as

$$y = y_1 + y_2 + y_3 + \cdots, \quad y_1 \in M, y_j/\|y_j\| \in M, \|y_j\| \le \gamma^j, j = 1, 2, \cdots \ .$$

   (a) We now bound $\|Sy\|_2$ as follows.

$$Sy = Sy_1 + Sy_2 + \cdots \ .$$

   Using triangle inequality, $\|a + b + c + \cdots\| \le \|a\| + \|b\| + \|c\| + \cdots$,

$$\|Sy\| \le \|Sy_1\| + \|Sy_2\| + \cdots$$

   S preserves the norms of vectors in M to within factors of $1 \pm \epsilon$. Therefore,

$$\begin{aligned} \|Sy\| &\le 1 + \epsilon)\|y_1\| + (1 + \epsilon)\|y_2\| + \cdots \\ &= (1 + \epsilon)\left[\|y_1\| + \|y_2\| + \cdots\right] \\ &\le \frac{1 + \epsilon}{1 - \gamma} && \text{complete this step} \\ &\le 1 + O(\epsilon) && \text{which range of values of } \gamma\text{?} \end{aligned}$$

   For the lower bound, use triangle inequality as follows. $\|a + b\| \ge \|a\| - \|b\|$, as well as $\|a + b\| \ge \|b\| - \|a\|$. (*Pf.* $\|a\| = \|(a + b) + (-b)\| \le \|a + b\| + \|-b\| = \|a + b\| + \|b\|$, and so, $\|a + b\| \ge \|a\| - \|b\|$. Reverse the roles of a and b to get the analogous statement.) Applying this to $Sy = Sy_1 + Sy_2 + \cdots$,

$$\begin{aligned} \|Sy\| &= \|Sy_1 + Sy_2 + Sy_3 + \cdots\| \\ &\ge \|Sy_1\| - \|Sy_2 + Sy_3 + \cdots\| \\ &\ge (1 - \epsilon)\|y_1\| - \left[\|Sy_2\| + \|Sy_3\| + \cdots\right] \\ &\ge (1 - \epsilon)\|y_1\| - (1 + \epsilon)\left[\|y_2\| + \|y_3\| + \cdots\right] \\ &= (1 - \epsilon) - \frac{(1 + \epsilon)\gamma}{1 - \gamma}, && \text{fill this step} \\ &= \frac{1 - \epsilon + \gamma}{1 - \gamma} \\ &= 1 - O(\epsilon) && \text{which range of values of } \gamma\text{?} \end{aligned}$$

   Show that γ must be $O(\epsilon)$ for this argument to go through.

   (b) Assuming $\gamma = \epsilon$, show that $|M| = \exp\left\{\Theta(d \log \frac{1}{\epsilon})\right\}$.

   (c) Hence, show that to approximately preserve the norm of all column space vectors of A to within factors of $1 \pm \epsilon$ with a probability of anything in the range $(0, 1 - \exp\left\{-\Theta(d \log \frac{1}{\epsilon})\right\})$, the number of rows of S should be

$$k = O(\frac{d}{\epsilon^2} \log \frac{1}{\epsilon}) \ .$$

   remark. Note that we have an extra factor of $\log \frac{1}{\epsilon}$ in the number of rows of S. This arises since $\gamma = O(\epsilon)$.

(d) Hence, argue that if $\epsilon = o(1)$, and the failure probability may of the order of $\leq e^{-O(d)}$ or higher, then, the argument covered in class gives an order lower value of k for S (by a factor of $\log \frac{1}{\epsilon}$). However, for failure probabilities smaller than $e^{-O(d \log \frac{1}{\epsilon})} = \epsilon^{O(d)}$ or lower, then both arguments give values of k of the same order.