IE531: Algorithms for Data Analytics
Spring, 2023
Homework 4: Random Projections
Due Date: March 24, 2023

©Prof. R.S. Sreenivas

Let X be a discrete random variable that can take on three values in the set $\{-\sqrt{3}, 0, \sqrt{3}\}$, where

$$Prob\{X = -\sqrt{3}\} = Prob\{X = \sqrt{3}\} = \frac{1}{6}, \text{ and } Prob\{X = 0\} = \frac{2}{3}.$$

Let $\{\mathbf{u}_i\}_{i=1}^k$ be a set of k-many, $(d \times 1)$-vectors that resulted from making d-many calls to code that generates i.i.d. samples of the RV X. Assume d is large, and $k \ll d$.

1. (*25 points*) Show that $||\mathbf{u}_1||_2 = \sqrt{d}$ with high-probability.

2. (*25 points*) Show that the members of the set $\{\mathbf{u}_i\}_{i=1}^k$ are mutually-perpendicular with high-probability.

3. (*25 points*) Let $\mathbf{x}$ be a $(d \times 1)$-vector and $\mathbf{y}$ be a $(k \times 1)$-vector, where

$$\mathbf{y} = \frac{1}{\sqrt{k}}\mathbf{A}^T\mathbf{x}, \text{ where } \underbrace{\mathbf{A}}_{d\times k} = \begin{pmatrix} \mathbf{u}_1 & \mathbf{u}_2 & \cdots & \mathbf{u}_k \end{pmatrix}$$

   Show that $E\{||\mathbf{y}||_2^2\} = ||\mathbf{x}||_2^2$.

4. (*25 points*) Assume there is a constant C such that

$$Prob\{\Big| \, ||\mathbf{y}||_2^2 - ||\mathbf{x}||_2^2 \, \Big| > \epsilon\} < e^{-Ck\epsilon^2}$$

   Show that any value of

$$k > \frac{1}{C\epsilon^2} \ln \frac{1}{(1-\delta)}$$

   is sufficient to guarantee that if two d-dimensional vectors are close to each other, then their k-dimensional "proxies" will be close to each other with probability δ.