Homework 9

Problem 1. *Explain how the volume of a ball in high dimensions can simultaneously be in a narrow slice at the equator and also be concentrated in a narrow annulus at the surface of the ball.*

Solution. Roughly speaking, points on the surface of the ball satisfy $x_1^2 + x_2^2 + \cdots + x_d^2 = 1$, so for each coordinate i, a typical value will be $O(1/\sqrt{d})$.

Formal proof can be found in the textbook. □

Problem 2. *Find the singular value decomposition of the following matrix*

$$A = \begin{pmatrix} 1 & 0 & 1 \\ 0 & 1 & -1 \end{pmatrix}.$$

Solution.

$$A = U\Sigma V^T = \begin{pmatrix} 1/\sqrt{2} & 1/\sqrt{2} \\ -1/\sqrt{2} & 1/\sqrt{2} \end{pmatrix} \begin{pmatrix} \sqrt{3} & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix} \begin{pmatrix} 1/\sqrt{6} & -1/\sqrt{6} & 2/\sqrt{6} \\ 1/\sqrt{2} & 1/\sqrt{2} & 0 \\ 1/\sqrt{3} & -1/\sqrt{3} & -1/\sqrt{3} \end{pmatrix}$$

□

Problem 3. *(Optional) Randomly generate 30 points inside the cube $[-\frac{1}{2}, \frac{1}{2}]^{100}$ and plot distance between points and the angle between the vectors from the origin to the points for all pairs of points.*

Problem 4. *(Optional) Randomly generate 10,0000 points in $\mathbb{R}^{1024}$ and use Johnson-Lindenstrauss projection to project them into dimension 15. List the top 10% of data that are closest to each other, both before and after the projection. Compare the resulting set.*

Randomly generate two separate sets of high dimensional data (for example, two separate high dimensional unit balls) in $\mathbb{R}^{1024}$, call them R_1 and R_2. Each R_i set contains 5,0000 data. Repeat the above random projection process and get R_1' and R_2' respectively. Check whether the union dataset $R_1' \cup R_2'$ is separable by the closest distance algorithm. If yes, try lower reduced dimension like 10, 5, 3, etc.