UC DAVIS, BIG DATA (MAT280), SPRING 2016

Homework #1

Sam Fleischer

April 18, 2016

Problem 1

Show that two random vectors in high dimensions are almost orthogonal.

Note: In your theorem you need to formalize what "almost orthogonal" means (what it means will come out of your proof). You first need to select a probability distribution of your choice and apply an appropriate concentration inequality (but keep in mind that if e.g. x and y are Gaussian random vectors, then the entries of the inner product $\langle x, y\rangle$ are no longer Gaussian).

Proof. Let a be a uniformly chosen random vector on the unit sphere in $\mathbb{R}^d$. Now let a be the last element of an orthonormal basis $\{a_1, \ldots, a_{d-1}, a\}$, extended via Gram-Schmidt. In polar coordinates,

$$a = \left(1, \frac{\pi}{2}, \frac{\pi}{2}, \ldots, \frac{\pi}{2}, \frac{\pi}{2}\right)$$

where the first entry is magnitude ρ, the last entry $\theta \in [0, 2\pi)$, and all other entries $\phi_i \in [0, \pi)$ for $i = 1, \ldots, d-2$. These coordinates are essentally a rotation such that a is parallel with the positive x_d-axis. In rectangular coordinates,

$$a = (0, 0, \ldots, 0, 1).$$

Now choose b uniformly at random on the unit sphere. That is, $\|b\| = 1$ and $\theta \in [0, 2\pi)$ is chosen uniformly at random and $\phi_1, \ldots, \phi_{d-2} \in [0, \pi)$ are each chosen uniformly at random. b can be represented in the polar coordinates as defined above as

$$b = \left(1, \phi_1, \phi_2, \ldots, \phi_{d-2}, \theta\right)$$

and in rectangular coordinates as

$$b = (b_1, b_2, \ldots, b_d)$$

where

$$\begin{aligned}
b_1 &= \cos(\phi_1) \\
b_2 &= \sin(\phi_1)\cos(\phi_2) \\
b_3 &= \sin(\phi_1)\sin(\phi_2)\cos(\phi_3) \\
&\vdots \\
b_{d-2} &= \sin(\phi_1)\sin(\phi_2)\sin(\phi_3)\ldots\sin(\phi_{d-3})\cos(\phi_{d-2}) \\
b_{d-1} &= \sin(\phi_1)\sin(\phi_2)\sin(\phi_3)\ldots\sin(\phi_{d-3})\sin(\phi_{d-2})\cos(\theta) \\
b_d &= \sin(\phi_1)\sin(\phi_2)\sin(\phi_3)\ldots\sin(\phi_{d-3})\sin(\phi_{d-2})\sin(\theta)
\end{aligned}$$

Then

$$\langle a, b\rangle = \langle (0, 0, \ldots, 0, 1), (b_1, b_2, \ldots, b_d)\rangle = b_d = \sin(\phi_1)\sin(\phi_2)\sin(\phi_3)\ldots\sin(\phi_{d-3})\sin(\phi_{d-2})\sin(\theta).$$

This is a product of $d-1$ numbers in $[-1, 1]$. Note that since $\phi_1, \ldots, \phi_{d-2}$ are independent indentical random variables uniformly distributed in $[0, \pi)$, and since θ is uniformly distributed in $[0, 2\pi)$ then

$$\mathbb{P}\left(\left|\sin(\phi_i)\right| > \frac{1}{2}\right) = \frac{2}{3} \qquad \text{and} \qquad \mathbb{P}\left(|\sin(\theta)| > \frac{1}{2}\right) = \frac{2}{3}$$

Thus,

$$\mathbb{P}\left(\langle a, b\rangle > \frac{1}{2^{d-1}}\right) = \mathbb{P}\left(\left|\sin(\theta)\prod_{i=1}^{d-2}\sin(\phi_i)\right| > \frac{1}{2^{d-1}}\right) = \left(\frac{2}{3}\right)^{d-1} \to 0 \qquad \text{as } d \to \infty$$

Thus two random vectors in high dimensions are almost orthogonal. $\square$

Problem 2

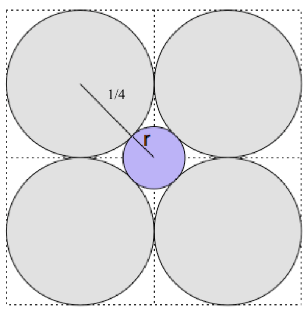
Consider the following setup. Given a square of side length 1, we place four circles in the square as depicted in Figure 0.1a (each of the gray circles has radius 1/4). We now place a circle at the center of the square (the blue circle in Figure 0.1a) such that this circle in the middle touches each of the four identical circles. Let r denote the radius of the blue circle.

We can do something analogous in three dimensions, see Figure 0.1b. We place eight spheres of radius 1/4 inside a cube of side length 1, and put a (blue) sphere in the middle such that it touches all eight (gray) spheres.

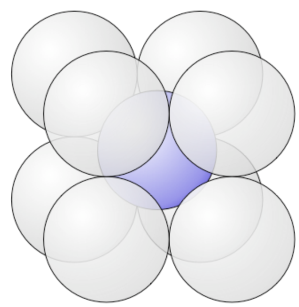
In four dimensions we can arrange 16 hyperspheres of radius 1/4 inside a hypercube of side length 1 and place a hypersphere in the middle, so that this hypersphere all the other 16 hyperspheres.

Obviously we can do this for increasing dimension d. What happens with the blue hypersphere in the middle as d increases? Will it shrink? Will it be of constant size? Will it grow outside the hypercube?

(Hint: Check the diameter of the blue hypersphere in comparison to the sidelength of the cube as d increases. This is actually not difficult to compute, it may sound more complicated than it is).



(a) 4 circles



(b) 8 spheres

Figure 0.1

Proof. In a given orthant, the hypersphere of radius $\frac{1}{4}$ is always tangent to the hypercube of apothem size $\frac{1}{4}$. However, the distance from each vertex of the hypercube of apothem size $\frac{1}{4}$ to the center that hypercube is $\frac{\sqrt{d}}{4}$. The origin (center of the blue hypersphere) is a vertex of each orthant's hypercube, and thus the radius of the blue sphere is at most $\frac{\sqrt{d}-1}{4}$. A worse upper bound is the distance from the origin to the point at which two hyperspheres meet, which is $\frac{\sqrt{d-1}}{4}$. Since two hyperspheres can only meet at one point, and $\frac{\sqrt{d}-1}{4} < \frac{\sqrt{d-1}}{4}$ for large d, they must meet when the blue hypersphere has radius $r_{\text{blue hypersphere}}$

$$r_{\text{blue hypersphere}} = \frac{\sqrt{d}-1}{4}.$$

Since the side length of the hypercube is always 1 (and the length from the origin to an apothem is $\frac{1}{2}$, the blue hypersphere extends beyond the hypercube for $d > 9$. $\square$

Problem 3

Show that for every fixed dimension reduction matrix A of size $k \times d$ with $k < d$, there exists vectors $x, y \in \mathbb{R}^d$ such that the distance $\|Ax - Ay\|$ (no matter which norm we use) is vastly different from $\|x - y\|$.

Proof. First note that $\dim \operatorname{ran} A \leq k < d$. Thus, by the Fundamental Theorem of Linear Algebra, $\dim \operatorname{null} A \geq 1$. Then let $x \in \operatorname{null} A$ and let $y = ax$. Thus $y \in \operatorname{null} A$ and $\|x - y\| \to \infty$ as $a \to \infty$. However $\|Ax - Ay\| = \|0 - 0\| = 0$ for all values of a. □

Problem 4

The Yale Face Database contains images from various individuals in different poses and under different lighting conditions. Some of the images are stored in the file `SomeYaleFaces.mat`.

Load this file into Matlab. The variable `X` is a matrix of size 1024×2414. Each column of `X` is an image of size 32×32 (in vectorized form). The 2414 columns are images of 38 different persons in about 64 poses each. You can easily covert the k-th column of `X` back to an image via the commands

```
xk = X(:,k);
xk = reshape(xk,32,32);
```

The command

```
imagesc(x1); colormap(gray);
```

will display the image.

You can conveniently display multiple images if you want with the file `showfaces.m`.

We want to compare three dimension reduction methods by comparing how well distances between the different images are preserved: (i) Johnson-Lindenstrauss projection, Fast Johnson Lindenstrauss projection and simple random sampling (i.e., randomly picking k indices).

Choose different values for the reduced dimension k and compare the dimension reduction ability of the three methods. You need to think about how to devise such an experiment. There are of course multiple options to do so.

Proof. The following sections briefly describe each method of dimension reduction.

Johnson-Lindenstrauss projection

First, form a Gaussian $d \times k$ matrix (`A = randn(d,k);`) and normalize it (`T = orth(A)';`). Then find the pairwise distances between each x_i and x_j (`PD_1 = pdist(X'); PD_2 = pdist((T*X)');`).

Fast Johnson-Lindenstrauss projection

First form a hadamard matrix of dimension d (`H = hadamard(d);`). Then form a sparse matrix (`S`) in which there is one non-zero element in each row, and that non-zero element is $\frac{1}{\sqrt{d}}$. Then form a diagonal matrix (`D`) whose diagonal elements are 1 or -1 each with probability $\frac{1}{2}$. Finally, the linear operator is `T = S*H*D`. Then find the pairwise distances between each x_i and x_j (`PD_1 = pdist(X'); PD_2 = pdist((T*X)');`).

Simple random sampling

First, form a sparse matrix $k \times d$ matrix in which there is one non-zero element in each row, and that non-zero element is 1. This sparse matrix `T` is the dimension-reducing linear operator. Then find the pairwise distances between each x_i and x_j (`PD_1 = pdist(X'); PD_2 = pdist((T*X)');`).

Error checking

Then, for each dimension reduction, I calculated the ratio

$$\gamma_{i,j} = \frac{\|Tx_i - Tx_j\|}{\|x_i - x_j\|}$$

for each pair of points and found the ε such that

$$(1-\varepsilon)\|x_i - x_j\| \leq \|Tx_i - Tx_j\| \leq (1+\varepsilon)\|x_i - x_j\|$$

by the following logic: If $\gamma_{i,j} > 1$, then $\varepsilon_{i,j} = \gamma_{i,j} - 1$, and if $\gamma_{i,j} < 1$, then $\varepsilon_{i,j} = 1 - \gamma_{i,j}$. Then I took the mean and maximum ε of each method and plotted the results against the reduced dimension k. Figure 0.2 shows these results. All three methods have approximately the same average ε for any dimension, however the Johnson Lindenstrauss dimension reduction technique is the best because of its low maximum ε.

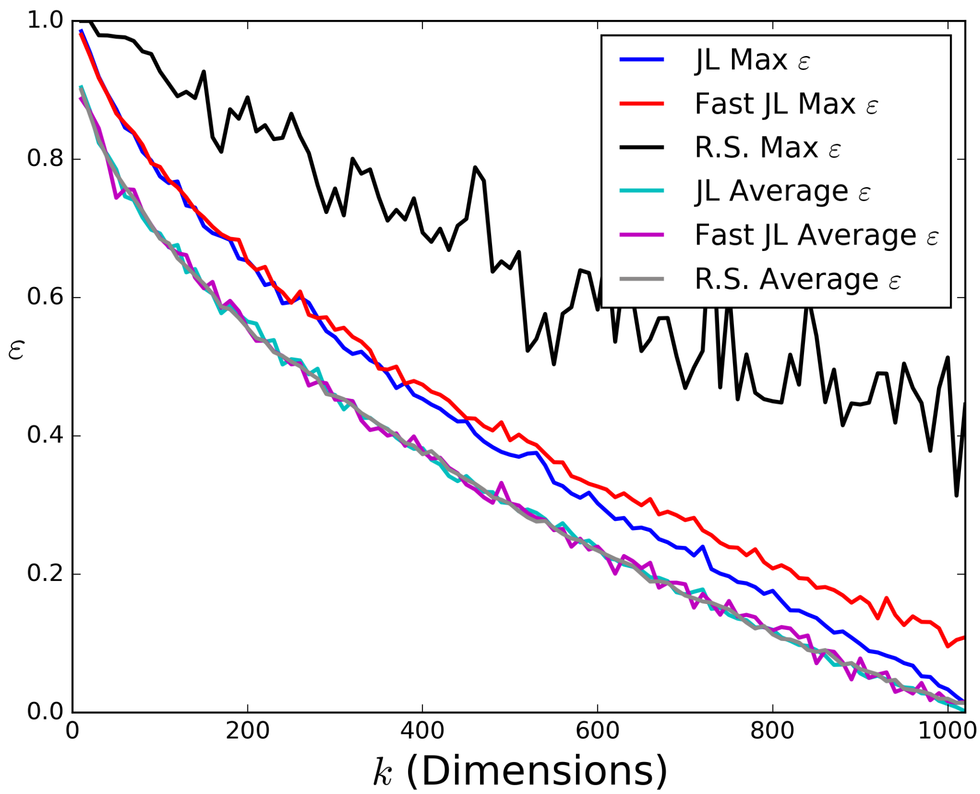


Figure 0.2: JL: Johnson-Lindenstrauss, R.S.: Random Sampling

□