Homework 1: due 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).

Problem 2: Consider the following setup. Given a square of side length 1, we place four circles in the square as depicted in Figure 1 (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 1) such that this circle in the middle touches each of the four identical circles. Let r denote the radius of the blue circle.

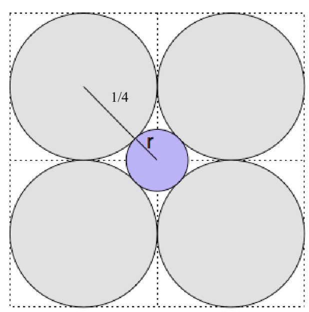

Figure 1: 4 circles

We can do something analogous in three dimensions, see Figure 2. 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?

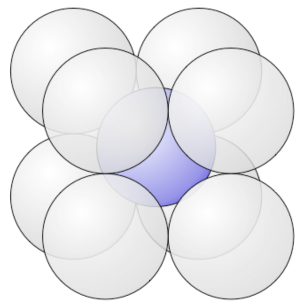

Figure 2: 8 spheres

(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).

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 $\|x - y\|$.

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 differen 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.