CSE 521: Design and Analysis of Algorithms I **Fall 2021**

Lecture 7: Curse of Dimensionality, Dimension Reduction

Lecturer: Shayan Oveis Gharan *Oct 20st*

Disclaimer: *These notes have not been subjected to the usual scrutiny reserved for formal publications.*

High dimensional vectors appear frequently in recent development in CS, examples of which are user-movie ratings of netflix, DNA strings of patients, and images pixel values. In this lecture we study higher dimensions geometry and see how randomization help us to design algorithm. We first start with a few definitions, then we discuss several properties of high dimensional geometry, and finally we investigate a way to map points from high dimensions to a low dimensional space preserving the ℓ_2 distances.

7.1 Introduction

Definition 7.1. *For a vector $v \in \mathbb{R}^d$, ℓ_2 norm of v is defined $\|v\|_2 = \sqrt{\sum_i v_i}$, and ℓ_∞ of v is defined $\|v\|_\infty = \max_i |v_i|$.*

Definition 7.2. *A d dimensional ℓ_2 ball is defined as $B_2(c, r) = \{(x_1, \cdots, x_d) : \|x - c\|_2 \leq r\}$, and similarly a d dimensional ℓ_∞ ball is $B_\infty(c, r) = \{x : \|x - d\|_\infty \leq r\}$. Note that an ℓ_∞ ball is just a d-dimensional cube. It can also be seen that a d-dimensional ℓ_1 ball is a simplex.*

In geometry, diameter of a circle is any line segment that passes through the center of the circle. And for a square it's the line segment that connects two opposite vertices. In more general form, the diameter of a shape (a set of points) is the supremum of distances between every two points in that set, $\sup_{x,y} d(x, y)$. The diameter of a d dimensional ball is always 2, and the diameter of a d dimensional cube is $2\sqrt{d}$, which is the distance between $(1, 1, \cdots, 1)$ and $(-1, -1, \cdots, -1)$.

Volume is a notion for the size of a shape D, and is formally defined as $\int_D 1dD$ where the integration is with respect to the Lebesgue measure. Volume of a d dimensional ball is $\frac{\pi^{\frac{d}{2}}}{(\frac{d}{2})!}$, and volume of a d dimensional cube is 2^d. It is interesting to observe even though a 2-dimensional ball covers most of the area of a 2-dimensional cube (see Figure 7.1), the ratio of the volumes of the in d dimensions is exponentially small in d, $\frac{\text{Val}(B_2(0,1))}{\text{Val}(B_\infty(0,1))} = \frac{1}{e^{\Omega(d)}}$. In other words, as d grows, the ball gets exponentially smaller than the cube.

Theorem 7.3. $\frac{\text{Val}(B_2(0,1))}{\text{Val}(B_\infty(0,1))} = \frac{1}{e^{\Omega(d)}}$

To estimate this ratio we use the Monte Carlo method. We choose a uniformly random point in $B_\infty(0, 1)$ and we compute the probability that this point lies inside the ℓ_2 ball $B_2(0, 1)$. First, observe that to choose a uniformly random point in the cube it is enough to choose $x_1, \ldots, x_d$ independently and uniformly in the range $[-1, 1]$.

Now, such a point x lies in the ℓ_2 ball if and only if $\sum_i x_i^2 \leq 1$. So, we need to compute

$$\mathbb{P}\left[\sum_i x_i^2 \leq 1\right].$$

We use the Hoeffding's inequality to bound the above quantity. First recall that for $X_1, \ldots, X_d$ independently chosen in the range $[a, b]$, the Hoeffding's inequality implies

$$\mathbb{P}\left[\left|\sum_i X_i - \mathbb{E}\sum_i X_i\right| > \epsilon\right] \leq 2\exp\left(\frac{-2\epsilon^2}{d(a-b)^2}\right). \tag{7.1}$$

Now, let $X_i = x_i^2$. First, observe that

$$\mathbb{E}\left[X\right]_i = \int_{t=-1}^{1} t^2/2 = \frac{t^3}{6}\Big|_{-1}^{1} = 2/6 = 1/3.$$

Also, observe that $X_i \in [0, 1]$. Therefore, by (7.1)

$$\mathbb{P}\left[\sum_i X_i \leq 1\right] \leq \mathbb{P}\left[\left|\sum_i X_i - \mathbb{E}\sum_i X_i\right| \geq d/3 - 1\right] \leq 2\exp\left(\frac{2(d/3-1)^2}{d(1-0)^2}\right) = e^{-\Omega(d)}$$

as desired.

7.2 Near Orthogonal Vectors

In $\mathbb{R}^d$, the maximum number of vectors that you can find such that *each* pair of them are orthogonal to each other is exactly d. One example of such vectors set is the standard basis vectors, $e^1, \ldots, e^d$ where for each i, e^i is the vector

$$e_j^i = \begin{cases} 1 & \text{if } j = i \\ 0 & \text{otherwise.} \end{cases} \tag{7.2}$$

In this section we want to study the maximum number of vectors one can choose in $\mathbb{R}^d$ such that the angle between each pair of them is close to $90°$ (say a number between $88°$ and $92°$)? For $d = 2$, the answer is still 2, but we are going to show that in $\mathbb{R}^d$ we can choose exponential in d many vectors such that each pair of them are almost orthogonal.

Theorem 7.4. *$\exists v^1, \cdots, v^m \in \mathbb{R}^d$ where $m \geq e^{\Omega(d)}$, such that $\forall i, j : \angle v^i, v^j \approx 90°$.*

Proof. Choose each vector v^i randomly where each coordinate of v^i is picked independently and uniformly from $\{\frac{+1}{\sqrt{d}}, \frac{-1}{\sqrt{d}}\}$. The factor $1/\sqrt{d}$ is chosen to make sure that each v^i has norm exactly 1. This implies that $\langle v^i, v^j\rangle = \cos(\angle v^i, v^j)$. So, to show that $\angle v^i, v^j \approx 90°$, it is enough to show that $\langle v^i, v^j\rangle$ is very close to 0. We show that with high probability the angle between any two vectors sampled this way is close to $90°$, and then we use a union bound to finish the proof.

Fix a unit vector u, and let v be a random vector chosen as described in the previous paragraph. We can easily calculate the expected value of their inner product. We have

$$\mathbb{E}\left[\langle v, u\rangle\right] = \mathbb{E}\left[\sum_i v_i \cdot u_i\right] = \sum_{i=1}^{d} u_i \cdot \mathbb{E}\left[v_i\right] = 0.$$

Now, for each i, let $X_i = v_i \cdot u_i$. Given that $\mathbb{E}\left[\sum_i X_i\right] = 0$, we use Hoeffding inequality to find a bound for the inner product. Observe that for each i, $\frac{-1}{d} \leq X_i \leq \frac{+1}{d}$. So, by (7.1)

$$\mathbb{P}\left[\left|\sum v_i \cdot u_i\right| > \epsilon\right] = \mathbb{P}\left[\left|\sum X_i\right| > \epsilon\right] \leq \exp\left(\frac{-2\epsilon^2}{d(1/d - (-1/d))^2}\right) = e^{-d\epsilon^2/2}.$$

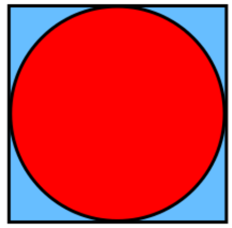

Figure 7.1: Illustration of why sampling coordinates uniformly random doesn't give a rotationally uniform vector.

We showed a probability bound for the inner product of two vectors, now we use union bound to prove a bound for the inner product of all pairs of vector.

$$\begin{aligned}\mathbb{P}\left[\forall i,j : |\langle v^i, v^j\rangle| \leq \epsilon\right] &= 1 - \mathbb{P}\left[\forall i,j |\langle v^i, v^j\rangle| > \epsilon\right] \\ &\geq 1 - \sum_{i<j} \mathbb{P}\left[|\langle v^i, v^j\rangle| > \epsilon\right] \\ &\geq 1 - \binom{m}{2} e^{\frac{-\epsilon^2 d}{2}} \geq 1 - m^2 e^{\frac{-\epsilon^2 d}{2}}\end{aligned}$$

If we set $\epsilon = \sqrt{\frac{5 \lg m}{d}}$, we have:

$$\mathbb{P}\left[\forall i,j : |\langle v^i, v^j\rangle| \leq \sqrt{\frac{5 \lg m}{d}}\right] \geq 1 - \frac{1}{\sqrt{m}}$$

Now if we set $m = e^{\frac{d}{10000}}$, then $\epsilon = \sqrt{\frac{5\frac{d}{10000}}{d}} \leq \frac{1}{40}$. This means that for every pair v^i, v^j of $e^{\frac{d}{10000}}$ vectors we have $-\frac{1}{40} \leq \cos(\angle v^i, v^j) \leq \frac{1}{40}$, so $85° \leq \angle v^i, v^j \leq 95°$. □

7.3 Sample Vectors with Uniform Direction

Let's say we want to sample a d dimensional vector v with a uniformly random direction. In other words, we would like to choose a uniformly random point on a d-dimensional ball. One simple way is to randomly sample the coordinates of a vector $v = \langle v_1, \cdots, v_d\rangle$ from range $[-1, 1]$. However, with this approach of sampling, the direction of the vectors would not be uniformly distributed. For example in $\mathbb{R}^2$, if we sample a vector by sampling each coordinate uniformly and independently, we're actually choosing a uniform random point from H_2 (square of with side of 2 around the origin). The point that gets picked is either inside B_2 (the red area of Figure 7.1) or outside of it. If the point is inside the ball, then the angle is uniformly distributed. If it's outside the ball (the blue area of Figure 7.1), it's angle is more biased towards 45°, 135°, 225°, and 315°.

We say a vector $g \in \mathbb{R}^d$ is a Guassian vector, if each coordinate of g is a uniformly and independently chosen $\mathcal{N}(0, 1)$ random variable. To choose a random direction in $\mathbb{R}^d$ it is enough to choose a Guassian vector. That vector points to a random direction. Similarly, to choose a uniformly random point on the surface of the d dimensional ball we can use $g/\|g\|$. To justify this claim, recall that the density function of $\mathcal{N}(0,1)$ is $\frac{1}{\sqrt{2\pi}}e^{-x^2/2}$. Since coordinates of g are chosen independently, the density function of g is

$$\frac{1}{(2\pi)^{d/2}}e^{-\frac{g_1^2+\cdots+g_d^2}{2}} = \frac{1}{(2\pi)^{d/2}}e^{-\|g\|_2^2/2},$$

i.e., it only depends on the length of g and it is independent of the direction that g points to.

Next, we show that a Gaussian vector g is rotationally invariant which implies the aforementioned properties.

First, let us state a nice fact about Gaussians.

Fact 7.5. *Given two Gaussian random variables $g_1 \sim \mathcal{N}(\mu_1, \sigma_1^2)$ and $g_2 \sim \mathcal{N}(\mu_2, \sigma_2^2)$, any linear combination of them, ag_1+bg_2, has the same distribution as a Gaussian random variable with mean $a\mu_1+b\mu_2$ and variance $a^2\sigma_1^2 + b^2\sigma_2^2$.*

Knowing the fact above, we can make the following claim.

Claim 7.6. *Given a d dimensional unit vector $\|v\| = 1$, and a gaussian vector $g = \langle g_1, \cdots, g_d\rangle$ whose coordinates are drawn iid from $\mathcal{N}(0,1)$, the inner product of v and g is a standard normal random variable, i.e. $\langle v, g\rangle \sim \mathcal{N}(0,1)$.*

Note that the claim is that the distribution of $\langle v, g\rangle$ doesn't depend on the direction of v. This means that the Gaussian vector g is equally directed towards any direction, meaning that it's rotationally uniform.

To prove the claim, we use Fact 7.5. In particular, we can write

$$\langle v, g\rangle = \sum_i v_i \cdot g_i \overset{\text{D}}{=} \mathcal{N}(0, v_1^2 + \cdots + v_d^2) = \mathcal{N}(0,1).$$

where we used that $\|v\| = 1$.

7.4 Rotationally Invariant Distributions

In this part we give a short introduction on Rotationally invariant distributions. First, we start by defining a rotation matrix. In 2-dimensions, a rotation matrix is matrix that rotates all of the points by an angle θ about the origin. We can display such a matrix as follows:

$$\begin{bmatrix}\cos\theta & -\sin\theta \\ \sin\theta & \cos\theta\end{bmatrix}$$

More generally, a rotation matrix is defined as follows:

Definition 7.7 (Rotation Matrix)**.** *A matrix $R \in \mathbb{R}^{n\times n}$ is a rotation matrix if for all $u \in \mathbb{R}^n$, $\|Ru\|_2 = \|u\|$.*

There is an equivalent way to define a rotation matrix. An $n \times n$ matrix

$$R = \begin{bmatrix} r^1 & \Big| & r^2 & \Big| & \dots & \Big| & r^n \end{bmatrix}$$

is a rotation matrix if it satisfies the following two properties

i) For all i, $\|r^i\| = 1$,

ii) For all $1 \le i < j \le n$, $\langle r^i, r^j \rangle = 0$.

In other words, $r^1, \dots, r^n$ form an *orthonormal* set of vectors in $\mathbb{R}^n$. Observe that if $r^1, \dots, r^n$ are orthonormal, then

$$R^T R = I$$

The reason being that the i, j-th entry of $R^T R$ is simply $\langle r^i, r^j \rangle$.

Next, we show that if $r^1, \dots, r^n$ are orthonormal, then R is a rotation matrix. Fix an arbitrary $u \in \mathbb{R}^n$. We need to show $\|Ru\| = \|u\|$ or equivalently, $\|Ru\|^2 = \|u\|^2$. But,

$$\|Ru\|^2 = (Ru)^\intercal Ru = u^\intercal R^\intercal Ru = u^\intercal I u = u^\intercal u = \|u\|^2 .$$

Definition 7.8 (Rotationally Invariant Distributions)**.** *We say a probability distribution D over $\mathbb{R}^n$ is rotationally invariant if for all $x \in \mathbb{R}^n$, and all rotation matrices R,*

$$\mathbb{P}[x] = \mathbb{P}[Rx] .$$

Next, we show that the distribution of n independent standard normal random variable $x_1, \dots, x_n$ is rotationally invariant. First, we need to recall that density function of the multivariate Gaussian distribution.

We say $X = (X_1, \dots, X_n)$ form a multivariate normal random variable when they have following density function:

$$\det(2\pi\Sigma)^{-1/2} e^{-(X-\mu)^\intercal \Sigma^{-1}(X-\mu)/2}$$

where for all i, $\mu_i = \mathbb{E}[X_i]$ and Σ is the covariance matrix of $X_1, \dots, X_n$. In particular, for all i, j,

$$\Sigma_{i,j} = \mathrm{Cov}(X_i, X_j) = \mathbb{E}[X_i - \mathbb{E}[X_i]] \, \mathbb{E}[X_j - \mathbb{E}[X_j]] = \mathbb{E}[X_i X_j] - \mathbb{E}[X_i] \, \mathbb{E}[X_j] .$$

As a special case, if $X_1, \dots, X_n$ are standard normals chosen independently then Σ is just the identity matrix. This is because by independence $\mathbb{E}[X_i X_j] = \mathbb{E}[X_i] \, \mathbb{E}[X_j]$ so the off-diagonal entries of Σ are 0, and $\mathbb{E}\left[X_i^2\right] = 1$ because X_i is standard normal, so all diagonal entries are 1.

Next, we show that if $x_1, \dots, x_n$ are independent standard normals and R is a rotation matrix, then $\mathbb{P}[x] = \mathbb{P}[Rx]$. Observe that $\mu = 0$ and $\Sigma = I$. So,

$$\mathbb{P}[Rx] \propto e^{-x^\intercal R^\intercal \Sigma^{-1} Rx/2} = e^{-x^\intercal R^\intercal I Rx/2} = e^{-x^\intercal R^\intercal Rx/2} = e^{-x^\intercal I x/2} \propto \mathbb{P}[x] ,$$

where we have not written down the normalizing constant $\det 2\pi\Sigma^{-1/2}$ to simplify the notation. This proves that by choosing n independent standard normals we obtain a rotationally invariant distribution.

7.5 Concentration of Measure for Gaussians

Let $X \sim \mathcal{N}(0, 1)$. Observe that $\mathbb{E}\left[X^2\right] = 1$. Concentration of measure for gaussians states that as we get more samples $X_1, \dots, X_k \sim \mathcal{N}(0, 1)$, the mean of the square of the samples $\frac{1}{k}\sum_i X_i^2$ gets exponentially closer to 1.

Theorem 7.9. *For k standard normal random variables $X_1, \cdots, X_k \sim \mathcal{N}(0, 1)$, and $0 < \epsilon < 1$, we have*

$$\mathbb{P}\left[\left|\frac{1}{k}\sum_i X_i^2 - 1\right| \ge \epsilon\right] \le 2e^{\frac{-k\epsilon^2}{8}}$$

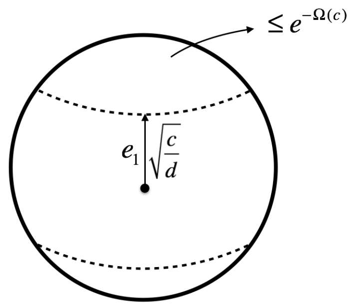

Figure 7.2: Most of the area of a d dimensional ball is on the belt around its equator.

We can use this bound to show that most of the area of a d dimensional ball is on the belt around its equator.

Theorem 7.10. *If we slice the ball B_d with the hyperplane $xe^1 = \sqrt{\frac{c}{d}}$, where e^1 is as defined in (7.2), the volume of the ball above (or below) the hyperplane is less than $e^{-\Omega(c)}$.*

Proof. To prove the claim it is enough to show that a uniformly random point $g/\|g\|$ on the surface of unit d-dimensional ball has a small inner product with e^1 vector. To be precise we show that

$$\mathbb{P}\left[|\langle \frac{g}{\|g\|}, e^1\rangle| \geq \sqrt{\frac{c}{d}}\right] \leq e^{-\Omega(c)}.$$

First, observe that

$$\mathbb{P}\left[|\langle \frac{g}{\|g\|}, e^1\rangle| \geq \sqrt{\frac{c}{d}}\right] = \mathbb{P}\left[\langle \frac{g}{\|g\|}, e^1\rangle^2 \geq \frac{c}{d}\right] = \mathbb{P}\left[\langle g, e^1\rangle^2 \geq \|g\|^2 \frac{c}{d}\right] = \mathbb{P}\left[g_1^2 \geq \|g\|^2 \frac{c}{d}\right]$$

We use the concentration of measure of independent Gaussians, to show that $\|g\|^2$ is highly concentrated around d.

$$\mathbb{P}\left[| \|g\|^2 - d| > \epsilon d\right] = \mathbb{P}\left[\left|\sum_i g_i^2 - d\right| > \epsilon d\right] = \mathbb{P}\left[\left|\frac{1}{d}\sum g_i^2 - 1\right| > \epsilon\right] \leq e^{\frac{-d\epsilon^2}{8}},$$

where in the last inequality we used Theorem 7.9. So, for $\epsilon = \frac{1}{2}$ we have:

$$\mathbb{P}\left[\|g\|^2 < \frac{d}{2}\right] \leq e^{-\Omega(d)}$$

On the other hand, g_1 is just a $\mathcal{N}(0,1)$ random variable, so

$$\mathbb{P}\left[g_1^2 \geq c/2\right] = \mathbb{P}\left[|g_i| \geq \sqrt{c/2}\right] \leq e^{-\Omega(c)},$$

where the latter simply follows from the Gaussian probability density function.

Now, we can finish the proof by a union bound argument. If $g_1^2 > \frac{c}{d}\|g\|^2$, then we must either have $\|g\|^2 < d/2$ or $g_1^2 > c/2$ (or both). So, by union bound,

$$\mathbb{P}\left[g_1^2 \geq \|g\|^2 \frac{c}{d}\right] \leq \mathbb{P}\left[g_1^2 \geq c/2\right] + \mathbb{P}\left[\|g\|^2 \geq d/2\right] \leq e^{-\Omega(c)} + e^{-\Omega(d)} \leq 2e^{-\Omega(c)}.$$

□

7.6 Dimension Reduction

Now we describe a central result of high dimensional geometry. Given n points $x^1, x^2, \cdots, x^n \in \mathbb{R}^d$. We would like to find n points $y^1, y^2, \cdots, y^n \in \mathbb{R}^k$, where $k \ll d$ such that for all i, j

$$(1-\epsilon)\left\|x^i - x^j\right\| \le \left\|y^i - y^j\right\| \le (1+\epsilon)\left\|x^i - x^j\right\|. \tag{7.3}$$

Ideally, we would like to have ϵ very close to zero and $k \ll d$.

Theorem 7.11 (Johnson-Lindenstrauss)**.** . *Given n points $x^1, x^2, \cdots, x^n \in \mathbb{R}^d$, there exists a linear mapping $Gx_i = y_i \in \mathbb{R}^k$ such that each i, j satisfy (7.3) and $k = O(\frac{\lg n}{\epsilon^2})$.*

The remarkable fact about this theorem is that k does not depend on d; it only depends on the number of points and ϵ.

One simple idea to prove the above theorem is to let the rows of G be k standard basis vectors chosen uniformly at random. However, this doesn't work very well in the worst case. A bad example is where all x_i's are zero on all coordinates except the last k. Such a mapping G with high probability distorts most pair of points.

The right mapping is to let G be a $k \times n$ matrix whose element are drawn iid from $\mathcal{N}(0, \frac{1}{k})$, i.e., each row of G is a d-dimensional Gaussian vector scaled by $1/\sqrt{k}$,

$$G = \begin{bmatrix} \sqrt{1/k} & g^1 \\ \sqrt{1/k} & g^2 \\ \cdots & \\ \sqrt{1/k} & g^k. \end{bmatrix}$$

Firstly, observe that for any vector $v \in \mathbb{R}^d$, $\mathbb{E}\left[\|Gv\|^2\right] = 1$. This is because

$$\mathbb{E}\left[\|Gv\|^2\right] = \mathbb{E}\sum_{i=1}^{k} \frac{1}{k}\langle g^i, v\rangle^2 = \frac{1}{k}\sum_{i=1}^{k}\mathbb{E}\left[\langle g^i, v\rangle^2\right] = 1,$$

where the last equality uses the rotation invariance property of Gaussians. In particular, let $X \sim \mathcal{N}(0,1)$. Then, since $\langle g^i, v\rangle \overset{\mathrm{D}}{=} \mathcal{N}(0,1)$, we can write

$$\mathbb{E}\left[\langle g^i, v\rangle^2\right] = \mathbb{E}X^2 = \mathrm{Var}(X) - \mathbb{E}\left[X\right]^2 = \mathrm{Var}(X) = 1.$$

Claim 7.12. *For any unit vector v, and $0 < \epsilon < 1$,*

$$\mathbb{P}\left[\left|\|Gv\|^2 - 1\right| > \epsilon\right] \le e^{-\epsilon^2 k/8}$$

Proof. By the rotation invariance property, Gv has the same distribution as Ge^1. We can write

$$Ge^1 = \begin{pmatrix} g_1^1/\sqrt{k} \\ g_1^2/\sqrt{k} \\ \vdots \\ g_1^k/\sqrt{k} \end{pmatrix}.$$

So,

$$\left\|Ge^1\right\|^2 = \sum_{i=1}^{k} \frac{1}{k}(g_1^i)^2.$$

The claim follows by Theorem 7.9. □

Now, we are ready to finish the proof of Theorem 7.11. Fix a pair $1 \leq i < j \leq n$. In the following claim we show that $\left\|y^i - y^j\right\| \approx \left\|x^i - x^j\right\|$ with high probability. Then, Theorem 7.11 follows by a union bound argument.

Claim 7.13. *For any $1 \leq i < j \leq n$, and $0 < \epsilon < 1$,*

$$\mathbb{P}\left[(1-\epsilon)\left\|x^i - x^j\right\| \leq \left\|y^i - y^j\right\| \leq (1+\epsilon)\left\|x^i - x^j\right\|\right] \geq 1 - e^{-\epsilon^2 k/8}.$$

Proof. The claim essentially follows from Claim 7.12 after doing some algebraic manipulations. The main nontrivial step is to use the linearity of the projection operator, i.e., $y^i - y^j = Gx^i - Gx^j = G(x^i - x^j)$.

$$\begin{aligned}
\mathbb{P}\left[(1-\epsilon)\left\|x^i - x^j\right\| \leq \left\|y^i - y^j\right\| \leq (1+\epsilon)\left\|x^i - x^j\right\|\right] &= \mathbb{P}\left[(1-\epsilon)^2\left\|x^i - x^j\right\|^2 \leq \left\|y^i - y^j\right\|^2 \leq (1+\epsilon)^2\left\|x^i - x^j\right\|^2\right] \\
&\geq \mathbb{P}\left[1-\epsilon \leq \frac{\left\|y^i - y^j\right\|^2}{\left\|x^i - x^j\right\|^2} \leq 1+\epsilon\right] \\
&= \mathbb{P}\left[\left|\frac{\left\|Gx^i - Gx^j\right\|^2}{\left\|x^i - x^j\right\|^2} - 1\right| \leq \epsilon\right] \\
&= \mathbb{P}\left[\left|\left\|G\frac{x^i - x^j}{\left\|x^i - x^j\right\|}\right\|^2 - 1\right| \leq \epsilon\right] \geq 1 - e^{-k\epsilon^2/8},
\end{aligned}$$

where the last inequality follows by Claim 7.12. □

To prove Theorem 7.9 it is enough to let $k = 24\log n/\epsilon^2$. Then, we get

$$\mathbb{P}\left[\forall i,j : (1-\epsilon)\left\|x^i - x^j\right\| \leq \left\|y^i - y^j\right\| \leq (1+\epsilon)\left\|x^i - x^j\right\|\right] \geq 1 - n^2 e^{-\epsilon^2 k/8} \geq 1 - 1/n.$$

This completes the proof of Theorem 7.11.

See the course website for several recent works related to dimension reduction and Johnson-Lindenstrauss theorem.