For points $x=(x_1,\dots,x_k)$ & $y=(y_1,\dots,y_k)$ where $x,y\in\mathbb{R}^k$,
their L_2 = Euclidean distance

$$\|x-y\| = \|x-y\|_2 = \sqrt{\sum_{i=1}^{k}(x_i-y_i)^2}$$

We saw fingerprinting where we transformed from all integers to modulo a prime.

Here we're going to look at the geometric version.

Suppose we have n points $x_1,\dots,x_n$ in $\mathbb{R}^k$ where k is large.

We want to project down into a small # of dimensions & preserve (approximately) the distances between all pairs of points.

Choose a <u>random vector</u> $r=(r_1,\dots,r_k)\in\mathbb{R}^k$
(Will specify later how to choose r)

Let $f_r:\mathbb{R}^k\to\mathbb{R}$ be defined as $f_r(x)=\langle x, r\rangle = x^T r = \sum_{j=1}^{k} x_j r_j$

(this is the projection of x onto r)

Gaussian = normal distribution:

for $X \sim N(\mu, \sigma^2)$, $p(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{\frac{-(x-\mu)^2}{2\sigma^2}}$

if $X_1 \sim N(\mu_1, \sigma_1^2)$ & $X_2 \sim N(\mu_2, \sigma_2^2)$

then $X_1 + X_2 \sim N(\mu_1 + \mu_2, \sigma_1^2 + \sigma_2^2)$

Choose $r_1, \ldots, r_k$ independently from standard Gaussian $= N(0,1)$

Then, for $x, y \in \mathbb{R}^k$,

$$f_r(x) - f_r(y) = \sum_{j=1}^{k} x_j r_j - \sum_{j=1}^{k} y_j r_j = \sum_{j=1}^{k} (x_j - y_j) r_j$$

For fixed x & y, since $r_j \sim N(0,1)$

then $(x_j - y_j) r_j \sim N(0, (x_j - y_j)^2)$

(for r.v. X with mean μ & variance σ^2, λX has var. $\lambda^2\sigma^2$) for constant λ

Since Gaussians add see

thus, $(f_r(x) - f_r(y)) \sim N(0, \sum_j (x_j - y_j)^2)$

$= N(0, \|x - y\|_2^2)$

Note for r.v. X with mean 0, $Var(X) = E[X^2]$.

Hence, for the r.v. $f_r(x) - f_r(y)$:

$$E[(f_r(x) - f_r(y))^2] = \|x-y\|_2^2$$

Therefore, $(f_r(x) - f_r(y))^2$ is an unbiased estimator for the squared distance b/w x & y.

Now that just preserves the distance in expectation, but we want to do this for all $\binom{n}{2}$ distances & we want to do this with high probability.

So instead of picking one r let's pick d vectors $r_1, \dots, r_d \in \mathbb{R}^k$ then we'll get d independent estimates of $\|x-y\|_2^2$ & take the average = mean of these estimates.

By a Chernoff-type inequality we'll get that $d = O(\frac{1}{\epsilon^2} \log n)$ suffices, to preserve all $\binom{n}{2}$ distances within $(1 \pm \epsilon)$-factor.

Theorem: [Johnson, Lindenstrauss '84]

For all $\epsilon > 0$, all $X = \{x_1, \ldots, x_n\}$ where $x_i \in \mathbb{R}^k$,
there is a (linear) map $f: \mathbb{R}^k \to \mathbb{R}^\ell$ with $\ell = O\left(\frac{\log n}{\epsilon^2}\right)$
where for all $x_i, x_j \in X$,

$$(1-\epsilon)\|x_i - x_j\|_2^2 \le \|f(x_i) - f(x_j)\|_2^2 \le (1+\epsilon)\|x_i - x_j\|_2^2$$

with probability $\geq 1 - \frac{1}{\text{poly}(n)}$

χ^2-distribution:

Let $Z_1, \ldots, Z_\ell$ be iid $N(0,1)$ random variables

& let $Q = \sum_{i=1}^{\ell} Z_i^2$

then $Q \sim \chi_\ell^2$

Charnoff-type bound for χ^2-distribution:

$$\Pr\left(\chi_\ell^2 \geq (1+\epsilon)\ell\right) \leq e^{-\frac{\ell}{4}(\epsilon^2 - \epsilon^3)}$$

$$\Pr\left(\chi_\ell^2 \leq (1-\epsilon)\ell\right) \leq e^{-\frac{\ell}{4}(\epsilon^2 - \epsilon^3)}$$

Proof:

$$\Pr\left(\chi_\ell^2 \geq (1+\epsilon)\ell\right) = \Pr\left(\sum_{i=1}^{\ell} Z_i^2 \geq (1+\epsilon)\ell\right)$$

$$= \Pr\left(e^{\lambda \sum_i Z_i^2} \geq e^{(1+\epsilon)\ell\lambda}\right)$$

$$\leq \frac{E[e^{\lambda \sum Z_i^2}]}{e^{(1+\epsilon)\ell\lambda}} \cancel{= E[e^{\lambda \sum Z_i^2}}$$

$$= \frac{\left(E[e^{\lambda Z_i^2}]\right)^\ell}{e^{(1+\epsilon)\ell\lambda}}$$

Claim: $E\left[e^{\lambda Z_i^2}\right] = \left(\frac{1}{1-2\lambda}\right)^{1/2}$ for $\lambda < 1/2$

Thus,

$$\Pr\left(\chi_\ell^2 \geq (1+\epsilon)\ell\right) \leq e^{-(1+\epsilon)\ell\lambda}\left(\frac{1}{1-2\lambda}\right)^{\ell/2}$$

Set $\lambda = \frac{\epsilon}{2(1+\epsilon)}$

then:

$$\leq \left(\frac{1+\epsilon}{e^{\epsilon}}\right)^{\ell/2}$$

$$\leq \left(\frac{e^{\epsilon - \frac{\epsilon^2}{2} + \frac{\epsilon^3}{2}}}{e^{\epsilon}}\right)^{\ell/2}$$ Since $1+\epsilon \leq e^{\epsilon - \frac{\epsilon^2}{2} + \epsilon^3/2}$

$$\leq e^{-\frac{\ell}{4}(\epsilon^2 - \epsilon^3)}$$

$\blacksquare$

This concentration for χ^2 distributions implies the following lemma.

Lemma: For $x \in \mathbb{R}^m$, let A be a $l \times m$ matrix where each entry is iid $N(0,1)$. Then,

$$\Pr\left((1-\epsilon)\|x\|^2 \leq \left\|\frac{1}{\sqrt{l}} Ax\right\|^2 \leq (1+\epsilon)\|x\|^2 \right) \geq 1 - 2e^{-\frac{l}{4}(\epsilon^2-\epsilon^3)}$$

Then to prove the JL-theorem

let $f = \frac{1}{\sqrt{l}} Ax$

& the theorem follows by a union bound.

Why $\frac{1}{\sqrt{l}}$ Scaling?

It gets squared so we get $\frac{1}{l}$

Also, if $\|x\|^2 = 1$ then we want $\left\|\frac{1}{\sqrt{l}} Ax\right\|^2 = 1$?

And this is easily checked for the toy case where $x = e_1 = (1, 0, \ldots, 0)$.