COMPSCI 514: ALGORITHMS FOR DATA SCIENCE

Andrew McGregor
Lecture 11
**The Johnson-Lindenstrauss Lemma**

Johnson-Lindenstrauss Lemma: For any set of points $\mathbf{x}_1, \ldots, \mathbf{x}_n \in \mathbb{R}^d$ and $\epsilon > 0$ there exists a linear map $M : \mathbb{R}^d \to \mathbb{R}^m$ such that $m = O\left(\frac{\log n}{\epsilon^2}\right)$ and letting $\tilde{x}_i = M\mathbf{x}_i$:

For all $i, j$:

$$(1 - \epsilon)\|\mathbf{x}_i - \mathbf{x}_j\|_2 \leq \|\tilde{x}_i - \tilde{x}_j\|_2 \leq (1 + \epsilon)\|\mathbf{x}_i - \mathbf{x}_j\|_2.$$  

Further, if $M \in \mathbb{R}^{m \times d}$ has each entry chosen independently from $\mathcal{N}(0, 1/m)$, it satisfies the guarantee with high probability.
The Johnson-Lindenstrauss Lemma is a direct consequence of:

**Distributional JL Lemma:** Let $M \in \mathbb{R}^{m \times d}$ have each entry chosen i.i.d. as $\mathcal{N}(0, 1/m)$. If we set $m = O\left(\frac{\log(1/\delta)}{\epsilon^2}\right)$, then for any $\vec{y} \in \mathbb{R}^d$, with probability $\geq 1 - \delta$

$$(1 - \epsilon)\|\vec{y}\|_2 \leq \|M\vec{y}\|_2 \leq (1 + \epsilon)\|\vec{y}\|_2$$

**M ∈ R^{m×d}:** random projection matrix. $d$: original dimension. $m$: compressed dimension, $\epsilon$: embedding error, $\delta$: embedding failure prob.
The Johnson-Lindenstrauss Lemma is a direct consequence of:

**Distributional JL Lemma:** Let $M \in \mathbb{R}^{m \times d}$ have each entry chosen i.i.d. as $\mathcal{N}(0, 1/m)$. If we set $m = O\left(\frac{\log(1/\delta)}{\epsilon^2}\right)$, then for any $\tilde{y} \in \mathbb{R}^d$, with probability $\geq 1 - \delta$

$$(1 - \epsilon)\|\tilde{y}\|_2 \leq \|M\tilde{y}\|_2 \leq (1 + \epsilon)\|\tilde{y}\|_2$$

I.e., applying a random matrix $M$ to any vector $\tilde{y}$ preserves the norm with high probability. Like a low-distortion embedding, but for the length of a compressed vector rather than distances between vectors.

$M \in \mathbb{R}^{m \times d}$: random projection matrix. $d$: original dimension. $m$: compressed dimension, $\epsilon$: embedding error, $\delta$: embedding failure prob.
Distributional JL Lemma: Let $M \in \mathbb{R}^{m \times d}$ have independent $\mathcal{N}(0, 1/m)$ entries. If we set $m = O\left(\frac{\log(1/\delta)}{\epsilon^2}\right)$, then for any $y \in \mathbb{R}^d$, with probability at least $1 - \delta$

\[(1 - \epsilon)\|y\|_2 \leq \|My\|_2 \leq (1 + \epsilon)\|y\|_2.\]
Distributional JL Lemma: Let $M \in \mathbb{R}^{m \times d}$ have independent $\mathcal{N}(0, 1/m)$ entries. If we set $m = O \left( \frac{\log(1/\delta)}{\epsilon^2} \right)$, then for any $y \in \mathbb{R}^d$, with probability at least $1 - \delta$

$$(1 - \epsilon)\|y\|_2 \leq \|My\|_2 \leq (1 + \epsilon)\|y\|_2.$$

- Let $\tilde{y} = My$ and $M_j$ be the $j^{th}$ row of $M$
Distributional JL Lemma: Let \( M \in \mathbb{R}^{m \times d} \) have independent \( \mathcal{N}(0, 1/m) \) entries. If we set \( m = O \left( \frac{\log(1/\delta)}{\epsilon^2} \right) \), then for any \( y \in \mathbb{R}^d \), with probability at least \( 1 - \delta \)

\[
(1 - \epsilon)\|y\|_2 \leq \|My\|_2 \leq (1 + \epsilon)\|y\|_2.
\]

- Let \( \tilde{y} = My \) and \( M_j \) be the \( j^{th} \) row of \( M \)
- For any \( j \), \( \tilde{y}_j = \langle M_j, y \rangle \)
**Distributional JL Lemma:** Let \( M \in \mathbb{R}^{m \times d} \) have independent \( \mathcal{N}(0, 1/m) \) entries. If we set \( m = O\left(\frac{\log(1/\delta)}{\epsilon^2}\right) \), then for any \( y \in \mathbb{R}^d \), with probability at least \( 1 - \delta \)

\[
(1 - \epsilon)\|y\|_2 \leq \|My\|_2 \leq (1 + \epsilon)\|y\|_2.
\]

- Let \( \tilde{y} = My \) and \( M_j \) be the \( j^{th} \) row of \( M \)
- For any \( j \), \( \tilde{y}_j = \langle M_j, y \rangle = \sum_{i=1}^d g_i \cdot y_i \) where \( g_i \sim \mathcal{N}(0, 1/m) \).
DISTRIBUTIONAL JL PROOF (PART 1 OF 3)

**Distributional JL Lemma:** Let $M \in \mathbb{R}^{m \times d}$ have independent $\mathcal{N}(0, 1/m)$ entries. If we set $m = O\left(\frac{\log(1/\delta)}{\epsilon^2}\right)$, then for any $y \in \mathbb{R}^d$, with probability at least $1 - \delta$

$$(1 - \epsilon)\|y\|_2 \leq \|My\|_2 \leq (1 + \epsilon)\|y\|_2.$$ 

- Let $\tilde{y} = My$ and $M_j$ be the $j^{th}$ row of $M$.
- For any $j$, $\tilde{y}_j = \langle M_j, y \rangle = \sum_{i=1}^{d} g_i \cdot y_i$ where $g_i \sim \mathcal{N}(0, 1/m)$.
- By linearity of expectation:
  $$\mathbb{E}[\tilde{y}_j] = \sum_{i=1}^{d} \mathbb{E}[g_i] \cdot y_i = 0.$$
**Distributional JL Lemma:** Let $M \in \mathbb{R}^{m \times d}$ have independent $\mathcal{N}(0, 1/m)$ entries. If we set $m = O\left(\frac{\log(1/\delta)}{\epsilon^2}\right)$, then for any $y \in \mathbb{R}^d$, with probability at least $1 - \delta$,

$$(1 - \epsilon)\|y\|_2 \leq \|My\|_2 \leq (1 + \epsilon)\|y\|_2.$$ 

- Let $\tilde{y} = My$ and $M_j$ be the $j^{th}$ row of $M$.
- For any $j$, $\tilde{y}_j = \langle M_j, y \rangle = \sum_{i=1}^{d} g_i \cdot y_i$ where $g_i \sim \mathcal{N}(0, 1/m)$.
- By linearity of expectation:

$$\mathbb{E}[\tilde{y}_j] = \sum_{i=1}^{d} \mathbb{E}[g_i] \cdot y_i = 0 .$$

- Since $\mathbb{E}[\tilde{y}_j] = 0$, we have $\mathbb{E}[\tilde{y}_j^2] = \text{Var}[\tilde{y}_j]$. Then, by linearity of variance:

$$\mathbb{E}[\tilde{y}_j^2] = \text{Var}[\tilde{y}_j] = \sum_{i=1}^{d} \text{Var}[g_i \cdot y_i] = \sum_{i} y_i^2 / m = \|y\|_2^2 / m .$$
**Distributional JL Lemma:** Let $M \in \mathbb{R}^{m \times d}$ have independent $\mathcal{N}(0, 1/m)$ entries. If we set $m = O\left(\frac{\log(1/\delta)}{\epsilon^2}\right)$, then for any $y \in \mathbb{R}^d$, with probability at least $1 - \delta$

$$(1 - \epsilon)\|y\|_2 \leq \|My\|_2 \leq (1 + \epsilon)\|y\|_2.$$

- Let $\tilde{y} = My$ and $M_j$ be the $j^{th}$ row of $M$.
- For any $j$, $\tilde{y}_j = \langle M_j, y \rangle = \sum_{i=1}^{d} g_i \cdot y_i$ where $g_i \sim \mathcal{N}(0, 1/m)$.
- By linearity of expectation:
  
  $$\mathbb{E}[\tilde{y}_j] = \sum_{i=1}^{d} \mathbb{E}[g_i] \cdot y_i = 0.$$  

- Since $\mathbb{E}[\tilde{y}_j] = 0$ we have $\mathbb{E}[\tilde{y}_j^2] = \text{Var}[\tilde{y}_j]$. Then, by linearity of variance:
  
  $$\mathbb{E}[\tilde{y}_j^2] = \text{Var}[\tilde{y}_j] = \sum_{i=1}^{d} \text{Var}[g_i \cdot y_i] = \sum_{i} y_i^2 / m = \|y\|_2^2 / m.$$  

- Hence $\mathbb{E}[\|\tilde{y}\|_2^2] = \mathbb{E}[\sum_j \tilde{y}_j^2] = \|y\|_2^2$. 
Distributional JL Lemma: Let $M \in \mathbb{R}^{m \times d}$ have independent $\mathcal{N}(0, 1/m)$ entries. If we set $m = O\left(\frac{\log(1/\delta)}{\epsilon^2}\right)$, then for any $y \in \mathbb{R}^d$, with probability at least $1 - \delta$

$$(1 - \epsilon)\|y\|_2 \leq \|M y\|_2 \leq (1 + \epsilon)\|y\|_2.$$

- Let $\tilde{y} = My$ and $M_j$ be the $j^{th}$ row of $M$.
- For any $j$, $\tilde{y}_j = \langle M_j, y \rangle = \sum_{i=1}^d g_i \cdot y_i$ where $g_i \sim \mathcal{N}(0, 1/m)$.
- By linearity of expectation:

$$\mathbb{E}[\tilde{y}_j] = \sum_{i=1}^d \mathbb{E}[g_i] \cdot y_i = 0.$$

- Since $\mathbb{E}[\tilde{y}_j] = 0$ we have $\mathbb{E}[\tilde{y}_j^2] = \text{Var}[\tilde{y}_j]$. Then, by linearity of variance:

$$\mathbb{E}[\tilde{y}_j^2] = \text{Var}[\tilde{y}_j] = \sum_{i=1}^d \text{Var}[g_i \cdot y_i] = \sum_{i} y_i^2 / m = \|y\|_2^2 / m.$$

- Hence $\mathbb{E}[\|\tilde{y}\|_2^2] = \mathbb{E}[\sum_j \tilde{y}_j^2] = \|y\|_2^2$. Remains to show $\|\tilde{y}\|_2$ is concentrated.
Letting $\tilde{y} = My$, we have $\tilde{y}_j = \langle M_j, y \rangle$ and:

$$\tilde{y}_j = \sum_{i=1}^{d} g_i \cdot y_i$$

where $g_i \cdot y_i \sim \mathcal{N}(0, y_i^2 / m)$.

**Stability of Gaussian Random Variables.** For independent $a \sim \mathcal{N}(\mu_1, \sigma_1^2)$ and $b \sim \mathcal{N}(\mu_2, \sigma_2^2)$ we have:

$$a + b \sim \mathcal{N}(\mu_1 + \mu_2, \sigma_1^2 + \sigma_2^2)$$

Thus, $\tilde{y}_j \sim \mathcal{N}(0, \sum_{i=1}^{d} y_i^2 / m) = \mathcal{N}(0, \|y\|_2^2 / m)$. 
So Far: Each entry of our compressed vector $\tilde{y}$ is Gaussian with:

$$\tilde{y}_j \sim \mathcal{N}(0, \|y\|^2_2/m) \text{ and } \mathbb{E}[\|\tilde{y}\|_2^2] = \|y\|^2_2$$
So Far: Each entry of our compressed vector $\tilde{y}$ is Gaussian with:

$$\tilde{y}_j \sim \mathcal{N}(0, \frac{\|y\|_2^2}{m}) \text{ and } \mathbb{E}[\|\tilde{y}\|_2^2] = \|y\|_2^2$$

$$\|\tilde{y}\|_2^2 = \sum_{i=1}^{m} \tilde{y}_j^2$$  a Chi-Squared random variable with $m$ degrees of freedom (a sum of $m$ squared independent Gaussians)
So Far: Each entry of our compressed vector $\tilde{y}$ is Gaussian with:

$$\tilde{y}_j \sim \mathcal{N}(0, \|y\|_2^2/m)$$

and

$$\mathbb{E}[\|\tilde{y}\|_2^2] = \|y\|_2^2$$

$$\|\tilde{y}\|_2^2 = \sum_{i=1}^{m} \tilde{y}_j^2$$

a Chi-Squared random variable with $m$ degrees of freedom (a sum of $m$ squared independent Gaussians)
**So Far:** Each entry of our compressed vector $\tilde{y}$ is Gaussian with:

$$\tilde{y}_j \sim \mathcal{N}(0, \|y\|_2^2/m) \text{ and } \mathbb{E}[\|\tilde{y}\|_2^2] = \|y\|_2^2$$

$$\|\tilde{y}\|_2^2 = \sum_{i=1}^{m} \tilde{y}_j^2 \quad \text{a Chi-Squared random variable with } m \text{ degrees of freedom (a sum of } m \text{ squared independent Gaussians)}$$

**Lemma:** (Chi-Squared Concentration) Letting $Z$ be a Chi-Squared random variable with $m$ degrees of freedom,

$$\Pr[|Z - \mathbb{E}Z| \geq \epsilon \mathbb{E}Z] \leq 2e^{-me^2/8}.$$
**Distributional JL Proof (Part 3 of 3)**

**So Far:** Each entry of our compressed vector $\tilde{y}$ is Gaussian with:

$$\tilde{y}_j \sim \mathcal{N}(0, \|y\|^2_2/m) \text{ and } \mathbb{E}[\|\tilde{y}\|^2_2] = \|y\|^2_2$$

$$\|\tilde{y}\|^2_2 = \sum_{i=1}^m \tilde{y}_j^2$$

is a Chi-Squared random variable with $m$ degrees of freedom (a sum of $m$ squared independent Gaussians)

**Lemma:** (Chi-Squared Concentration) Letting $Z$ be a Chi-Squared random variable with $m$ degrees of freedom,

$$\Pr[|Z - \mathbb{E}Z| \geq \epsilon \mathbb{E}Z] \leq 2e^{-me^2/8}.$$ 

If we set $m = O\left(\frac{\log(1/\delta)}{\epsilon^2}\right)$, with probability $1 - O(e^{-\log(1/\delta)}) \geq 1 - \delta$:

$$(1 - \epsilon)\|y\|^2_2 \leq \|\tilde{y}\|^2_2 \leq (1 + \epsilon)\|y\|^2_2.$$
DISTRIBUTIONAL JL PROOF (PART 3 OF 3)

So Far: Each entry of our compressed vector $\tilde{y}$ is Gaussian with:

$$\tilde{y}_j \sim \mathcal{N}(0, \|y\|_2^2/m) \text{ and } \mathbb{E}[\|\tilde{y}\|_2^2] = \|y\|_2^2$$

$$\|\tilde{y}\|_2^2 = \sum_{i=1}^{m} \tilde{y}_j^2$$ is a Chi-Squared random variable with $m$ degrees of freedom (a sum of $m$ squared independent Gaussians)

**Lemma:** (Chi-Squared Concentration) Letting $Z$ be a Chi-Squared random variable with $m$ degrees of freedom,

$$\Pr[|Z - \mathbb{E}Z| \geq \epsilon \mathbb{E}Z] \leq 2e^{-me^2/8}.$$

If we set $m = O\left(\frac{\log(1/\delta)}{\epsilon^2}\right)$, with probability $1 - O(e^{-\log(1/\delta)}) \geq 1 - \delta$:

$$(1 - \epsilon)\|y\|_2^2 \leq \|\tilde{y}\|_2^2 \leq (1 + \epsilon)\|y\|_2^2.$$

Gives the distributional JL Lemma and thus the classic JL Lemma!
**Goal:** Separate $n$ points in $d$ dimensional space into $k$ groups $C_1, \ldots, C_k$. 

**k-means Objective:**

$$\text{Cost}(C_1, \ldots, C_k) = \sum_{j=1}^{k} \sum_{\vec{x} \in C_j} \|\vec{x} - \mu_j\|^2$$

where $\mu_j = \frac{1}{|C_j|} \sum_{\vec{x} \in C_j} \vec{x}$ is the average of the points in $C_j$. 

**Exercise:** Can be rewritten as

$$\text{Cost}(C_1, \ldots, C_k) = \sum_{j=1}^{k} \sum_{\vec{x}_1, \vec{x}_2 \in C_j} \|\vec{x}_1 - \vec{x}_2\|^2$$

where $\mu_j$ is the average of the points in $C_j$. 

---

**EXAMPLE APPLICATION: k-MEANS CLUSTERING**
**Goal:** Separate $n$ points in $d$ dimensional space into $k$ groups $C_1, \ldots, C_k$.

**k-means Objective:** $\text{Cost}(C_1, \ldots, C_k) = \sum_{j=1}^{k} \sum_{\vec{x} \in C_j} \| \vec{x} - \mu_j \|^2$ where 

$$\mu_j = \frac{1}{|C_j|} \sum_{\vec{x} \in C_j} \vec{x}$$

is the average of the points in $C_j$. 

*Example application: k-means clustering*
**Goal:** Separate $n$ points in $d$ dimensional space into $k$ groups $C_1, \ldots, C_k$.

**k-means Objective:** $Cost(C_1, \ldots, C_k) = \sum_{j=1}^{k} \sum_{\vec{x} \in C_j} \|\vec{x} - \vec{\mu}_j\|_2^2$ where

$$\vec{\mu}_j = \frac{1}{|C_j|} \sum_{\vec{x} \in C_j} \vec{x}$$

is the average of the points in $C_j$.

**Exercise:** Can be rewritten as $Cost(C_1, \ldots, C_k) = \sum_{j=1}^{k} \sum_{\vec{x}_1, \vec{x}_2 \in C_j} \frac{\|\vec{x}_1 - \vec{x}_2\|_2^2}{|C_j|}$
**k-means Objective:** $\text{Cost}(C_1, \ldots, C_k) = \sum_{j=1}^{k} \sum_{\bar{x}_1, \bar{x}_2 \in C_j} \frac{||\bar{x}_1 - \bar{x}_2||^2}{|C_j|}$
**EXAMPLE APPLICATION: k-MEANS CLUSTERING**

**k-means Objective:** $\text{Cost}(C_1, \ldots, C_k) = \sum_{j=1}^{k} \sum_{\vec{x}_1, \vec{x}_2 \in C_j} \frac{\|\vec{x}_1 - \vec{x}_2\|_2^2}{|C_j|}$

If we randomly project to $m = O\left(\epsilon^{-2} \log n\right)$ dimensions, for all pairs $\vec{x}_1, \vec{x}_2$,

$$(1 - \epsilon)\|\vec{x}_1 - \vec{x}_2\|_2^2 \leq \|\tilde{\vec{x}}_1 - \tilde{\vec{x}}_2\|_2^2 \leq (1 + \epsilon)\|\vec{x}_1 - \vec{x}_2\|_2^2$$
**EXAMPLE APPLICATION: k-MEANS CLUSTERING**

**k-means Objective:** \( \text{Cost}(C_1, \ldots, C_k) = \sum_{j=1}^{k} \sum \bar{x}_1, \bar{x}_2 \in C_j \frac{\|\bar{x}_1 - \bar{x}_2\|^2}{|C_j|} \)

If we randomly project to \( m = O(\epsilon^{-2} \log n) \) dimensions, for all pairs \( \bar{x}_1, \bar{x}_2 \),

\[
(1 - \epsilon)\|\bar{x}_1 - \bar{x}_2\|_2^2 \leq \|\bar{x}_1 - \bar{x}_2\|_2^2 \leq (1 + \epsilon)\|\bar{x}_1 - \bar{x}_2\|_2^2
\]

Letting \( \overline{\text{Cost}}(C_1, \ldots, C_k) = \sum_{j=1}^{k} \sum \bar{x}_1, \bar{x}_2 \in C_j \frac{\|\bar{x}_1 - \bar{x}_2\|^2}{|C_j|} \)

\[
(1 - \epsilon)\text{Cost}(C_1, \ldots, C_k) \leq \overline{\text{Cost}}(C_1, \ldots, C_k) \leq (1 + \epsilon)\text{Cost}(C_1, \ldots, C_k).
\]
**Example Application: k-Means Clustering**

**k-means Objective:** \( \text{Cost}(C_1, \ldots, C_k) = \sum_{j=1}^{k} \sum_{x_1, x_2 \in C_j} \frac{\|x_1 - x_2\|_2^2}{|C_j|} \)

If we randomly project to \( m = O(\epsilon^{-2} \log n) \) dimensions, for all pairs \( x_1, x_2 \),

\[
(1 - \epsilon)\|x_1 - x_2\|_2^2 \leq \|\tilde{x}_1 - \tilde{x}_2\|_2^2 \leq (1 + \epsilon)\|x_1 - x_2\|_2^2
\]

Letting \( \overline{\text{Cost}}(C_1, \ldots, C_k) = \sum_{j=1}^{k} \sum_{x_1, x_2 \in C_j} \frac{\|\tilde{x}_1 - \tilde{x}_2\|_2^2}{|C_j|} \)

\[
(1 - \epsilon)\text{Cost}(C_1, \ldots, C_k) \leq \overline{\text{Cost}}(C_1, \ldots, C_k) \leq (1 + \epsilon)\text{Cost}(C_1, \ldots, C_k).
\]

**Upshot:** Can cluster in \( m \) dimensional space (much more efficiently) and minimize \( \overline{\text{Cost}}(C_1, \ldots, C_k) \).
JL LEMMA IS ALMOST OPTIMAL
• Recall that we say two vectors $x, y$ are orthogonal if $\langle x, y \rangle = 0$. 

• What is the largest set of mutually orthogonal unit vectors in $d$-dimensional space? Answer: $d$. 

• How large can a set of unit vectors in $d$-dimensional space be that have all pairwise dot products $|\langle x, y \rangle| \leq \epsilon$? Answer: $2\Omega(\epsilon^2 d)$. An exponentially large set of random vectors will be nearly pairwise orthogonal with high probability!
• Recall that we say two vectors $x, y$ are orthogonal if $\langle x, y \rangle = 0$.

• What is the largest set of mutually orthogonal unit vectors in $d$-dimensional space?

Answer: $d$.

• How large can a set of unit vectors in $d$-dimensional space be that have all pairwise dot products $|\langle x, y \rangle| \leq \epsilon$?

Answer: $2\Omega(\epsilon^2 d)$.

An exponentially large set of random vectors will be nearly pairwise orthogonal with high probability!
• Recall that we say two vectors $x, y$ are orthogonal if $\langle x, y \rangle = 0$.
• What is the largest set of mutually orthogonal unit vectors in $d$-dimensional space? Answer: $d$. 
• Recall that we say two vectors $x, y$ are orthogonal if $\langle x, y \rangle = 0$.

• What is the largest set of mutually orthogonal unit vectors in $d$-dimensional space? Answer: $d$.

• How large can a set of unit vectors in $d$-dimensional space be that have all pairwise dot products $|\langle x, y \rangle| \leq \epsilon$?
• Recall that we say two vectors $x, y$ are **orthogonal** if $\langle x, y \rangle = 0$.

• What is the largest set of mutually orthogonal unit vectors in $d$-dimensional space? Answer: $d$.

• How large can a set of unit vectors in $d$-dimensional space be that have all pairwise dot products $|\langle x, y \rangle| \leq \epsilon$? Answer: $2^{\Omega(\epsilon^2 d)}$.

An exponentially large set of **random vectors** will be nearly pairwise orthogonal with high probability!
**Claim:** $2^{O(\epsilon^2 d)}$ random $d$-dimensional unit vectors will have all pairwise dot products $|\langle x, y \rangle| \leq \epsilon$ (be nearly orthogonal).

**Proof:** Let $x_1, \ldots, x_t \in \mathbb{R}^d$ have independent random entries $\pm \frac{1}{\sqrt{d}}$. 
**Claim:** $2^{O(\epsilon^2 d)}$ random $d$-dimensional unit vectors will have all pairwise dot products $|\langle x, y \rangle| \leq \epsilon$ (be nearly orthogonal).

**Proof:** Let $x_1, \ldots, x_t \in \mathbb{R}^d$ have independent random entries $\pm \frac{1}{\sqrt{d}}$.

- What is $\|x_i\|_2$?
Claim: $2^{O(\epsilon^2 d)}$ random $d$-dimensional unit vectors will have all pairwise dot products $|\langle x, y \rangle| \leq \epsilon$ (be nearly orthogonal).

Proof: Let $x_1, \ldots, x_t \in \mathbb{R}^d$ have independent random entries $\pm \frac{1}{\sqrt{d}}$.

- What is $\|x_i\|_2$? Every $x_i$ is always a unit vector.
Claim: \( 2^{O(\epsilon^2 d)} \) random \( d \)-dimensional unit vectors will have all pairwise dot products \( |\langle x, y \rangle| \leq \epsilon \) (be nearly orthogonal).

Proof: Let \( x_1, \ldots, x_t \in \mathbb{R}^d \) have independent random entries \( \pm \frac{1}{\sqrt{d}} \).

• What is \( \|x_i\|_2 \)? Every \( x_i \) is always a unit vector.

• What is \( \mathbb{E}[\langle x_i, x_j \rangle] \)?
Claim: $2^{O(\epsilon^2 d)}$ random $d$-dimensional unit vectors will have all pairwise dot products $|\langle x, y \rangle| \leq \epsilon$ (be nearly orthogonal).

Proof: Let $x_1, \ldots, x_t \in \mathbb{R}^d$ have independent random entries $\pm \frac{1}{\sqrt{d}}$.

- What is $\|x_i\|_2$? Every $x_i$ is always a unit vector.
- What is $\mathbb{E}[\langle x_i, x_j \rangle]$? $\mathbb{E}[\langle x_i, x_j \rangle] = 0$
Claim: $2^{O(\epsilon^2 d)}$ random $d$-dimensional unit vectors will have all pairwise dot products $|\langle x, y \rangle| \leq \epsilon$ (be nearly orthogonal).

Proof: Let $x_1, \ldots, x_t \in \mathbb{R}^d$ have independent random entries $\pm \frac{1}{\sqrt{d}}$.

- What is $\|x_i\|_2$? Every $x_i$ is always a unit vector.
- What is $\mathbb{E}[\langle x_i, x_j \rangle]$? $\mathbb{E}[\langle x_i, x_j \rangle] = 0$
- By a Bernstein bound, $\Pr[|\langle x_i, x_j \rangle| \geq \epsilon] \leq 2e^{-\epsilon^2 d/6}$. 
Claim: $2^{O(\epsilon^2d)}$ random $d$-dimensional unit vectors will have all pairwise dot products $|\langle x, y \rangle| \leq \epsilon$ (be nearly orthogonal).

Proof: Let $x_1, \ldots, x_t \in \mathbb{R}^d$ have independent random entries $\pm \frac{1}{\sqrt{d}}$.

- **What is $\|x_i\|_2$?** Every $x_i$ is always a unit vector.

- **What is $\mathbb{E}[\langle x_i, x_j \rangle]$?** $\mathbb{E}[\langle x_i, x_j \rangle] = 0$

- **By a Bernstein bound,** $\Pr[|\langle x_i, x_j \rangle| \geq \epsilon] \leq 2e^{-\epsilon^2d/6}$.

- **If $t = \frac{1}{2}e^{\epsilon^2d/12}$, using a union bound over $\binom{t}{2} \leq \frac{1}{8}e^{\epsilon^2d/6}$ possible pairs,** with probability $\geq 3/4$ all will be nearly orthogonal.
Claim: \(2^{O(\epsilon^2 d)}\) random \(d\)-dimensional unit vectors will have all pairwise dot products \(|\langle x, y \rangle| \leq \epsilon\) (be nearly orthogonal).

Proof: Let \(x_1, \ldots, x_t \in \mathbb{R}^d\) have independent random entries \(\pm \frac{1}{\sqrt{d}}\).

- **What is \(\|x_i\|_2\)?** Every \(x_i\) is always a unit vector.
- **What is \(E[\langle x_i, x_j \rangle]\)?** \(E[\langle x_i, x_j \rangle] = 0\)
- **By a Bernstein bound,** \(Pr[|\langle x_i, x_j \rangle| \geq \epsilon] \leq 2e^{-\epsilon^2 d/6}\).
- **If** \(t = \frac{1}{2}e^{\epsilon^2 d/12}\), using a union bound over \(\binom{t}{2} \leq \frac{1}{8}e^{\epsilon^2 d/6}\) possible pairs, with probability \(\geq 3/4\) all will be nearly orthogonal.

We won’t prove it but this is essentially optimal: In \(d\) dimensions, there can be at most \(2^{O(\epsilon^2 d)}\) nearly orthogonal unit vectors.
Recall: The Johnson Lindenstrauss lemma states that if $M \in \mathbb{R}^{m \times d}$ is a random matrix (linear map) with $m = O \left( \frac{\log n}{\epsilon^2} \right)$, for $x_1, \ldots, x_n \in \mathbb{R}^d$ with high probability, for all $i, j$:

$$(1 - \epsilon) \| x_i - x_j \|_2^2 \leq \| Mx_i - Mx_j \|_2^2 \leq (1 + \epsilon) \| x_i - x_j \|_2^2.$$
Recall: The Johnson Lindenstrauss lemma states that if $M \in \mathbb{R}^{m \times d}$ is a random matrix (linear map) with $m = O\left(\frac{\log n}{\epsilon^2}\right)$, for $x_1, \ldots, x_n \in \mathbb{R}^d$ with high probability, for all $i, j$:

$$(1 - \epsilon)\|x_i - x_j\|_2^2 \leq \|Mx_i - Mx_j\|_2^2 \leq (1 + \epsilon)\|x_i - x_j\|_2^2.$$

Implies: If $x_1, \ldots, x_n$ are nearly orthogonal unit vectors in $d$-dimensions (with pairwise dot products bounded by $\epsilon/8$), then

$$\frac{Mx_1}{\|Mx_1\|_2}, \ldots, \frac{Mx_n}{\|Mx_n\|_2}$$

are nearly orthogonal unit vectors in $m$-dimensions (with pairwise dot products bounded by $\epsilon$).
Recall: The Johnson Lindenstrauss lemma states that if \( M \in \mathbb{R}^{m \times d} \) is a random matrix (linear map) with \( m = O\left(\frac{\log n}{\epsilon^2}\right) \), for \( x_1, \ldots, x_n \in \mathbb{R}^d \) with high probability, for all \( i, j \):

\[
(1 - \epsilon) \|x_i - x_j\|_2^2 \leq \|Mx_i - Mx_j\|_2^2 \leq (1 + \epsilon) \|x_i - x_j\|_2^2.
\]

Implies: If \( x_1, \ldots, x_n \) are nearly orthogonal unit vectors in \( d \)-dimensions (with pairwise dot products bounded by \( \epsilon/8 \)), then

\[
\begin{align*}
\frac{Mx_1}{\|Mx_1\|_2}, & \quad \cdots, & \quad \frac{Mx_n}{\|Mx_n\|_2}
\end{align*}
\]

are nearly orthogonal unit vectors in \( m \)-dimensions (with pairwise dot products bounded by \( \epsilon \)). Algebra is a bit messy but a good exercise to partially work through. Proof uses the fact that

\[
\|x - y\|_2^2 = \|x\|_2^2 + \|y\|_2^2 - 2\langle x, y \rangle.
\]
Claim 1: $n$ nearly orthogonal unit vectors can be projected to $m = O \left( \frac{\log n}{\epsilon^2} \right)$ dimensions and still be nearly orthogonal.

Claim 2: In $m$ dimensions, there can be at most $2^{O(\epsilon^2 m)}$ nearly orthogonal unit vectors.
**Claim 1:** $n$ nearly orthogonal unit vectors can be projected to $m = O\left(\frac{\log n}{\epsilon^2}\right)$ dimensions and still be nearly orthogonal.

**Claim 2:** In $m$ dimensions, there can be at most $2^{O(\epsilon^2 m)}$ nearly orthogonal unit vectors.

- For both of these to hold it must be that $n \leq 2^{O(\epsilon^2 m)}$. 
Claim 1: $n$ nearly orthogonal unit vectors can be projected to $m = O \left( \frac{\log n}{\epsilon^2} \right)$ dimensions and still be nearly orthogonal.

Claim 2: In $m$ dimensions, there can be at most $2^{O(\epsilon^2 m)}$ nearly orthogonal unit vectors.

- For both of these to hold it must be that $n \leq 2^{O(\epsilon^2 m)}$.
- I.e., $n = 2^{\log n} \leq 2^{O(\epsilon^2 m)}$ and so $m = \Omega \left( \frac{\log n}{\epsilon^2} \right)$. 

Tells us that the JL lemma is optimal up to constants.
Claim 1: $n$ nearly orthogonal unit vectors can be projected to $m = O \left( \frac{\log n}{\epsilon^2} \right)$ dimensions and still be nearly orthogonal.

Claim 2: In $m$ dimensions, there can be at most $2^{O(\epsilon^2 m)}$ nearly orthogonal unit vectors.

- For both of these to hold it must be that $n \leq 2^{O(\epsilon^2 m)}$.
- I.e., $n = 2^{\log n} \leq 2^{O(\epsilon^2 m)}$ and so $m = \Omega \left( \frac{\log n}{\epsilon^2} \right)$.
- Tells us that the JL lemma is optimal up to constants.