• Problem Set 2 was due yesterday.
• The midterm exam will be held next Tuesday-Wednesday. Let me know ASAP if you have accommodations (e.g., extended time) via Disability Services.
Last Class: The Johnson-Lindenstrauss Lemma

- Low-distortion embeddings for any set of points via random projection.
- Started on proof of the JL Lemma via the Distributional JL Lemma.
Last Class: The Johnson-Lindenstrauss Lemma

• Low-distortion embeddings for any set of points via random projection.
• Started on proof of the JL Lemma via the Distributional JL Lemma.

This Class:

• Finish Up proof of the JL lemma.
• Example applications to classification and clustering.
• Discuss connections to high dimensional geometry.
**Johnson-Lindenstrauss Lemma:** For any set of points \( \vec{x}_1, \ldots, \vec{x}_n \in \mathbb{R}^d \) and \( \epsilon > 0 \) there exists a linear map \( \Pi : \mathbb{R}^d \rightarrow \mathbb{R}^m \) such that \( m = O \left( \frac{\log n}{\epsilon^2} \right) \) and letting \( \tilde{x}_i = \Pi \vec{x}_i \):

For all \( i, j : (1 - \epsilon) \| \vec{x}_i - \vec{x}_j \|_2 \leq \| \tilde{x}_i - \tilde{x}_j \|_2 \leq (1 + \epsilon) \| \vec{x}_i - \vec{x}_j \|_2 \).

Further, if \( \Pi \in \mathbb{R}^{m \times d} \) has each entry chosen i.i.d. from \( \mathcal{N}(0, 1/m) \) and \( m = O \left( \frac{\log n/\delta}{\epsilon^2} \right) \), \( \Pi \) satisfies the guarantee with probability \( \geq 1 - \delta \).
Johnson-Lindenstrauss Lemma: For any set of points $\vec{x}_1, \ldots, \vec{x}_n \in \mathbb{R}^d$ and $\epsilon > 0$ there exists a linear map $\Pi : \mathbb{R}^d \to \mathbb{R}^m$ such that $m = O\left(\frac{\log n}{\epsilon^2}\right)$ and letting $\tilde{x}_i = \Pi \vec{x}_i$:

For all $i, j$:

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

Further, if $\Pi \in \mathbb{R}^{m \times d}$ has each entry chosen i.i.d. from $\mathcal{N}(0, 1/m)$ and $m = O\left(\frac{\log n/\delta}{\epsilon^2}\right)$, $\Pi$ satisfies the guarantee with probability $\geq 1 - \delta$. 

![Diagram](image.png)
We showed that the Johnson-Lindenstrauss Lemma follows from:

**Distributional JL Lemma:** Let $\Pi \in \mathbb{R}^{m \times d}$ 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 $\geq 1 - \delta$

\[(1 - \epsilon)\|\vec{y}\|_2 \leq \|\Pi \vec{y}\|_2 \leq (1 + \epsilon)\|\vec{y}\|_2.\]
We showed that the Johnson-Lindenstrauss Lemma follows from:

**Distributional JL Lemma**: Let $\Pi \in \mathbb{R}^{m \times d}$ 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 $\geq 1 - \delta$

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

**Main Idea**: Union bound over $\binom{n}{2}$ difference vectors $\vec{y}_{ij} = \vec{x}_i - \vec{x}_j$. 
**Distributional JL Lemma:** Let \( \Pi \in \mathbb{R}^{m \times d} \) 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 \( \geq 1 - \delta \)

\[
(1 - \epsilon)\|\vec{y}\|_2 \leq \|\Pi\vec{y}\|_2 \leq (1 + \epsilon)\|\vec{y}\|_2.
\]

\( \vec{y} \in \mathbb{R}^d \): arbitrary vector, \( y \in \mathbb{R}^m \): compressed vector, \( \Pi \in \mathbb{R}^{m \times d} \): random projection. \( d \): original dim. \( m \): compressed dim, \( \epsilon \): error, \( \delta \): failure prob.
Distributional JL Lemma: Let $\Pi \in \mathbb{R}^{m \times d}$ 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 $\geq 1 - \delta$

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

- Let $y$ denote $\Pi \vec{y}$ and let $\Pi(j)$ denote the $j^{th}$ row of $\Pi$. 

\[\vec{y} \in \mathbb{R}^d: \text{arbitrary vector}, \ y \in \mathbb{R}^m: \text{compressed vector}, \ \Pi \in \mathbb{R}^{m \times d}: \text{random projection.} \ d: \text{original dim.} \ m: \text{compressed dim,} \ \epsilon: \text{error,} \ \delta: \text{failure prob.}\]
**Distributional JL Lemma:** Let \( \Pi \in \mathbb{R}^{m \times d} \) have each entry chosen i.i.d. as \( \mathcal{N}(0, 1/m) \). If we set \( m = \mathcal{O}\left(\frac{\log(1/\delta)}{\epsilon^2}\right) \), then for any \( \bar{y} \in \mathbb{R}^d \), with probability \( \geq 1 - \delta \)

\[
(1 - \epsilon) \|\bar{y}\|_2 \leq \|\Pi \bar{y}\|_2 \leq (1 + \epsilon) \|\bar{y}\|_2.
\]

- Let \( y \) denote \( \Pi \bar{y} \) and let \( \Pi(j) \) denote the \( j^{th} \) row of \( \Pi \).
- For any \( j \), \( y(j) = \langle \Pi(j), \bar{y} \rangle \).

\( \bar{y} \in \mathbb{R}^d \): arbitrary vector, \( y \in \mathbb{R}^m \): compressed vector, \( \Pi \in \mathbb{R}^{m \times d} \): random projection. \( d \): original dim. \( m \): compressed dim, \( \epsilon \): error, \( \delta \): failure prob.
**Distributional JL Lemma:** Let \( \Pi \in \mathbb{R}^{m \times d} \) 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 \( \geq 1 - \delta \)

\[
(1 - \epsilon)\|\vec{y}\|_2 \leq \|\Pi \vec{y}\|_2 \leq (1 + \epsilon)\|\vec{y}\|_2.
\]

- Let \( y \) denote \( \Pi \vec{y} \) and let \( \Pi(j) \) denote the \( j \)th row of \( \Pi \).
- For any \( j \), \( y(j) = \langle \Pi(j), \vec{y} \rangle \).

\( \vec{y} \in \mathbb{R}^d \): arbitrary vector, \( y \in \mathbb{R}^m \): compressed vector, \( \Pi \in \mathbb{R}^{m \times d} \): random projection. \( d \): original dim. \( m \): compressed dim, \( \epsilon \): error, \( \delta \): failure prob.
Distributional JL Lemma: Let $\Pi \in \mathbb{R}^{m \times d}$ 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 $\geq 1 - \delta$

\[(1 - \epsilon)\|\vec{y}\|_2 \leq \|\Pi \vec{y}\|_2 \leq (1 + \epsilon)\|\vec{y}\|_2.\]

- Let $y$ denote $\Pi \vec{y}$ and let $\Pi(j)$ denote the $j^{th}$ row of $\Pi$.
- For any $j$, $y(j) = \langle \Pi(j), \vec{y} \rangle = \sum_{i=1}^{d} g_i \cdot \vec{y}(i)$ where $g_i \sim \mathcal{N}(0, 1/m)$.

$\vec{y} \in \mathbb{R}^d$: arbitrary vector, $y \in \mathbb{R}^m$: compressed vector, $\Pi \in \mathbb{R}^{m \times d}$: random projection. $d$: original dim. $m$: compressed dim, $\epsilon$: error, $\delta$: failure prob.
**Distributional JL Lemma:** Let $\Pi \in \mathbb{R}^{m \times d}$ 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 $\geq 1 - \delta$

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

- Let $y$ denote $\Pi \vec{y}$ and let $\Pi(j)$ denote the $j^{th}$ row of $\Pi$.
- For any $j$, $y(j) = \langle \Pi(j), \vec{y} \rangle = \frac{1}{\sqrt{m}} \sum_{i=1}^{d} g_i \cdot \vec{y}(i)$ where $g_i \sim \mathcal{N}(0, 1)$.

$\vec{y} \in \mathbb{R}^d$: arbitrary vector, $y \in \mathbb{R}^m$: compressed vector, $\Pi \in \mathbb{R}^{m \times d}$: random projection. $d$: original dim. $m$: compressed dim, $\epsilon$: error, $\delta$: failure prob.
Let $y$ denote $\Pi \tilde{y}$ and let $\Pi(j)$ denote the $j^{th}$ row of $\Pi$.

For any $j$, $y(j) = \langle \Pi(j), \tilde{y} \rangle = \frac{1}{\sqrt{m}} \sum_{i=1}^{d} g_i \cdot \tilde{y}(i)$ where $g_i \sim \mathcal{N}(0, 1)$.
• Let $\mathbf{y}$ denote $\Pi\vec{y}$ and let $\Pi(j)$ denote the $j^{th}$ row of $\Pi$.
• For any $j$, $\mathbf{y}(j) = \langle \Pi(j), \vec{y} \rangle = \frac{1}{\sqrt{m}} \sum_{i=1}^{d} \mathbf{g}_i \cdot \vec{y}(i)$ where $\mathbf{g}_i \sim \mathcal{N}(0, 1)$.
• $\mathbf{g}_i \cdot \vec{y}(i) \sim \mathcal{N}(0, \vec{y}(i)^2)$: a normal distribution with variance $\vec{y}(i)^2$.

$\vec{y} \in \mathbb{R}^d$: arbitrary vector, $\mathbf{y} \in \mathbb{R}^m$: compressed vector, $\Pi \in \mathbb{R}^{m \times d}$: random projection mapping $\vec{y} \rightarrow \mathbf{y}$. $\Pi(j)$: $j^{th}$ row of $\Pi$, $d$: original dimension. $m$: compressed dimension, $\mathbf{g}_i$: normally distributed random variable.
• Let $\mathbf{y}$ denote $\mathbf{\Pi} \vec{y}$ and let $\mathbf{\Pi}(j)$ denote the $j^{th}$ row of $\mathbf{\Pi}$.

• For any $j$, $\mathbf{y}(j) = \langle \mathbf{\Pi}(j), \vec{y} \rangle = \frac{1}{\sqrt{m}} \sum_{i=1}^{d} \mathbf{g}_i \cdot \vec{y}(i)$ where $\mathbf{g}_i \sim \mathcal{N}(0, 1)$.

• $\mathbf{g}_i \cdot \vec{y}(i) \sim \mathcal{N}(0, \vec{y}(i)^2)$: a normal distribution with variance $\vec{y}(i)^2$.

---

$\vec{y} \in \mathbb{R}^d$: arbitrary vector, $\mathbf{y} \in \mathbb{R}^m$: compressed vector, $\mathbf{\Pi} \in \mathbb{R}^{m \times d}$: random projection mapping $\vec{y} \rightarrow \mathbf{y}$. $\mathbf{\Pi}(j)$: $j^{th}$ row of $\mathbf{\Pi}$, $d$: original dimension. $m$: compressed dimension, $\mathbf{g}_i$: normally distributed random variable.
• Let $\mathbf{y}$ denote $\mathbf{\Pi}\mathbf{\tilde{y}}$ and let $\mathbf{\Pi}(j)$ denote the $j^{th}$ row of $\mathbf{\Pi}$.
• For any $j$, $\mathbf{y}(j) = \langle \mathbf{\Pi}(j), \mathbf{\tilde{y}} \rangle = \frac{1}{\sqrt{m}} \sum_{i=1}^{d} \mathbf{g}_i \cdot \mathbf{\tilde{y}}(i)$ where $\mathbf{g}_i \sim \mathcal{N}(0, 1)$.
• $\mathbf{g}_i \cdot \mathbf{\tilde{y}}(i) \sim \mathcal{N}(0, \mathbf{\tilde{y}}(i)^2)$: a normal distribution with variance $\mathbf{\tilde{y}}(i)^2$.

$$\mathbf{\tilde{y}}(j) = \frac{1}{\sqrt{m}} [\mathbf{g}_1 \cdot \mathbf{y}(1) + \mathbf{g}_2 \cdot \mathbf{y}(2) + \ldots + \mathbf{g}_n \cdot \mathbf{y}(d)]$$

$\mathbf{\tilde{y}} \in \mathbb{R}^d$: arbitrary vector, $\mathbf{y} \in \mathbb{R}^m$: compressed vector, $\mathbf{\Pi} \in \mathbb{R}^{m \times d}$: random projection mapping $\mathbf{\tilde{y}} \rightarrow \mathbf{y}$. $\mathbf{\Pi}(j)$: $j^{th}$ row of $\mathbf{\Pi}$, $d$: original dimension. $m$: compressed dimension, $\mathbf{g}_i$: normally distributed random variable.
Let \( y \) denote \( \Pi \vec{y} \) and let \( \Pi(j) \) denote the \( j^{th} \) row of \( \Pi \).

For any \( j \),
\[
y(j) = \langle \Pi(j), \vec{y} \rangle = \frac{1}{\sqrt{m}} \sum_{i=1}^{d} g_i \cdot \vec{y}(i) \text{ where } g_i \sim \mathcal{N}(0, 1).
\]

\( g_i \cdot \vec{y}(i) \sim \mathcal{N}(0, \vec{y}(i)^2) \): a normal distribution with variance \( \vec{y}(i)^2 \).

\[
\vec{y}(j) = \frac{1}{\sqrt{m}} [g_1 \cdot y(1) + g_2 \cdot y(2) + \ldots + g_n \cdot y(d)]
\]

What is the distribution of \( y(j) \)?

\( \vec{y} \in \mathbb{R}^d \): arbitrary vector, \( y \in \mathbb{R}^m \): compressed vector, \( \Pi \in \mathbb{R}^{m \times d} \): random projection mapping \( \vec{y} \to y \). \( \Pi(j) \): \( j^{th} \) row of \( \Pi \), \( d \): original dimension. \( m \): compressed dimension, \( g_i \): normally distributed random variable.
Let \( y \) denote \( \Pi \vec{y} \) and let \( \Pi(j) \) denote the \( j^{th} \) row of \( \Pi \).

For any \( j \),
\[
y(j) = \langle \Pi(j), \vec{y} \rangle = \frac{1}{\sqrt{m}} \sum_{i=1}^{d} g_i \cdot \vec{y}(i) \text{ where } g_i \sim \mathcal{N}(0, 1).
\]

\( g_i \cdot \vec{y}(i) \sim \mathcal{N}(0, \vec{y}(i)^2) \): a normal distribution with variance \( \vec{y}(i)^2 \).

\[
\vec{y}(j) = \frac{1}{\sqrt{m}} [g_1 \cdot y(1) + g_2 \cdot y(2) + ... + g_n \cdot y(d)]
\]

What is the distribution of \( y(j) \)? Also Gaussian!

\( \vec{y} \in \mathbb{R}^d \): arbitrary vector, \( y \in \mathbb{R}^m \): compressed vector, \( \Pi \in \mathbb{R}^{m \times d} \): random projection mapping \( \vec{y} \rightarrow y \). \( \Pi(j) \): \( j^{th} \) row of \( \Pi \), \( d \): original dimension. \( m \): compressed dimension, \( g_i \): normally distributed random variable.
Letting $\mathbf{y} = \Pi \bar{\mathbf{y}}$, we have $\mathbf{y}(j) = \langle \Pi(j), \bar{\mathbf{y}} \rangle$ and:

$$\mathbf{y}(j) = \frac{1}{\sqrt{m}} \sum_{i=1}^{d} g_i \cdot \bar{\mathbf{y}}(i) \text{ where } g_i \cdot \bar{\mathbf{y}}(i) \sim \mathcal{N}(0, \bar{\mathbf{y}}(i)^2).$$

$\bar{\mathbf{y}} \in \mathbb{R}^d$: arbitrary vector, $\mathbf{y} \in \mathbb{R}^m$: compressed vector, $\Pi \in \mathbb{R}^{m \times d}$: random projection mapping $\bar{\mathbf{y}} \rightarrow \mathbf{y}$. $\Pi(j)$: $j^{th}$ row of $\Pi$, $d$: original dimension. $m$: compressed dimension, $g_i$: normally distributed random variable.
Letting $\mathbf{y} = \Pi \tilde{\mathbf{y}}$, we have $\mathbf{y}(j) = \langle \Pi(j), \tilde{\mathbf{y}} \rangle$ and:

$$\mathbf{y}(j) = \frac{1}{\sqrt{m}} \sum_{i=1}^{d} \mathbf{g}_i \cdot \tilde{\mathbf{y}}(i) \text{ where } \mathbf{g}_i \cdot \tilde{\mathbf{y}}(i) \sim \mathcal{N}(0, \tilde{\mathbf{y}}(i)^2).$$

**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:

$$a + b \sim \mathcal{N}(\mu_1 + \mu_2, \sigma_1^2 + \sigma_2^2)$$
Letting $\mathbf{y} = \Pi \tilde{\mathbf{y}}$, we have $\mathbf{y}(j) = \langle \Pi(j), \tilde{\mathbf{y}} \rangle$ and:

$$\mathbf{y}(j) = \frac{1}{\sqrt{m}} \sum_{i=1}^{d} \mathbf{g}_i \cdot \tilde{\mathbf{y}}(i) \text{ where } \mathbf{g}_i \cdot \tilde{\mathbf{y}}(i) \sim \mathcal{N}(0, \tilde{\mathbf{y}}(i)^2).$$

**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:

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

$\tilde{\mathbf{y}} \in \mathbb{R}^d$: arbitrary vector, $\mathbf{y} \in \mathbb{R}^m$: compressed vector, $\Pi \in \mathbb{R}^{m \times d}$: random projection mapping $\tilde{\mathbf{y}} \rightarrow \mathbf{y}$. $\Pi(j)$: $j^{th}$ row of $\Pi$, $d$: original dimension. $m$: compressed dimension, $\mathbf{g}_i$: normally distributed random variable.
Letting $y = \Pi \bar{y}$, we have $y(j) = \langle \Pi(j), \bar{y} \rangle$ and:

$$y(j) = \frac{1}{\sqrt{m}} \sum_{i=1}^{d} g_i \cdot \bar{y}(i) \text{ where } g_i \cdot \bar{y}(i) \sim \mathcal{N}(0, \bar{y}(i)^2).$$

**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, $y(j) \sim \frac{1}{\sqrt{m}} \mathcal{N}(0, \bar{y}(1)^2 + \bar{y}(2)^2 + \ldots + \bar{y}(d)^2)$

$\bar{y} \in \mathbb{R}^d$: arbitrary vector, $y \in \mathbb{R}^m$: compressed vector, $\Pi \in \mathbb{R}^{m \times d}$: random projection mapping $\bar{y} \rightarrow y$. $\Pi(j)$: $j^{th}$ row of $\Pi$, $d$: original dimension. $m$: compressed dimension, $g_i$: normally distributed random variable.
Letting $\mathbf{y} = \Pi \mathbf{\bar{y}}$, we have $\mathbf{y}(j) = \langle \Pi(j), \mathbf{\bar{y}} \rangle$ and:

$$
\mathbf{y}(j) = \frac{1}{\sqrt{m}} \sum_{i=1}^{d} \mathbf{g}_i \cdot \mathbf{\bar{y}}(i) \text{ where } \mathbf{g}_i \cdot \mathbf{\bar{y}}(i) \sim \mathcal{N}(0, \mathbf{\bar{y}}(i)^2).
$$

**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, $\mathbf{y}(j) \sim \frac{1}{\sqrt{m}} \mathcal{N}(0, \|\mathbf{\bar{y}}\|_2^2)$

$\mathbf{\bar{y}} \in \mathbb{R}^d$: arbitrary vector, $\mathbf{y} \in \mathbb{R}^m$: compressed vector, $\Pi \in \mathbb{R}^{m \times d}$: random projection mapping $\mathbf{\bar{y}} \rightarrow \mathbf{y}$. $\Pi(j)$: $j^{th}$ row of $\Pi$, $d$: original dimension. $m$: compressed dimension, $\mathbf{g}_i$: normally distributed random variable.
Letting $\mathbf{y} = \mathbf{\Pi}\tilde{\mathbf{y}}$, we have $\mathbf{y}(j) = \langle \mathbf{\Pi}(j), \tilde{\mathbf{y}} \rangle$ and:

$$\mathbf{y}(j) = \frac{1}{\sqrt{m}} \sum_{i=1}^{d} g_i \cdot \tilde{\mathbf{y}}(i) \text{ where } g_i \cdot \tilde{\mathbf{y}}(i) \sim \mathcal{N}(0, \tilde{\mathbf{y}}(i)^2).$$

**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, $\mathbf{y}(j) \sim \mathcal{N}(0, \|\tilde{\mathbf{y}}\|_2^2/m)$.

\(\tilde{\mathbf{y}} \in \mathbb{R}^d: \) arbitrary vector, $\mathbf{y} \in \mathbb{R}^m$: compressed vector, $\mathbf{\Pi} \in \mathbb{R}^{m \times d}$: random projection mapping $\tilde{\mathbf{y}} \rightarrow \mathbf{y}$. $\mathbf{\Pi}(j)$: $j^{th}$ row of $\mathbf{\Pi}$, $d$: original dimension. $m$: compressed dimension, $g_i$: normally distributed random variable.
Letting \( y = \Pi \tilde{y} \), we have 
\[
y(j) = \langle \Pi(j), \tilde{y} \rangle
\]
and:
\[
y(j) = \frac{1}{\sqrt{m}} \sum_{i=1}^{d} g_i \cdot \tilde{y}(i)
\]
where 
\[
g_i \cdot \tilde{y}(i) \sim \mathcal{N}(0, \tilde{y}(i)^2).
\]

**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, 
\[
y(j) \sim \mathcal{N}(0, \|\tilde{y}\|_2^2/m).
\]
I.e., \( y \) itself is a random Gaussian vector.

Rotational invariance of the Gaussian distribution.

\( \tilde{y} \in \mathbb{R}^d \): arbitrary vector, \( y \in \mathbb{R}^m \): compressed vector, \( \Pi \in \mathbb{R}^{m \times d} \): random projection mapping \( \tilde{y} \rightarrow y \). \( \Pi(j) \): \( j \)th row of \( \Pi \), \( d \): original dimension. \( m \): compressed dimension, \( g_i \): normally distributed random variable
DISTRIBUTIONAL JL PROOF

So far: Letting $\mathbf{\Pi} \in \mathbb{R}^{d \times m}$ have each entry chosen i.i.d. as $\mathcal{N}(0, 1/m)$, for any $\vec{y} \in \mathbb{R}^d$, letting $\mathbf{y} = \mathbf{\Pi}\vec{y}$:

$$y(j) \sim \mathcal{N}(0, \|\vec{y}\|_2^2/m).$$

$\vec{y} \in \mathbb{R}^d$: arbitrary vector, $\mathbf{y} \in \mathbb{R}^m$: compressed vector, $\mathbf{\Pi} \in \mathbb{R}^{m \times d}$: random projection mapping $\vec{y} \rightarrow \mathbf{y}$. $\mathbf{\Pi}(j)$: $j^{th}$ row of $\mathbf{\Pi}$, $d$: original dimension. $m$: compressed dimension, $g_i$: normally distributed random variable
**DISTRIBUTIONAL JL PROOF**

**So far:** Letting $\Pi \in \mathbb{R}^{d \times m}$ have each entry chosen i.i.d. as $\mathcal{N}(0, 1/m)$, for any $\tilde{y} \in \mathbb{R}^d$, letting $y = \Pi \tilde{y}$:

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

What is $\mathbb{E}[\|y\|_2^2]$?

---

$\tilde{y} \in \mathbb{R}^d$: arbitrary vector, $y \in \mathbb{R}^m$: compressed vector, $\Pi \in \mathbb{R}^{m \times d}$: random projection mapping $\tilde{y} \rightarrow y$. $\Pi(j)$: $j^{th}$ row of $\Pi$, $d$: original dimension. $m$: compressed dimension, $g_i$: normally distributed random variable
**So far:** Letting $\Pi \in \mathbb{R}^{d \times m}$ have each entry chosen i.i.d. as $\mathcal{N}(0, 1/m)$, for any $\tilde{y} \in \mathbb{R}^d$, letting $y = \Pi \tilde{y}$:

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

What is $\mathbb{E}[\|y\|_2^2]$?

$$\mathbb{E}[\|y\|_2^2] = \mathbb{E} \left[ \sum_{j=1}^{m} y(j)^2 \right]$$

---

$\tilde{y} \in \mathbb{R}^d$: arbitrary vector, $y \in \mathbb{R}^m$: compressed vector, $\Pi \in \mathbb{R}^{m \times d}$: random projection mapping $\tilde{y} \rightarrow y$. $\Pi(j)$: $j^{th}$ row of $\Pi$, $d$: original dimension. $m$: compressed dimension, $g_i$: normally distributed random variable
**DISTRIBUTIONAL JL PROOF**

So far: Letting $\Pi \in \mathbb{R}^{d \times m}$ have each entry chosen i.i.d. as $\mathcal{N}(0, 1/m)$, for any $\tilde{y} \in \mathbb{R}^d$, letting $y = \Pi \tilde{y}$:

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

What is $\mathbb{E}[\|y\|_2^2]$?

$$\mathbb{E}[\|y\|_2^2] = \mathbb{E} \left[ \sum_{j=1}^{m} y(j)^2 \right] = \sum_{j=1}^{m} \mathbb{E}[y(j)^2]$$

---

$\tilde{y} \in \mathbb{R}^d$: arbitrary vector, $y \in \mathbb{R}^m$: compressed vector, $\Pi \in \mathbb{R}^{m \times d}$: random projection mapping $\tilde{y} \rightarrow y$. $\Pi(j)$: $j^{th}$ row of $\Pi$, $d$: original dimension. $m$: compressed dimension, $g_i$: normally distributed random variable.
So far: Letting $\Pi \in \mathbb{R}^{d \times m}$ have each entry chosen i.i.d. as $\mathcal{N}(0, 1/m)$, for any $\tilde{y} \in \mathbb{R}^d$, letting $y = \Pi \tilde{y}$:

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

What is $\mathbb{E}[\|y\|_2^2]$?

$$\mathbb{E}[\|y\|_2^2] = \mathbb{E} \left[ \sum_{j=1}^{m} y(j)^2 \right] = \sum_{j=1}^{m} \mathbb{E}[y(j)^2]$$

$\tilde{y} \in \mathbb{R}^d$: arbitrary vector, $y \in \mathbb{R}^m$: compressed vector, $\Pi \in \mathbb{R}^{m \times d}$: random projection mapping $\tilde{y} \rightarrow y$. $\Pi(j)$: $j^{th}$ row of $\Pi$, $d$: original dimension. $m$: compressed dimension, $g_i$: normally distributed random variable.
**DISTRIBUTIONAL JL PROOF**

**So far:** Letting $\Pi \in \mathbb{R}^{d \times m}$ have each entry chosen i.i.d. as $\mathcal{N}(0, 1/m)$, for any $\bar{y} \in \mathbb{R}^d$, letting $y = \Pi \bar{y}$:

$$y(j) \sim \mathcal{N}(0, \|\bar{y}\|_2^2 / m).$$

What is $\mathbb{E}[\|y\|_2^2]$?

$$\mathbb{E}[\|y\|_2^2] = \mathbb{E}[\sum_{j=1}^{m} y(j)^2] = \sum_{j=1}^{m} \mathbb{E}[y(j)^2]$$

$$= \sum_{j=1}^{m} \frac{\|\bar{y}\|_2^2}{m}$$

---

$\bar{y} \in \mathbb{R}^d$: arbitrary vector, $y \in \mathbb{R}^m$: compressed vector, $\Pi \in \mathbb{R}^{m \times d}$: random projection mapping $\bar{y} \rightarrow y$. $\Pi(j)$: $j^{th}$ row of $\Pi$, $d$: original dimension. $m$: compressed dimension, $g_i$: normally distributed random variable.
So far: Letting $\Pi \in \mathbb{R}^{d \times m}$ have each entry chosen i.i.d. as $\mathcal{N}(0, 1/m)$, for any $\vec{y} \in \mathbb{R}^d$, letting $y = \Pi \vec{y}$:

$$y(j) \sim \mathcal{N}(0, \|\vec{y}\|_2^2/m).$$

What is $\mathbb{E}[\|y\|_2^2]$?

$$\mathbb{E}[\|y\|_2^2] = \mathbb{E}
\left[
\sum_{j=1}^{m} y(j)^2
\right] = \sum_{j=1}^{m} \mathbb{E}[y(j)^2]
= \sum_{j=1}^{m} \frac{\|\vec{y}\|_2^2}{m} = \|\vec{y}\|_2^2$$

$\vec{y} \in \mathbb{R}^d$: arbitrary vector, $y \in \mathbb{R}^m$: compressed vector, $\Pi \in \mathbb{R}^{m \times d}$: random projection mapping $\vec{y} \rightarrow y$. $\Pi(j)$: $j^{th}$ row of $\Pi$, $d$: original dimension. $m$: compressed dimension, $g_i$: normally distributed random variable
**DISTRIBUTIONAL JL PROOF**

**So far:** Letting $\Pi \in \mathbb{R}^{d \times m}$ have each entry chosen i.i.d. as $\mathcal{N}(0, 1/m)$, for any $\tilde{y} \in \mathbb{R}^{d}$, letting $y = \Pi \tilde{y}$:

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

**What is $\mathbb{E}[\|y\|_2^2]$?**

$$\mathbb{E}[\|y\|_2^2] = \mathbb{E}\left[\sum_{j=1}^{m} y(j)^2\right] = \sum_{j=1}^{m} \mathbb{E}[y(j)^2]$$

$$= \sum_{j=1}^{m} \frac{\|\tilde{y}\|_2^2}{m} = \|\tilde{y}\|_2^2$$

So $y$ has the right norm in expectation.

$\tilde{y} \in \mathbb{R}^{d}$: arbitrary vector, $y \in \mathbb{R}^{m}$: compressed vector, $\Pi \in \mathbb{R}^{m \times d}$: random projection mapping $\tilde{y} \rightarrow y$. $\Pi(j)$: $j^{th}$ row of $\Pi$, $d$: original dimension. $m$: compressed dimension, $g_{ij}$: normally distributed random variable.
**DISTRIBUTIONAL JL PROOF**

**So far:** Letting $\Pi \in \mathbb{R}^{d \times m}$ have each entry chosen i.i.d. as $\mathcal{N}(0, 1/m)$, for any $\tilde{y} \in \mathbb{R}^d$, letting $y = \Pi \tilde{y}$:

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

What is $\mathbb{E}[\|y\|_2^2]$?

$$\mathbb{E}[\|y\|_2^2] = \mathbb{E} \left[ \sum_{j=1}^{m} y(j)^2 \right] = \sum_{j=1}^{m} \mathbb{E}[y(j)^2] = \sum_{j=1}^{m} \frac{\|\tilde{y}\|_2^2}{m} = \|\tilde{y}\|_2^2$$

So $y$ has the right norm in expectation.

How is $\|y\|_2^2$ distributed? Does it concentrate?

---

$\tilde{y} \in \mathbb{R}^d$: arbitrary vector, $y \in \mathbb{R}^m$: compressed vector, $\Pi \in \mathbb{R}^{m \times d}$: random projection mapping $\tilde{y} \rightarrow y$. $\Pi(j)$: $j^{th}$ row of $\Pi$, $d$: original dimension. $m$: compressed dimension, $g_i$: normally distributed random variable
So Far: Each entry of our compressed vector $y$ is Gaussian with:

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

and

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

---

$\vec{y} \in \mathbb{R}^d$: arbitrary vector, $y \in \mathbb{R}^m$: compressed vector, $\Pi \in \mathbb{R}^{m \times d}$: random projection mapping $\vec{y} \rightarrow y$. $d$: original dimension. $m$: compressed dimension, $\epsilon$: embedding error, $\delta$: embedding failure prob.
So Far: Each entry of our compressed vector $\mathbf{y}$ is Gaussian with:

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

$$\|\mathbf{y}\|_2^2 = \sum_{i=1}^{m} y(j)^2$$

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 $\mathbf{y}$ is Gaussian with:

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

and

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

$$\|\mathbf{y}\|_2^2 = \sum_{i=1}^{m} y(j)^2$$

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

$\mathbf{y} \in \mathbb{R}^d$: arbitrary vector, $\mathbf{y} \in \mathbb{R}^m$: compressed vector, $\Pi \in \mathbb{R}^{m \times d}$: random projection mapping $\mathbf{y} \rightarrow \mathbf{y}$. $d$: original dimension. $m$: compressed dimension, $\epsilon$: embedding error, $\delta$: embedding failure prob.
So Far: Each entry of our compressed vector $\mathbf{y}$ is Gaussian with:

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

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

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

$$\Pr[|\mathbf{Z} - \mathbb{E}\mathbf{Z}| \geq \epsilon \mathbb{E}\mathbf{Z}] \leq 2e^{-m\epsilon^2/8}.$$
So Far: Each entry of our compressed vector $y$ is Gaussian with:

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

and

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

$$\|y\|_2^2 = \sum_{i=1}^{m} y(j)^2$$

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,

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

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

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

$\vec{y} \in \mathbb{R}^d$: arbitrary vector, $y \in \mathbb{R}^m$: compressed vector, $\Pi \in \mathbb{R}^{m \times d}$: random projection mapping $\vec{y} \rightarrow y$. $d$: original dimension. $m$: compressed dimension, $\epsilon$: embedding error, $\delta$: embedding failure prob.
So Far: Each entry of our compressed vector $y$ is Gaussian with:

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

$$\|y\|_2^2 = \sum_{i=1}^{m} y(j)^2$$

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,

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

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

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

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

**k-means Objective:** $\text{Cost}(C_1, \ldots, C_k) = \min_{C_1, \ldots, C_k} \sum_{j=1}^{k} \sum_{\vec{x} \in C_k} \|\vec{x} - \mu_j\|^2_2$. 
**Example Application:** *k*-Means Clustering

**Goal:** Separate \( n \) points in \( d \) dimensional space into \( k \) groups.

**k-means Objective:**

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

**Write in terms of distances:**

\[
\text{Cost}(C_1, \ldots, C_k) = \min_{C_1, \ldots, C_k} \sum_{j=1}^{k} \sum_{\vec{x}_1, \vec{x}_2 \in C_k} ||\vec{x}_1 - \vec{x}_2||^2
\]
**k-means Objective:** \(\text{Cost}(C_1, \ldots, C_k) = \min_{C_1, \ldots, C_k} \sum_{j=1}^{k} \sum_{\bar{x}_1, \bar{x}_2 \in C_k} \|\bar{x}_1 - \bar{x}_2\|_2^2\)
**k-means Objective:** \( \text{Cost}(C_1, \ldots, C_k) = \min_{C_1, \ldots, C_k} \sum_{j=1}^{k} \sum_{x_1, x_2 \in C_k} \|x_1 - x_2\|_2^2 \) If we randomly project to \( m = O\left(\frac{\log n}{\epsilon^2}\right) \) dimensions, for all pairs \( \vec{x}_1, \vec{x}_2 \),

\[ (1 - \epsilon)\|\vec{x}_1 - \vec{x}_2\|_2^2 \leq \|\vec{x}_1 - \vec{x}_2\|_2^2 \leq (1 + \epsilon)\|\vec{x}_1 - \vec{x}_2\|_2^2 \]
**k-means Objective:** \( \text{Cost}(C_1, \ldots, C_k) = \min_{C_1, \ldots, C_k} \sum_{j=1}^{k} \sum_{\bar{x}_1, \bar{x}_2 \in C_k} \| \bar{x}_1 - \bar{x}_2 \|^2 \)

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

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

Letting \( \overline{\text{Cost}}(C_1, \ldots, C_k) = \min_{C_1, \ldots, C_k} \sum_{j=1}^{k} \sum_{x_1, x_2 \in C_k} \| x_1 - x_2 \|^2 \)

\[
(1 - \epsilon)\text{Cost}(C_1, \ldots, C_k) \leq \overline{\text{Cost}}(C_1, \ldots, C_k) \leq (1 + \epsilon)\text{Cost}(C_1, \ldots, C_k).
\]
k-means Objective: $\text{Cost}(C_1, \ldots, C_k) = \min_{C_1, \ldots, C_k} \sum_{j=1}^{k} \sum_{x_1, x_2 \in C_k} \|x_1 - x_2\|_2^2$

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

$$(1 - \epsilon)\|x_1 - x_2\|_2^2 \leq \|\bar{x}_1 - \bar{x}_2\|_2^2 \leq (1 + \epsilon)\|x_1 - x_2\|_2^2 \implies$$

Letting $\overline{\text{Cost}}(C_1, \ldots, C_k) = \min_{C_1, \ldots, C_k} \sum_{j=1}^{k} \sum_{x_1, x_2 \in C_k} \|x_1 - x_2\|_2^2$

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

**Upshot:** Can cluster in $m$ dimensional space (much more efficiently) and minimize $\overline{\text{Cost}}(C_1, \ldots, C_k)$. 
The Johnson-Lindenstrauss Lemma and High Dimensional Geometry
The Johnson-Lindenstrauss Lemma and High Dimensional Geometry

• High-dimensional Euclidean space looks *very different* from low-dimensional space. So how can JL work?
• Is Euclidean distance in high-dimensional meaningless, making JL useless? (The curse of dimensionality)
What is the largest set of mutually orthogonal unit vectors in $d$-dimensional space?

a) 1    b) $\log d$    c) $\sqrt{d}$    d) $d$
What is the largest set of mutually orthogonal unit vectors in $d$-dimensional space?

a) 1  
b) $\log d$  
c) $\sqrt{d}$  
d) $d$
What is the largest set of unit vectors in $d$-dimensional space that have all pairwise dot products $|\langle \mathbf{x}, \mathbf{y} \rangle| \leq \epsilon$? (think $\epsilon = .01$)

a) $d$  b) $\Theta(d)$  c) $\Theta(d^2)$  d) $2^{\Theta(d)}$
What is the largest set of unit vectors in $d$-dimensional space that have all pairwise dot products $|\langle \vec{x}, \vec{y} \rangle| \leq \epsilon$? (think $\epsilon = .01$)

a) $d$  b) $\Theta(d)$  c) $\Theta(d^2)$  d) $2^{\Theta(d)}$
What is the largest set of unit vectors in $d$-dimensional space that have all pairwise dot products $|\langle \vec{x}, \vec{y} \rangle| \leq \epsilon$? (think $\epsilon = .01$)

a) $d$  b) $\Theta(d)$  c) $\Theta(d^2)$  d) $2^{\Theta(d)}$

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

Proof: Let \(\vec{x}_1, \ldots, \vec{x}_t\) each have independent random entries set to \(\pm 1/\sqrt{d}\).
Claim: $2^{\Theta(\epsilon^2 d)}$ random $d$-dimensional unit vectors will have all pairwise dot products $|\langle \vec{x}, \vec{y} \rangle| \leq \epsilon$ (be nearly orthogonal).

Proof: Let $\vec{x}_1, \ldots, \vec{x}_t$ each have independent random entries set to $\pm 1/\sqrt{d}$.

- What is $\|\vec{x}_i\|_2$?
- What is $\mathbb{E}[\langle \vec{x}_i, \vec{x}_j \rangle]$?
**Claim:** $2^{\Theta(\epsilon^2 d)}$ random $d$-dimensional unit vectors will have all pairwise dot products $|\langle \vec{x}, \vec{y} \rangle| \leq \epsilon$ (be nearly orthogonal).

**Proof:** Let $\vec{x}_1, \ldots, \vec{x}_t$ each have independent random entries set to $\pm 1/\sqrt{d}$.

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

Proof: Let $\vec{x}_1, \ldots, \vec{x}_t$ each have independent random entries set to $\pm 1/\sqrt{d}$.

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

Proof: Let $\vec{x}_1, \ldots, \vec{x}_t$ each have independent random entries set to $\pm 1/\sqrt{d}$.

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

Proof: Let $\vec{x}_1, \ldots, \vec{x}_t$ each have independent random entries set to $\pm 1/\sqrt{d}$.

- What is $\|\vec{x}_i\|_2$? Every $\vec{x}_i$ is always a unit vector.
- What is $\mathbb{E}[\langle \vec{x}_i, \vec{x}_j \rangle]$? $\mathbb{E}[\langle \vec{x}_i, \vec{x}_j \rangle] = 0$
- By a Chernoff bound, $\Pr[|\langle \vec{x}_i, \vec{x}_j \rangle| \geq \epsilon] \leq 2e^{-\epsilon^2 d/6}$ (great exercise).
- If we chose $t = \frac{1}{2}e^{\epsilon^2 d/12}$, using a union bound over all $\binom{t}{2} \leq \frac{1}{8}e^{\epsilon^2 d/6}$ possible pairs, with probability $\geq 3/4$ all will be nearly orthogonal.
**Up Shot:** In $d$-dimensional space, a set of $2^{\Theta(\epsilon^2 d)}$ random unit vectors have all pairwise dot products at most $\epsilon$ (think $\epsilon = .01$).
Up Shot: In $d$-dimensional space, a set of $2^{\Theta(\epsilon^2 d)}$ random unit vectors have all pairwise dot products at most $\epsilon$ (think $\epsilon = .01$)

$$\|\vec{x}_i - \vec{x}_j\|_2^2$$
**Up Shot:** In $d$-dimensional space, a set of $2^{\Theta(\epsilon^2 d)}$ random unit vectors have all pairwise dot products at most $\epsilon$ (think $\epsilon = .01$)

$$
\| \vec{x}_i - \vec{x}_j \|_2^2 = \| \vec{x}_i \|_2^2 + \| \vec{x}_j \|_2^2 - 2 \vec{x}_i^T \vec{x}_j
$$
Curse of Dimensionality

**Up Shot:** In $d$-dimensional space, a set of $2^{\Theta(\epsilon^2 d)}$ random unit vectors have all pairwise dot products at most $\epsilon$ (think $\epsilon = .01$)

$$
\|\vec{x}_i - \vec{x}_j\|_2^2 = \|\vec{x}_i\|_2^2 + \|\vec{x}_j\|_2^2 - 2\vec{x}_i^T\vec{x}_j \geq 1.98.
$$
Up Shot: In $d$-dimensional space, a set of $2^\Theta(\epsilon^2 d)$ random unit vectors have all pairwise dot products at most $\epsilon$ (think $\epsilon = .01$)

$$\|\vec{x}_i - \vec{x}_j\|_2^2 = \|\vec{x}_i\|_2^2 + \|\vec{x}_j\|_2^2 - 2\vec{x}_i^T \vec{x}_j \geq 1.98.$$  

Even with an exponential number of random vector samples, we don’t see any nearby vectors.
**Up Shot:** In $d$-dimensional space, a set of $2^{\Theta(\epsilon^2 d)}$ random unit vectors have all pairwise dot products at most $\epsilon$ (think $\epsilon = .01$)

$$
\|\vec{x}_i - \vec{x}_j\|_2^2 = \|\vec{x}_i\|_2^2 + \|\vec{x}_j\|_2^2 - 2\vec{x}_i^T \vec{x}_j \geq 1.98.
$$

Even with an exponential number of random vector samples, we don’t see any nearby vectors.

- Can make methods like nearest neighbor classification or clustering useless.
The curse of dimensionality refers to various phenomena that arise when analyzing and organizing data in high-dimensional spaces (often with hundreds or thousands of dimensions) that do not occur in low-dimensional settings. It affects many fields, including statistics, machine learning, information retrieval, and data mining.

**Up Shot:** In $d$-dimensional space, a set of $2^\Theta(\epsilon^2 d)$ random unit vectors have all pairwise dot products at most $\epsilon$ (think $\epsilon = 0.01$)

$$\|\vec{x}_i - \vec{x}_j\|_2^2 = \|\vec{x}_i\|_2^2 + \|\vec{x}_j\|_2^2 - 2\vec{x}_i^T\vec{x}_j \geq 1.98.$$ 

Even with an exponential number of random vector samples, we don’t see any nearby vectors.

- Can make methods like nearest neighbor classification or clustering useless.

**Curse of dimensionality** for sampling/learning functions in high-dimensional space – samples are very ‘sparse’ unless we have a huge amount of data.
Up Shot: In $d$-dimensional space, a set of $2^{\Theta(\epsilon^2 d)}$ random unit vectors have all pairwise dot products at most $\epsilon$ (think $\epsilon = 0.01$)

$$\|\mathbf{x}_i - \mathbf{x}_j\|_2^2 = \|\mathbf{x}_i\|_2^2 + \|\mathbf{x}_j\|_2^2 - 2\mathbf{x}_i^T \mathbf{x}_j \geq 1.98.$$ 

Even with an exponential number of random vector samples, we don’t see any nearby vectors.

- Can make methods like nearest neighbor classification or clustering useless.

Curse of dimensionality for sampling/learning functions in high-dimensional space – samples are very ‘sparse’ unless we have a huge amount of data.

- Only hope is if we lots of structure (which we typically do...)

**Curse of dimensionality**
Distances for MNIST Digits:

Another Interpretation:
Tells us that random data can be a very bad model for actual input data.
Distances for MNIST Digits:

Distances for Random Images:

Another Interpretation: Tells us that random data can be a very bad model for actual input data.
Recall: The Johnson Lindenstrauss lemma states that if \( \Pi \in \mathbb{R}^{m \times d} \) is a random matrix (linear map) with \( m = O \left( \frac{\log n}{\epsilon^2} \right) \), for \( \vec{x}_1, \ldots, \vec{x}_n \in \mathbb{R}^d \) with high probability, for all \( i, j \):

\[
(1 - \epsilon) \| \vec{x}_i - \vec{x}_j \|_2^2 \leq \| \Pi \vec{x}_i - \Pi \vec{x}_j \|_2^2 \leq (1 + \epsilon) \| \vec{x}_i - \vec{x}_j \|_2^2.
\]
Recall: The Johnson Lindenstrauss lemma states that if $\Pi \in \mathbb{R}^{m \times d}$ is a random matrix (linear map) with $m = O \left( \frac{\log n}{\epsilon^2} \right)$, for $\vec{x}_1, \ldots, \vec{x}_n \in \mathbb{R}^d$ with high probability, for all $i, j$:

$$(1 - \epsilon) \| \vec{x}_i - \vec{x}_j \|_2^2 \leq \| \Pi \vec{x}_i - \Pi \vec{x}_j \|_2^2 \leq (1 + \epsilon) \| \vec{x}_i - \vec{x}_j \|_2^2.$$ 

Implies: If $\vec{x}_1, \ldots, \vec{x}_n$ are nearly orthogonal unit vectors in $d$-dimensions (with pairwise dot products bounded by $\epsilon/8$), then $\frac{\Pi \vec{x}_1}{\| \Pi \vec{x}_1 \|_2}, \ldots, \frac{\Pi \vec{x}_n}{\| \Pi \vec{x}_n \|_2}$ are nearly orthogonal unit vectors in $m$-dimensions (with pairwise dot products bounded by $\epsilon$).
Recall: The Johnson Lindenstrauss lemma states that if \( \Pi \in \mathbb{R}^{m \times d} \) is a random matrix (linear map) with \( m = O\left(\frac{\log n}{\epsilon^2}\right) \), for \( \vec{x}_1, \ldots, \vec{x}_n \in \mathbb{R}^d \) with high probability, for all \( i, j \):

\[
(1 - \epsilon) \|\vec{x}_i - \vec{x}_j\|_2^2 \leq \|\Pi \vec{x}_i - \Pi \vec{x}_j\|_2^2 \leq (1 + \epsilon) \|\vec{x}_i - \vec{x}_j\|_2^2.
\]

Implies: If \( \vec{x}_1, \ldots, \vec{x}_n \) are nearly orthogonal unit vectors in \( d \)-dimensions (with pairwise dot products bounded by \( \epsilon/8 \)), then \( \frac{\Pi \vec{x}_1}{\|\Pi \vec{x}_1\|_2}, \ldots, \frac{\Pi \vec{x}_n}{\|\Pi \vec{x}_n\|_2} \) 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.
**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 are at most $2^{O(\epsilon^2 m)}$ nearly orthogonal 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 are at most $2^{O(\epsilon^2 m)}$ nearly orthogonal vectors.

- For both these to hold it might 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 are at most $2^{O(\epsilon^2 m)}$ nearly orthogonal vectors.

- For both these to hold it might be that $n \leq 2^{O(\epsilon^2 m)}$.
- $2^{O(\epsilon^2 m)} = 2^{O(\log n)} \geq n$. 

This tells us that the JL lemma is optimal up to constants. $m$ is chosen just large enough so that the odd geometry of $d$-dimensional space still holds on the $n$ points in question after projection to a much lower dimensional space.
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 are at most $2^{O(\epsilon^2 m)}$ nearly orthogonal vectors.

- For both these to hold it might be that $n \leq 2^{O(\epsilon^2 m)}$.
- $2^{O(\epsilon^2 m)} = 2^{O(\log n)} \geq n$. 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 are at most $2^{O(\epsilon^2 m)}$ nearly orthogonal vectors.

- For both these to hold it might be that $n \leq 2^{O(\epsilon^2 m)}$.
- $2^{O(\epsilon^2 m)} = 2^{O(\log n)} \geq n$. Tells us that the JL lemma is optimal up to constants.
- $m$ is chosen just large enough so that the odd geometry of $d$-dimensional space still holds on the $n$ points in question after projection to a much lower dimensional space.
Let $B_d$ be the unit ball in $d$ dimensions. $B_d = \{ x \in \mathbb{R}^d : \|x\|_2 \leq 1 \}$. 

What percentage of the volume of $B_d$ falls within $\epsilon$ distance of its surface?

Answer: all but a $(1 - \epsilon^d)$ fraction. Exponentially small in the dimension $d$!
Let $B_d$ be the unit ball in $d$ dimensions. $B_d = \{ x \in \mathbb{R}^d : \|x\|_2 \leq 1 \}$.

What percentage of the volume of $B_d$ falls within $\epsilon$ distance of its surface?
Let $\mathcal{B}_d$ be the unit ball in $d$ dimensions. $\mathcal{B}_d = \{x \in \mathbb{R}^d : \|x\|_2 \leq 1\}$.

What percentage of the volume of $\mathcal{B}_d$ falls within $\epsilon$ distance of its surface?

**Volume of a radius $R$ ball is** $\frac{\pi^{d/2}}{(d/2)!} \cdot R^d$. 
Let $B_d$ be the unit ball in $d$ dimensions. $B_d = \{ x \in \mathbb{R}^d : \| x \|_2 \leq 1 \}$.

What percentage of the volume of $B_d$ falls within $\epsilon$ distance of its surface? Answer: all but a $(1 - \epsilon)^d \leq e^{-\epsilon d}$ fraction. Exponentially small in the dimension $d$!

Volume of a radius $R$ ball is $\frac{\pi^{\frac{d}{2}}}{(d/2)!} \cdot R^d$. 
All but an $e^{-εd}$ fraction of a unit ball’s volume is within $ε$ of its surface.
All but an $e^{-\epsilon d}$ fraction of a unit ball’s volume is within $\epsilon$ of its surface. If we randomly sample points with $\|x\|_2 \leq 1$, nearly all will have $\|x\|_2 \geq 1 - \epsilon$. 
All but an $e^{-\epsilon d}$ fraction of a unit ball’s volume is within $\epsilon$ of its surface. If we randomly sample points with $\|x\|_2 \leq 1$, nearly all will have $\|x\|_2 \geq 1 - \epsilon$.

- **Isoperimetric inequality**: the ball has the minimum surface area/volume ratio of any shape.
All but an $e^{-\epsilon d}$ fraction of a unit ball’s volume is within $\epsilon$ of its surface. If we randomly sample points with $\|x\|_2 \leq 1$, nearly all will have $\|x\|_2 \geq 1 - \epsilon$.

- **Isoperimetric inequality**: the ball has the minimum surface area/volume ratio of any shape.

- If we randomly sample points from any high-dimensional shape, nearly all will fall near its surface.
All but an $e^{-\epsilon^d}$ fraction of a unit ball’s volume is within $\epsilon$ of its surface. If we randomly sample points with $\|x\|_2 \leq 1$, nearly all will have $\|x\|_2 \geq 1 - \epsilon$.

- **Isoperimetric inequality**: the ball has the minimum surface area/volume ratio of any shape.

- If we randomly sample points from any high-dimensional shape, nearly all will fall near its surface.
- ‘All points are outliers.’
What fraction of the cubes are visible on the surface of the cube?
What fraction of the cubes are visible on the surface of the cube?

\[
\frac{10^3 - 8^3}{10^3} = \frac{1000 - 512}{1000} = .488.
\]
What percentage of the volume of $B_d$ falls within $\epsilon$ distance of its equator?

Formally: volume of set $S = \{x \in B_d : |x(1)| \leq \epsilon\}$.
What percentage of the volume of $B_d$ falls within $\epsilon$ distance of its equator? Answer: all but a $2^{\Theta(-\epsilon^2 d)}$ fraction.

Formally: volume of set $S = \{ x \in B_d : |x(1)| \leq \epsilon \}$. 
BIZARRE SHAPE OF HIGH-DIMENSIONAL BALLS

What percentage of the volume of $B_d$ falls within $\varepsilon$ distance of its equator? Answer: all but a $2^{\Theta(-\varepsilon^2 d)}$ fraction.

Formally: volume of set $S = \{ x \in B_d : |x(1)| \leq \varepsilon \}$.

By symmetry, all but a $2^{\Theta(-\varepsilon^2 d)}$ fraction of the volume falls within $\varepsilon$ of any equator! $S = \{ x \in B_d : |\langle x, t \rangle| \leq \varepsilon \}$
Claim 1: All but a $2^{\Theta(-\epsilon^2 d)}$ fraction of the volume of a ball falls within $\epsilon$ of any equator.

Claim 2: All but a $2^{\Theta(-\epsilon d)}$ fraction falls within $\epsilon$ of its surface.
Claim 1: All but a $2^{\Theta(-\epsilon^2 d)}$ fraction of the volume of a ball falls within $\epsilon$ of any equator.

Claim 2: All but a $2^{\Theta(-\epsilon d)}$ fraction falls within $\epsilon$ of its surface.
**Claim 1:** All but a $2^{\Theta(-\epsilon^2d)}$ fraction of the volume of a ball falls within $\epsilon$ of any equator.

**Claim 2:** All but a $2^{\Theta(-\epsilon d)}$ fraction falls within $\epsilon$ of its surface.
**Claim 1:** All but a $2^{\Theta(-\epsilon^2 d)}$ fraction of the volume of a ball falls within $\epsilon$ of any equator.

**Claim 2:** All but a $2^{\Theta(-\epsilon d)}$ fraction falls within $\epsilon$ of its surface.
Claim 1: All but a $2^{\Theta(-\epsilon^2 d)}$ fraction of the volume of a ball falls within $\epsilon$ of any equator.

Claim 2: All but a $2^{\Theta(-\epsilon d)}$ fraction falls within $\epsilon$ of its surface.

How is this possible?
**Claim 1:** All but a $2^{\Theta(-\epsilon^2d)}$ fraction of the volume of a ball falls within $\epsilon$ of any equator.

**Claim 2:** All but a $2^{\Theta(-\epsilon d)}$ fraction falls within $\epsilon$ of its surface.

How is this possible? High-dimensional space looks nothing like this picture!