COMPSCI 514: ALGORITHMS FOR DATA SCIENCE

Cameron Musco
University of Massachusetts Amherst. Fall 2021.
Lecture 12
• Problem Set 2 is due Friday, 11:59pm.
• Quiz 6 is due today at 8pm.
• The exam will be held next Tuesday in class. Let me know ASAP if you need accommodations (e.g., extended time).
• We will do some midterm review in class on Thursday. I will also hold additional office hours for midterm prep, next Monday, 4-6pm, and potentially Friday afternoon as well.

-Practice problems on schedule.
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.
SUMMARY

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 \( \bar{x}_1, \ldots, \bar{x}_n \in \mathbb{R}^d \) 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 \( \tilde{x}_i = \Pi \bar{x}_i \):

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

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 \( \geq 1 - \delta \).
We showed that the Johnson-Lindenstrauss Lemma follows from:

**Distributional JL Lemma:** Let $\mathbf{W} \in \mathbb{R}^{m \times d}$ 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 $\hat{\mathbf{y}} \in \mathbb{R}^d$, with probability $\geq 1 - \delta$

$$(1 - \epsilon)\|\hat{\mathbf{y}}\|_2 \leq \|\mathbf{W}\hat{\mathbf{y}}\|_2 \leq (1 + \epsilon)\|\hat{\mathbf{y}}\|_2.$$
We showed that the Johnson-Lindenstrauss Lemma follows from:

**Distributional JL Lemma:** Let \( \mathbf{P} \in \mathbb{R}^{m \times d} \) 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 \( \geq 1 - \delta \)

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

**Main Idea:** Union bound over \( \binom{n}{2} \) difference vectors \( \vec{y}_{ij} = \vec{x}_i - \vec{x}_j \).

\[
S = \frac{1}{\binom{n}{2}}
\]

\[
\log(1/\delta) \approx \log(n^2) = \log(n).
\]
Distributional JL Lemma: Let $\mathbf{P} \in \mathbb{R}^{m \times d}$ 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 $\mathbf{y} \in \mathbb{R}^d$, with probability $\geq 1 - \delta$,

$$(1 - \epsilon)\|\mathbf{y}\|_2 \leq \|\mathbf{P}\mathbf{y}\|_2 \leq (1 + \epsilon)\|\mathbf{y}\|_2$$

$\mathbf{y} \in \mathbb{R}^d$: arbitrary vector, $\hat{\mathbf{y}} \in \mathbb{R}^m$: compressed vector, $\mathbf{P} \in \mathbb{R}^{m \times d}$: random projection. $d$: original dim. $m$: compressed dim, $\epsilon$: error, $\delta$: failure prob.
**DISTRIBUTIONAL JL PROOF**

**Distributional JL Lemma:** Let \( \mathbf{P} \in \mathbb{R}^{m \times d} \) 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 \( \mathbf{y} \in \mathbb{R}^d \), with probability \( \geq 1 - \delta \)

\[
(1 - \epsilon) \|\mathbf{y}\|_2 \leq \|\mathbf{PY}\|_2 \leq (1 + \epsilon) \|\mathbf{y}\|_2
\]

- Let \( \hat{\mathbf{y}} \) denote \( \mathbf{PY} \) and let \( \mathbf{P}(j) \) denote the \( j^{th} \) row of \( \mathbf{P} \).

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

\[
(1 - \epsilon)\|\tilde{\mathbf{y}}\|_2 \leq \|\mathbf{P}\tilde{\mathbf{y}}\|_2 \leq (1 + \epsilon)\|\tilde{\mathbf{y}}\|_2
\]

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

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

\[
(1 - \epsilon) \| \mathbf{y} \|_2 \leq \| \mathbf{P} \mathbf{y} \|_2 \leq (1 + \epsilon) \| \mathbf{y} \|_2
\]

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

\( \mathbf{\tilde{y}} \in \mathbb{R}^d \): arbitrary vector, \( \mathbf{\tilde{y}} \in \mathbb{R}^m \): compressed vector, \( \mathbf{P} \in \mathbb{R}^{m \times d} \): random projection. \( d \): original dim. \( m \): compressed dim, \( \epsilon \): error, \( \delta \): failure prob.
**DISTRIBUTIONAL JL PROOF**

### Distributional JL Lemma:

Let $\mathbf{\Pi} \in \mathbb{R}^{m \times d}$ 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 $\mathbf{\tilde{y}} \in \mathbb{R}^d$, with probability $\geq 1 - \delta$

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

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

$\tilde{\mathbf{y}} \in \mathbb{R}^d$: arbitrary vector, $\mathbf{\tilde{y}} \in \mathbb{R}^m$: compressed vector, $\mathbf{\Pi} \in \mathbb{R}^{m \times d}$: random projection. $d$: original dim. $m$: compressed dim, $\epsilon$: error, $\delta$: failure prob.
• Let $\tilde{y}$ denote $\Pi \tilde{Y}$ and let $\Pi(j)$ denote the $j^{th}$ row of $\Pi$.
• For any $j$, $\tilde{y}(j) = \langle \Pi(j), \tilde{y} \rangle = \sum_{i=1}^{d} g_i \cdot \tilde{y}(i)$ where $g_i \sim \mathcal{N}(0, 1/m)$.

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

For any $j$, $\tilde{y}(j) = \langle \mathbf{P}(j), \tilde{y} \rangle = \sum_{i=1}^{d} g_i \cdot \tilde{y}(i)$ where $g_i \sim \mathcal{N}(0, 1/m)$.

$g_i \cdot \tilde{y}(i) \sim \mathcal{N}(0, \frac{\tilde{y}(i)^2}{m})$: normally distributed with variance $\frac{\tilde{y}(i)^2}{m}$.

---

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

- Let $\tilde{y}$ denote $\Pi \tilde{y}$ and let $\Pi(j)$ denote the $j^{th}$ row of $\Pi$.
- For any $j$, $\tilde{y}(j) = \langle \Pi(j), \tilde{y} \rangle = \sum_{i=1}^{d} g_i \cdot \tilde{y}(i)$ where $g_i \sim \mathcal{N}(0, 1/m)$.
- $g_i \cdot \tilde{y}(i) \sim \mathcal{N}(0, \frac{\tilde{y}(i)^2}{m})$: normally distributed with variance $\frac{\tilde{y}(i)^2}{m}$.

\[\text{variance } \frac{1}{m}\]

\[\text{variance } \frac{\tilde{y}(i)^2}{m}\]

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

- Let $\tilde{y}$ denote $\Pi \tilde{y}$ and let $\Pi(j)$ denote the $j^{th}$ row of $\Pi$.
- For any $j$, $\tilde{y}(j) = \langle \Pi(j), \tilde{y} \rangle = \sum_{i=1}^{d} g_i \cdot \tilde{y}(i)$ where $g_i \sim \mathcal{N}(0, 1/m)$.
- $g_i \cdot \tilde{y}(i) \sim \mathcal{N}(0, \frac{\tilde{y}(i)^2}{m})$: normally distributed with variance $\frac{\tilde{y}(i)^2}{m}$.

\[
\tilde{y}(j) = \sqrt{\frac{\gamma(1)^2}{m}} + \sqrt{\frac{\gamma(2)^2}{m}} + \ldots + \sqrt{\frac{\gamma(d)^2}{m}}
\]

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

- Let $\tilde{y}$ denote $\Pi\tilde{y}$ and let $\Pi(j)$ denote the $j^{th}$ row of $\Pi$.
- For any $j$, $\tilde{y}(j) = \langle \Pi(j), \tilde{y} \rangle = \sum_{i=1}^{d} g_i \cdot \tilde{y}(i)$ where $g_i \sim \mathcal{N}(0, 1/m)$.
- $g_i \cdot \tilde{y}(i) \sim \mathcal{N}(0, \frac{\tilde{y}(i)^2}{m})$: normally distributed with variance $\frac{\tilde{y}(i)^2}{m}$.

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

What is the distribution of $\tilde{y}(j)$?

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

- For any \( j \), \( \tilde{y}(j) = \langle \Pi(j), \tilde{y} \rangle = \sum_{i=1}^{d} g_i \cdot \tilde{y}(i) \) where \( g_i \sim \mathcal{N}(0, 1/m) \).
- \( g_i \cdot \tilde{y}(i) \sim \mathcal{N}(0, \frac{\tilde{y}(i)^2}{m}) \): normally distributed with variance \( \frac{\tilde{y}(i)^2}{m} \).

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

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

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

$$
\tilde{y}(j) = \sum_{i=1}^{d} g_i \cdot \tilde{y}(i) \text{ where } g_i \cdot \tilde{y}(i) \sim \mathcal{N} \left( 0, \frac{\tilde{y}(i)^2}{m} \right).
$$
Letting $\tilde{y} = \Pi \hat{y}$, we have $\tilde{y}(j) = \langle \Pi(j), \hat{y} \rangle$ and:

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

**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:

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

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

\[
\tilde{y}(j) = \sum_{i=1}^{d} g_i \cdot \tilde{y}(i) \text{ where } g_i \cdot \tilde{y}(i) \sim \mathcal{N}\left(0, \frac{\tilde{y}(i)^2}{m}\right).
\]

**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:

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

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

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

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

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

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

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

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

**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, \( \tilde{y}(j) \sim \mathcal{N}(0, \frac{\|\hat{y}\|^2}{m}) \) i.e., \( \tilde{y} \) itself is a random Gaussian vector.

Rotational invariance of the Gaussian distribution.

\( \tilde{y} \in \mathbb{R}^d \): arbitrary vector, \( \hat{y} \in \mathbb{R}^m \): compressed vector, \( \Pi \in \mathbb{R}^{m \times d} \): random projection mapping \( \tilde{y} \to \hat{y} \). \( \Pi(j) \): \( j^{th} \) 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 \( \mathbf{\tilde{y}} \in \mathbb{R}^d \), letting \( \tilde{y} = \mathbf{\Pi}\mathbf{y} \):

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

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

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

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

---

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

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

\[
\tilde{y}(j) \sim \mathcal{N}(0, \|\mathbf{y}\|_{2}^2/m).
\]

What is \( \mathbb{E}[\|\tilde{\mathbf{y}}\|_2^2] \)?

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

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

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

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

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

$\tilde{y} \in \mathbb{R}^d$: arbitrary vector, $\mathbf{y} \in \mathbb{R}^m$: compressed vector, $\mathbf{P} \in \mathbb{R}^{m \times d}$: random projection mapping $\tilde{y} \rightarrow \mathbf{y}$. $\mathbf{P}(j)$: $j^{th}$ row of $\mathbf{P}$, $d$: 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 $\tilde{\tilde{y}} = \Pi \tilde{y}$:

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

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

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

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

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

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

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

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

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

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

What is \( \mathbb{E}[\|\tilde{y}\|_2^2] \)?

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

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

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

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

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

What is \( \mathbb{E}[\|\tilde{\mathbf{y}}\|_2^2] \)?

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

So \( \tilde{\mathbf{y}} \) has the right norm in expectation.

\( \tilde{\mathbf{y}} \in \mathbb{R}^d \): arbitrary vector, \( \mathbf{\hat{y}} \in \mathbb{R}^m \): compressed vector, \( \mathbf{\Pi} \in \mathbb{R}^{m \times d} \): random projection mapping \( \mathbf{\tilde{y}} \rightarrow \mathbf{\hat{y}} \). \( \mathbf{\Pi}(j) \): \( j \)th row of \( \mathbf{\Pi} \), \( d \): 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 $\tilde{y} = \Pi \tilde{y}$:

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

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

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

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

So $\tilde{y}$ has the right norm in expectation.

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

$\tilde{y} \in \mathbb{R}^d$: arbitrary vector, $\tilde{y} \in \mathbb{R}^m$: compressed vector, $\Pi \in \mathbb{R}^{m \times d}$: random projection mapping $\tilde{y} \to \tilde{y}$. $\Pi(j)$: $j^{th}$ 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 $\frac{1}{\sqrt{m}} \cdot \mathcal{N}(0, 1)$, for any $\vec{y} \in \mathbb{R}^d$, letting $\tilde{y} = \Pi \vec{y}$:

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

$\vec{y} \in \mathbb{R}^d$: arbitrary vector, $\tilde{y} \in \mathbb{R}^m$: compressed vector, $\Pi \in \mathbb{R}^{m \times d}$: random projection mapping $\vec{y} \rightarrow \tilde{y}$. $\Pi(j)$: $j^{th}$ row of $\Pi$, $d$: original dimension. $m$: compressed dimension, $\epsilon$: embedding error, $\delta$: embedding failure prob.
DISTRIBUTIONAL JL PROOF

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

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

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

$\tilde{y} \in \mathbb{R}^d$: arbitrary vector, $\tilde{y} \in \mathbb{R}^m$: compressed vector, $\Pi \in \mathbb{R}^{m \times d}$: random projection mapping $\tilde{y} \to \tilde{y}$. $\Pi(j)$: $j^{th}$ row of $\Pi$, $d$: original dimension. $m$: compressed dimension, $\epsilon$: embedding error, $\delta$: embedding failure prob.
DISTRIBUTIONAL JL PROOF

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

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

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

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

$\tilde{\mathbf{y}} \in \mathbb{R}^d$: arbitrary vector, $\tilde{\mathbf{y}} \in \mathbb{R}^m$: compressed vector, $\mathbf{\Pi} \in \mathbb{R}^{m \times d}$: random projection mapping $\mathbf{y} \rightarrow \tilde{\mathbf{y}}$. $\mathbf{\Pi}(j)$: $j^{th}$ row of $\mathbf{\Pi}$, $d$: original dimension. $m$: compressed dimension, $\epsilon$: embedding error, $\delta$: embedding failure prob.
So far: Letting $\mathbf{\Pi} \in \mathbb{R}^{d \times m}$ have each entry chosen i.i.d. as $\frac{1}{\sqrt{m}} \cdot \mathcal{N}(0, 1)$, for any $\mathbf{\hat{y}} \in \mathbb{R}^d$, letting $\mathbf{\tilde{y}} = \mathbf{\Pi} \mathbf{\hat{y}}$:

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

$$\|\mathbf{\tilde{y}}\|_2^2 = \sum_{i=1}^{m} \mathbf{\tilde{y}}(j)^2$$ 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: Letting $\Pi \in \mathbb{R}^{d \times m}$ have each entry chosen i.i.d. as 

$$\frac{1}{\sqrt{m}} \cdot \mathcal{N}(0, 1),$$

for any $\tilde{y} \in \mathbb{R}^d$, letting $\tilde{y} = \Pi \tilde{y}$:

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

and

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

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

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

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

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

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

$$m = 80 \cdot \frac{\log(1/\delta)}{\epsilon^2}, \quad (1 - \epsilon)\|\tilde{y}\|_2^2 \leq \|\tilde{y}\|_2^2 \leq (1 + \epsilon)\|\tilde{y}\|_2^2.$$

$\tilde{y} \in \mathbb{R}^d$: arbitrary vector, $\tilde{y} \in \mathbb{R}^m$: compressed vector, $\Pi \in \mathbb{R}^{m \times d}$: random projection mapping $\tilde{y} \rightarrow \tilde{y}$. $\Pi(j)$: $j^{th}$ row of $\Pi$, $d$: original dimension. $m$: compressed dimension, $\epsilon$: embedding error, $\delta$: embedding failure prob.
So far: Letting $\mathbf{P} \in \mathbb{R}^{d \times m}$ have each entry chosen i.i.d. as
\[
\frac{1}{\sqrt{m}} \cdot \mathcal{N}(0, 1),
\]
for any $\mathbf{y} \in \mathbb{R}^d$, letting $\tilde{\mathbf{y}} = \mathbf{P}\mathbf{y}$:
\[
\tilde{\mathbf{y}}(j) \sim \mathcal{N}(0, \|\mathbf{y}\|_2^2/m) \quad \text{and} \quad \mathbb{E}[\|\tilde{\mathbf{y}}\|_2^2] = \|\mathbf{y}\|_2^2
\]

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

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

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

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

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

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

Goal: Separate $n$ points in $d$ dimensional space into $k$ groups.

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

\[
\text{Cost}(C_1, \ldots, C_k) = \min_{C_1, \ldots, C_k} \sum_{j=1}^{k} \sum_{\vec{x}_1, \vec{x}_2 \in C_k} ||\vec{x}_1 - \vec{x}_2||^2
\]
EXAMPLE APPLICATION: $k$-MEANS CLUSTERING

k-means Objective: $Cost(C_1, \ldots, C_k) = \min_{C_1, \ldots, C_k} \sum_{j=1}^{k} \sum_{\bar{x}_1, \bar{x}_2 \in C_k} ||\bar{x}_1 - \bar{x}_2||^2_2$
**EXAMPLE APPLICATION: \(k\)-MEANS CLUSTERING**

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

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

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

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

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

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

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

$$(1 - \epsilon)\text{Cost}(C_1, \ldots, C_k) \leq \text{Cost}(C_1, \ldots, C_k) \leq (1 + \epsilon)\text{Cost}(C_1, \ldots, C_k).$$
**EXAMPLE APPLICATION: $k$-MEANS CLUSTERING**

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

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

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

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

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

**Upshot:** Can cluster in \( m \) dimensional space (much more efficiently) and minimize \( \overline{\text{Cost}}(C_1, \ldots, C_k) \). The optimal set of clusters will have true cost within \( 1 + c\epsilon \) times the true optimal. **Good exercise to prove this.**
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) $d$
NEARLY ORTHOGONAL VECTORS

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) $2\Theta(d)$
NEARLY ORTHOGONAL VECTORS

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) $2^{\Theta(d)}$
NEARLY ORTHOGONAL VECTORS

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: $2^{\Theta(e^2 d)}$ random $d$-dimensional unit vectors will have all pairwise dot products $|\langle \tilde{x}, \tilde{y} \rangle| \leq \epsilon$ (be nearly orthogonal) with high probability.
Claim: \(2^{\Theta(\epsilon^2 d)}\) random \(d\)-dimensional unit vectors will have all pairwise dot products \(|\langle \tilde{x}, \tilde{y} \rangle| \leq \epsilon\) (be nearly orthogonal) with high probability.

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

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

$\begin{bmatrix}
\frac{1}{\sqrt{d}} & -\frac{1}{\sqrt{d}} & -\frac{1}{\sqrt{d}} & \cdots & -\frac{1}{\sqrt{d}}
\end{bmatrix}$

- What is $||\vec{x}_i||_2^2$?
  $||\vec{x}_i||_2^2 = \frac{\epsilon}{d}$

- What is $\mathbb{E}[\langle \vec{x}_i, \vec{x}_j \rangle]$?
  $\frac{1}{d}$
ORTHOGONAL VECTORS PROOF

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) with high probability.

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 $\mathbb{E}[\langle \vec{x}_i, \vec{x}_j \rangle]$?
ORTHOGONAL VECTORS PROOF

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) with high probability.

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 $\mathbb{E}[\langle \vec{x}_i, \vec{y} \rangle]$? $\mathbb{E}[\langle \vec{x}_i, \vec{x}_j \rangle] = 0$
ORTHOGONAL VECTORS PROOF

\[ \| \mathbf{x}_i \|_2 = \sqrt{\sum_{j=1}^{d} x_i(j)^2} = \sqrt{1} = 1 \]

Claim: \(2^{\Theta(e^2d)}\) random \(d\)-dimensional unit vectors will have all pairwise dot products \(|\langle \mathbf{x}, \mathbf{y} \rangle| \leq \epsilon\) (be nearly orthogonal) with high probability.

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

\[ \| \mathbf{x}_i \|_2 = \sqrt{\sum_{j=1}^{d} x_i(j)^2} = \sum_{j=1}^{d} x_i(j) \]

- What is \(\| \mathbf{x}_i \|_2\)? Every \(\mathbf{x}_i\) is always a unit vector.

\[ \langle \mathbf{x}_i, \mathbf{x}_j \rangle = \sum_{k=1}^{d} x_i(k) \cdot x_j(k) \]

- What is \(\mathbb{E}[\langle \mathbf{x}_i, \mathbf{x}_j \rangle]\)? \(\mathbb{E}[\langle \mathbf{x}_i, \mathbf{x}_j \rangle] = 0\)

\[ \langle \mathbf{x}_i, \mathbf{x}_i \rangle = \sum_{j=1}^{d} x_i(j)^2 = \| \mathbf{x}_i \|_2^2 = 1 \]

- By a Chernoff bound, \(\Pr[|\langle \mathbf{x}_i, \mathbf{x}_j \rangle| \geq \epsilon] \leq 2e^{-\epsilon^2d/6}\) (great exercise).
ORTHOGONAL VECTORS PROOF

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

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, $\Pr[|\langle \vec{x}_i, \vec{x}_j \rangle| \geq \epsilon] \leq 2e^{-\epsilon^2d/6}$ (great exercise).

- If we chose $t = \frac{1}{2}e^{\epsilon^2d/12}$, using a union bound over all $\binom{t}{2} \leq \frac{1}{8}e^{\epsilon^2d/6}$ possible pairs, with probability $\geq 3/4$ all will be nearly orthogonal.

$$\leq 2e^{-\epsilon^2d/6} \cdot \frac{1}{8} \cdot e^{\epsilon^2d/6} = \frac{1}{4}$$
**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 $\epsilon = .01$)

$$
\|\vec{x}_i - \vec{x}_j\|_2^2 = \|\vec{x}_i\|_2^2 + \|\vec{x}_j\|_2^2 - 2\vec{x}_i^T\vec{x}_j
$$
**Up Shot:** In $d$-dimensional space, a set of $2^{\Theta(\epsilon^2d)}$ random unit vectors have all pairwise dot products at most $\epsilon$ (think $\epsilon = 0.01$)

$$\|\bar{x}_i - \bar{x}_j\|^2 = \|\bar{x}_i\|^2 + \|\bar{x}_j\|^2 - 2\bar{x}_i^T\bar{x}_j \in [1.98, 2.02].$$
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 $\epsilon = .01$)

$$
\|\vec{x}_i - \vec{x}_j\|_2^2 = \|\vec{x}_i\|_2^2 + \|\vec{x}_j\|_2^2 - 2\vec{x}_i^T\vec{x}_j \in [1.98, 2.02].
$$

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 $2^\Theta(e^2d)$ random unit vectors have all pairwise dot products at most $\epsilon$ (think $\epsilon = .01$)

\[
\hat{\|X_i - X_j\|_2^2 = \|X_i\|_2^2 + \|X_j\|_2^2 - 2X_i^T X_j \in [1.98, 2.02].}
\]

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

- One version of the ‘curse of dimensionality’.

- If all your distances are roughly the same, distance based methods (k-means clustering, nearest neighbors, SVMs, etc.) aren’t going to work well.

- Distances are only meaningful if we have lots of structure and our data isn’t just independent random vectors.
Distances for MNIST Digits:

Distances for Random Images:
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 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 are at most $2^{O(\epsilon^2 m)}$ nearly orthogonal vectors.

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