COMPSCI 514: ALGORITHMS FOR DATA SCIENCE

Cameron Musco
University of Massachusetts Amherst. Spring 2020.
Lecture 10
LOGISTICS

- Problem Set 2 is due Sunday 3/8.
- Midterm on Thursday, 3/12. Will cover material through today.
- I have posted a study guide and practice questions on the course schedule.
- Next Tuesday I can’t do office hours after class. I will hold them before class on Tuesday (10:00am - 11:15am) and after class on Thursday (12:45pm-2:00pm).
Last Class: Dimensionality Reduction
Last Class: Dimensionality Reduction

- Finished up Count-Min Sketch and Frequent Items.
- Applications and examples of dimensionality reduction in data science (PCA, LSA, autoencoders, etc.)
- Low-distortion embeddings and some simple cases of when no-distortion embeddings are possible.
SUMMARY

Last Class: Dimensionality Reduction

- Finished up Count-Min Sketch and Frequent Items.
- Applications and examples of dimensionality reduction in data science (PCA, LSA, autoencoders, etc.)
- Low-distortion embeddings and some simple cases of when no-distortion embeddings are possible.

The Johnson-Lindenstrauss Lemma.

- Any data set can be embedded with low distortion into low-dimensional space.
- Prove the JL Lemma.
- Discuss algorithmic considerations, connections to other methods (SimHash), etc.
Low Distortion Embedding: Given \( \tilde{x}_1, \ldots, \tilde{x}_n \in \mathbb{R}^d \), distance function \( D \), and error parameter \( \epsilon \geq 0 \), find \( \tilde{x}_1, \ldots, \tilde{x}_n \in \mathbb{R}^m \) (where \( m \ll d \)) and distance function \( \tilde{D} \) such that for all \( i, j \in [n] \):

\[
(1 - \epsilon)D(\tilde{x}_i, \tilde{x}_j) \leq \tilde{D}(\tilde{x}_i, \tilde{x}_j) \leq (1 + \epsilon)D(\tilde{x}_i, \tilde{x}_j).
\]
Low Distortion Embedding: Given $\tilde{x}_1, \ldots, \tilde{x}_n \in \mathbb{R}^d$, distance function $D$, and error parameter $\epsilon \geq 0$, find $\tilde{x}_1, \ldots, \tilde{x}_n \in \mathbb{R}^m$ (where $m \ll d$) and distance function $\tilde{D}$ such that for all $i, j \in [n]$:

$$(1 - \epsilon)D(\tilde{x}_i, \tilde{x}_j) \leq \tilde{D}(\tilde{x}_i, \tilde{x}_j) \leq (1 + \epsilon)D(\tilde{x}_i, \tilde{x}_j).$$

Euclidean Low Distortion Embedding: Given $\tilde{x}_1, \ldots, \tilde{x}_n \in \mathbb{R}^d$ and error parameter $\epsilon \geq 0$, find $\tilde{x}_1, \ldots, \tilde{x}_n \in \mathbb{R}^m$ (where $m \ll d$) such that for all $i, j \in [n]$:

$$(1 - \epsilon)\|\tilde{x}_i - \tilde{x}_j\|_2 \leq \|\tilde{x}_i - \tilde{x}_j\|_2 \leq (1 + \epsilon)\|\tilde{x}_i - \tilde{x}_j\|_2.$$
Euclidean Low Distortion Embedding: Given $\tilde{x}_1, \ldots, \tilde{x}_n \in \mathbb{R}^d$ and error parameter $\epsilon \geq 0$, find $\tilde{x}_1, \ldots, \tilde{x}_n \in \mathbb{R}^m$ (where $m \ll d$) such that for all $i, j \in [n]$:

$$(1 - \epsilon)\|\tilde{x}_i - \tilde{x}_j\|_2 \leq \|\tilde{x}_i - \tilde{x}_j\|_2 \leq (1 + \epsilon)\|\tilde{x}_i - \tilde{x}_j\|_2.$$
Assume that $\tilde{x}_1, \ldots, \tilde{x}_n$ all lie on the 1st axis in $\mathbb{R}^d$.

Set $m = 1$ and $\tilde{x}_i = \bar{x}_i(1)$ (i.e., $\tilde{x}_i$ is just a single number).

- $\|\tilde{x}_i - \tilde{x}_j\|_2 = \sqrt{[\bar{x}_i(1) - \bar{x}_j(1)]^2} = |\bar{x}_i(1) - \bar{x}_j(1)| = \|\bar{x}_i - \bar{x}_j\|_2$.
- An embedding with no distortion from any $d$ into $m = 1$. 
Assume that $\tilde{x}_1, \ldots, \tilde{x}_n$ all lie on the unit circle in $\mathbb{R}^2$.

- Admits a low-distortion embedding to 1 dimension by letting $\tilde{x}_i = \theta(\tilde{x}_i)$.
- Does it admit a low-distortion Euclidean embedding?
Assume that $\tilde{x}_1, \ldots, \tilde{x}_n$ all lie on the unit circle in $\mathbb{R}^2$.

- Admits a low-distortion embedding to 1 dimension by letting $\tilde{x}_i = \theta(\tilde{x}_i)$.
- Does it admit a low-distortion Euclidean embedding?
Assume that $\tilde{x}_1, \ldots, \tilde{x}_n$ all lie on the unit circle in $\mathbb{R}^2$.

- Admits a low-distortion embedding to 1 dimension by letting $\tilde{x}_i = \theta(\bar{x}_i)$.
- Does it admit a low-distortion Euclidean embedding? No! Send me a proof on Piazza for 3 bonus points on Problem Set 2.
Another easy case: Assume that $\tilde{x}_1, \ldots, \tilde{x}_n$ lie in any $k$-dimensional subspace $\mathcal{V}$ of $\mathbb{R}^d$. 
Another easy case: Assume that $\tilde{x}_1, \ldots, \tilde{x}_n$ lie in any $k$-dimensional subspace $\mathcal{V}$ of $\mathbb{R}^d$.

- Let $\vec{v}_1, \vec{v}_2, \ldots, \vec{v}_k$ be an orthonormal basis for $\mathcal{V}$ and let $V \in \mathbb{R}^{d \times k}$ be the matrix with these vectors as its columns.
Another easy case: Assume that $\tilde{x}_1, \ldots, \tilde{x}_n$ lie in any $k$-dimensional subspace $\mathcal{V}$ of $\mathbb{R}^d$.

Let $\vec{v}_1, \vec{v}_2, \ldots, \vec{v}_k$ be an orthonormal basis for $\mathcal{V}$ and let $\mathbf{V} \in \mathbb{R}^{d \times k}$ be the matrix with these vectors as its columns.
Another easy case: Assume that $\tilde{x}_1, \ldots, \tilde{x}_n$ lie in any $k$-dimensional subspace $\mathcal{V}$ of $\mathbb{R}^d$.

- Let $\tilde{v}_1, \tilde{v}_2, \ldots, \tilde{v}_k$ be an orthonormal basis for $\mathcal{V}$ and let $V \in \mathbb{R}^{d \times k}$ be the matrix with these vectors as its columns.
- If we set $\tilde{x}_i \in \mathbb{R}^k$ to $\tilde{x}_i = V^T \tilde{x}_i$ we have:

$$
\|\tilde{x}_i - \tilde{x}_j\|_2 = \|V^T (\tilde{x}_i - \tilde{x}_j)\|_2 = \|\tilde{x}_i - \tilde{x}_j\|_2.
$$
Another easy case: Assume that $\tilde{x}_1, \ldots, \tilde{x}_n$ lie in any $k$-dimensional subspace $\mathcal{V}$ of $\mathbb{R}^d$.

- Let $\tilde{v}_1, \tilde{v}_2, \ldots, \tilde{v}_k$ be an orthonormal basis for $\mathcal{V}$ and let $V \in \mathbb{R}^{d \times k}$ be the matrix with these vectors as its columns.
- If we set $\tilde{x}_i \in \mathbb{R}^k$ to $\tilde{x}_i = V^T \tilde{x}_i$ we have:
  $$\|\tilde{x}_i - \tilde{x}_j\|_2 = \|V^T(\tilde{x}_i - \tilde{x}_j)\|_2 = \|\tilde{x}_i - \tilde{x}_j\|_2.$$  
- An embedding with no distortion from any $d$ into $m = k$. 

Another easy case: Assume that \( \tilde{x}_1, \ldots, \tilde{x}_n \) lie in any \( k \)-dimensional subspace \( \mathcal{V} \) of \( \mathbb{R}^d \).

\[
\begin{bmatrix}
    x_1 \\
    x_2 \\
    \vdots \\
    x_n
\end{bmatrix} \rightarrow \begin{bmatrix}
    \tilde{x}_1 \\
    \tilde{x}_2 \\
    \vdots \\
    \tilde{x}_n
\end{bmatrix}
\]

- Let \( \tilde{v}_1, \tilde{v}_2, \ldots, \tilde{v}_k \) be an orthonormal basis for \( \mathcal{V} \) and let \( V \in \mathbb{R}^{d \times k} \) be the matrix with these vectors as its columns.

- If we set \( \tilde{x}_i \in \mathbb{R}^k \) to \( \tilde{x}_i = V^T \tilde{x}_i \) we have:

\[
\| \tilde{x}_i - \tilde{x}_j \|_2 = \| V^T (\tilde{x}_i - \tilde{x}_j) \|_2 = \| \tilde{x}_i - \tilde{x}_j \|_2.
\]

- An embedding with no distortion from any \( d \) into \( m = k \).

- \( V^T : \mathbb{R}^d \rightarrow \mathbb{R}^k \) is a linear map giving our embedding.
What about when we don’t make any assumptions on $\tilde{x}_1, \ldots, \tilde{x}_n$. I.e., they can be scattered arbitrarily around $d$-dimensional space?

- Can we find a no-distortion embedding into $m \ll d$ dimensions?
What about when we don’t make any assumptions on $\tilde{x}_1, \ldots, \tilde{x}_n$. I.e., they can be scattered arbitrarily around $d$-dimensional space?

- Can we find a no-distortion embedding into $m \ll d$ dimensions? **No. Require $m = d$.**
What about when we don’t make any assumptions on $\tilde{x}_1, \ldots, \tilde{x}_n$. I.e., they can be scattered arbitrarily around $d$-dimensional space?

- Can we find a no-distortion embedding into $m \ll d$ dimensions? No. Require $m = d$.
- Can we find an $\epsilon$-distortion embedding into $m \ll d$ dimensions for $\epsilon > 0$?

For all $i, j$:

\[(1 - \epsilon)\|\tilde{x}_i - \tilde{x}_j\|_2 \leq \|\tilde{x}_i - \tilde{x}_j\|_2 \leq (1 + \epsilon)\|\tilde{x}_i - \tilde{x}_j\|_2.\]
What about when we don’t make any assumptions on $\tilde{x}_1, \ldots, \tilde{x}_n$. I.e., they can be scattered arbitrarily around $d$-dimensional space?

- Can we find a no-distortion embedding into $m \ll d$ dimensions? No. Require $m = d$.
- Can we find an $\epsilon$-distortion embedding into $m \ll d$ dimensions for $\epsilon > 0$? Yes! Always, with $m$ depending on $\epsilon$.

For all $i, j : (1 - \epsilon)\|\tilde{x}_i - \tilde{x}_j\|_2 \leq \|\tilde{x}_i - \tilde{x}_j\|_2 \leq (1 + \epsilon)\|\tilde{x}_i - \tilde{x}_j\|_2$. 
THE JOHNSON-LINDENSTRAUSS LEMMA

\[ \sqrt{T} \]

**Johnson-Lindenstrauss Lemma:** For any set of points \( \tilde{x}_1, \ldots, \tilde{x}_n \in \mathbb{R}^d \) 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 \( \tilde{x}_i = \Pi \tilde{x}_i \):

For all \( i, j \):

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

Further, if \( \Pi \in \mathbb{R}^{m \times d} \) has each entry chosen i.i.d. from \( \mathcal{N}(0, 1/m) \), it satisfies the guarantee with high probability.
**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 $\mathcal{N}(0, 1/m)$, it satisfies the guarantee with high probability.

For $d = 1$ trillion, $\epsilon = .05$, and $n = 100,000$, $m \approx 6600$. 
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 $\mathcal{N}(0, 1/m)$, it satisfies the guarantee with high probability.

For $d = 1$ trillion, $\epsilon = .05$, and $n = 100,000$, $m \approx 6600$.

Very surprising! Powerful result with a simple construction: applying a random linear transformation to a set of points preserves distances between all those points with high probability.
For any \( \mathbf{x}_1, \ldots, \mathbf{x}_n \) and \( \Pi \in \mathbb{R}^{m \times d} \) with each entry chosen i.i.d. from \( \mathcal{N}(0, 1/m) \), with high probability, letting \( \tilde{\mathbf{x}}_i = \Pi \mathbf{x}_i \):

For all \( i, j \) :

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

\[
m = O\left(\frac{\log n}{\epsilon^2}\right)
\]
RANDOM PROJECTION

For any $\tilde{x}_1, \ldots, \tilde{x}_n$ and $\Pi \in \mathbb{R}^{m \times d}$ with each entry chosen i.i.d. from $\mathcal{N}(0, 1/m)$, with high probability, letting $\tilde{x}_i = \Pi \tilde{x}_i$:

For all $i, j$:

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

- $\Pi$ is known as a random projection. It is a random linear function, mapping length $d$ vectors to length $m$ vectors.
For any \( \bar{x}_1, \ldots, \bar{x}_n \) and \( \Pi \in \mathbb{R}^{m \times d} \) with each entry chosen i.i.d. from \( \mathcal{N}(0, 1/m) \), with high probability, letting \( \tilde{x}_i = \Pi \bar{x}_i \):

For all \( i, j \) :

\[
(1 - \epsilon) \| \bar{x}_i - \bar{x}_j \|_2 \leq \| \tilde{x}_i - \tilde{x}_j \|_2 \leq (1 + \epsilon) \| \bar{x}_i - \bar{x}_j \|_2.
\]

\( m = O \left( \frac{\log n}{\epsilon^2} \right) \)

- \( \Pi \) is known as a random projection. It is a random linear function, mapping length \( d \) vectors to length \( m \) vectors.
- \( \Pi \) is data oblivious. Stark contrast to methods like PCA.
ALGORITHMIC CONSIDERATIONS

- Many alternative constructions: ±1 entries, sparse (most entries 0), Fourier structured (Problem Set 2), etc. $\implies$ more efficient computation of $\tilde{x}_i = \Pi \tilde{x}_i$. 
• Many alternative constructions: ±1 entries, sparse (most entries 0), Fourier structured (Problem Set 2), etc. $\implies$ more efficient computation of $\tilde{x}_i = \Pi \tilde{x}_i$.

• Data oblivious property means that once $\Pi$ is chosen, $\tilde{x}_1, \ldots, \tilde{x}_n$ can be computed in a stream with little memory.

• Memory needed is just $O(d + nm)$ vs. $O(nd)$ to store the full data set.
ALGORITHMIC CONSIDERATIONS

- Many alternative constructions: ±1 entries, sparse (most entries 0), Fourier structured (Problem Set 2), etc. $\implies$ more efficient computation of $\tilde{x}_i = \Pi \tilde{x}_i$.
- Data oblivious property means that once $\Pi$ is chosen, $\tilde{x}_1, \ldots, \tilde{x}_n$ can be computed in a stream with little memory.
- Memory needed is just $O(d + nm)$ vs. $O(nd)$ to store the full data set.
- Compression can also be easily performed in parallel on different servers.
ALGORITHMIC CONSIDERATIONS

• Many alternative constructions: ±1 entries, sparse (most entries 0), Fourier structured (Problem Set 2), etc. \( \implies \) more efficient computation of \( \tilde{x}_i = \Pi \tilde{x}_i \).

• Data oblivious property means that once \( \Pi \) is chosen, \( \tilde{x}_1, \ldots, \tilde{x}_n \) can be computed in a stream with little memory.

• Memory needed is just \( O(d + nm) \) vs. \( O(nd) \) to store the full data set.

• Compression can also be easily performed in parallel on different servers.

• When new data points are added, can be easily compressed, without updating existing points.
Compression operation is $\tilde{x}_i = \mathbf{p} x_i$, so for any $j$,

$$\tilde{x}_i(j) = \langle \mathbf{p}(j), \tilde{x}_i \rangle = \sum_{k=1}^{d} \mathbf{p}(j, k) \cdot \tilde{x}_i(k).$$

$\tilde{x}_1, \ldots, \tilde{x}_n$: original points ($d$ dims.), $\tilde{x}_1, \ldots, \tilde{x}_n$: compressed points ($m < d$ dims.), $\mathbf{p} \in \mathbb{R}^{m \times d}$: random projection (embedding function)
Compression operation is $\tilde{x}_i = \Pi \vec{x}_i$, so for any $j$, 

$$\tilde{x}_i(j) = \langle \Pi(j), \vec{x}_i \rangle = \sum_{k=1}^{d} \Pi(j, k) \cdot \vec{x}_i(k).$$

$\Pi(j)$ is a vector with independent random Gaussian entries.

$\tilde{x}_1, \ldots, \tilde{x}_n$: original points ($d$ dims.), $\hat{x}_1, \ldots, \hat{x}_n$: compressed points ($m < d$ dims.), $\Pi \in \mathbb{R}^{m \times d}$: random projection (embedding function)
Compression operation is $\tilde{x}_i = \mathbf{P}\vec{x}_i$, so for any $j$,

$$\tilde{x}_i(j) = \langle \mathbf{p}(j), \vec{x}_i \rangle = \sum_{k=1}^{d} \mathbf{p}(j, k) \cdot \vec{x}_i(k).$$

$\mathbf{p}(j)$ is a vector with independent random Gaussian entries.

$\vec{x}_1, \ldots, \vec{x}_n$: original points ($d$ dims.), $\tilde{x}_1, \ldots, \tilde{x}_n$: compressed points ($m < d$ dims.), $\mathbf{P} \in \mathbb{R}^{m \times d}$: random projection (embedding function)
Compression operation is $\tilde{x}_i = \Pi \tilde{x}_i$, so for any $j$,

$$\tilde{x}_i(j) = \langle \Pi(j), \tilde{x}_i \rangle = \sum_{k=1}^{d} \Pi(j, k) \cdot \tilde{x}_i(k).$$

$\Pi(j)$ is a vector with independent random Gaussian entries.

$\tilde{x}_1, \ldots, \tilde{x}_n$: original points ($d$ dims.), $\tilde{x}_1, \ldots, \tilde{x}_n$: compressed points ($m < d$ dims.), $\Pi \in \mathbb{R}^{m \times d}$: random projection (embedding function)
Compression operation is \( \tilde{x}_i = \Pi \tilde{x}_i \), so for any \( j \),

\[
\tilde{x}_i(j) = \langle \Pi(j), \tilde{x}_i \rangle = \sum_{k=1}^{d} \Pi(j, k) \cdot \tilde{x}_i(k).
\]

\( \Pi(j) \) is a vector with independent random Gaussian entries.

\( \tilde{x}_1, \ldots, \tilde{x}_n \): original points (\( d \) dims.), \( \tilde{x}_1, \ldots, \tilde{x}_n \): compressed points (\( m < d \) dims.), \( \Pi \in \mathbb{R}^{m \times d} \): random projection (embedding function)
Compression operation is \( \tilde{x}_i = \Pi \tilde{x}_i \), so for any \( j \),

\[
\tilde{x}_i(j) = \langle \Pi(j), \tilde{x}_i \rangle = \sum_{k=1}^{d} \Pi(j,k) \cdot \tilde{x}_i(k).
\]

\( \Pi(j) \) is a vector with independent random Gaussian entries.
Compression operation is \( \tilde{x}_i = \mathbf{\Pi} \tilde{x}_i \), so for any \( j \),

\[
\tilde{x}_i(j) = \langle \mathbf{\Pi}(j), \tilde{x}_i \rangle = \sum_{k=1}^{d} \mathbf{\Pi}(j, k) \cdot \tilde{x}_i(k).
\]

\( \mathbf{\Pi}(j) \) is a vector with independent random Gaussian entries.

\( \tilde{x}_1, \ldots, \tilde{x}_n \): original points \((d \text{ dims.})\), \( \tilde{x}_1, \ldots, \tilde{x}_n \): compressed points \((m < d \text{ dims.})\), \( \mathbf{\Pi} \in \mathbb{R}^{m \times d} \): random projection (embedding function)
Compression operation is $\tilde{x}_i = \Pi \tilde{x}_i$, so for any $j$,

$$\tilde{x}_i(j) = \langle \Pi(j), \tilde{x}_i \rangle = \sum_{k=1}^{d} \Pi(j, k) \cdot \tilde{x}_i(k).$$

$\Pi(j)$ is a vector with independent random Gaussian entries.

Points with high cosine similarity have similar random projections.

Computing a length $m$ SimHash signature $SH_1(\tilde{x}_i), \ldots, SH_m(\tilde{x}_i)$ is identical to computing $\tilde{x}_i = \Pi \tilde{x}_i$ and then taking $\text{sign}(\tilde{x}_i)$. 
The Johnson-Lindenstrauss Lemma is a direct consequence of a closely related lemma:

**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 $\tilde{y} \in \mathbb{R}^d$, with probability $\geq 1 - \delta$

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

$\mathbf{P} \in \mathbb{R}^{m \times d}$: random projection matrix. $d$: original dimension. $m$: compressed dimension, $\epsilon$: embedding error, $\delta$: embedding failure prob.
The Johnson-Lindenstrauss Lemma is a direct consequence of a closely related lemma:

**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 \( \geq 1 - \delta \)

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

Applying a random matrix \( \mathbf{P} \) to any vector \( \mathbf{y} \) preserves \( \mathbf{y} \)'s norm with high probability.

\[ \mathbf{P} \in \mathbb{R}^{m \times d} \]: random projection matrix. \( d \): original dimension. \( m \): compressed dimension, \( \epsilon \): embedding error, \( \delta \): embedding failure prob.
The Johnson-Lindenstrauss Lemma is a direct consequence of a closely related lemma:

**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 $\tilde{\mathbf{y}} \in \mathbb{R}^d$, with probability $\geq 1 - \delta$

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

Applying a random matrix $\mathbf{\Pi}$ to any vector $\tilde{\mathbf{y}}$ preserves $\tilde{\mathbf{y}}$'s norm with high probability.

- Like a low-distortion embedding, but for the length of a compressed vector rather than distances between vectors.

$\mathbf{\Pi} \in \mathbb{R}^{m \times d}$: random projection matrix. $d$: original dimension. $m$: compressed dimension, $\epsilon$: embedding error, $\delta$: embedding failure prob.
The Johnson-Lindenstrauss Lemma is a direct consequence of a closely related lemma:

**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 $\tilde{y} \in \mathbb{R}^d$, with probability $\geq 1 - \delta$

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

Applying a random matrix $\mathbf{P}$ to any vector $\tilde{y}$ preserves $\tilde{y}$'s norm with high probability.

- Like a low-distortion embedding, but for the length of a compressed vector rather than distances between vectors.
- Can be proven from first principles. Will see next.

$\mathbf{P} \in \mathbb{R}^{m \times d}$: random projection matrix. $d$: original dimension. $m$: compressed dimension, $\epsilon$: embedding error, $\delta$: embedding failure prob.
Distributional JL Lemma $\implies$ JL Lemma: Distributional JL show that a random projection $\Pi$ preserves the norm of any $y$. The main JL Lemma says that $\Pi$ preserves distances between vectors.

$\bar{x}_1, \ldots, \bar{x}_n$: original points, $\tilde{x}_1, \ldots, \tilde{x}_n$: compressed points, $\Pi \in \mathbb{R}^{m \times d}$: random projection matrix. $d$: original dimension. $m$: compressed dimension, $\varepsilon$: embedding error, $\delta$: embedding failure prob.
Distributional JL Lemma $\implies$ JL Lemma: Distributional JL show that a random projection $\Pi$ preserves the norm of any $y$. The main JL Lemma says that $\Pi$ preserves distances between vectors.

Since $\Pi$ is linear these are the same thing!

$\vec{x}_1, \ldots, \vec{x}_n$: original points, $\tilde{\vec{x}}_1, \ldots, \tilde{\vec{x}}_n$: compressed points, $\Pi \in \mathbb{R}^{m \times d}$: random projection matrix. $d$: original dimension. $m$: compressed dimension, $\epsilon$: embedding error, $\delta$: embedding failure prob.
Distributional JL Lemma $\Rightarrow$ JL Lemma: Distributional JL show that a random projection $\Pi$ preserves the norm of any $y$. The main JL Lemma says that $\Pi$ preserves distances between vectors.

Since $\Pi$ is linear these are the same thing!

Proof: Given $\tilde{x}_1, \ldots, \tilde{x}_n$, define $\binom{n}{2}$ vectors $\tilde{y}_{ij}$ where $\tilde{y}_{ij} = \tilde{x}_i - \tilde{x}_j$. 

$\tilde{x}_1, \ldots, \tilde{x}_n$: original points, $\tilde{x}_1, \ldots, \tilde{x}_n$: compressed points, $\Pi \in \mathbb{R}^{m \times d}$: random projection matrix. $d$: original dimension. $m$: compressed dimension, $\epsilon$: embedding error, $\delta$: embedding failure prob.
Distributional JL Lemma $\implies$ JL Lemma: Distributional JL show that a random projection $\Pi$ preserves the norm of any $y$. The main JL Lemma says that $\Pi$ preserves distances between vectors.

Since $\Pi$ is linear these are the same thing!

Proof: Given $\tilde{x}_1, \ldots, \tilde{x}_n$, define $\binom{n}{2}$ vectors $\tilde{y}_{ij}$ where $\tilde{y}_{ij} = \tilde{x}_i - \tilde{x}_j$. 

$x_1, \ldots, x_n$: original points, $\tilde{x}_1, \ldots, \tilde{x}_n$: compressed points, $\Pi \in \mathbb{R}^{m \times d}$: random projection matrix. $d$: original dimension. $m$: compressed dimension, $\epsilon$: embedding error, $\delta$: embedding failure prob.
DISTRIBUTIONAL JL $\implies$ JL

**Distributional JL Lemma $\implies$ JL Lemma:** Distributional JL show that a random projection $\Pi$ preserves the norm of any $y$. The main JL Lemma says that $\Pi$ preserves distances between vectors.

Since $\Pi$ is linear these are the same thing!

**Proof:** Given $\tilde{x}_1, \ldots, \tilde{x}_n$, define $\binom{n}{2}$ vectors $\tilde{y}_{ij}$ where $\tilde{y}_{ij} = \tilde{x}_i - \tilde{x}_j$.

- If we choose $\Pi$ with $m = O\left(\frac{\log 1/\delta}{\epsilon^2}\right)$, for each $\tilde{y}_{ij}$ with probability $\geq 1 - \delta$ we have:

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

\[\forall i \neq j\]

\[\tilde{x}_1, \ldots, \tilde{x}_n: \text{original points}, \; \tilde{x}_1, \ldots, \tilde{x}_n: \text{compressed points}, \; \Pi \in \mathbb{R}^{m \times d}: \text{random projection matrix.} \; d: \text{original dimension.} \; m: \text{compressed dimension,} \; \epsilon: \text{embedding error,} \; \delta: \text{embedding failure prob.}\]
**DISTRIBUTIONAL JL  \[\implies\] JL**

**Distributional JL Lemma  \[\implies\] JL Lemma:** Distributional JL show that a random projection $\Pi$ preserves the norm of any $y$. The main JL Lemma says that $\Pi$ preserves distances between vectors.

Since $\Pi$ is linear these are the same thing!

**Proof:** Given $\vec{x}_1, \ldots, \vec{x}_n$, define $\binom{n}{2}$ vectors $\vec{y}_{ij}$ where $\vec{y}_{ij} = \vec{x}_i - \vec{x}_j$.

- If we choose $\Pi$ with $m = O\left(\frac{\log 1/\delta}{\epsilon^2}\right)$, for each $\vec{y}_{ij}$ with probability $\geq 1 - \delta$ we have:

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

\(\vec{x}_1, \ldots, \vec{x}_n\): original points, $\hat{\vec{x}}_1, \ldots, \hat{\vec{x}}_n$: compressed points, $\Pi \in \mathbb{R}^{m \times d}$: random projection matrix. $d$: original dimension. $m$: compressed dimension, $\epsilon$: embedding error, $\delta$: embedding failure prob.
DISTRIBUTIONAL JL  ➔  JL

Distributional JL Lemma  ➔  JL Lemma: Distributional JL show that a random projection $\Pi$ preserves the norm of any $y$. The main JL Lemma says that $\Pi$ preserves distances between vectors.

Since $\Pi$ is linear these are the same thing!

Proof: Given $\vec{x}_1, \ldots, \vec{x}_n$, define $\binom{n}{2}$ vectors $\vec{y}_{ij}$ where $\vec{y}_{ij} = \vec{x}_i - \vec{x}_j$.

- If we choose $\Pi$ with $m = O \left( \frac{\log 1/\delta}{\epsilon^2} \right)$, for each $\vec{y}_{ij}$ with probability $\geq 1 - \delta$ we have:

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

$\vec{x}_1, \ldots, \vec{x}_n$: original points, $\tilde{\vec{x}}_1, \ldots, \tilde{\vec{x}}_n$: compressed points, $\Pi \in \mathbb{R}^{m \times d}$: random projection matrix. $d$: original dimension. $m$: compressed dimension, $\epsilon$: embedding error, $\delta$: embedding failure prob.
**Claim:** If we choose $\mathbf{\Pi}$ with i.i.d. $\mathcal{N}(0, 1/m)$ entries and 
$m = O\left(\frac{\log(1/\delta')}{\epsilon^2}\right)$, letting $\tilde{x}_i = \mathbf{\Pi}x_i$, for each pair $\tilde{x}_i, \tilde{x}_j$ with probability $\geq 1 - \delta'$ we have:

$$(1 - \epsilon)\|\tilde{x}_i - \tilde{x}_j\|_2 \leq \|\tilde{x}_i - \tilde{x}_j\|_2 \leq (1 + \epsilon)\|\tilde{x}_i - \tilde{x}_j\|_2.$$
**Claim:** If we choose $\Pi$ with i.i.d. $\mathcal{N}(0, 1/m)$ entries and $m = O\left(\frac{\log(1/\delta')}{\epsilon^2}\right)$, letting $\tilde{x}_i = \Pi x_i$, for each pair $\tilde{x}_i, \tilde{x}_j$ with probability $\geq 1 - \delta'$ we have:

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

With what probability are all pairwise distances preserved?

---

$\tilde{x}_1, \ldots, \tilde{x}_n$: original points, $\tilde{x}_1, \ldots, \tilde{x}_n$: compressed points, $\Pi \in \mathbb{R}^{m \times d}$: random projection matrix. $d$: original dimension. $m$: compressed dimension, $\epsilon$: embedding error, $\delta$: embedding failure prob.
**Claim:** If we choose \( \Pi \) with i.i.d. \( \mathcal{N}(0, 1/m) \) entries and \( m = O \left( \frac{\log(1/\delta')}{\epsilon^2} \right) \), letting \( \tilde{x}_i = \Pi \hat{x}_i \), for each pair \( \tilde{x}_i, \tilde{x}_j \) with probability \( \geq 1 - \delta' \) we have:

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

With what probability are all pairwise distances preserved?

**Union bound:** With probability \( \geq 1 - \binom{n}{2} \cdot \delta' \) all pairwise distances are preserved.

\[
\Pr(\|\tilde{x}_i - \tilde{x}_j\| \neq \|x_i - x_j\|) \leq \delta^1
\]

\[
\Pr(\text{at least one } \|\tilde{x}_i - \tilde{x}_j\| \neq \|x_i - x_j\|) \leq \binom{n}{2} \delta^1
\]

\( \tilde{x}_1, \ldots, \tilde{x}_n \): original points, \( \hat{x}_1, \ldots, \hat{x}_n \): compressed points, \( \Pi \in \mathbb{R}^{m \times d} \): random projection matrix. \( d \): original dimension. \( m \): compressed dimension, \( \epsilon \): embedding error, \( \delta \): embedding failure prob.
Claim: If we choose $\mathbf{\Pi}$ with i.i.d. $\mathcal{N}(0, 1/m)$ entries and $m = O\left( \frac{\log(1/\delta')}{\epsilon^2} \right)$, letting $\tilde{x}_i = \mathbf{\Pi} x_i$, for each pair $\tilde{x}_i, \tilde{x}_j$ with probability $\geq 1 - \delta'$ we have:

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

With what probability are all pairwise distances preserved?

Union bound: With probability $\geq 1 - \binom{n}{2} \cdot \delta'$ all pairwise distances are preserved.

Apply the claim with $\delta' = \delta / \binom{n}{2}$.

$\tilde{x}_1, \ldots, \tilde{x}_n$: original points, $\bar{x}_1, \ldots, \bar{x}_n$: compressed points, $\mathbf{\Pi} \in \mathbb{R}^{m \times d}$: random projection matrix. $d$: original dimension. $m$: compressed dimension, $\epsilon$: embedding error, $\delta$: embedding failure prob.
**Claim:** If we choose \( \mathbf{P} \) with i.i.d. \( \mathcal{N}(0, 1/m) \) entries and \( m = O \left( \frac{\log(1/\delta')}{\epsilon^2} \right) \), letting \( \tilde{x}_i = \mathbf{P}x_i \), for each pair \( \tilde{x}_i, \tilde{x}_j \) with probability \( \geq 1 - \delta' \) we have:

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

With what probability are all pairwise distances preserved?

**Union bound:** With probability \( \geq 1 - \binom{n}{2} \cdot \delta' \) all pairwise distances are preserved.

Apply the claim with \( \delta' = \delta / \binom{n}{2} \). \( \implies \) for \( m = O \left( \frac{\log(1/\delta')}{\epsilon^2} \right) \), all pairwise distances are preserved with probability \( \geq 1 - \delta \).

\( \tilde{x}_1, \ldots, \tilde{x}_n \): original points, \( \tilde{x}_1, \ldots, \tilde{x}_n \): compressed points, \( \mathbf{P} \in \mathbb{R}^{m \times d} \): random projection matrix. \( d \): original dimension. \( m \): compressed dimension, \( \epsilon \): embedding error, \( \delta \): embedding failure prob.
**Claim:** If we choose $\mathbf{Π}$ with i.i.d. $\mathcal{N}(0, 1/m)$ entries and $m = O\left(\frac{\log(1/\delta')}{\epsilon^2}\right)$, letting $\tilde{x}_i = \mathbf{Π} \tilde{x}_i$, for each pair $\tilde{x}_i, \tilde{x}_j$ with probability $\geq 1 - \delta'$ we have:

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

With what probability are all pairwise distances preserved?

**Union bound:** With probability $\geq 1 - \binom{n}{2} \cdot \delta'$ all pairwise distances are preserved.

Apply the claim with $\delta' = \delta / \binom{n}{2}$. $\implies$ for $m = O\left(\frac{\log(1/\delta')}{\epsilon^2}\right)$, all pairwise distances are preserved with probability $\geq 1 - \delta$.

$m = O\left(\frac{\log(1/\delta')}{\epsilon^2}\right)$

$\tilde{x}_1, \ldots, \tilde{x}_n$: original points, $\tilde{x}_1, \ldots, \tilde{x}_n$: compressed points, $\mathbf{Π} \in \mathbb{R}^{m \times d}$: random projection matrix. $d$: original dimension. $m$: compressed dimension, $\epsilon$: embedding error, $\delta$: embedding failure prob.
Claim: If we choose $\Pi$ with i.i.d. $\mathcal{N}(0, 1/m)$ entries and $m = O\left(\frac{\log(1/\delta')}{\epsilon^2}\right)$, letting $\tilde{x}_i = \Pi \tilde{x}_i$, for each pair $\tilde{x}_i, \tilde{x}_j$ with probability $\geq 1 - \delta'$ we have:

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

With what probability are all pairwise distances preserved?

Union bound: With probability $\geq 1 - \binom{n}{2} \cdot \delta'$ all pairwise distances are preserved.

Apply the claim with $\delta' = \delta/\binom{n}{2}$. $\implies$ for $m = O\left(\frac{\log(1/\delta')}{\epsilon^2}\right)$, all pairwise distances are preserved with probability $\geq 1 - \delta$.

$$m = O\left(\frac{\log(1/\delta')}{\epsilon^2}\right) = O\left(\frac{\log(\binom{n}{2}/\delta)}{\epsilon^2}\right)$$

$\tilde{x}_1, \ldots, \tilde{x}_n$: original points, $\tilde{x}_1, \ldots, \tilde{x}_n$: compressed points, $\Pi \in \mathbb{R}^{m \times d}$: random projection matrix. $d$: original dimension. $m$: compressed dimension, $\epsilon$: embedding error, $\delta$: embedding failure prob.
Claim: If we choose \( \mathbf{\Pi} \) with i.i.d. \( \mathcal{N}(0, 1/m) \) entries and \( m = O \left( \frac{\log(1/\delta')}{\epsilon^2} \right) \), letting \( \tilde{x}_i = \mathbf{\Pi} \tilde{X}_i \), for each pair \( \tilde{x}_i, \tilde{x}_j \) with probability \( \geq 1 - \delta' \) we have:

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

With what probability are all pairwise distances preserved?

Union bound: With probability \( \geq 1 - \binom{n}{2} \cdot \delta' \) all pairwise distances are preserved.

Apply the claim with \( \delta' = \delta / \binom{n}{2} \). \( \implies \) for \( m = O \left( \frac{\log(1/\delta')}{\epsilon^2} \right) \), all pairwise distances are preserved with probability \( \geq 1 - \delta \).

\[
m = O \left( \frac{\log(1/\delta')}{\epsilon^2} \right) = O \left( \frac{\log(\binom{n}{2}/\delta)}{\epsilon^2} \right) = O \left( \frac{\log(n^2/\delta)}{\epsilon^2} \right)
\]

\( \tilde{x}_1, \ldots, \tilde{x}_n: \) original points, \( \tilde{x}_1, \ldots, \tilde{x}_n: \) compressed points, \( \mathbf{\Pi} \in \mathbb{R}^{m \times d}: \) random projection matrix. \( d: \) original dimension. \( m: \) compressed dimension, \( \epsilon: \) embedding error, \( \delta: \) embedding failure prob.
Claim: If we choose \( \mathbf{\Pi} \) with i.i.d. \( \mathcal{N}(0, 1/m) \) entries and 
\[ m = O \left( \frac{\log(1/\delta')}{\epsilon^2} \right), \]
letting \( \tilde{x}_i = \mathbf{\Pi} \tilde{x}_i \), for each pair \( \tilde{x}_i, \tilde{x}_j \) with probability 
\( \geq 1 - \delta' \) we have:
\[ (1 - \epsilon)\|\tilde{x}_i - \tilde{x}_j\|_2 \leq \|\tilde{x}_i - \tilde{x}_j\|_2 \leq (1 + \epsilon)\|\tilde{x}_i - \tilde{x}_j\|_2. \]

With what probability are all pairwise distances preserved?

Union bound: With probability \( \geq 1 - \binom{n}{2} \cdot \delta' \) all pairwise distances are preserved.

Apply the claim with \( \delta' = \delta / \binom{n}{2} \). \( \forall \) for \( m = O \left( \frac{\log(1/\delta')}{\epsilon^2} \right) \), all pairwise distances are preserved with probability \( \geq 1 - \delta \).

\[ m = O \left( \frac{\log(1/\delta')}{\epsilon^2} \right) = O \left( \frac{\log(\binom{n}{2}/\delta)}{\epsilon^2} \right) = O \left( \frac{\log(n^2/\delta)}{\epsilon^2} \right) = O \left( \frac{\log(n/\delta)}{\epsilon^2} \right) \]

\( \tilde{x}_1, \ldots, \tilde{x}_n \): original points, \( \tilde{x}_1, \ldots, \tilde{x}_n \): compressed points, \( \mathbf{\Pi} \in \mathbb{R}^{m \times d} \): random projection matrix. \( d \): original dimension. \( m \): compressed dimension, \( \epsilon \): embedding error, \( \delta \): embedding failure prob.
Claim: If we choose $\mathbf{\Pi}$ with i.i.d. $\mathcal{N}(0, 1/m)$ entries and
$m = O \left( \frac{\log(1/\delta')}{\epsilon^2} \right)$, letting $\tilde{x}_i = \mathbf{\Pi}x_i$, for each pair $\tilde{x}_i, \tilde{x}_j$ with probability
$\geq 1 - \delta'$ we have:

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

With what probability are all pairwise distances preserved?

Union bound: With probability $\geq 1 - \binom{n}{2} \cdot \delta'$ all pairwise distances are preserved.

Apply the claim with $\delta' = \delta'/(\binom{n}{2})$. \implies for $m = O \left( \frac{\log(1/\delta')}{\epsilon^2} \right)$, all
pairwise distances are preserved with probability $\geq 1 - \delta$.

$$m = O \left( \frac{\log(1/\delta')}{\epsilon^2} \right) = O \left( \frac{\log(\binom{n}{2}/\delta)}{\epsilon^2} \right) = O \left( \frac{\log(n^2/\delta)}{\epsilon^2} \right) = O \left( \frac{\log(n/\delta)}{\epsilon^2} \right)$$

Yields the JL lemma.
**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 \( \tilde{y} \in \mathbb{R}^d \), with probability \( \geq 1 - \delta \)

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

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

**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 $\tilde{y} \in \mathbb{R}^d$, with probability $\geq 1 - \delta$

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

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

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

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

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

\( \mathbf{\tilde{y}} \in \mathbb{R}^d \): arbitrary vector, \( \mathbf{\tilde{y}} \in \mathbb{R}^m \): compressed vector, \( \mathbf{\Pi} \in \mathbb{R}^{m \times d} \): random projection. \( d \): 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 $\tilde{y} \in \mathbb{R}^d$, with probability $\geq 1 - \delta$

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

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

\[\tilde{y} \in \mathbb{R}^d: \text{arbitrary vector, } \tilde{y} \in \mathbb{R}^m: \text{compressed vector, } \Pi \in \mathbb{R}^{m \times d}: \text{random projection. } d: \text{original dim. } m: \text{compressed dim, } \epsilon: \text{error, } \delta: \text{failure prob.}\]
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 \( m = O \left( \frac{\log(1/\delta)}{\epsilon^2} \right) \), then for any \( \tilde{\mathbf{y}} \in \mathbb{R}^{d} \), with probability \( \geq 1 - \delta \)

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

- Let \( \tilde{\mathbf{y}} \) denote \( \Pi \tilde{\mathbf{y}} \) and let \( \Pi(j) \) denote the \( j^{th} \) row of \( \Pi \).
- For any \( j \), \( \tilde{\mathbf{y}}(j) = \langle \Pi(j), \tilde{\mathbf{y}} \rangle = \sum_{i=1}^{d} g_i \cdot \tilde{\mathbf{y}}(i) \) where \( g_i \sim \mathcal{N}(0, 1/m) \).
**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 \( \geq 1 - \delta \)
\[
(1 - \epsilon) \| \mathbf{\tilde{y}} \|_2 \leq \| \mathbf{\Pi} \mathbf{\tilde{y}} \|_2 \leq (1 + \epsilon) \| \mathbf{\tilde{y}} \|_2
\]

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

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

• For any \( j \), \( \tilde{y}(j) = \langle \Pi(j), \tilde{y} \rangle = \frac{1}{\sqrt{m}} \sum_{i=1}^{d} g_i \cdot \tilde{y}(i) \) where \( g_i \sim \mathcal{N}(0, 1) \).
DISTRIBUTIONAL JL PROOF

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

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

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

\[
\sum_{i=1}^{\sqrt{m}} \mathcal{N}(0, \tilde{y}(i)^2)
\]

\( \sqrt{m} \) variance 1

\( \tilde{y} \in \mathbb{R}^d \): 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 \( j \), \( \tilde{y}(j) = \langle \Pi(j), \tilde{y} \rangle = \frac{1}{\sqrt{m}} \sum_{i=1}^{d} g_i \cdot \tilde{y}(i) \) where \( g_i \sim \mathcal{N}(0, 1) \).

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

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

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

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

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

$\tilde{y} \in \mathbb{R}^d$: 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 \( j \), \( \tilde{y}(j) = \langle \Pi(j), \tilde{y} \rangle = \frac{1}{\sqrt{m}} \sum_{i=1}^{d} g_i \cdot \tilde{y}(i) \) where \( g_i \sim \mathcal{N}(0, 1) \).

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

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

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 $\tilde{y} = \mathbf{p} \tilde{y}$, we have $\tilde{y}(j) = \langle \mathbf{p}(j), \tilde{y} \rangle$ and:

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

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

**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:

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

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

**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:

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

\( \tilde{y} \in \mathbb{R}^d \): 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:

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

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

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

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

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

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

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

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

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

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

### 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, $\tilde{y}(j) \sim \mathcal{N}(0, \|\vec{y}\|_2^2/m)$. I.e., $\vec{y}$ itself is a random Gaussian vector. Rotational invariance of the Gaussian distribution.

$\vec{y} \in \mathbb{R}^d$: arbitrary vector, $\vec{y} \in \mathbb{R}^m$: compressed vector, $\Pi \in \mathbb{R}^{m \times d}$: random projection mapping $\vec{y} \rightarrow \vec{y}$. $\Pi(j)$: $j^{th}$ 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:

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

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

Stability is another explanation for the central limit theorem.

\( \tilde{y} \in \mathbb{R}^d \): 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
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 $\tilde{y} = \Pi \tilde{y}$:

$$\tilde{y}(j) \sim \mathcal{N}(0, \|\tilde{y}\|_2^2/m).$$
**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 $\mathbf{y} \in \mathbb{R}^d$, letting $\mathbf{\tilde{y}} = \mathbf{\Pi} \mathbf{y}$:

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

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

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

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

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

$\tilde{\mathbf{y}} \in \mathbb{R}^d$: 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, $\mathbf{g}_i$: 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 $\mathbf{y} \in \mathbb{R}^d$, letting $\mathbf{\tilde{y}} = \mathbf{\Pi} \mathbf{y}$:

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

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

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

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

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

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

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

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

\( \mathbf{\tilde{y}} \in \mathbb{R}^d \): arbitrary vector, \( \mathbf{\tilde{y}} \in \mathbb{R}^m \): compressed vector, \( \mathbf{\Pi} \in \mathbb{R}^{m \times d} \): random projection mapping \( \mathbf{\tilde{y}} \to \mathbf{\tilde{y}} \). \( \mathbf{\Pi}(j) \): \( j \text{th} \) 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 \( \mathbb{E}[||\mathbf{\tilde{y}}||_2^2] \)?

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

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

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

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

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

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

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

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

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

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

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

$\tilde{\mathbf{y}} \in \mathbb{R}^d$: 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, $\mathbf{g}_i$: 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 $\tilde{\mathbf{y}} = \mathbf{P}\mathbf{y}$:

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

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

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

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

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

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

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

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

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

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

$$||\mathbf{\tilde{y}}||_2^2 = \sum_{i=1}^{m} \mathbf{\tilde{y}}(j)^2 \text{ a Chi-Squared random variable with } m \text{ degrees of freedom (a sum of } m \text{ squared independent Gaussians)}$$
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 \( \vec{y} \in \mathbb{R}^d \), letting \( \vec{\tilde{y}} = \mathbf{\Pi} \vec{y} \):

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

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

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

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

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

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

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

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

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

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

$\tilde{y} \in \mathbb{R}^d$: 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.
**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)

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

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

Gives the distributional JL Lemma and thus the classic JL Lemma!
Support Vector Machines: A classic ML algorithm, where data is classified with a hyperplane.
Support Vector Machines: A classic ML algorithm, where data is classified with a hyperplane.
Support Vector Machines: A classic ML algorithm, where data is classified with a hyperplane.
Support Vector Machines: A classic ML algorithm, where data is classified with a hyperplane.

- For any point $a$ in A, $\langle a, w \rangle \geq c + m$
- For any point $b$ in B, $\langle b, w \rangle \leq c - m$.
- Assume all vectors have unit norm.
Support Vector Machines: A classic ML algorithm, where data is classified with a hyperplane.

- For any point \( a \) in \( A \), \( \langle a, w \rangle \geq c + m \)
- For any point \( b \) in \( B \), \( \langle b, w \rangle \leq c - m \).
- Assume all vectors have unit norm.

JL Lemma implies that after projection into \( O \left( \frac{\log n}{m^2} \right) \) dimensions, still have \( \langle \tilde{a}, \tilde{w} \rangle \geq c + m/4 \) and \( \langle \tilde{b}, \tilde{w} \rangle \leq c - m/4 \).
Support Vector Machines: A classic ML algorithm, where data is classified with a hyperplane.

- For any point $a$ in A, $\langle a, w \rangle \geq c + m$
- For any point $b$ in B, $\langle b, w \rangle \leq c - m$.
- Assume all vectors have unit norm.

JL Lemma implies that after projection into $O\left(\frac{\log n}{m^2}\right)$ dimensions, still have $\langle \tilde{a}, \tilde{w} \rangle \geq c + m/4$ and $\langle \tilde{b}, \tilde{w} \rangle \leq c - m/4$.

**Upshot:** Can random project and run SVM (much more efficiently) in the lower dimensional space to find separator $\tilde{w}$. 
Questions?