COMPSCI 514: ALGORITHMS FOR DATA SCIENCE

Andrew McGregor

Lecture 11

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 $\boldsymbol{M} : \mathbb{R}^d \to \mathbb{R}^m$ such that $m = O\left(\frac{\log n}{\epsilon^2}\right)$ and letting $\tilde{x}_i = \boldsymbol{M}\vec{x}_i$:

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

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

The Johnson-Lindenstrauss Lemma is a direct consequence of:

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

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

$\boldsymbol{M} \in \mathbb{R}^{m \times d}$: random projection matrix. d: original dimension. m: compressed dimension, ϵ: embedding error, δ: embedding failure prob.

The Johnson-Lindenstrauss Lemma is a direct consequence of:

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

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

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

$\boldsymbol{M} \in \mathbb{R}^{m \times d}$: random projection matrix. d: original dimension. m: compressed dimension, ϵ: embedding error, δ: embedding failure prob.

DISTRIBUTIONAL JL PROOF (PART 1 OF 3)

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

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

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

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

- Let $\tilde{y} = \boldsymbol{M}y$ and $\boldsymbol{M}_j$ be the j^{th} row of $\boldsymbol{M}$

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

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

- Let $\tilde{y} = \boldsymbol{M}y$ and $\boldsymbol{M}_j$ be the j^{th} row of $\boldsymbol{M}$
- For any j, $\tilde{y}_j = \langle \boldsymbol{M}_j, y \rangle$

DISTRIBUTIONAL JL PROOF (PART 1 OF 3)

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

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

- Let $\tilde{y} = \boldsymbol{M}y$ and $\boldsymbol{M}_j$ be the j^{th} row of $\boldsymbol{M}$
- For any j, $\tilde{y}_j = \langle \boldsymbol{M}_j, y \rangle = \sum_{i=1}^{d} \mathbf{g}_i \cdot y_i$ where $\mathbf{g}_i \sim \mathcal{N}(0, 1/m)$.

DISTRIBUTIONAL JL PROOF (PART 1 OF 3)

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

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

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

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

DISTRIBUTIONAL JL PROOF (PART 1 OF 3)

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

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

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

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

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

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

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

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

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

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

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

$$\mathbb{E}[\tilde{y}_j^2] = \mathrm{Var}[\tilde{y}_j] = \sum_{i=1}^d \mathrm{Var}[\mathbf{g}_i \cdot y_i] = \sum_i y_i^2/m = \|y\|_2^2/m \ .$$

- Hence $\mathbb{E}[\|\tilde{y}\|_2^2] = \mathbb{E}[\sum_j \tilde{y}_j^2] = \|y\|_2^2$.

DISTRIBUTIONAL JL PROOF (PART 1 OF 3)

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

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

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

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

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

$$\mathbb{E}[\tilde{y}_j^2] = \mathrm{Var}[\tilde{y}_j] = \sum_{i=1}^{d} \mathrm{Var}[\mathbf{g}_i \cdot y_i] = \sum_i y_i^2/m = \|y\|_2^2/m \ .$$

- Hence $\mathbb{E}[\|\tilde{y}\|_2^2] = \mathbb{E}[\sum_j \tilde{y}_j^2] = \|y\|_2^2$. Remains to show $\|\tilde{y}\|_2^2$ is concentrated.

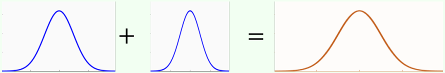
DISTRIBUTIONAL JL PROOF (PART 2 OF 3)

Letting $\tilde{y} = \boldsymbol{M}y$, we have $\tilde{y}_j = \langle \boldsymbol{M}_j, y \rangle$ and:

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

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

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



Thus, $\tilde{y}_j \sim \mathcal{N}(0, \sum_{i=1}^{d} y_i^2/m) = \mathcal{N}(0, \|y\|_2^2/m)$.

DISTRIBUTIONAL JL PROOF (PART 3 OF 3)

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

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

DISTRIBUTIONAL JL PROOF (PART 3 OF 3)

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

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

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

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

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

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

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

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

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

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

DISTRIBUTIONAL JL PROOF (PART 3 OF 3)

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

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

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

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

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

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

DISTRIBUTIONAL JL PROOF (PART 3 OF 3)

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

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

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

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

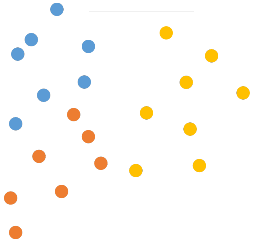
If we set $m = O\left(\frac{\log(1/\delta)}{\epsilon^2}\right)$, with probability $1 - O(e^{-\log(1/\delta)}) \geq 1 - \delta$:

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

Gives the distributional JL Lemma and thus the classic JL Lemma!

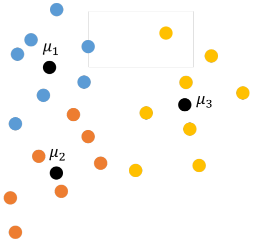
EXAMPLE APPLICATION: k-MEANS CLUSTERING

Goal: Separate n points in d dimensional space into k groups $\mathcal{C}_1, \ldots, \mathcal{C}_k$.



EXAMPLE APPLICATION: k-MEANS CLUSTERING

Goal: Separate n points in d dimensional space into k groups $\mathcal{C}_1, \ldots, \mathcal{C}_k$.

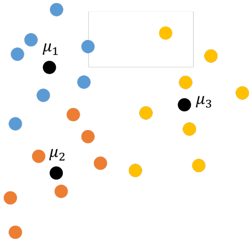


k-means Objective: $Cost(\mathcal{C}_1, \ldots, \mathcal{C}_k) = \sum_{j=1}^{k} \sum_{\vec{x} \in \mathcal{C}_j} \|\vec{x} - \mu_j\|_2^2$ where

$$\mu_j = \frac{1}{|\mathcal{C}_j|} \sum_{\vec{x} \in \mathcal{C}_j} \vec{x}$$

is the average of the points in $\mathcal{C}_j$.

Goal: Separate n points in d dimensional space into k groups $\mathcal{C}_1, \ldots, \mathcal{C}_k$.



k-means Objective: $Cost(\mathcal{C}_1, \ldots, \mathcal{C}_k) = \sum_{j=1}^{k} \sum_{\vec{x} \in \mathcal{C}_j} \|\vec{x} - \mu_j\|_2^2$ where

$$\mu_j = \frac{1}{|\mathcal{C}_j|} \sum_{\vec{x} \in \mathcal{C}_j} \vec{x}$$

is the average of the points in $\mathcal{C}_j$.

Exercise: Can be rewritten as $Cost(\mathcal{C}_1, \ldots, \mathcal{C}_k) = \sum_{j=1}^{k} \sum_{\vec{x}_1, \vec{x}_2 \in \mathcal{C}_j} \frac{\|\vec{x}_1 - \vec{x}_2\|_2^2}{|\mathcal{C}_j|}$

EXAMPLE APPLICATION: k-MEANS CLUSTERING

k-means Objective: $Cost(\mathcal{C}_1, \ldots, \mathcal{C}_k) = \sum_{j=1}^{k} \sum_{\vec{x}_1, \vec{x}_2 \in \mathcal{C}_j} \frac{\|\vec{x}_1 - \vec{x}_2\|_2^2}{|\mathcal{C}_j|}$

k-means Objective: $Cost(\mathcal{C}_1, \ldots, \mathcal{C}_k) = \sum_{j=1}^{k} \sum_{\vec{x}_1, \vec{x}_2 \in \mathcal{C}_j} \frac{\|\vec{x}_1 - \vec{x}_2\|_2^2}{|\mathcal{C}_j|}$

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

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

k-means Objective: $Cost(\mathcal{C}_1, \ldots, \mathcal{C}_k) = \sum_{j=1}^{k} \sum_{\vec{x}_1, \vec{x}_2 \in \mathcal{C}_j} \frac{\|\vec{x}_1 - \vec{x}_2\|_2^2}{|\mathcal{C}_j|}$

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

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

Letting $\overline{Cost}(\mathcal{C}_1, \ldots, \mathcal{C}_k) = \sum_{j=1}^{k} \sum_{\vec{x}_1, \vec{x}_2 \in \mathcal{C}_j} \frac{\|\tilde{x}_1 - \tilde{x}_2\|_2^2}{|\mathcal{C}_j|}$

$$(1-\epsilon)Cost(\mathcal{C}_1, \ldots, \mathcal{C}_k) \leq \overline{Cost}(\mathcal{C}_1, \ldots, \mathcal{C}_k) \leq (1+\epsilon)Cost(\mathcal{C}_1, \ldots, \mathcal{C}_k).$$

EXAMPLE APPLICATION: k-MEANS CLUSTERING

k-means Objective: $Cost(\mathcal{C}_1, \ldots, \mathcal{C}_k) = \sum_{j=1}^k \sum_{\vec{x}_1, \vec{x}_2 \in \mathcal{C}_j} \frac{\|\vec{x}_1 - \vec{x}_2\|_2^2}{|\mathcal{C}_j|}$

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

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

Letting $\overline{Cost}(\mathcal{C}_1, \ldots, \mathcal{C}_k) = \sum_{j=1}^k \sum_{\vec{x}_1, \vec{x}_2 \in \mathcal{C}_j} \frac{\|\tilde{x}_1 - \tilde{x}_2\|_2^2}{|\mathcal{C}_j|}$

$$(1-\epsilon) Cost(\mathcal{C}_1, \ldots, \mathcal{C}_k) \leq \overline{Cost}(\mathcal{C}_1, \ldots, \mathcal{C}_k) \leq (1+\epsilon) Cost(\mathcal{C}_1, \ldots, \mathcal{C}_k).$$

Upshot: Can cluster in m dimensional space (much more efficiently) and minimize $\overline{Cost}(\mathcal{C}_1, \ldots, \mathcal{C}_k)$.

JL LEMMA IS ALMOST OPTIMAL

ORTHOGONAL VECTORS

- Recall that we say two vectors x, y are orthogonal if $\langle x, y \rangle = 0$.

ORTHOGONAL VECTORS

- Recall that we say two vectors x, y are orthogonal if $\langle x, y \rangle = 0$.
- What is the largest set of mutually orthogonal unit vectors in d-dimensional space?

ORTHOGONAL VECTORS

- Recall that we say two vectors x, y are orthogonal if $\langle x, y \rangle = 0$.
- What is the largest set of mutually orthogonal unit vectors in d-dimensional space? Answer: d.

ORTHOGONAL VECTORS

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

ORTHOGONAL VECTORS

- Recall that we say two vectors x, y are orthogonal if $\langle x, y \rangle = 0$.
- What is the largest set of mutually orthogonal unit vectors in d-dimensional space? Answer: d.
- How large can a set of unit vectors in d-dimensional space be that have all pairwise dot products $|\langle x, y \rangle| \leq \epsilon$? Answer: $2^{\Omega(\epsilon^2 d)}$.

An exponentially large set of random vectors will be nearly pairwise orthogonal with high probability!

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

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

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

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

- What is $\|x_i\|_2$?

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

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

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

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

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

- What is $\|x_i\|_2$? Every x_i is always a unit vector.
- What is $\mathbb{E}[\langle x_i, x_j\rangle]$?

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

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

- What is $\|x_i\|_2$? Every x_i is always a unit vector.
- What is $\mathbb{E}[\langle x_i, x_j \rangle]$? $\mathbb{E}[\langle x_i, x_j \rangle] = 0$

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

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

- What is $\|x_i\|_2$? Every x_i is always a unit vector.
- What is $\mathbb{E}[\langle x_i, x_j\rangle]$? $\mathbb{E}[\langle x_i, x_j\rangle] = 0$
- By a Bernstein bound, $\Pr[|\langle x_i, x_j\rangle| \geq \epsilon] \leq 2e^{-\epsilon^2 d/6}$.

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

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

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

ORTHOGONAL VECTORS PROOF

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

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

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

We won't prove it but this is essentially optimal: In d dimensions, there can be at most $2^{O(\epsilon^2 d)}$ nearly orthogonal unit vectors.

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

$$(1-\epsilon)\|x_i - x_j\|_2^2 \leq \|\boldsymbol{M}x_i - \boldsymbol{M}x_j\|_2^2 \leq (1+\epsilon)\|x_i - x_j\|_2^2.$$

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

$$(1-\epsilon)\|x_i - x_j\|_2^2 \leq \|\boldsymbol{M}x_i - \boldsymbol{M}x_j\|_2^2 \leq (1+\epsilon)\|x_i - x_j\|_2^2.$$

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

$$\frac{\boldsymbol{M}x_1}{\|\boldsymbol{M}x_1\|_2}, \ldots, \frac{\boldsymbol{M}x_n}{\|\boldsymbol{M}x_n\|_2}$$

are nearly orthogonal unit vectors in m-dimensions (with pairwise dot products bounded by ϵ).

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

$$(1-\epsilon)\|x_i - x_j\|_2^2 \leq \|\boldsymbol{M}x_i - \boldsymbol{M}x_j\|_2^2 \leq (1+\epsilon)\|x_i - x_j\|_2^2.$$

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

$$\frac{\boldsymbol{M}x_1}{\|\boldsymbol{M}x_1\|_2}, \ldots, \frac{\boldsymbol{M}x_n}{\|\boldsymbol{M}x_n\|_2}$$

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

$$\|x - y\|_2^2 = \|x\|_2^2 + \|y\|_2^2 - 2\langle x, y \rangle \ .$$

CONNECTION TO DIMENSIONALITY REDUCTION

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

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

CONNECTION TO DIMENSIONALITY REDUCTION

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

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

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

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

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

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

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

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

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