COMPSCI 514: Algorithms for Data Science

Cameron Musco

University of Massachusetts Amherst. Fall 2022.
Lecture 14

Logistics

- We will be grading the exams this week.
- We will release solutions, but still have some students taking make up exams, so are holding off.
- Feel free to ask about the questions in office hours.

Summary

Last Few Classes: The Johnson-Lindenstrauss Lemma

- Dimensionality reduction and low-distortion embeddings.
- Statement of the JL Lemma: we can obtain low-distortion embeddings for any set of points via random projection.
- Reduction of the JL Lemma to the 'distributional JL Lemma'.
- Proof of the distributional JL lemma.

Summary

Last Few Classes: The Johnson-Lindenstrauss Lemma

- Dimensionality reduction and low-distortion embeddings.
- Statement of the JL Lemma: we can obtain low-distortion embeddings for any set of points via random projection.
- Reduction of the JL Lemma to the 'distributional JL Lemma'.
- Proof of the distributional JL lemma.

This Class:

- Connections between the JL Lemma, low-distortion embeddings, and high dimensional geometry.

Summary

Last Few Classes: The Johnson-Lindenstrauss Lemma

- Dimensionality reduction and low-distortion embeddings.
- Statement of the JL Lemma: we can obtain low-distortion embeddings for any set of points via random projection.
- Reduction of the JL Lemma to the 'distributional JL Lemma'.
- Proof of the distributional JL lemma.

This Class:

- Connections between the JL Lemma, low-distortion embeddings, and high dimensional geometry.

Next Few Classes:

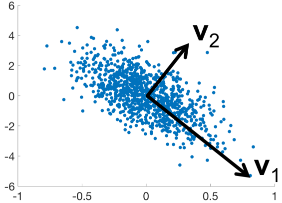
- Data-dependent dimensionality reduction via PCA. Formulation as low-rank matrix approximation.

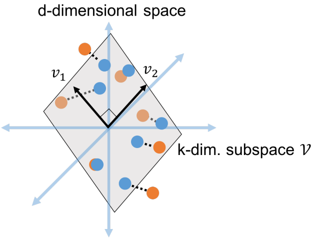
The Johnson-Lindenstrauss Lemma and High Dimensional Geometry

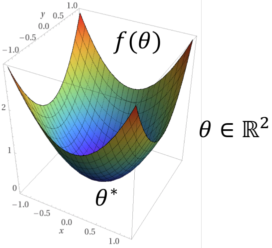
The Johnson-Lindenstrauss Lemma and High Dimensional Geometry

- High-dimensional Euclidean space looks *very different* from low-dimensional space. So how can JL work?
- Is Euclidean distance in high-dimensional meaningless, making JL useless? (The curse of dimensionality)

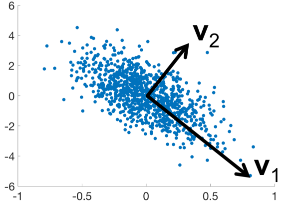
Low-Dimensional Intuition

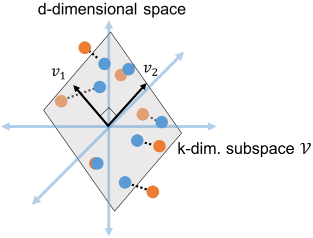


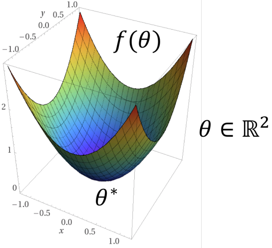




Low-Dimensional Intuition





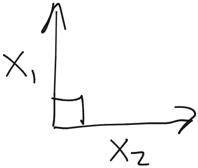


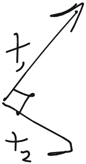
This can be a bit dangerous as in reality high-dimensional space is very different from low-dimensional space.

Orthogonal Vectors

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

a) 1 b) $\log d$ c) $\sqrt{d}$ [✓ d) d]





Orthogonal Vectors

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

a) 1 b) $\log d$ c) $\sqrt{d}$ d) d

Nearly Orthogonal Vectors

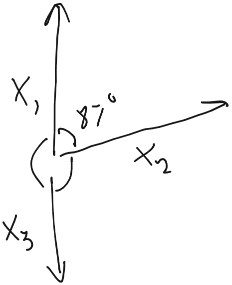
What is the largest set of unit vectors in d-dimensional space that have all pairwise dot products $|\langle \vec{x}, \vec{y} \rangle| \leq \epsilon$? (think $\epsilon = .01$)

a) d

b) $\Theta(d)$ 50%

c) $\Theta(d^2)$ 5

d) $2^{\Theta(d)}$ 5



Nearly Orthogonal Vectors

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

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

$X_1 \ldots\ldots X_t$ $t = 2^{\Theta(d)}$

for all i, j $|\langle X_i, X_j \rangle| \leq \varepsilon$ $(\varepsilon = .01)$

Nearly Orthogonal Vectors

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

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

In fact, an exponentially large set of random vectors will be nearly pairwise orthogonal with high probability!

Orthogonal Vectors Proof

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

Orthogonal Vectors Proof

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

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

$$X_i = \left[\frac{1}{\sqrt{d}}, -\frac{1}{\sqrt{d}} \quad -\frac{1}{\sqrt{d}} \cdots \right]$$

Orthogonal Vectors Proof

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

$\in \mathbb{R}^d$

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

- What is $\|\vec{x}_i\|_2$? $= 1 \quad = \sqrt{\sum_{i=1}^{d} x(i)^2} = \sqrt{d \cdot \frac{1}{d}} = 1$
- What is $\mathbb{E}[\langle \vec{x}_i, \vec{x}_j \rangle]$?

$$= \mathbb{E}\left[\sum_{k=1}^{d} x_i(k) x_j(k)\right] = \sum_{k=1}^{d} \mathbb{E}\, \underbrace{x_i(k)\, x_j(k)}_{0}$$

Orthogonal Vectors Proof

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

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

- What is $\|\vec{x}_i\|_2$? Every $\vec{x}_i$ is always a unit vector.
- What is $\mathbb{E}[\langle \vec{x}_i, \vec{x}_j \rangle]$?

Orthogonal Vectors Proof

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

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

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

Orthogonal Vectors Proof

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

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

$\sum_{k=1}^{d} \underbrace{\overset{\pm 1/\sqrt{d}}{x_i(k)}\ \overset{\pm 1/\sqrt{d}}{x_j(k)}}$ $\frac{1}{d}$ w.p $\frac{1}{2}$ $\frac{-1}{d}$ w.p $\frac{1}{2}$

- What is $\|\vec{x}_i\|_2$? Every $\vec{x}_i$ is always a unit vector.
- What is $\mathbb{E}[\langle \vec{x}_i, \vec{x}_j \rangle]$? $\mathbb{E}[\langle \vec{x}_i, \vec{x}_j \rangle] = 0$
- By a Chernoff bound, $\underline{\Pr[|\langle \vec{x}_i, \vec{x}_j \rangle| \geq \epsilon]} \leq \underline{2e^{-\epsilon^2 d/6}}$ (great exercise).

$\{0,1\}$ $\mathbb{1}_k = 1$ if $x_i(k)x_j(k) = \frac{1}{d}$

$\mathbb{1}_k = 0$ if $x_i(k)x_j(k) = -\frac{1}{d}$

$Z = \sum \mathbb{1}_k$ $\mathbb{E}[Z]$

$|\langle x_i, x_j \rangle| \geq \varepsilon \iff Z \geq \frac{d}{2} + \tilde{\varepsilon d}$

$Z \leq \frac{d}{2} - \varepsilon d$

Orthogonal Vectors Proof

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

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

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

$$\binom{t}{2} \leq \frac{t^2}{2} \leq \frac{1}{2} \cdot \frac{1}{2^2} \left(e^{\epsilon^2 d/12}\right)^2 \qquad \leq \frac{1}{4}$$

Curse of Dimensionality

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

Curse of Dimensionality

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

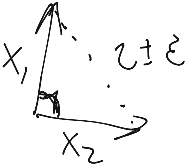
$$\underbrace{\|\vec{x}_i - \vec{x}_j\|_2^2}$$

Curse of Dimensionality

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

$$\|\vec{x}_i - \vec{x}_j\|_2^2 = \underbrace{\|\vec{x}_i\|_2^2}_{1} + \underbrace{\|\vec{x}_j\|_2^2}_{1} - \underbrace{2\vec{x}_i^T\vec{x}_j}_{\pm 2\varepsilon}$$

$$= \underline{2 \pm \varepsilon}$$



Curse of Dimensionality

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

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

Curse of Dimensionality

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

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

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

Curse of Dimensionality

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

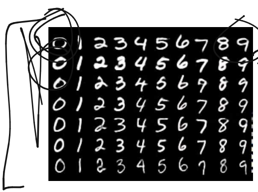
$$\|\vec{x}_i - \vec{x}_j\|_2^2 = \|\vec{x}_i\|_2^2 + \|\vec{x}_j\|_2^2 - 2\vec{x}_i^T\vec{x}_j \in [1.98, 2.02].$$

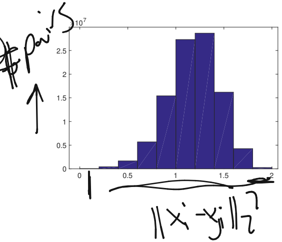
Even with an exponential number of random vector samples, we don't see any nearby vectors.

- One version of the 'curse of dimensionality'.
- If all your distances are roughly the same, distance based methods (k-means clustering, nearest neighbors, SVMs, etc.) aren't going to work well.
- Distances are only meaningful if we have lots of structure and our data isn't just independent random vectors.

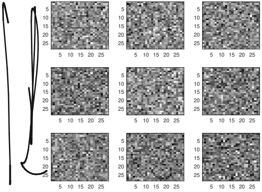
Curse of Dimensionality

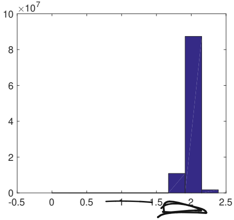
Distances for MNIST Digits:





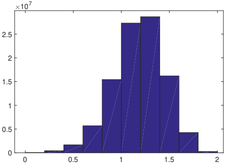
Distances for Random Images:



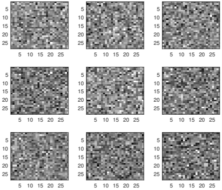


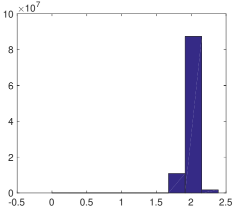
Curse of Dimensionality

Distances for MNIST Digits:



Distances for Random Images:





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

Connection to Dimensionality Reduction

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

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

Connection to Dimensionality Reduction

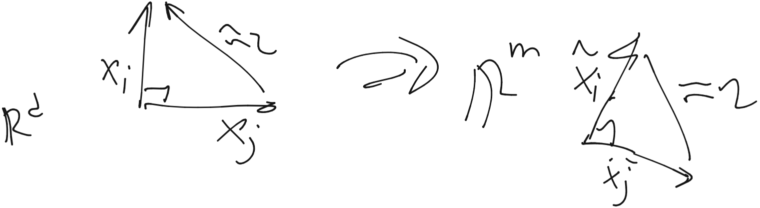
n is # datapoints initial dim d final dim m $m = \frac{\log n}{\epsilon}$

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

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

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

$\frac{\log n}{\epsilon^2}$



Connection to Dimensionality Reduction

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

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

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

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 are at most $2^{O(\epsilon^2 m)}$ nearly orthogonal 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 are at most $2^{O(\epsilon^2 m)}$ nearly orthogonal vectors.

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

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 are at most $2^{O(\epsilon^2 m)}$ nearly orthogonal vectors.

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

$$\epsilon^2 m \geq \log n$$

$$\underline{m \geq \frac{\log n}{\epsilon^2}}$$

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 are at most $2^{O(\epsilon^2 m)}$ nearly orthogonal vectors.

- For both these to hold it must be that $n \leq 2^{O(\epsilon^2 m)}$.
- $2^{O(\epsilon^2 m)} \geq 2^{O(\log n)} = n.$ Tells us that the JL lemma is optimal up to constants.

Connection to Dimensionality Reduction

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

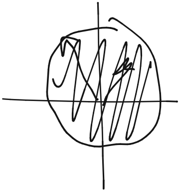
Claim 2: In m dimensions, there are at most $2^{O(\epsilon^2 m)}$ nearly orthogonal vectors.

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

Bizarre Shape of High-Dimensional Balls

Let $\mathcal{B}_d$ be the unit ball in d dimensions. $\mathcal{B}_d = \{x \in \mathbb{R}^d : \|x\|_2 \leq 1\}$.

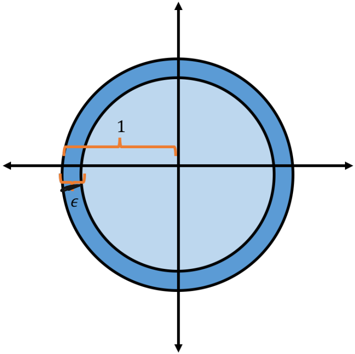
$\mathcal{B}_2$



Bizarre Shape of High-Dimensional Balls

Let $\mathcal{B}_d$ be the unit ball in d dimensions. $\mathcal{B}_d = \{x \in \mathbb{R}^d : \|x\|_2 \leq 1\}$.

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



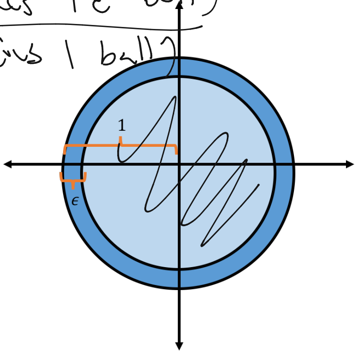
Bizarre Shape of High-Dimensional Balls

Let $\mathcal{B}_d$ be the unit ball in d dimensions. $\mathcal{B}_d = \{x \in \mathbb{R}^d : \|x\|_2 \leq 1\}$.

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

$$\frac{\text{Volume (radius } 1-\varepsilon \text{ ball)}}{\text{volume (radius 1 ball)}}$$

$$\frac{(1-\varepsilon)^d}{1^d}$$



Volume of a radius R ball is $\underline{\frac{\pi^{\frac{d}{2}}}{(d/2)!}} \cdot \underline{R^d}$.

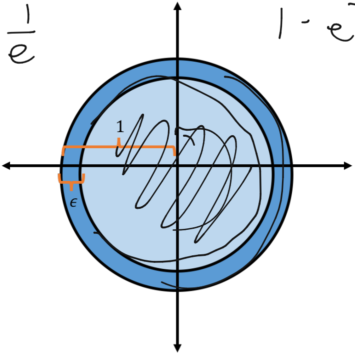
Bizarre Shape of High-Dimensional Balls

Let $\mathcal{B}_d$ be the unit ball in d dimensions. $\mathcal{B}_d = \{x \in \mathbb{R}^d : \|x\|_2 \leq 1\}$.

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

$(1-\varepsilon)^{1/\varepsilon} \leq \frac{1}{e}$

$1 - e^{-\varepsilon d}$



$1 - e^{-\frac{1}{10}\cdot 100} = 1 - e^{-10}$

$= .9999$

Volume of a radius R ball is $\frac{\pi^{\frac{d}{2}}}{(d/2)!} \cdot R^d$.

Bizarre Shape of High-Dimensional Balls

All but an $e^{-\epsilon d}$ fraction of a unit ball's volume is within ϵ of its surface.

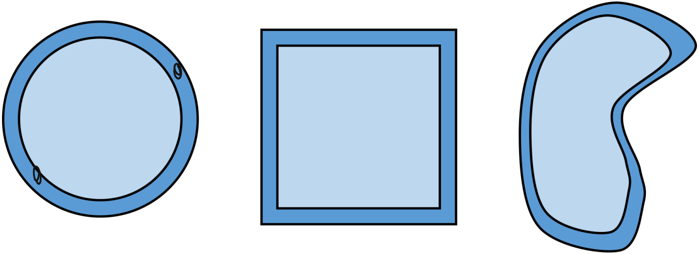
Bizarre Shape of High-Dimensional Balls

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

Bizarre Shape of High-Dimensional Balls

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

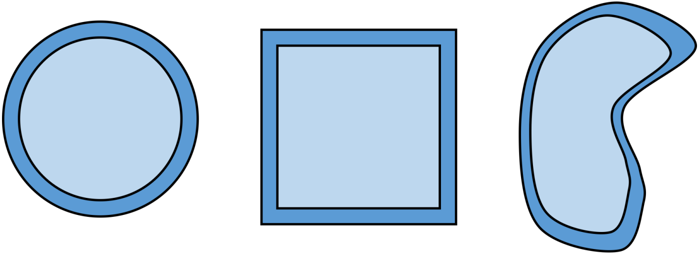
- **Isoperimetric inequality**: the ball has the minimum surface area/volume ratio of any shape.



Bizarre Shape of High-Dimensional Balls

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

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

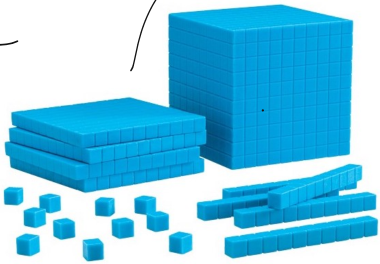


- If we randomly sample points from any high-dimensional shape, nearly all will fall near its surface.
- 'All points are outliers.'

Bizarre Shape of High-Dimensional Balls

What fraction of the small cubes are visible on the surface of the larger 10 × 10 × 10 cube?

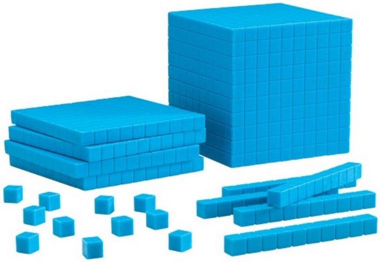
1/2



$\frac{8^3}{10^3} =$

Bizarre Shape of High-Dimensional Balls

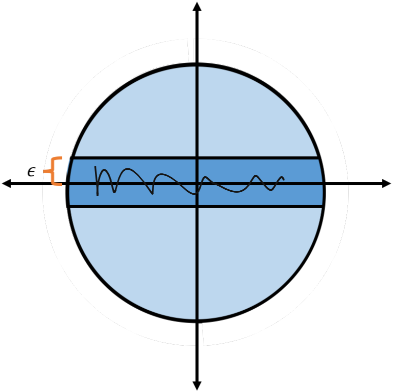
What fraction of the small cubes are visible on the surface of the larger 10 × 10 × 10 cube?



$$\frac{\underline{10}^3 - \underline{8}^3}{\underline{10}^3} = \frac{1000 - 512}{1000} = \underline{.488.}$$

Bizarre Shape of High-Dimensional Balls

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

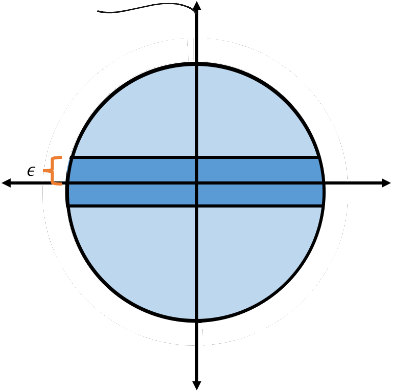


$\frac{1}{\sqrt{d}}$

Formally: volume of set $S = \{x \in \mathcal{B}_d : |x(1)| \leq \epsilon\}$.

Bizarre Shape of High-Dimensional Balls

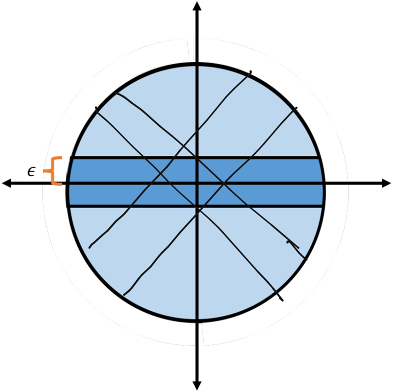
What percentage of the volume of $\mathcal{B}_d$ falls within ϵ distance of its equator? Answer: all but a $2^{\Theta(-\epsilon^2 d)}$ fraction.



Formally: volume of set $S = \{x \in \mathcal{B}_d : |x(1)| \leq \epsilon\}$.

Bizarre Shape of High-Dimensional Balls

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



Formally: volume of set $S = \{x \in \mathcal{B}_d : |x(1)| \leq \epsilon\}$.

By symmetry, all but a $2^{\Theta(-\epsilon^2 d)}$ fraction of the volume falls within ϵ of any equator! $S = \{x \in \mathcal{B}_d : |\langle x, t\rangle| \leq \epsilon\}$

Bizarre Shape of High-Dimensional Balls

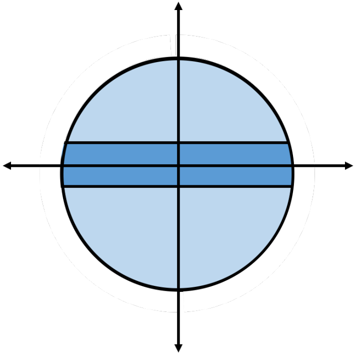
Claim 1: All but a $2^{\Theta(-\epsilon^2 d)}$ fraction of the volume of a ball falls within ϵ of any equator.

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

Bizarre Shape of High-Dimensional Balls

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

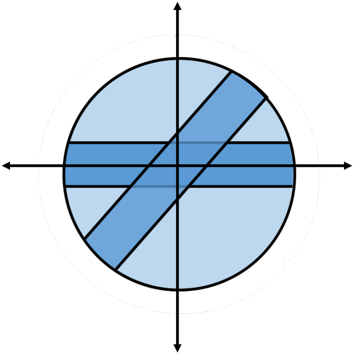
Claim 2: All but a $2^{\Theta(-\epsilon d)}$ fraction falls within ϵ of its surface.



Bizarre Shape of High-Dimensional Balls

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

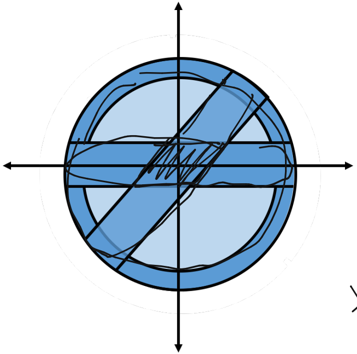
Claim 2: All but a $2^{\Theta(-\epsilon d)}$ fraction falls within ϵ of its surface.



Bizarre Shape of High-Dimensional Balls

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

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



$[x(1)\ x(2) \ldots x(d)]$

$\sum x(i)^2 = 1$

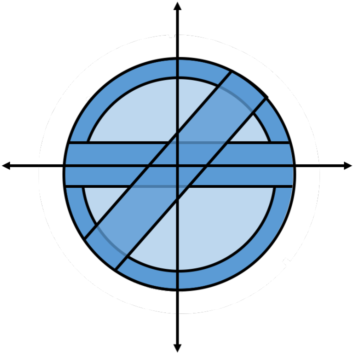
$x(i)^2 = \frac{1}{d}$ on avg

$x(i) \approx \frac{1}{\sqrt{d}}$

Bizarre Shape of High-Dimensional Balls

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

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

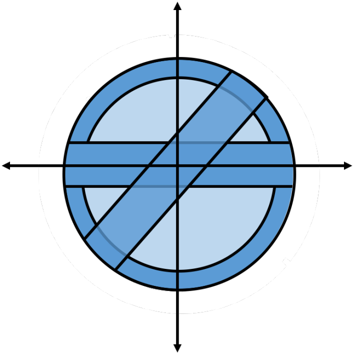


How is this possible?

Bizarre Shape of High-Dimensional Balls

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

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



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

Take-aways

- High-dimensional space behaves very differently from low-dimensional space.
- Random projection (i.e., the JL Lemma) reduces to a much lower-dimensional space that is still large enough to capture this behavior on a subset of n points.
- Need to be careful when using low-dimensional intuition for high-dimensional vectors.
- Need to be careful when modeling data as random vectors in high-dimensions.

Additional Material

Concentration of Volume At Equator

Claim: All but a $2^{\Theta(-\epsilon^2 d)}$ fraction of the volume of a ball falls within ϵ of its equator. I.e., in $S = \{x \in \mathcal{B}_d : |x(1)| \leq \epsilon\}$.

Concentration of Volume At Equator

Claim: All but a $2^{\Theta(-\epsilon^2 d)}$ fraction of the volume of a ball falls within ϵ of its equator. I.e., in $S = \{x \in \mathcal{B}_d : |x(1)| \leq \epsilon\}$.

Proof Sketch:

- Let x have independent Gaussian $\mathcal{N}(0, 1)$ entries and let $\bar{x} = \frac{x}{\|x\|_2}$. $\bar{x}$ is selected uniformly at random from the surface of the ball.

Concentration of Volume At Equator

Claim: All but a $2^{\Theta(-\epsilon^2 d)}$ fraction of the volume of a ball falls within ϵ of its equator. I.e., in $S = \{x \in \mathcal{B}_d : |x(1)| \leq \epsilon\}$.

Proof Sketch:

- Let x have independent Gaussian $\mathcal{N}(0, 1)$ entries and let $\bar{x} = \frac{x}{\|x\|_2}$. $\bar{x}$ is selected uniformly at random from the surface of the ball.
- Suffices to show that $\Pr[|\bar{x}(1)| > \epsilon] \leq 2^{\Theta(-\epsilon^2 d)}$. Why?

Concentration of Volume At Equator

Claim: All but a $2^{\Theta(-\epsilon^2 d)}$ fraction of the volume of a ball falls within ϵ of its equator. I.e., in $S = \{x \in \mathcal{B}_d : |x(1)| \leq \epsilon\}$.

Proof Sketch:

- Let x have independent Gaussian $\mathcal{N}(0,1)$ entries and let $\bar{x} = \frac{x}{\|x\|_2}$. $\bar{x}$ is selected uniformly at random from the surface of the ball.
- Suffices to show that $\Pr[|\bar{x}(1)| > \epsilon] \leq 2^{\Theta(-\epsilon^2 d)}$. Why?
- $\bar{x}(1) = \frac{x(1)}{\|x\|_2}$. What is $\mathbb{E}[\|x\|_2^2]$?

Concentration of Volume At Equator

Claim: All but a $2^{\Theta(-\epsilon^2 d)}$ fraction of the volume of a ball falls within ϵ of its equator. I.e., in $S = \{x \in \mathcal{B}_d : |x(1)| \leq \epsilon\}$.

Proof Sketch:

- Let x have independent Gaussian $\mathcal{N}(0, 1)$ entries and let $\bar{x} = \frac{x}{\|x\|_2}$. $\bar{x}$ is selected uniformly at random from the surface of the ball.
- Suffices to show that $\Pr[|\bar{x}(1)| > \epsilon] \leq 2^{\Theta(-\epsilon^2 d)}$. Why?
- $\bar{x}(1) = \frac{x(1)}{\|x\|_2}$. $\mathbb{E}[\|x\|_2^2] = \sum_{i=1}^{d} \mathbb{E}[x(i)^2] = d$.

Concentration of Volume At Equator

Claim: All but a $2^{\Theta(-\epsilon^2 d)}$ fraction of the volume of a ball falls within ϵ of its equator. I.e., in $S = \{x \in \mathcal{B}_d : |x(1)| \leq \epsilon\}$.

Proof Sketch:

- Let x have independent Gaussian $\mathcal{N}(0,1)$ entries and let $\bar{x} = \frac{x}{\|x\|_2}$. $\bar{x}$ is selected uniformly at random from the surface of the ball.
- Suffices to show that $\Pr[|\bar{x}(1)| > \epsilon] \leq 2^{\Theta(-\epsilon^2 d)}$. Why?
- $\bar{x}(1) = \frac{x(1)}{\|x\|_2}$. $\mathbb{E}[\|x\|_2^2] = \sum_{i=1}^{d} \mathbb{E}[x(i)^2] = d$. $\Pr[\|x\|_2^2 \leq d/2] \leq 2^{-\Theta(d)}$

Concentration of Volume At Equator

Claim: All but a $2^{\Theta(-\epsilon^2 d)}$ fraction of the volume of a ball falls within ϵ of its equator. I.e., in $S = \{x \in \mathcal{B}_d : |x(1)| \leq \epsilon\}$.

Proof Sketch:

- Let x have independent Gaussian $\mathcal{N}(0,1)$ entries and let $\bar{x} = \frac{x}{\|x\|_2}$. $\bar{x}$ is selected uniformly at random from the surface of the ball.
- Suffices to show that $\Pr[|\bar{x}(1)| > \epsilon] \leq 2^{\Theta(-\epsilon^2 d)}$. Why?
- $\bar{x}(1) = \frac{x(1)}{\|x\|_2}$. $\mathbb{E}[\|x\|_2^2] = \sum_{i=1}^d \mathbb{E}[x(i)^2] = d$. $\Pr[\|x\|_2^2 \leq d/2] \leq 2^{-\Theta(d)}$
- Conditioning on $\|x\|_2^2 \geq d/2$, since $x(1)$ is normally distributed,

$$\Pr[|\bar{x}(1)| > \epsilon] = \Pr[|x(1)| > \epsilon \cdot \|x\|_2]$$

Concentration of Volume At Equator

Claim: All but a $2^{\Theta(-\epsilon^2 d)}$ fraction of the volume of a ball falls within ϵ of its equator. I.e., in $S = \{x \in \mathcal{B}_d : |x(1)| \leq \epsilon\}$.

Proof Sketch:

- Let x have independent Gaussian $\mathcal{N}(0,1)$ entries and let $\bar{x} = \frac{x}{\|x\|_2}$. $\bar{x}$ is selected uniformly at random from the surface of the ball.
- Suffices to show that $\Pr[|\bar{x}(1)| > \epsilon] \leq 2^{\Theta(-\epsilon^2 d)}$. Why?
- $\bar{x}(1) = \frac{x(1)}{\|x\|_2}$. $\mathbb{E}[\|x\|_2^2] = \sum_{i=1}^{d} \mathbb{E}[x(i)^2] = d$. $\Pr[\|x\|_2^2 \leq d/2] \leq 2^{-\Theta(d)}$
- Conditioning on $\|x\|_2^2 \geq d/2$, since $x(1)$ is normally distributed,

$$\Pr[|\bar{x}(1)| > \epsilon] = \Pr[|x(1)| > \epsilon \cdot \|x\|_2]$$

$$\leq Pr[|x(1)| > \epsilon \cdot \sqrt{d/2}]$$

Concentration of Volume At Equator

Claim: All but a $2^{\Theta(-\epsilon^2 d)}$ fraction of the volume of a ball falls within ϵ of its equator. I.e., in $S = \{x \in \mathcal{B}_d : |x(1)| \leq \epsilon\}$.

Proof Sketch:

- Let x have independent Gaussian $\mathcal{N}(0,1)$ entries and let $\bar{x} = \frac{x}{\|x\|_2}$. $\bar{x}$ is selected uniformly at random from the surface of the ball.
- Suffices to show that $\Pr[|\bar{x}(1)| > \epsilon] \leq 2^{\Theta(-\epsilon^2 d)}$. Why?
- $\bar{x}(1) = \frac{x(1)}{\|x\|_2}$. $\mathbb{E}[\|x\|_2^2] = \sum_{i=1}^d \mathbb{E}[x(i)^2] = d$. $\Pr[\|x\|_2^2 \leq d/2] \leq 2^{-\Theta(d)}$
- Conditioning on $\|x\|_2^2 \geq d/2$, since $x(1)$ is normally distributed,

$$\Pr[|\bar{x}(1)| > \epsilon] = \Pr[|x(1)| > \epsilon \cdot \|x\|_2]$$
$$\leq Pr[|x(1)| > \epsilon \cdot \sqrt{d/2}] = 2^{\Theta(-(\epsilon\sqrt{d/2})^2)} = 2^{\Theta(-\epsilon^2 d)}.$$

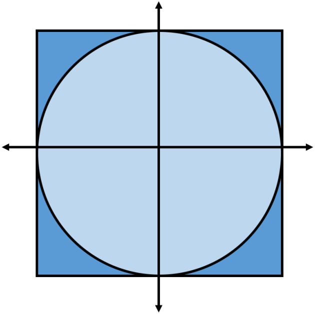
High-Dimensional Cubes

Let $\mathcal{C}_d$ be the d-dimensional cube: $\mathcal{C}_d = \{x \in \mathbb{R}^d : \underline{|x(i)|} \leq 1 \; \forall \; i\}$.

High-Dimensional Cubes

Let $\mathcal{C}_d$ be the d-dimensional cube: $\mathcal{C}_d = \{x \in \mathbb{R}^d : |x(i)| \leq 1 \; \forall \; i\}$.

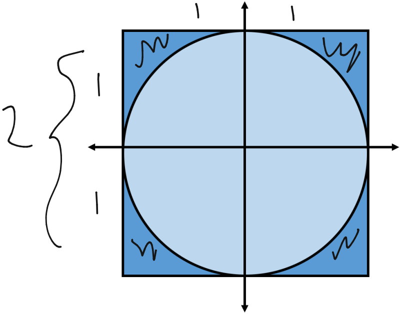
In low-dimensions, the cube is not that different from the ball.



High-Dimensional Cubes

Let $\mathcal{C}_d$ be the d-dimensional cube: $\mathcal{C}_d = \{x \in \mathbb{R}^d : |x(i)| \leq 1 \; \forall \; i\}$.

In low-dimensions, the cube is not that different from the ball.

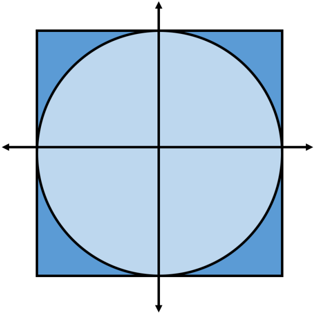


But volume of $\mathcal{C}_d$ is $\underline{2^d}$ while volume of $\mathcal{B}^d$ is $\underline{\frac{\pi^{\frac{d}{2}}}{(d/2)!}} = \underline{\frac{1}{d^{\Theta(d)}}}$. A huge gap!

High-Dimensional Cubes

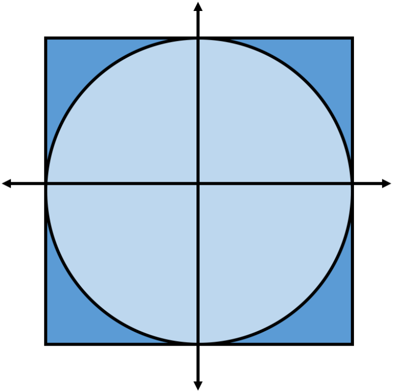
Let $\mathcal{C}_d$ be the d-dimensional cube: $\mathcal{C}_d = \{x \in \mathbb{R}^d : |x(i)| \leq 1 \, \forall \, i\}$.

In low-dimensions, the cube is not that different from the ball.

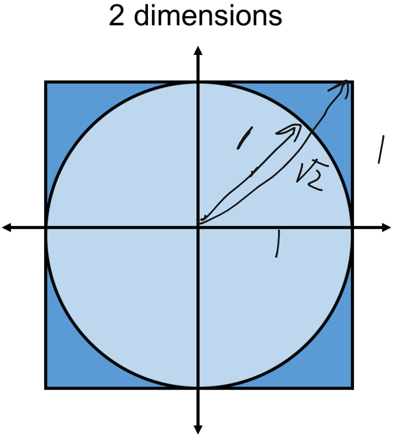


But volume of $\mathcal{C}_d$ is 2^d while volume of $\mathcal{B}^d$ is $\frac{\pi^{\frac{d}{2}}}{(d/2)!} = \frac{1}{d^{\Theta(d)}}$. A huge gap! So something is very different about these shapes...

High-Dimensional Cubes


2 dimensions

High-Dimensional Cubes



Corners of cube are $\sqrt{d}$ times further away from the origin than the surface of the ball.

High-Dimensional Cubes

in m dimensions

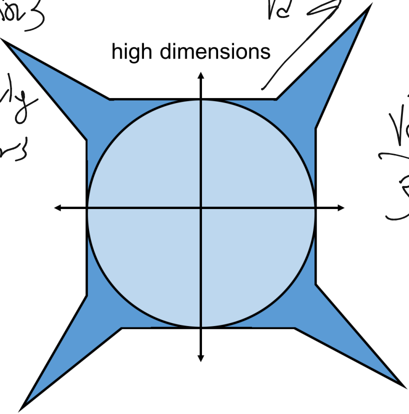
are $\geq 2^{\varepsilon^2 m}$ nearly

orth. unit vectors

$n \leq 2^{\varepsilon^2 m}$

n n.o.v.

$m \geq \frac{\log n}{\varepsilon^2}$



$\sqrt{d}$

$\boxed{\sqrt{d}}$

$\frac{\sqrt{d}}{3}$

Corners of cube are $\sqrt{d}$ times further away from the origin than the surface of the ball.

High-Dimensional Cubes

Data generated from the ball $\mathcal{B}_d$ will behave very differently than data generated from the cube $\mathcal{C}_d$.

High-Dimensional Cubes

Data generated from the ball $\mathcal{B}_d$ will behave very differently than data generated from the cube $\mathcal{C}_d$.

- $x \sim \mathcal{B}_d$ has $\|x\|_2^2 \leq 1$.
- $x \sim \mathcal{C}_d$ has $\mathbb{E}[\|x\|_2^2] = ?$,

High-Dimensional Cubes

Data generated from the ball $\mathcal{B}_d$ will behave very differently than data generated from the cube $\mathcal{C}_d$.

- $x \sim \mathcal{B}_d$ has $\|x\|_2^2 \leq 1$.
- $x \sim \mathcal{C}_d$ has $\mathbb{E}[\|x\|_2^2] = d/3$,

High-Dimensional Cubes

Data generated from the ball $\mathcal{B}_d$ will behave very differently than data generated from the cube $\mathcal{C}_d$.

- $x \sim \mathcal{B}_d$ has $\|x\|_2^2 \leq 1$.
- $x \sim \mathcal{C}_d$ has $\mathbb{E}[\|x\|_2^2] = d/3$, and $\Pr[\|x\|_2^2 \leq d/6] \leq 2^{-\Theta(d)}$.

High-Dimensional Cubes

Data generated from the ball $\mathcal{B}_d$ will behave very differently than data generated from the cube $\mathcal{C}_d$.

- $x \sim \mathcal{B}_d$ has $\|x\|_2^2 \leq 1$.
- $x \sim \mathcal{C}_d$ has $\mathbb{E}[\|x\|_2^2] = d/3$, and $\Pr[\|x\|_2^2 \leq d/6] \leq 2^{-\Theta(d)}$.
- Almost all the volume of the unit cube falls far away from the origin – i.e., far outside the unit ball.