COMPSCI 514: ALGORITHMS FOR DATA SCIENCE

Cameron Musco

University of Massachusetts Amherst. Fall 2019.

Lecture 22

LOGISTICS

- Problem Set 4 released last night. Due Sunday 12/15 at 8pm.
- Final Exam Thursday 12/19 at 10:30am in Thompson 104.
- Exam prep materials (list of topics, practice problems) coming in next couple of days.

Before Break:

- Finished discussion of SGD.
- Gradient descent and SGD as applied to least squares regression.

Before Break:

- Finished discussion of SGD.
- Gradient descent and SGD as applied to least squares regression.

This Class:

- A quick tour of the counterintuitive properties of high-dimensional space.
- Many connections to concentration inequalities.
- Implications for working with high-dimensional data (curse of dimensionality).

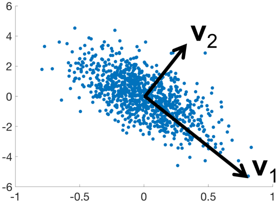
Modern data analysis often involves very high-dimensional data points.

- Websites record (tens of) thousands of measurements per user: who they follow, when they visit the site, timestamps for specific iteractions, etc.
- A 3 minute, 500×500 pixel video clip at 15 FPS has ≥ 2 billion pixel values.
- The human genome has 3 billion+ base pairs.

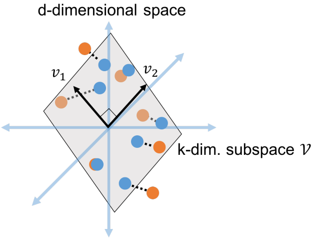
Modern data analysis often involves very high-dimensional data points.

- Websites record (tens of) thousands of measurements per user: who they follow, when they visit the site, timestamps for specific iteractions, etc.
- A 3 minute, 500×500 pixel video clip at 15 FPS has ≥ 2 billion pixel values.
- The human genome has 3 billion+ base pairs.

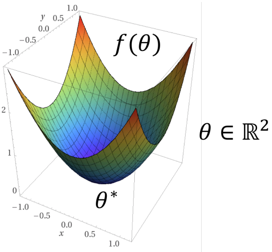
Typically when discussing algorithm design we imagine data in much lower (usually 3) dimensional space.



6
4
2
0
-2
-4
-6
-1
-0.5
0
0.5
1
$\mathbf{v}_2$
$\mathbf{v}_1$

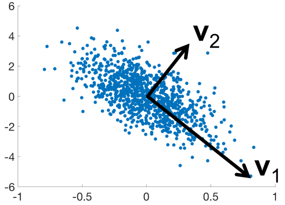


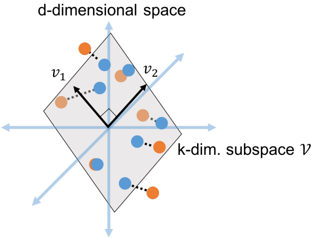
d-dimensional space
v_1
v_2
k-dim. subspace $\mathcal{V}$

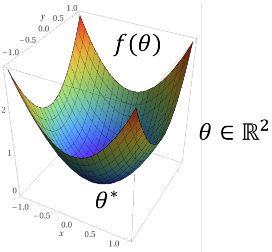


$f(\theta)$
θ^*
$\theta \in \mathbb{R}^2$

LOW-DIMENSIONAL INTUITION







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

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

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

NEARLY ORTHOGONAL VECTORS

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

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 $\epsilon = .01$)

1. d
2. $\Theta(d)$
3. $\Theta(d^2)$
4. $2^{\Theta(d)}$

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

1. d
2. $\Theta(d)$
3. $\Theta(d^2)$
4. $2^{\Theta(d)}$

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

1. d
2. $\Theta(d)$
3. $\Theta(d^2)$
4. $2^{\Theta(d)}$

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

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 $\epsilon = .01$)

1. d
2. $\Theta(d)$
3. $\Theta(d^2)$
4. $2^{\Theta(d)}$

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

Proof: Let $x_1, \ldots, x_t$ each have independent random entries set to $\pm 1/\sqrt{d}$.

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

1. d
2. $\Theta(d)$
3. $\Theta(d^2)$
4. $2^{\Theta(d)}$

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

Proof: Let $x_1, \ldots, x_t$ each have independent random entries set to $\pm 1/\sqrt{d}$.

- x_i is always a unit vector.

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 $\epsilon = .01$)

1. d
2. $\Theta(d)$
3. $\Theta(d^2)$
4. $2^{\Theta(d)}$

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

Proof: Let $x_1, \ldots, x_t$ each have independent random entries set to $\pm 1/\sqrt{d}$.

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

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 $\epsilon = .01$)

1. d
2. $\Theta(d)$
3. $\Theta(d^2)$
4. $2^{\Theta(d)}$

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

Proof: Let $x_1, \ldots, x_t$ each have independent random entries set to $\pm 1/\sqrt{d}$.

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

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 $\epsilon = .01$)

1. d
2. $\Theta(d)$
3. $\Theta(d^2)$
4. $2^{\Theta(d)}$

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

Proof: Let $x_1, \ldots, x_t$ each have independent random entries set to $\pm 1/\sqrt{d}$.

- x_i is always a unit vector.
- $\mathbb{E}[\langle x_i, x_j \rangle] = 0$.
- By a Chernoff bound, $\Pr[|\langle x_i, x_j \rangle| \geq \epsilon] \leq 2e^{-\epsilon^2 d/3}$.

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 $\epsilon = .01$)

1. d 2. $\Theta(d)$ 3. $\Theta(d^2)$ 4. $2^{\Theta(d)}$

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

Proof: Let $x_1, \ldots, x_t$ each have independent random entries set to $\pm 1/\sqrt{d}$.

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

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

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

$$\|x_i - x_j\|_2^2$$

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

$$\|x_i - x_j\|_2^2 = \|x_i\|_2^2 + \|x_j\|_2^2 - 2x_i^T x_j$$

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 $\epsilon = .01$)

$$\|x_i - x_j\|_2^2 = \|x_i\|_2^2 + \|x_j\|_2^2 - 2x_i^T x_j \geq 1.98.$$

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

$$\|x_i - x_j\|_2^2 = \|x_i\|_2^2 + \|x_j\|_2^2 - 2x_i^T x_j \geq 1.98.$$

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

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

$$\|x_i - x_j\|_2^2 = \|x_i\|_2^2 + \|x_j\|_2^2 - 2x_i^T x_j \geq 1.98.$$

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

- Can make methods like k-nearest neighbor classification or kernel regression useless.

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 $\epsilon = .01$)

$$\|x_i - x_j\|_2^2 = \|x_i\|_2^2 + \|x_j\|_2^2 - 2x_i^T x_j \geq 1.98.$$

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

- Can make methods like k-nearest neighbor classification or kernel regression useless.

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

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

$$\|x_i - x_j\|_2^2 = \|x_i\|_2^2 + \|x_j\|_2^2 - 2x_i^T x_j \geq 1.98.$$

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

- Can make methods like k-nearest neighbor classification or kernel regression useless.

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

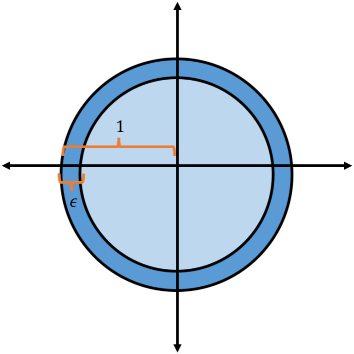
- Only hope is if we have strong low-dimensional structure.

BIZARRE SHAPE OF HIGH-DIMENSIONAL BALLS

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

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

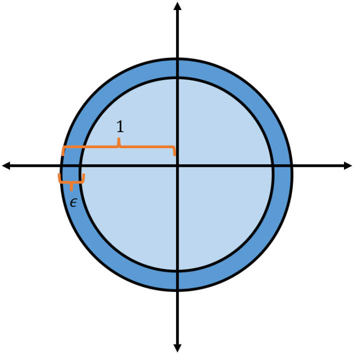
What percentage of the volume of $\mathcal{B}_d$ falls within ϵ distance of its surface?



BIZARRE SHAPE OF HIGH-DIMENSIONAL BALLS

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

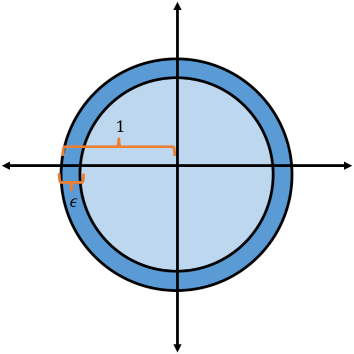
What percentage of the volume of $\mathcal{B}_d$ falls within ϵ distance of its surface?



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

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 $(1-\epsilon)^d \leq e^{-\epsilon d}$ fraction. Exponentially small in the dimension d!

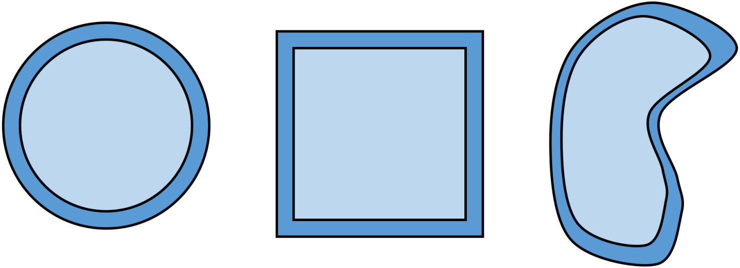


Volume of a radius R ball is $\frac{\pi^{\frac{d}{2}}}{(d/2)!} \cdot R^d$.

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

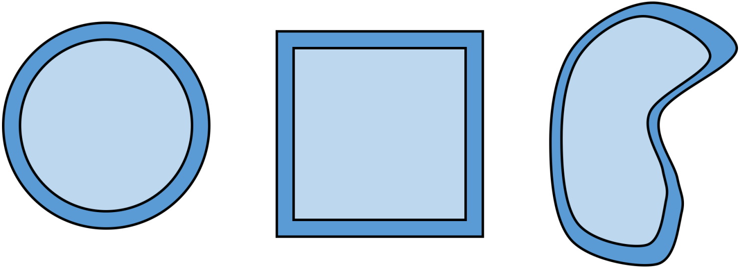
All but an $e^{-\epsilon d}$ fraction of a unit ball's volume is within ϵ of its surface.

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



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

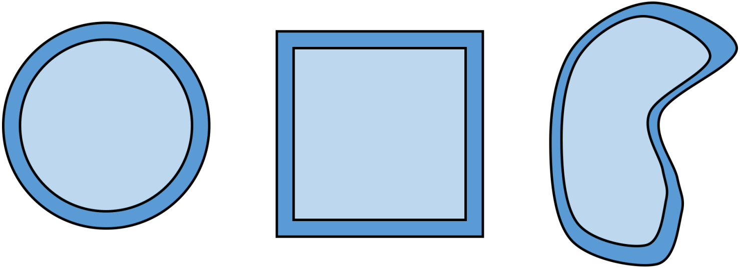
- **Isoperimetric inequality**: the ball has the maximum surface area/volume ratio of any shape.



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

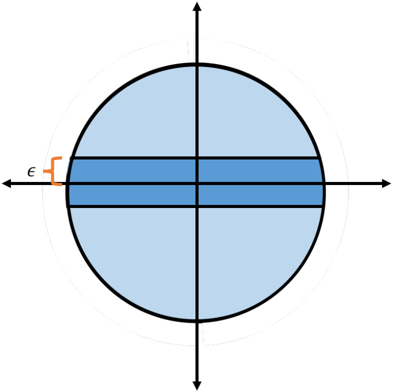
All but an $e^{-\epsilon d}$ fraction of a unit ball's volume is within ϵ of its surface.

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



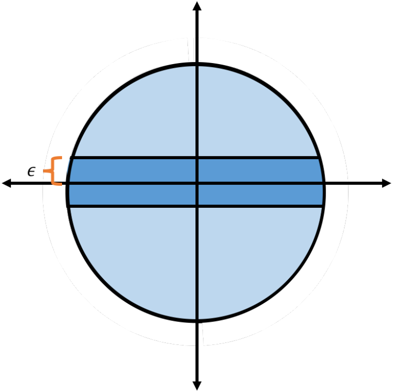
- If we randomly sample points from any high-dimensional shape, nearly all will fall near its surface.
- 'All points are outliers.'

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



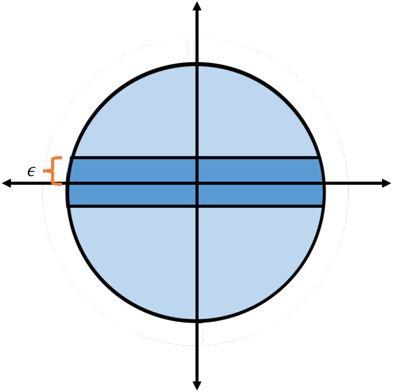
Formally: volume of set $S = \{x \in \mathcal{B}_d : |x(1)| \leq \epsilon\}$.

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

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

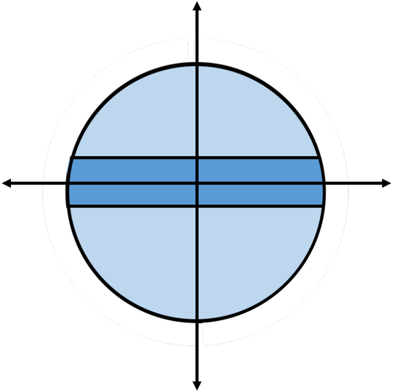
By symmetry, all but a $2^{\Theta(-\epsilon^2 d)}$ fraction of the volume falls within ϵ of any equator! $S = \{x \in \mathcal{B}_d : |\langle x, t\rangle| \leq \epsilon\}$

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.

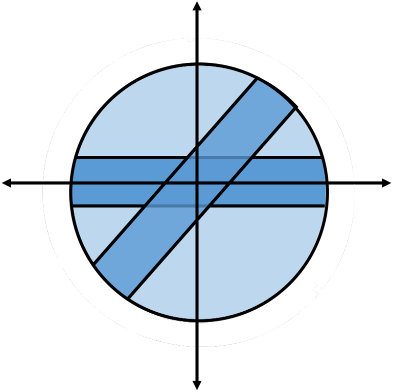
Claim 1: All but a $2^{\Theta(-\epsilon^2 d)}$ fraction of the volume of a ball falls within ϵ of any equator.

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



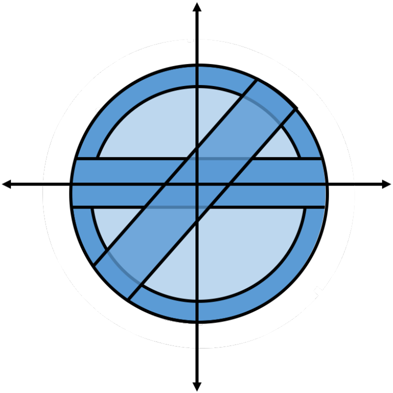
Claim 1: All but a $2^{\Theta(-\epsilon^2 d)}$ fraction of the volume of a ball falls within ϵ of any equator.

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



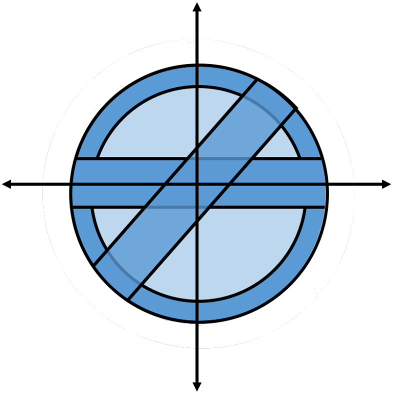
Claim 1: All but a $2^{\Theta(-\epsilon^2 d)}$ fraction of the volume of a ball falls within ϵ of any equator.

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



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

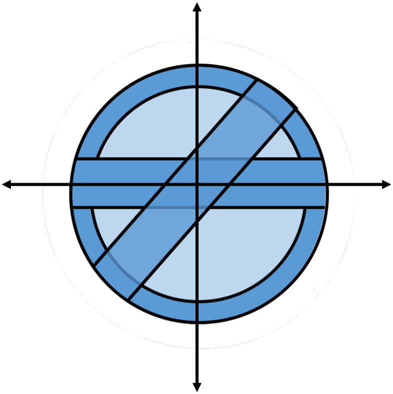
Claim 2: All but a $2^{\Theta(-\epsilon d)}$ fraction falls within ϵ of its surface.



How is this possible?

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!

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

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

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 entries set to independent Gaussians $\mathcal{N}(0,1)$ 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?

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 entries set to independent Gaussians $\mathcal{N}(0, 1)$ 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 $\mathbb{E}[\|x\|_2^2]$?

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 entries set to independent Gaussians $\mathcal{N}(0,1)$ 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?
- $\bar{x}(1) = \frac{x(1)}{\|x\|_2}$. $\mathbb{E}[\|x\|_2^2] = \sum_{i=1}^d \mathbb{E}[x(i)^2] = d$.

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 entries set to independent Gaussians $\mathcal{N}(0,1)$ 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 entries set to independent Gaussians $\mathcal{N}(0,1)$ 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,

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

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 entries set to independent Gaussians $\mathcal{N}(0,1)$ 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,

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

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:

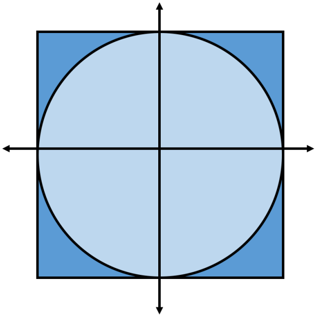
- Let x have entries set to independent Gaussians $\mathcal{N}(0,1)$ and let $\bar{x} = \frac{x}{\|x\|_2}$. $\bar{x}$ is selected uniformly at random from the surface of the ball.
- Suffices to show that $\Pr[|\bar{x}(1)| > \epsilon] \leq 2^{\Theta(-\epsilon^2 d)}$. Why?
- $\bar{x}(1) = \frac{x(1)}{\|x\|_2}$.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

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

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

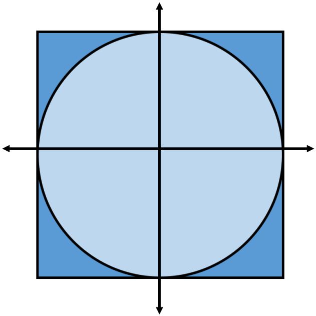
Let $\mathcal{C}_d$ be the d-dimensional cube: $\mathcal{C}_d = \{x \in \mathbb{R}^d : |x(i)| \leq 1 \; \forall \; i\}$.

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



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

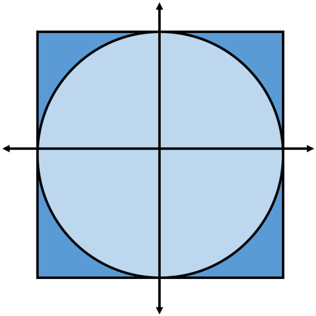
In low-dimensions, the cube is not that different from the ball.



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!

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...

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

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 $\mathbb{E}[\|x\|_2^2] = d/3$,

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

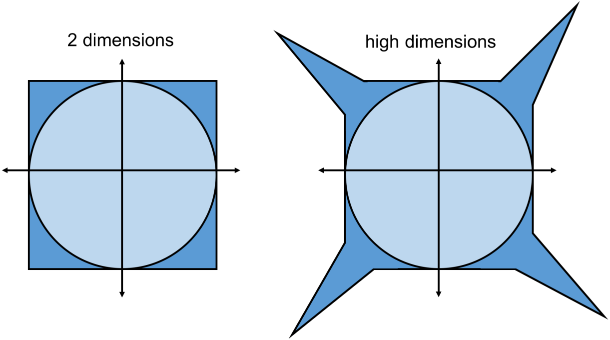
- $x \sim \mathcal{B}_d$ has $\|x\|_2^2 \leq 1$.
- $x \sim \mathcal{C}_d$ has $\mathbb{E}[\|x\|_2^2] = d/3$, and $\Pr[\|x\|_2^2 \leq d/6] \leq 2^{-\Theta(d)}$.

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 in its corners, and these corners lie far outside the unit ball.

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 in its corners, and these corners lie far outside the unit ball.



If high-dimensional geometry is so different from low-dimensional geometry, how is dimensionality reduction (e.g., the Johnson-Lindenstrauss lemma) possible?

If high-dimensional geometry is so different from low-dimensional geometry, how is dimensionality reduction (e.g., the Johnson-Lindenstrauss lemma) possible?

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 $x_1, \ldots, x_n \in \mathbb{R}^d$ with high probability, for all i, j:

$$(1-\epsilon)\|x_i - x_j\|_2 \leq \|\mathbf{\Pi} x_i - \mathbf{\Pi} x_j\|_2 \leq (1+\epsilon)\|x_i - x_j\|_2.$$

If high-dimensional geometry is so different from low-dimensional geometry, how is dimensionality reduction (e.g., the Johnson-Lindenstrauss lemma) possible?

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 $x_1, \ldots, x_n \in \mathbb{R}^d$ with high probability, for all i, j:

$$(1-\epsilon)\|x_i - x_j\|_2 \leq \|\mathbf{\Pi} x_i - \mathbf{\Pi} x_j\|_2 \leq (1+\epsilon)\|x_i - x_j\|_2.$$

If $x_1, \ldots, x_n$ are random unit vectors in d-dimensions, can show that $\mathbf{\Pi} x_1, \ldots, \mathbf{\Pi} x_n$ are essentially random unit vectors in m-dimensions.

If high-dimensional geometry is so different from low-dimensional geometry, how is dimensionality reduction (e.g., the Johnson-Lindenstrauss lemma) possible?

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 $x_1, \ldots, x_n \in \mathbb{R}^d$ with high probability, for all i, j:

$$(1-\epsilon)\|x_i - x_j\|_2 \leq \|\mathbf{\Pi} x_i - \mathbf{\Pi} x_j\|_2 \leq (1+\epsilon)\|x_i - x_j\|_2.$$

If $x_1, \ldots, x_n$ are random unit vectors in d-dimensions, can show that $\mathbf{\Pi} x_1, \ldots, \mathbf{\Pi} x_n$ are essentially random unit vectors in m-dimensions.

But these different dimensional spaces have very different geometries, so how is this possible?

$x_1, \ldots, x_n$ are sampled from the surface of $\mathcal{B}_d$ and $\mathbf{\Pi}x_1, \ldots, \mathbf{\Pi}x_n$ are (approximately) sampled from the surface of $\mathcal{B}_m$.

$x_1, \ldots, x_n$ are sampled from the surface of $\mathcal{B}_d$ and $\mathbf{\Pi}x_1, \ldots, \mathbf{\Pi}x_n$ are (approximately) sampled from the surface of $\mathcal{B}_m$.

- In d dimensions, $2^{\epsilon^2 d}$ random unit vectors will have all pairwise dot products at most ϵ with high probability

$x_1, \ldots, x_n$ are sampled from the surface of $\mathcal{B}_d$ and $\mathbf{\Pi} x_1, \ldots, \mathbf{\Pi} x_n$ are (approximately) sampled from the surface of $\mathcal{B}_m$.

- In d dimensions, $2^{\epsilon^2 d}$ random unit vectors will have all pairwise dot products at most ϵ with high probability
- After JL projection, $\mathbf{\Pi} x_1, \ldots, \mathbf{\Pi} x_n$ will still have pairwise dot products at most $O(\epsilon)$ with high probability.

$x_1, \ldots, x_n$ are sampled from the surface of $\mathcal{B}_d$ and $\mathbf{\Pi} x_1, \ldots, \mathbf{\Pi} x_n$ are (approximately) sampled from the surface of $\mathcal{B}_m$.

- In d dimensions, $2^{\epsilon^2 d}$ random unit vectors will have all pairwise dot products at most ϵ with high probability
- After JL projection, $\mathbf{\Pi} x_1, \ldots, \mathbf{\Pi} x_n$ will still have pairwise dot products at most $O(\epsilon)$ with high probability.
- In $m = O\left(\frac{\log n}{\epsilon^2}\right)$ dimensions, $2^{\epsilon^2 m} = 2^{O(\log n)} >> n$ random unit vectors will have all pairwise dot products at most ϵ with high probability.

$x_1, \ldots, x_n$ are sampled from the surface of $\mathcal{B}_d$ and $\mathbf{\Pi} x_1, \ldots, \mathbf{\Pi} x_n$ are (approximately) sampled from the surface of $\mathcal{B}_m$.

- In d dimensions, $2^{\epsilon^2 d}$ random unit vectors will have all pairwise dot products at most ϵ with high probability
- After JL projection, $\mathbf{\Pi} x_1, \ldots, \mathbf{\Pi} x_n$ will still have pairwise dot products at most $O(\epsilon)$ with high probability.
- In $m = O\left(\frac{\log n}{\epsilon^2}\right)$ dimensions, $2^{\epsilon^2 m} = 2^{O(\log n)} >> n$ random unit vectors will have all pairwise dot products at most ϵ with high probability.
- m is chosen just large enough so that the odd geometry of d-dimensional space will still hold on the n points in question.

Questions?