CS-GY 6763: Lecture 3
High Dimensional Geometry, the Johnson-Lindenstrauss Lemma

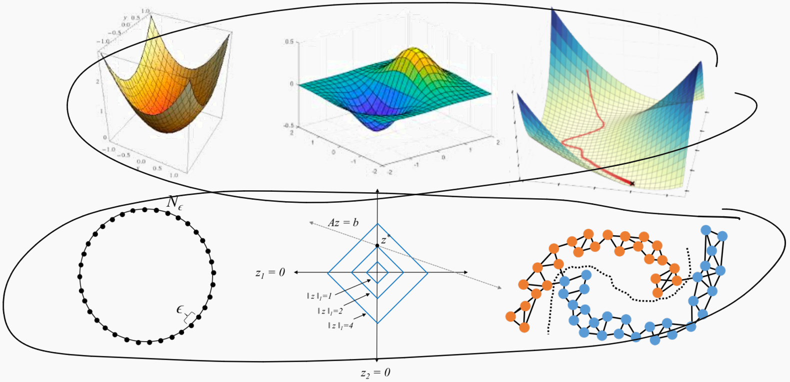
NYU Tandon School of Engineering, Prof. Christopher Musco

UNIFYING THEME OF THE COURSE

How do we deal with data (vectors) in high dimensions?

- Locality sensitive hashing for similarity search.
- Iterative methods for optimizing functions that depend on many variables.
- SVD + low-rank approximation to find and visualize low-dimensional structure.
- Convert large graphs to high dimensional vector data.

Often visualize data and algorithms in 1,2, or 3 dimensions.

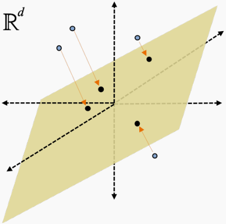


First part of lecture: Prove that high-dimensional space looks very different from low-dimensional space. These images are rarely very informative!

Second part of lecture: Ignore our own advice.

Learn about **sketching, aka dimensionality reduction** techniques that seek to approximate high-dimensional vectors with much lower dimensional vectors.

- Johnson-Lindenstrauss lemma for ℓ_2 space.
- MinHash for binary vectors.

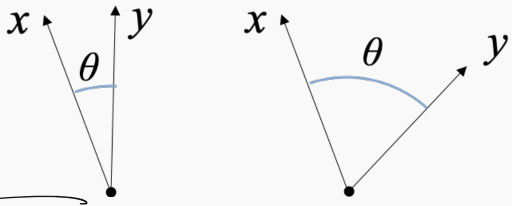


First part of lecture should help you understand the potential and limitations of these methods.

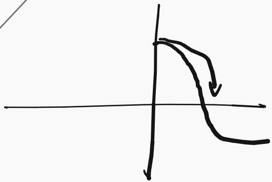
ORTHOGONAL VECTORS

Recall the inner product between two d dimensional vectors:

$$\langle \mathbf{x}, \mathbf{y} \rangle = \mathbf{x}^T\mathbf{y} = \mathbf{y}^T\mathbf{x} = \sum_{i=1}^{d} x[i]y[i]$$



$$\|x\|_2 = \sqrt{\sum_{i=1}^{d} x[i]^2}$$

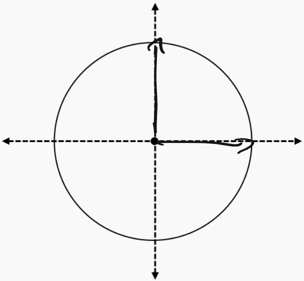


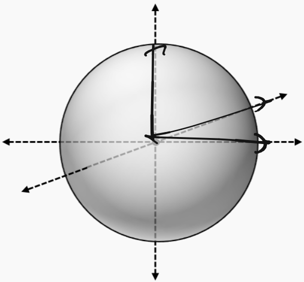
$$\langle \mathbf{x}, \mathbf{y} \rangle = \cos(\theta) \cdot \|\mathbf{x}\|_2 \cdot \|\mathbf{y}\|_2$$

normal

What is the largest set of **mutually ~~orthogonal~~** unit vectors $\mathbf{x}_1, \ldots, \mathbf{x}_t$ in d-dimensional space? I.e. with inner product

$$|\mathbf{x}_i^T \mathbf{x}_j| = 0 \text{ for all } i, j.$$





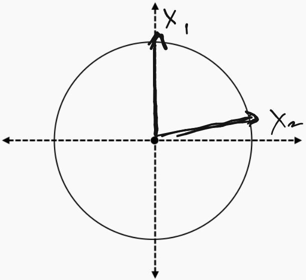
ORTHOGONAL VECTORS

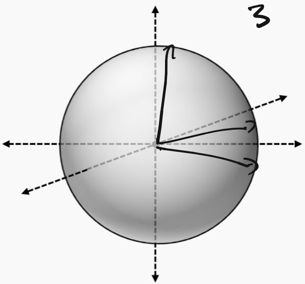
$\epsilon = .1$

normal

What is the largest set **nearly ortho~~gonal~~** unit vectors $\mathbf{x}_1, \ldots, \mathbf{x}_t$ in d-dimensional space. I.e., with inner product $\underline{|\mathbf{x}_i^T \mathbf{x}_j| \leq \epsilon}$ for all i, j.

2



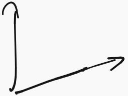


What is the largest set **nearly orthogonal** unit vectors $\mathbf{x}_1, \ldots, \mathbf{x}_t$ in d-dimensional space. I.e., with inner product $|\mathbf{x}_i^T \mathbf{x}_j| \leq \epsilon$ for all i, j.

$\downarrow$

$.1$

$|x_i^T x_j| = 0$



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

Claim: There is an exponential number (i.e., $\sim 2^d$) of nearly orthogonal unit vectors in d dimensional space.

Proof strategy: Use the Probabilistic Method! For $t = O(2^d)$, define a random process which generates random vectors $\mathbf{x}_1, \ldots, \mathbf{x}_t$ that are unlikely to have large inner product.

1. Claim that, with non-zero probability, $|\mathbf{x}_i^T \mathbf{x}_j| \leq \epsilon$ for all i, j.
2. Conclude that there must exists <u>some</u> set of t unit vectors with all pairwise inner-products bounded by ϵ.

$2^{\Omega(d)}$

Claim: There is an exponential number (i.e., $\sim 2^d$) of nearly orthogonal unit vectors in d dimensional space.

$\binom{t}{2} \approx t^2$

Proof: Let $\mathbf{x}_1, \ldots, \mathbf{x}_t$ all have independent random entries, each set to $\pm\frac{1}{\sqrt{d}}$ with equal probability.

$$x_i = \begin{bmatrix} +1/\sqrt{d} \\ +1/\sqrt{d} \\ -1/\sqrt{d} \\ \vdots \end{bmatrix}$$

- $\|\mathbf{x}_i\|_2 = \sqrt{\sum_{j=1}^{d} x_i[j]^2} = \sqrt{\sum_{j=1}^{d} 1/d} = \boxed{1}$
- $\mathbb{E}[\mathbf{x}_i^T\mathbf{x}_j] = \sum_{k=1}^{d} \underbrace{\mathbb{E}[x_i[k]]}_{0}\,\mathbb{E}[x_j[k]] = 0.$
- $\mathrm{Var}[\mathbf{x}_i^T\mathbf{x}_j] = 1/d$ $= \mathrm{Var}\left[\sum_{k=1}^{d} \underbrace{\overset{\pm 1/\sqrt{d}}{x_i[k]}\ \overset{\pm 1/\sqrt{d}}{x_j[k]}}\right]$

$= \sum_{k=1}^{d} \mathrm{Var}[C_k]$ $\qquad C_k = \begin{cases} +1/d & \text{w.p. } 1/2 \\ -1/d & \text{w.p. } 1/2 \end{cases}$

$\mathbb{E}[(C_k - 0)^2] = 1/d^2$

Let $Z = \mathbf{x}_i^T \mathbf{x}_j = \sum_{i=1}^{d} C_i$ where each C_i is $+\frac{1}{d}$ or $-\frac{1}{d}$ with equal probability.

$\mathbb{E}[Z] = 0$ $\quad$ $\mathrm{Var}[Z] = 1/d$ $\quad$ $\sigma = 1/\sqrt{d}$

Z is a sum of many i.i.d. random variables, so looks approximately Gaussian. Roughly, we expect that:

$|Z| > \varepsilon$

$\varphi = \varepsilon\sqrt{d}$

$$\Pr[|Z - \overset{=0}{\mathbb{E}Z}| \geq \alpha \cdot \overset{1/\sqrt{d}}{\sigma}] \leq O(e^{-\alpha^2})$$

$$\leq O(e^{-\varepsilon^2 d})$$

Note that we can transform to binary random variable: $(0, 1)$

$$Z = \sum_{i=1}^{d} C_i = \frac{2}{d} \sum_{i=1}^{d} \left(\frac{d}{2} \cdot C_i\right)$$

$+\frac{1}{2}$ $\quad$ $-1/2$ $\quad$ $= \{0, 1\} - 1/2$

$+1/d$

$-1/d$

$$= \frac{2}{d} \cdot \left(\sum_{i=1}^{d} B_i - 1/2\right) = \frac{2}{d} \cdot \left(-\frac{d}{2} - \sum_{i=1}^{d} B_i\right)$$

where each B_i is uniform in $\{0, 1\}$.

$Z = \mathbf{x}_i^T \mathbf{x}_j$.

$$Z = \frac{2}{d} \cdot \left(-\frac{d}{2} + \sum_{i=1}^{d} B_i \right)$$

where each B_i is uniform in $\{0, 1\}$.

$$\begin{aligned}
\Pr[|Z| \geq \epsilon] &= \Pr\left[\sum_{i=1}^{d} B \geq \frac{d}{2} + \frac{\epsilon d}{2}\right] + \Pr\left[\sum_{i=1}^{d} B \leq \frac{d}{2} - \frac{\epsilon d}{2}\right] \\
&= \Pr\left[\sum_{i=1}^{d} B \geq (1+\epsilon)\mathbb{E}[B]\right] + \Pr\left[\sum_{i=1}^{d} B \leq (1-\epsilon)\mathbb{E}[B]\right]
\end{aligned}$$

Theorem (Chernoff Bound)

Let $X_1, X_2, \ldots, X_d$ be independent $\{0,1\}$-valued random variables and let $S = \sum_{i=1}^{d} X_i$. We have for any $\epsilon < 1$:

$$\Pr[|S - \mathbb{E}[S]| \geq \epsilon \mathbb{E}[S]] \leq 2e^{\frac{-\epsilon^2 \mathbb{E}[S]}{3}}.$$

$$\Pr[|B - \mathbb{E}[B]| \geq \qquad] \leq$$

Formally, using a Chernoff bound:

$x_1, \ldots x_t$

$$\Pr[|Z - \mathbb{E}Z| \geq \epsilon] \leq 2e^{-\epsilon^2 d/6}$$

$\to 0$

$t = e^{O(\epsilon^2 d)}$

$= e^{O(d)}$

For any i, j pair, $\Pr[|\mathbf{x}_i^T \mathbf{x}_j| < \epsilon] \geq 1 - 2e^{-\epsilon^2 d/6}$.

By a union bound:

For all i, j pairs simultaneously, $\Pr[|\mathbf{x}_i^T \mathbf{x}_j| < \epsilon] \geq \underbrace{1 - \binom{t}{2} \cdot 2e^{-\epsilon^2 d/6}}_{\geq 0}$.

$\binom{t}{2} \cdot 2e^{-\epsilon^2 d/6} \leq 1$

$\frac{t \cdot (t-1)}{2} \leq \frac{1}{2} e^{\epsilon^2 d/6}$

$t^2 \leq e^{\epsilon^2 d/6}$

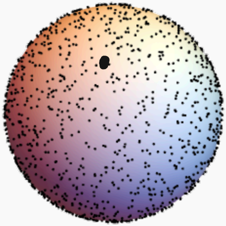
$\boxed{t \leq e^{\epsilon^2 d/12}}$

Final result: In d-dimensional space, there are $2^{\theta(\epsilon^2 d)}$ unit vectors with all pairwise inner products $\leq \epsilon$.

Corollary of proof: Random vectors tend to be far apart in high-dimensions.

$$\|x-y\|_2^2 = (x-y)^T(x-y) = x^Tx + y^Ty - 2x^Ty$$

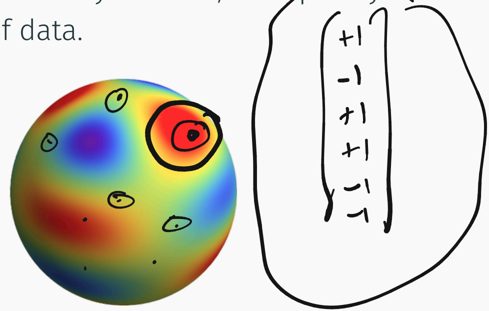
$$t = 2^{\theta(\epsilon^2 d)}$$



$$= 2 - 2x^Ty$$

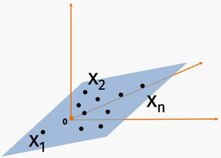
$$2 - \epsilon$$

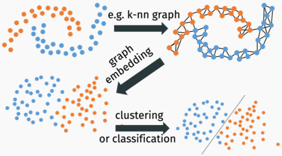
Curse of dimensionality: Suppose we want to use e.g. k-nearest neighbors to learn a function or classify points in $\mathbb{R}^d$. If our data distribution is truly random, we typically need an exponential amount of data.

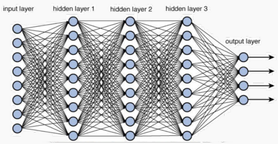


The existence of lower dimensional structure is our data is often the only reason we can hope to learn.

Low-dimensional structure.





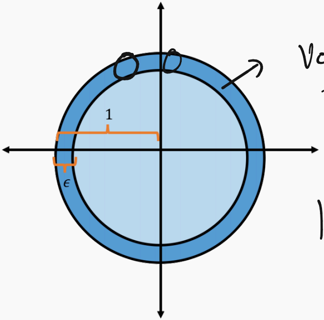


For example, data lies on low-dimensional subspace, or does so after transformation. Or function can be represented by a restricted class of functions, like neural net with specific structure.

Let $\mathcal{B}_d$ be the unit ball in d dimensions:

$$\mathcal{B}_d = \{\mathbf{x} \in \mathbb{R}^d : \|\mathbf{x}\|_2 \leq 1\}.$$

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



$$\frac{Vol(B_d(1)) - Vol(B_d(1-\varepsilon))}{Vol(B_d(1))}$$

$$1 - \frac{\cancel{\pi^{d/2}/(d/2)!} \cdot (1-\varepsilon)^d}{\cancel{\pi^{d/2}/(d/2)!} \cdot 1^d}$$

$$1 - (1-\varepsilon)^d$$

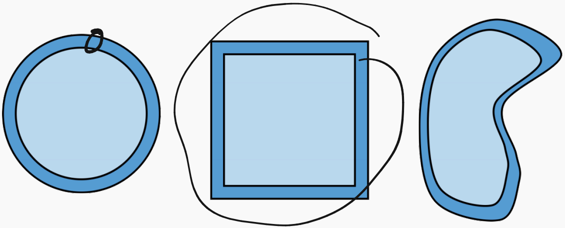
$$1 - \Big(\underbrace{(1-\varepsilon)^{1/\varepsilon}}_{1/e}\Big)^{\varepsilon d}$$

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

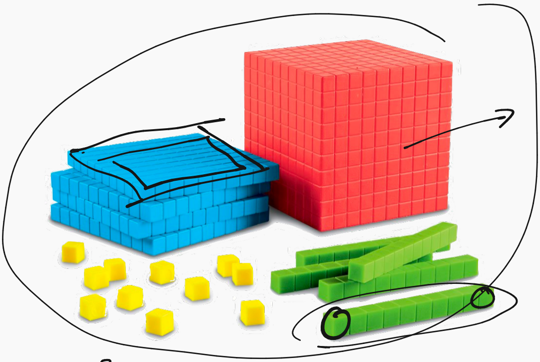
$$1 - \frac{1}{e}^{\varepsilon d}$$

All but an $\frac{1}{2}^{\Theta(\epsilon d)}$ 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.'



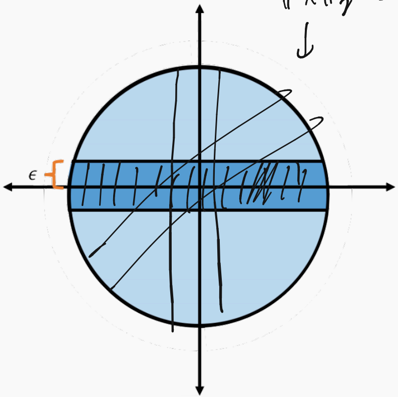
1D: $\frac{\text{surface cubes}}{\text{total cubes}} = \frac{2}{10} = .2$

2D: $\frac{\text{surface cubes}}{\text{total cubes}} = \frac{100-8^2}{100} = \frac{100-64}{100} = .36$

3D: $\frac{\text{surface cubes}}{\text{total cubes}} = \frac{1000-8^3}{1000} = \frac{488}{1000} = .488$

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

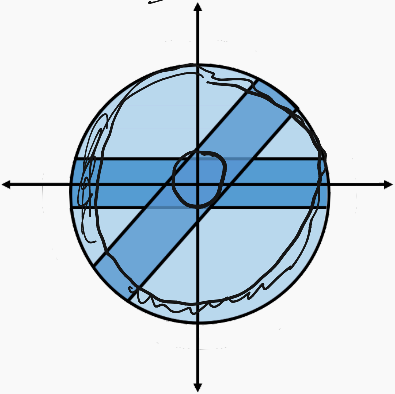
$\|x\|_2^2 \leq 1$



$$S = \{\mathbf{x} \in \mathcal{B}_d : |x_1| \leq \epsilon\}$$

What percentage of the volume of $\mathcal{B}_d$ falls within ϵ of its equator? **Answer**: all but a $\frac{1}{2}^{\Theta(\epsilon^2 d)}$ fraction. $1 - \frac{1}{2}^{\Theta(d)}$

$1 - \frac{1}{2}^{\Theta(d)}$

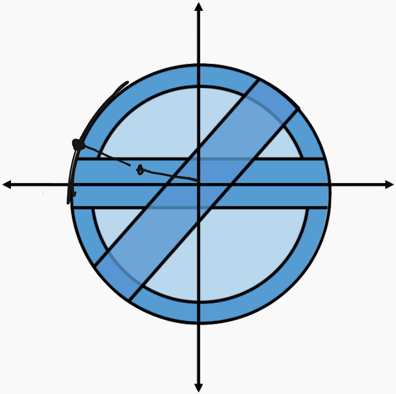


By symmetry, this is true for any equator:

$$S_{\mathbf{t}} = \{\mathbf{x} \in \mathcal{B}_d : \mathbf{x}^T\mathbf{t} \leq \epsilon\}.$$

BIZARRE SHAPE OF UNIT BALL

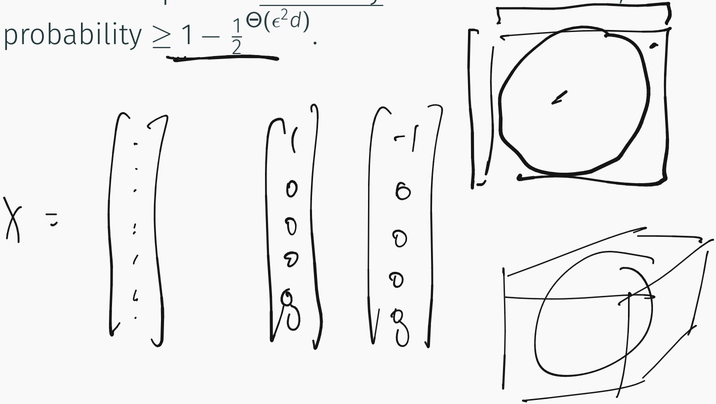
1. $\left(1 - \frac{1}{2}^{\Theta(\epsilon d)}\right)$ fraction of volume lies ϵ close to surface.
2. $\left(1 - \frac{1}{2}^{\Theta(\epsilon^2 d)}\right)$ fraction of volume lies ϵ close to any equator.



High-dimensional ball looks nothing like 2D ball!

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

Equivalent: If we draw a point $\mathbf{x}$ randomly from the unit ball, $|x_1| \leq \epsilon$ with probability $\geq 1 - \frac{1}{2}^{\Theta(\epsilon^2 d)}$.



Let $\mathbf{w} = \frac{\mathbf{x}}{\|\mathbf{x}\|_2}$. Because $\|\mathbf{x}\|_2 \leq 1$,

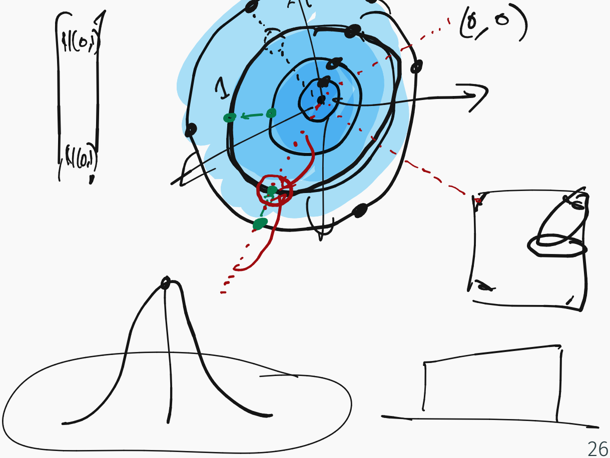
$$\Pr[|x_1| \leq \epsilon] \geq \Pr[|w_1| \leq \epsilon] . \geq 1 - \frac{1}{2}^{\Theta(\epsilon^2 d)}$$

Claim: $|w_1| \leq \epsilon$ with probability $\geq 1 - \frac{1}{2}^{\Theta(\epsilon^2 d)}$, which then proves our statement from the previous slide.

How can we generate w, which is a random vector taken from the unit sphere (the surface of the ball)?

Rotational Invariance of Gaussian distribution: Let $\mathbf{g}$ be a random Gaussian vector, with each entry drawn from $\mathcal{N}(0,1)$. Then $\mathbf{w} = \mathbf{g}/\|\mathbf{g}\|_2$ is distributed uniformly on the unit sphere.

Proof:



$\sim e^{-\|x\|_2^2}$

$g_i \sim N(0,1)$

$\Pr[X_1 \geq \epsilon]$

Let $\mathbf{g}$ be a random Gaussian vector and $\mathbf{w} = \mathbf{g}/\|\mathbf{g}\|_2$.

$\Pr[[g/\|g\|_2]_1 \geq \epsilon]$

- $\mathbb{E}[\|\mathbf{g}\|_2^2] = \mathbb{E} \sum_{i=1}^{d} g_i^2$

$= \sum_{i=1}^{d} \mathbb{E}[g_i^2] = \boxed{d}$

$\|g\|_2$

- $\mathrm{Var}[\|\mathbf{g}\|_2^2] =$

$$\begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}$$

- $\Pr\left[\|\mathbf{g}\|_2^2 \leq \frac{1}{2}\mathbb{E}[\|\mathbf{g}\|_2^2]\right] \leq \frac{1}{2}^{\theta(d)}$

$\leq \frac{1}{2} d$

$\Pr\left[\|g\|_2 \leq \frac{1}{\sqrt{2}}\sqrt{d}\right] \leq \frac{1}{2}^{\theta(d)}$

CONCENTRATION AT EQUATOR

For $1 - \frac{1}{2}^{\theta(d)}$ fraction of vectors $\mathbf{g}$, $\|\mathbf{g}\|_2 \geq \sqrt{d/2}$. Condition on the event that we get a random vector in this set.

Given this event:

$$\begin{aligned}
\Pr\left[|w_1| \leq \epsilon\right] &= \Pr\left[|w_1| \cdot \sqrt{d/2} \leq \epsilon \cdot \sqrt{d/2}\right] \\
&\geq \Pr\left[|g_1| \leq \epsilon \cdot \sqrt{d/2}\right] \\
&\geq 1 - \frac{1}{2}^{\theta\left((\epsilon \cdot \sqrt{d/2})^2\right)}
\end{aligned}$$

$= 1 - \frac{1}{2}^{O(\epsilon^2 d)}$

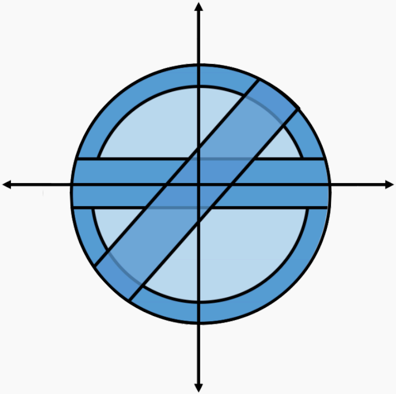
By union bound, overall we have:

$$\Pr\left[|w_1| \leq \epsilon\right] \geq 1 - \frac{1}{2}^{\theta\left((\epsilon \cdot \sqrt{d/2})^2\right)} - \frac{1}{2}^{\theta(d)}$$

$|w_1|\sqrt{d/2} \leq |g_1|$

Recall: $\mathbf{w} = \frac{\mathbf{g}}{\|\mathbf{g}\|_2}$. So after conditioning, we have $|w_1| \leq \frac{|g_1|}{\sqrt{d/2}}$.

1. $(1 - \frac{1}{2}^{\Theta(\epsilon d)})$ fraction of volume lies ϵ close to surface.
2. $(1 - \frac{1}{2}^{\Theta(\epsilon^2 d)})$ fraction of volume lies ϵ close to any equator.



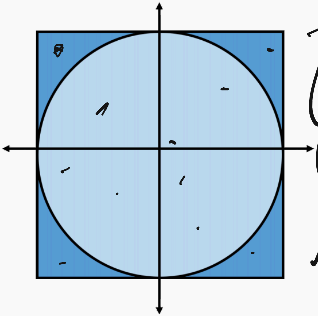
High-dimensional ball looks nothing like 2D ball!

HIGH DIMENSIONAL CUBE

Let $\mathcal{C}_d$ be the d-dimensional cube:

$$\mathcal{C}_d = \{\mathbf{x} \in \mathbb{R}^d : |\mathbf{x}(i)| \leq 1 \ \forall i\}.$$

$$\frac{2^d}{\sqrt{\pi}^d \, (d/2)!}$$



2

$$\frac{(.5)^d}{(d/2)!}$$

$$\frac{d}{2} \cdot \left(\frac{d}{2} - 1\right) \cdots$$

$$O(d)^{O(d)}$$

In two dimensions, the cube is pretty similar to the ball.

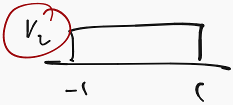
But volume of $\mathcal{C}_d$ is 2^d while volume of unit ball is $\frac{\sqrt{\pi}^d}{(d/2)!}$.

This is a huge gap! Cube has $O(d)^{O(d)}$ more volume.

Some other ways to see these shapes are very different:

- $\max_{\mathbf{x} \in \mathcal{B}_d} \|\mathbf{x}\|_2^2 =$ 1
- $\max_{\mathbf{x} \in \mathcal{C}_d} \|\mathbf{x}\|_2^2 =$ d

[1 1 1 1 1 1 1 1]



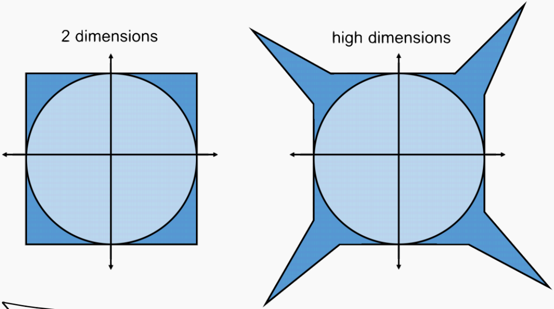
Some other ways to see these shapes are very different:

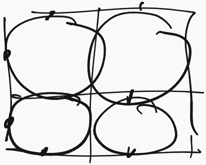
- $\mathbb{E}_{\mathbf{x}\sim\mathcal{B}_d}\|\mathbf{x}\|_2^2 \leq 1$
- $\mathbb{E}_{\mathbf{x}\sim\mathcal{C}_d}\|\mathbf{x}\|_2^2 = \mathbb{E}\sum_{i=1}^{d} x_i^2 = \sum_{i=1}^{d} \mathbb{E}(x_i^2) = d/3$

$x_i \sim \text{Unif}[-1,1]$

$$\int_{-1}^{1} x^2 \frac{1}{2}\, dx = \frac{1}{3}$$

Almost all of the volume of the unit cube falls in its corners, and these corners lie far outside the unit ball.





See The Journey to Define Dimension from Quanta Magazine for another fun example comparing cubes to balls!

Despite **all this** warning that low-dimensional space looks nothing like high-dimensional space, next we are going to learn about how to **compress high dimensional vectors to low dimensions.**

We will be very careful not to compress things too far. An extremely simple method known as Johnson-Lindenstrauss Random Projection pushes right up to the edge of how much compression is possible.