Plan

Logistics

Review

High-dimensional Geometry

Game night!
Wednesday @ 6

Project Proposal: Choose topic

Resources:
↳ Typed notes before lecture
↳ stay for office hours
↳ Ask questions!

Problem Set:
↳ Please read comments
↳ Want explanation ≥ solution
↳ Write up separately
(will start taking off points)
↳ come visit!! :)
↳ Will switch to gradescope for consistency

Importance of explaining for your own understanding
↳ This is my fifth experience with this class
↓
New groups!

Concentration Inequalities

Markov's $\quad X \geq 0 \quad \Pr(X \geq t) \leq \frac{\mathbb{E}[X]}{t}$

linear

Chebyshev's $\quad X$ with $\sigma^2 = Var(X) \quad \Pr(|X - \mu| \geq t \cdot \sigma) \leq \frac{1}{t^2}$

quadratic

Chernoff's $\quad X_1, \ldots, X_n$ independent binary

$$\mu = \sum_{i=1}^{n} \mathbb{E}[X_i]$$

$$\Pr(|X - \mu| \geq \epsilon \cdot \mu) \leq 2 \exp\left(- \frac{\epsilon^2 \mu}{3}\right)$$

exponential

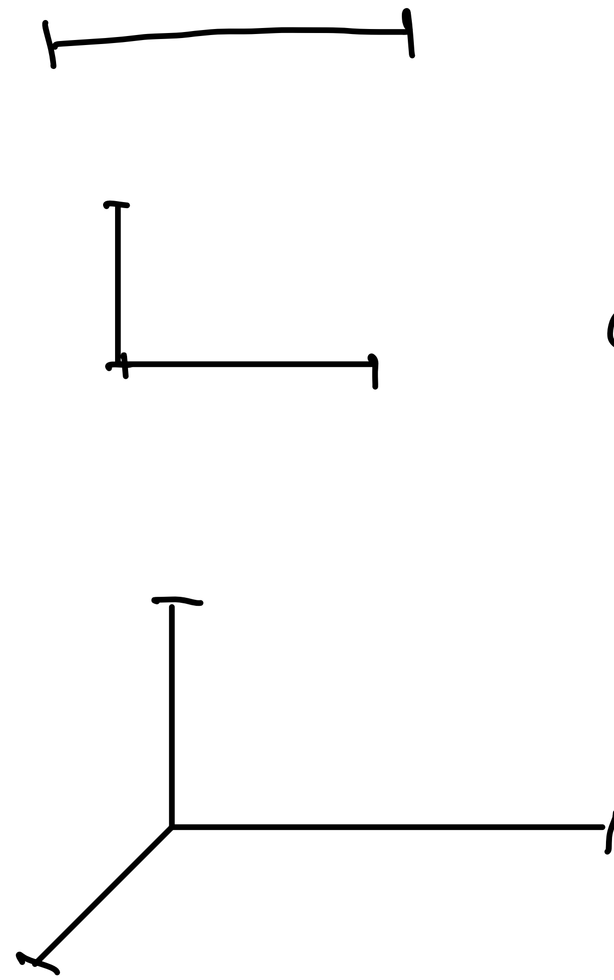
We want to work with high-dimensional data

↳ Find similar items

↳ Low-rank approximations

↳ Graphs!

High-Dimensional Geometry is Weird



$d=1$

$d=2$

$d=3$

What do high dimensions look like?

Vectors

$x, y \in \mathbb{R}^d$

e.g., $\begin{bmatrix}1\\3\\7\end{bmatrix}, \begin{bmatrix}2\\1\\2\end{bmatrix} \in \mathbb{R}^3$

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

? $\left\langle \begin{bmatrix}1\\3\\7\end{bmatrix}, \begin{bmatrix}2\\1\\2\end{bmatrix} \right\rangle = 2+3+14 = 19$

$$\langle x,x\rangle = x^T x$$
$$= \sum_{i=1}^{d} x[i]\, x[i]$$
$$= \sum_{i=1}^{d} (x[i])^2$$
$$= \|x\|_2^2 \geq 0$$?

$\langle x,y\rangle \approx$ similarity

$\langle x,y\rangle = \|x\|_2 \|y\|_2 \cos\theta$

If $\langle x,y\rangle = 0$,
x, y are orthogonal

Orthogonal Vectors

What's the largest set of mutually orthogonal vectors in d?

$x_1, \ldots, x_t$ so that

$\langle x_i, x_j \rangle = 0$ for $i \neq j$

what's the largest value of t?

$t \geq d$ because of standard basis vectors

$$e_i = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 1 \\ 0 \end{bmatrix} \leftarrow i^{th} \text{ dimension}$$

Suppose for contradiction $t \geq d+1$ so $x_1, \ldots, x_d, x_{d+1}, \ldots$

Because d orthogonal span d dimensions,

$$x_{d+1} = \sum_{i=1}^{d} \alpha_i x_i$$

$\alpha_j \neq 0$ for some j

$$\begin{aligned}
\langle x_j, x_{d+1} \rangle &= \langle x_j, \sum_{i=1}^{d} \alpha_i x_i \rangle \\
&= \sum_{i=1}^{d} \langle x_j, \alpha_i x_i \rangle \\
&= \alpha_j \langle x_j, x_j \rangle \\
&= \alpha_j \|x_j\|_2^2 \\
&\neq 0
\end{aligned}$$

Nearly orthogonal

Whats the largest set of nearly mutually orthogonal vectors in d?

$x_1, \dots, x_t$ so that

$|\langle x_i, x_j \rangle| \le \epsilon$ for $i \ne j$

whats the largest value of t?

Probabilistic Method

We'll construct a random process that generates $x_1, \dots, x_t$ so that

$Pr(x_1, \dots, x_t$ nearly ortho$) > 0$

=> Then there must be one nearly ortho set

Random Process

$x_i \in \mathbb{R}^d$

$$x_i[k] = \begin{cases} \frac{1}{\sqrt{d}} & \text{wp } 1/2 \\ -\frac{1}{\sqrt{d}} & \text{wp } 1/2 \end{cases}$$

$\|x_i\|_2^2 =$

$i \neq j$

$\mathbb{E}[\langle x_i, x_j \rangle] =$

$\mathrm{Var}(\langle x_i, x_j \rangle] =$

$$Z = x_i^T x_j = \sum_{k=1}^{d} \underbrace{x_i[k]\, x_j[k]}$$

$$C_k = \begin{cases} 1/d & \text{wp } 1/2 \\ -1/d & \text{wp } 1/2 \end{cases}$$

$Z \approx$ Gaussian for large d

Chernoff? But C_k not binary...

$$B_k = \begin{cases} 1 & \text{wp } 1/2 \\ 0 & \text{wp } 1/2 \end{cases}$$

$$C_k = \frac{2}{d} \cdot \frac{d}{2} C_k$$

$$\overset{*}{=} \frac{2}{d}\left(-\frac{1}{2} + B_k\right)$$

$$\frac{d}{2} \cdot C_k = \begin{cases} \frac{1}{2} & \text{wp } 1/2 \\ -\frac{1}{2} & \text{wp } 1/2 \end{cases}$$

$$\frac{d}{2} C_k + \frac{1}{2} = \begin{cases} 1 & \text{wp } 1/2 \\ 0 & \text{wp } 1/2 \end{cases}$$

$$\frac{d}{2} C_k + \frac{1}{2} \overset{*}{=} B_k$$

$$Z = \sum_{k=1}^{d} C_k = \sum_{k=1}^{d} \frac{2}{d}\left(-\frac{1}{2} + B_k\right)$$

$$= \frac{2}{d} \sum_{k=1}^{d} -\frac{1}{2} + \frac{2}{d} \sum_{k=1}^{d} B_k$$

$$= -1 + \frac{2}{d} \sum_{k=1}^{d} B_k$$

$$Z = x_i^T x_j$$

$$= -1 + \frac{2}{d}\sum_{k=1}^{d} B_k > \epsilon$$

$$\Longleftrightarrow$$

$$\sum_{k=1}^{d} B_k > (\epsilon+1)\frac{d}{2}$$

$$Z < -\epsilon$$

$$\Longleftrightarrow$$

$$\sum_{k=1}^{d} B_k < (1-\epsilon)\frac{d}{2}$$

$$\mathbb{E}\left[\sum_{k=1}^{d} B_k\right] = \mu = \frac{d}{2}$$

Chernoff! $0 < \epsilon < 1$

$$\Pr(|S-\mu| \geq \epsilon\mu) \leq 2\exp\left(-\frac{\epsilon^2\mu}{3}\right)$$

$$\Pr(|Z| > \epsilon)$$

$$= \Pr\left(\left|\sum_{k=1}^{d} B_k - \frac{d}{2}\right| \geq \epsilon\frac{d}{2}\right)$$

$$= \Pr\left(\left|\sum_{k=1}^{d} B_k - \mu\right| \geq \epsilon\mu\right)$$

$$\leq 2\exp\left(-\frac{\epsilon^2\mu}{3}\right)$$

$$= 2\exp\left(-\frac{\epsilon^2}{3}\cdot\frac{d}{2}\right)$$

$$\Pr(\nexists\ i \neq j : |x_i^T x_j| > \epsilon)$$

\# pairs

$$= \binom{t}{2} \Pr(|x_i^T x_j| > \epsilon)$$

want

$$\leq \frac{t(t-1)}{2}\, 2 \exp\left(-\frac{\epsilon^2}{3} \cdot \frac{d}{2}\right) < 1$$

$$t(t-1) < \exp\left(\frac{\epsilon^2 . d}{6}\right)$$

$$t < \exp\left(\frac{\epsilon^2 d}{12}\right) = e^{\epsilon^2 d/12}$$

$$= 2^{c\epsilon^2 d} \qquad c = \frac{\log_2(e)}{12}$$

t **exponential** !!

Corollary:

Random vectors tend to be far apart.

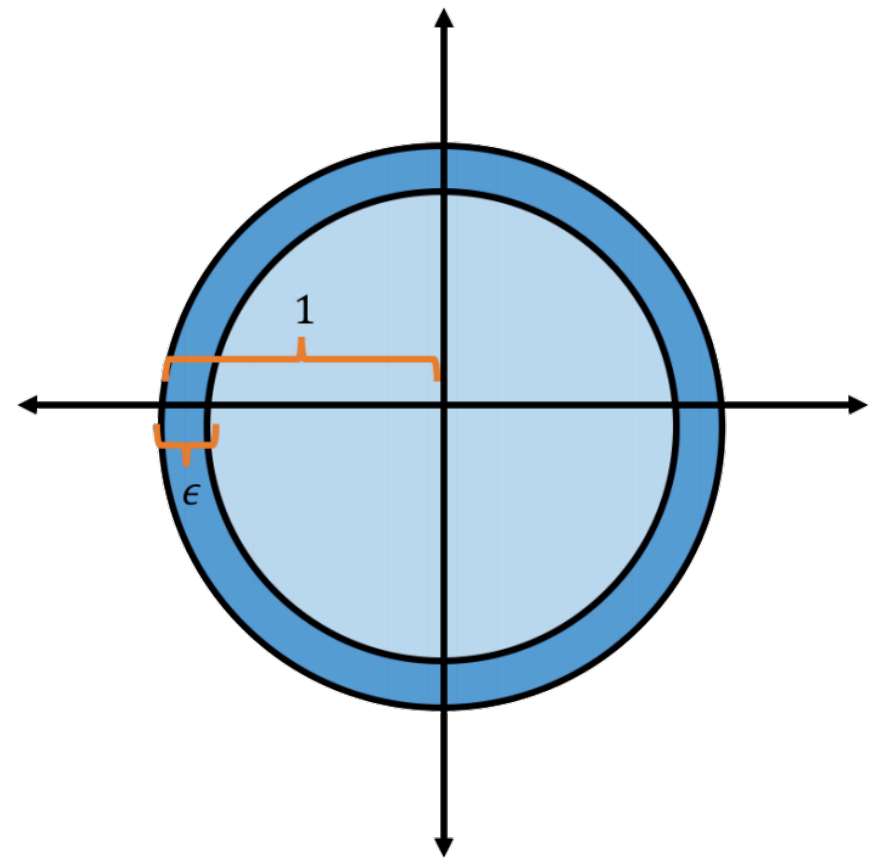
How do we find patterns?

structure

Where Points Live

$$B_d(R) = \{ x \in \mathbb{R}^d : \|x\|_2 \leq R \}$$

↑ Ball in d-dimensions with radius R



What fraction of volume ϵ close to surface?

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

$$Vol(B_d(R)) = \frac{\pi^{d/2}}{(d/2)!} R^d \quad \text{when } d \text{ even}$$

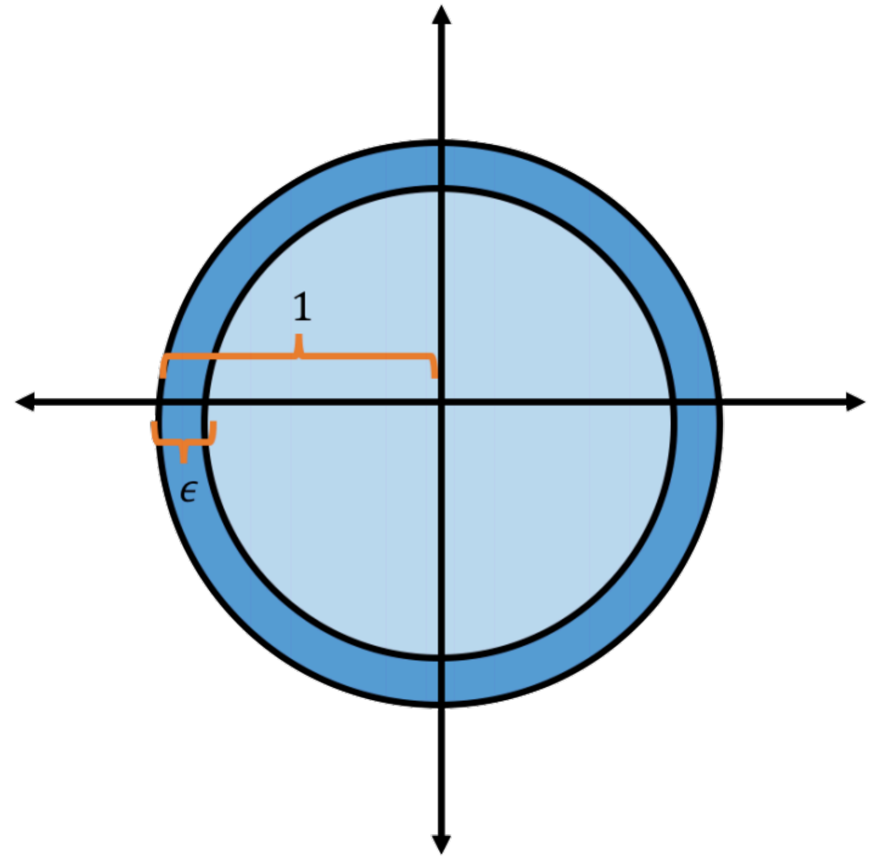
$$\frac{Vol(B_d(1)) - Vol(B_d(1-\epsilon))}{Vol(B_d(1))}$$

$$= 1 - \frac{\frac{\pi^{(d/2)}}{(d/2)!}(1-\epsilon)^d}{\frac{\pi^{(d/2)}}{(d/2)!}1^d}$$

$$= 1-(1-\epsilon)^d$$

$$= 1-\left[(1-\epsilon)^{1/\epsilon}\right]^{\epsilon d}$$

$$\approx 1-\left(\frac{1}{e}\right)^{\epsilon d} = 1-\frac{1}{e^{\epsilon d}}$$

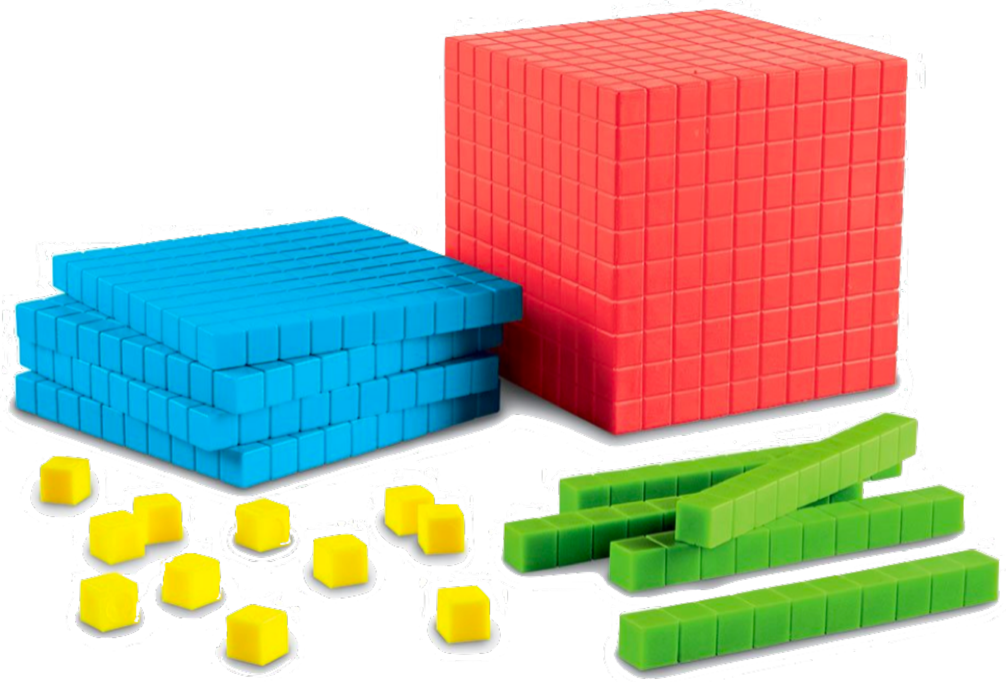


All but very small fraction! near surface of sphere

And other shapes in high dimensions?

In d-D,

?



In 1D, $\dfrac{\#\text{ cubes on surface}}{\#\text{ total}} = \dfrac{10 \cdot 2}{10}$

In 2D,

In 3D,

Sphere vs cube

B_d C_d

$$\max_{x \in B_d} \|x\|_2^2 =$$

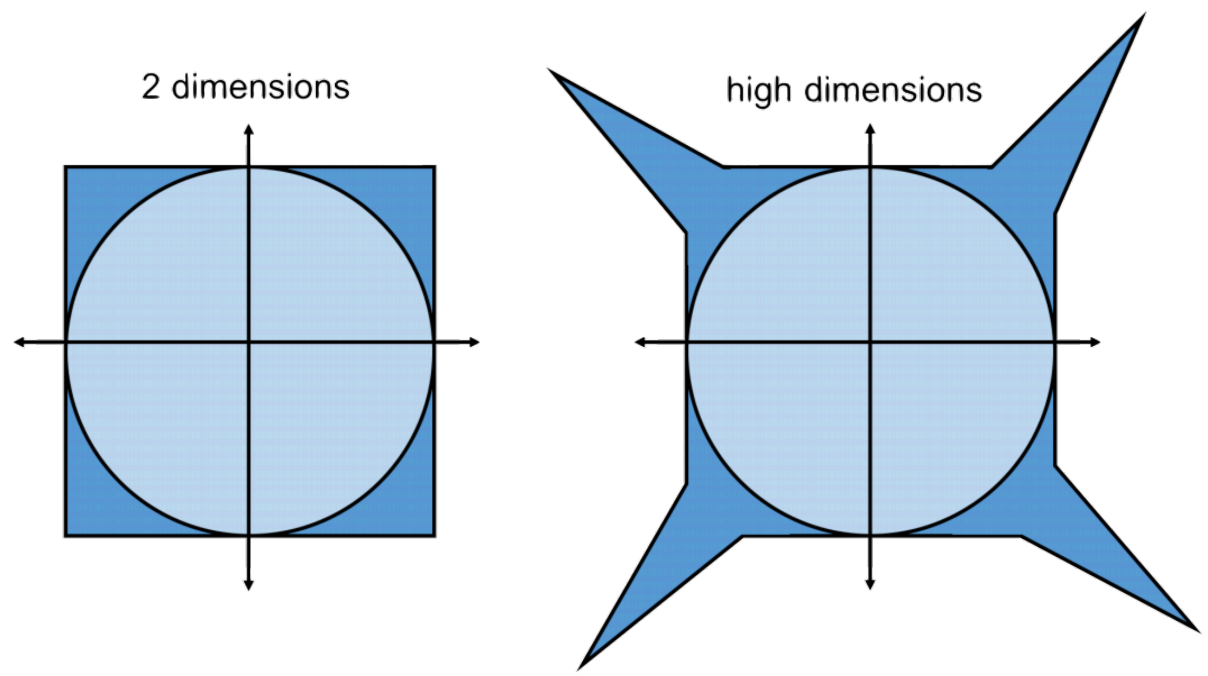
$$\max_{x \in C_d} \|x\|_2^2 =$$

What about expectation?

$$\mathbb{E}_{x \sim B_d}\left[\|x\|_2^2 \right] \leq$$

$$\mathbb{E}_{x \sim C_d}\left[\|x\|_2^2 \right] =$$

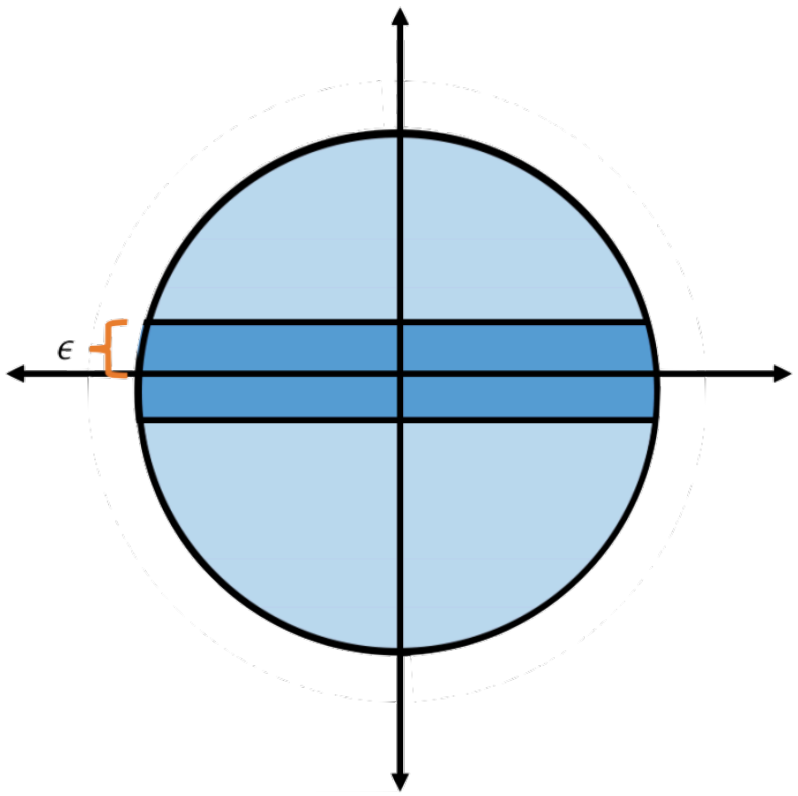
$$\sum_{i=1}^{d} \mathbb{E}_{x_i \sim [-1,1]} [x_i]^2 =$$



$$\frac{\mathrm{Vol}(C_d)}{\mathrm{Vol}(B_d)} = \frac{2^d}{\left(\frac{\pi^{(d/2)}}{(d/2)!}\right)} = \frac{2^d \, (d/2)!}{\pi^{(d/2)}} \approx d^d$$

huge!!

Spheres are weird



How much volume ϵ close to equator?

We know all but $\frac{1}{e^{\epsilon d}}$ ϵ-close to surface

AND all but $\frac{1}{e^{\epsilon^2 d}}$ ϵ-close to equator

for <u>any</u> equator

by symmetry

Prove using vectors
uniformly sampled from
sphere (in notes,
property for homework)