Plan

Logistics

Review

High-dimensional Geometry

Game night!
Wednesday@6
75 Shannon 202

Tea time!
Friday @ 2
Bihall 6th Floor

Project Proposal: Choose topic

Resources
- ↳ Typed notes
- ↳ Stay office hours
- ↳ Ask questions!

Problem Set
- ↳ Please read comments
- ↳ Explanation on par with solns.
- ↳ Write separately
- ↳ Come visit!! :)
- ↳ Gradescope

Concentration Inequalities

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

Chebyshev's X with $\sigma^2 = Var(X)$, $\mu = \mathbb{E}[X]$ $\quad Pr(|X-\mu| \geq k \cdot \sigma) \leq \frac{1}{k^2}$ quadratic

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

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

$$Pr(|X-\mu| \geq \epsilon \cdot \mu) \leq 2 \exp\left(-\frac{\epsilon^2 \mu}{3}\right) \qquad 0 < \epsilon < 1$$

exponentially

$X \geq \mu + \epsilon\mu$ or $X \leq \mu - \epsilon\mu$

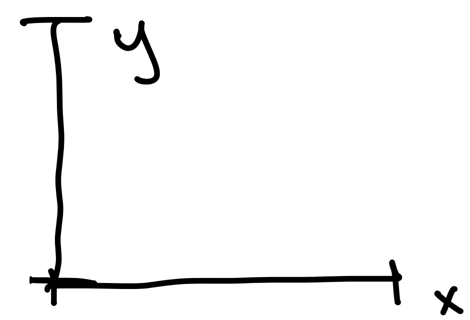
$X \geq (1+\epsilon)\mu$ or $X \leq (1-\epsilon)\mu$

High dimensional data

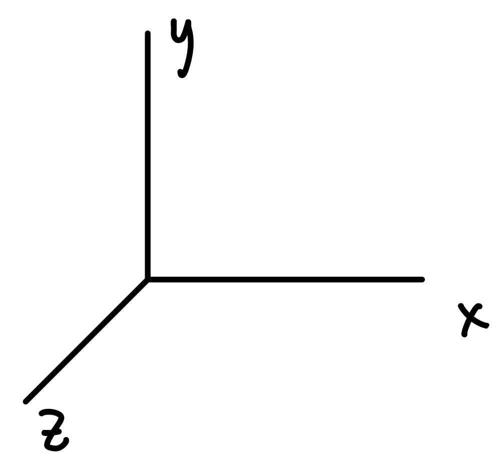
- ↳ Find similar items
- ↳ Low-rank approximation
- ↳ Graphs!

High-dimensional geometry is weird

$d=1$



$d=2$



$d=3$

Vectors

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

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

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

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

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

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

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

angle between x, y

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

than x, y are orthogonal

Orthogonal vectors

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

$x_1, \dots, x_t$ so that

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

What's the largest value of t?

$t \geq d$ because standard basis vectors

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

Suppose for contradiction

$t \geq d+1$ $\quad x_1, \dots, x_d, x_{d+1}, \dots$

Because d ortho span $\mathbb{R}^d$,

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

$J =$ set of $j : \alpha_j \neq 0$

$$\langle x_{d+1}, x_j \rangle = \langle x_j, \sum_{i=1}^{d} \alpha_i x_i \rangle$$

$$= \sum_{i=1}^{d} \alpha_i \langle x_j, x_i \rangle$$

$$= \sum_{j \in J} \alpha_j \langle x_j, x_j \rangle$$

$$= \sum_{j \in J} \alpha_j \, \|x_j\|_2^2 \neq 0$$

Nearly Orthogonal

What's the largest set of nearly orthogonal unit vector in d?

$x_1, \dots, x_t$ so that

$|\langle x_i, x_j \rangle| < \epsilon$ for $i \neq j$

What's the largest value of t?

Probalistic Method

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

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

$\Rightarrow$ There exists at least one set that is nearly ortho.

Random Process

$x_1, \dots, x_t \in \mathbb{R}^d$

$$\|x_i\|_2^2 = \sum_{k=1}^{d} (x_i[k])^2 = \sum_{k=1}^{d} \frac{1}{d} = 1$$

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

$i \neq j$

$$\mathbb{E}[\langle x_i, x_j \rangle] \overset{\text{linearity}}{=} \sum_{k=1}^{d} \mathbb{E}[x_i[k]\, x_j[k]] \overset{\text{indep}}{=} \sum_{k=1}^{d} \mathbb{E}[x_i[k]]\, \mathbb{E}[x_j[k]]$$

$$= 0$$

$$\mathrm{Var}(\langle x_i, x_j \rangle] \overset{\text{linearity}}{=} \sum_{k=1}^{d} \mathrm{Var}(\underbrace{x_i[k]\, x_j[k]}_{\cdot}) = \sum_{k=1}^{d} \mathbb{E}[\cdot^2] - \mathbb{E}[\cdot]^2$$

$$\overset{\text{indep}}{=} \sum_{k=1}^{d} \mathbb{E}[x_i[k]^2] \cdot \mathbb{E}[x_j[k]^2] = \sum_{k=1}^{d} \frac{1}{d} \cdot \frac{1}{d} = \frac{1}{d}$$

$$Z = \langle X_i, X_j \rangle = \sum_{k=1}^{d} \underbrace{x_i[k]\, x_j[k]}$$

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

$Z \approx$ Gaussian for large d

Chernoff? But C_k are not binary...

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

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

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

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

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

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

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

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

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

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

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

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

$$Z < -\epsilon$$

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

Chernoff $\quad 0<\epsilon<1$

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

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

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

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

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

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

$\Pr($ not nearly orthogonal$)$

$= \Pr(\exists\ i \neq j: |\langle x_i, x_j \rangle| > \epsilon)$

$\overset{\text{union}}{\leq} \binom{t}{2} \Pr(|\langle x_i, x_j \rangle| > \epsilon)$

$\overset{\text{last slide}}{\leq} \binom{t}{2} 2\exp\left(\frac{-\epsilon^2}{3} \cdot \frac{d}{2}\right) \overset{\text{want}}{<} 1$

$\frac{t(t-1)}{\cancel{2}} \cancel{2} \overset{\text{want}}{<} \exp\left(\frac{\epsilon^2 d}{6}\right)$

$\Uparrow$

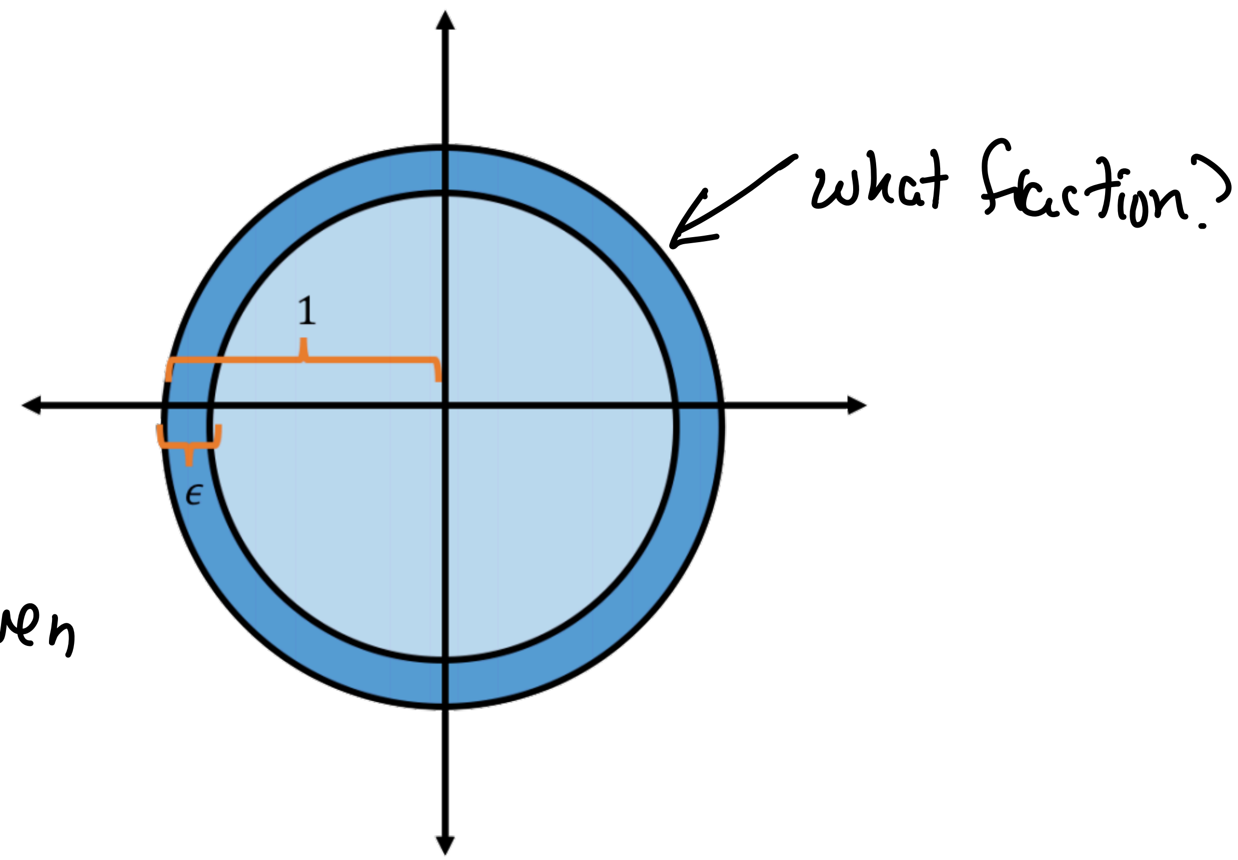
$t < \exp\left(\frac{\epsilon^2 d}{12}\right) = e^{\epsilon^2 d/12}$

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

Corollary:

Random vectors tend to be far apart

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



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

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

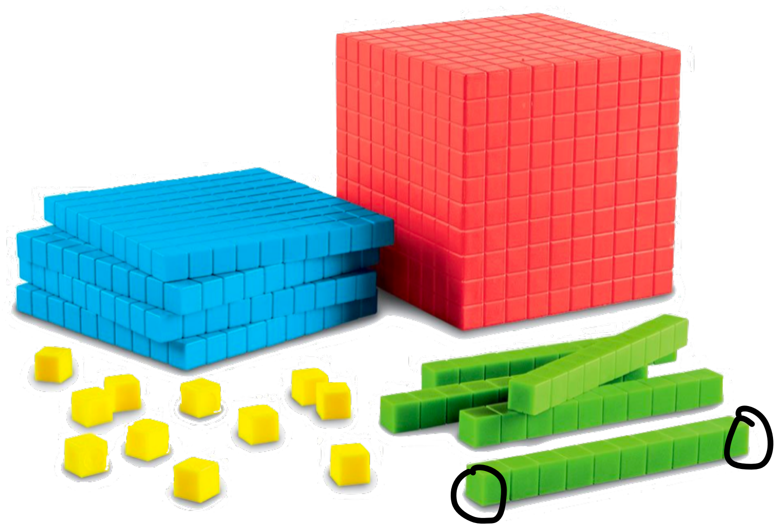
$$= 1 - \frac{\pi^{d/2}}{(d/2)!}(1-\epsilon)^d \Big/ \frac{\pi^{d/2}}{(d/2)!} 1^d$$

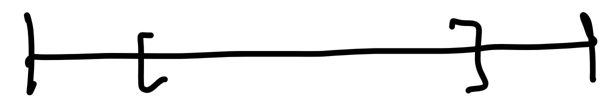
$$= 1 - (1-\epsilon)^d$$

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

$$\approx 1 - \frac{1}{e^{\epsilon d}}$$

And other shapes?





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

In 2D, $\dfrac{10^2 - 8^2}{10^2} = \dfrac{36}{100}$

In 3D, $\dfrac{10^3 - 8^3}{10^3} = .488$

sphere vs cube

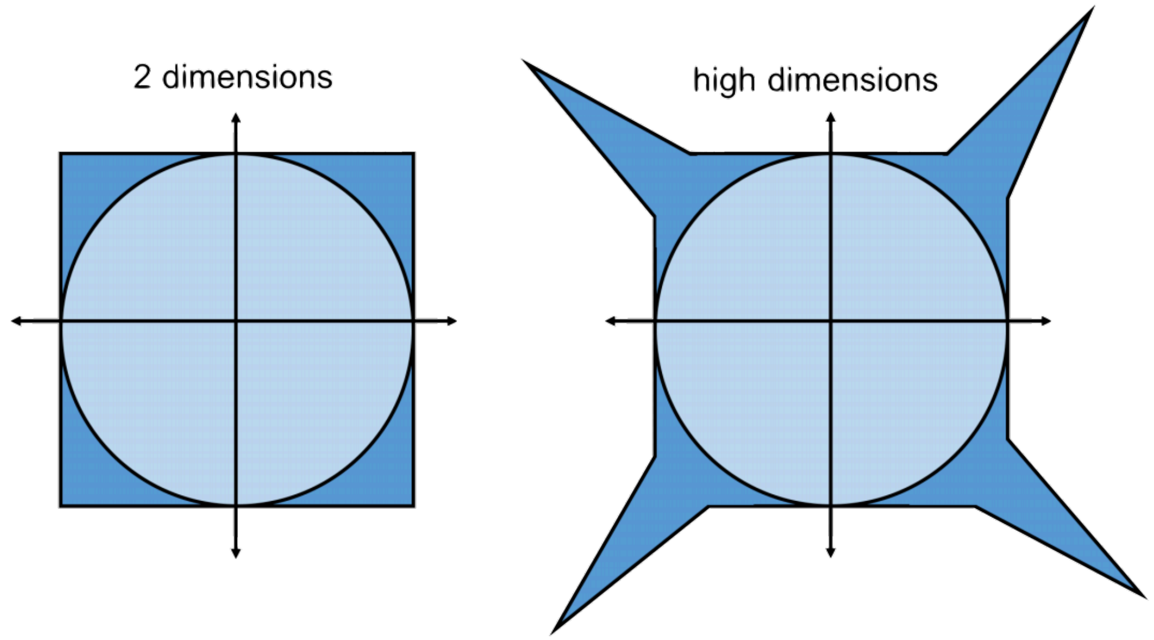
$B_d(1)$ $\quad$ C_d with "radius" 1

$$\max_{x \in B_d} \|x\|_2^2 = 1 \qquad \max_{x \in C_d} \|x\|_2^2 = d$$

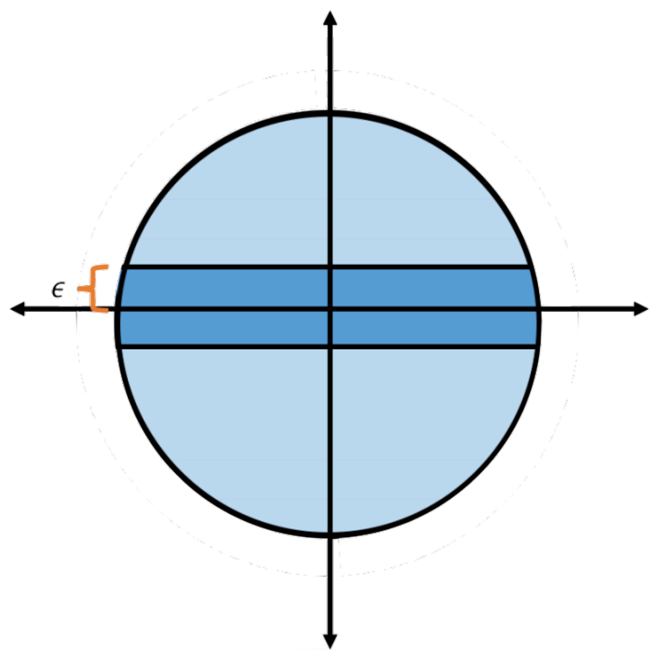
$$\int_{x=-1}^{1} \frac{1}{2} x^2 dx = \frac{1}{2} \cdot \frac{1}{3} x^3 \Big]_{-1}^{1} = \frac{1}{2}\left(\frac{1}{3} + \frac{1}{3}\right) = 1/3$$

$$\mathbb{E}_{x \sim B_d(1)} \left[\|x\|_2^2 \right] \le 1$$

$$\mathbb{E}_{x \sim C_d} \left[\|x\|_2^2 \right] = \sum_{i=1}^{d} \mathbb{E}_{x_i \sim [-1,1]} \left[x_i^2 \right] = \frac{d}{3}$$



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



ϵ