Tuesday, Feb 17

- Exam 3/5

Plan

- Revisiting Load Balancing
- High dimensional geometry

Exponential Concentration Inequalities

$X_1, \ldots, X_n$ indep

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

↳ Make assumptions on X_i:
stronger assumption = stronger bound

↳ Proved using clever applications of Markov's

Chernoff Bound

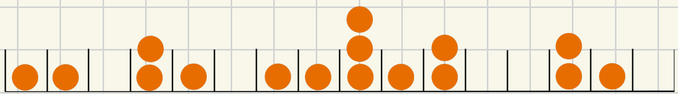
Binary $X_i \in \{0,1\}$. For $0 < \epsilon < 1$,

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

For $\epsilon > 0$,

$$\Pr(S \geq (1+\epsilon)\mu) \leq \exp\left(\frac{-\epsilon^2\mu}{2+\epsilon}\right)$$

Revisiting Load Balancing



$$S_i = \sum_{j=1}^{m} \mathbb{1}[j \text{ to } i]$$

$$S = \max_i S_i$$

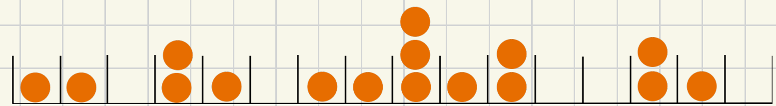
$$\Pr(S \geq c) \leq \frac{1}{10}$$

$$\Leftarrow \quad \Pr(S_i \geq c) \leq 1/10n$$

with Chebyshev's, $c = O(\sqrt{n})$

Q: Can we do better?

Power of Two Choices



use two hash functions, choose least occupied

Then $\Pr(S \geq c) \leq 1/10$ for $c = \log n$

$c = \log \log n$

$c = \log \log \log n$

How about power of three choices?

High dimensional geometry

Unifying theme...

- random projections
- locality sensitive hashing
- low rank approximation
- graph representations

Setup

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

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

$\|x\|_2^2 = \langle x, x \rangle = \sum_{i=1}^{d} x_i^2$

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

High-dimensional geometry is weird

Q: In d dimensions, how many orthogonal vectors are there?

i.e., orthogonal if $\langle x, y \rangle = 0$

Q: In d dimensions, how many nearly orthogonal vectors are there?

i.e., nearly orthogonal if $|\langle x, y \rangle| < \epsilon$

Probabilistic Method

Let $t = 2^{c\epsilon^2 d}$, for constant c.

We'll prove $\exists$ $x_1, \dots, x_t$ nearly orthogonal.

Strategy: Define random process and show, with non-zero probability,

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

$$x_i = \begin{bmatrix} +1/\sqrt{d} \\ -1/\sqrt{d} \\ -1/\sqrt{d} \\ \vdots \end{bmatrix} \qquad x_i[j] = \begin{cases} 1/\sqrt{d} & \text{wp } 1/2 \\ -1/\sqrt{d} & \text{wp } 1/2 \end{cases}$$

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

$$\mathbb{E}[\langle x_i, x_j \rangle] = \sum_{k=1}^{d} \mathbb{E}[x_i[k]\, x_j[k]] \overset{\text{indep}}{=} \sum_{k=1}^{d} \mathbb{E}[x_i[k]]\, \mathbb{E}[x_j[k]] = 0$$

$$\mathrm{Var}(\langle x_i, x_j \rangle) \overset{\text{indep}}{=} \sum_{k=1}^{d} \mathrm{Var}(x_i[k]\, x_j[k]) = \sum_{k=1}^{d} \mathbb{E}[(x_i[k]x_j[k] - 0)^2] = \sum_{k=1}^{d} \frac{1}{d^2} = \frac{1}{d}$$

Thursday, February 19

- High dimensional geometry

Probabilistic Method

Let $t = 2^{c\epsilon^2 d}$, for constant c.

We'll prove $\exists$ $x_1, \ldots, x_t$ nearly orthogonal.

Strategy: Define random process and show, with non-zero probability,

$$|\langle x_i, x_j \rangle| < \epsilon \text{ for all } i \neq j$$

$$x_i = \begin{bmatrix} +1/\sqrt{d} \\ -1/\sqrt{d} \\ -1/\sqrt{d} \\ \vdots \end{bmatrix} \qquad x_i[j] = \begin{cases} 1/\sqrt{d} & \text{wp } 1/2 \\ -1/\sqrt{d} & \text{wp } 1/2 \end{cases}$$

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

$$\mathbb{E}[\langle x_i, x_j \rangle] = \sum_{k=1}^{d} \mathbb{E}[x_i[k]\, x_j[k]] \overset{\text{indep}}{=} \sum_{k=1}^{d} \mathbb{E}[x_i[k]]\, \mathbb{E}[x_j[k]] = 0$$

$$\mathrm{Var}(\langle x_i, x_j \rangle) \overset{\text{indep}}{=} \sum_{k=1}^{d} \mathrm{Var}(x_i[k]\, x_j[k]) = \sum_{k=1}^{d} \mathbb{E}[(x_i[k] x_j[k] - 0)^2] = \sum_{k=1}^{d} \frac{1}{d^2} = \frac{1}{d}$$

Fix i, j. Let $z = \langle x_i, x_j \rangle = \sum_{k=1}^{d} c_k$

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

Since z iid sum, expect $z \sim N$.

If z <u>were</u> Gaussian,

$$\Pr\left(|z| \geq \alpha \frac{1}{\sqrt{d}} \right) = \Pr\left(|z - \mathbb{E} z| \geq \alpha \sigma \right) \leq O(e^{-\alpha^2})$$

and done by setting $\alpha = \epsilon \sqrt{d}$.

But z is <u>not</u> Gaussian so more work :)

Chernoff? Write z as sum of <u>binary</u>.

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

$$c_k = \frac{2}{d} (B_k - 1/2)$$

$$z = \sum_{k=1}^{d} c_k = \frac{2}{d} \sum_{k=1}^{d} \frac{d}{2} c_k$$

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

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

$z > \epsilon \iff$

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

$z < \epsilon \iff$

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

$$B = \sum_{k=1}^{d} B_k, \quad \mathbb{E}[B] = \frac{d}{2}$$

$$\Pr(|Z| \geq \epsilon) = \Pr(|B - \mathbb{E}B| \geq \epsilon\, \mathbb{E}B)$$

$$\leq 2\exp\left(\frac{-\epsilon^2\, \mathbb{E}B}{3}\right)$$

$$= 2\exp\left(-\frac{\epsilon^2 d}{6}\right)$$

$$\Pr(\exists\, i \neq j : |\langle x_i, x_j \rangle| \geq \epsilon) \leq \binom{t}{2}\, 2\exp\left(\frac{-\epsilon^2 d}{6}\right)$$

Choose t so failure prob $< t$.

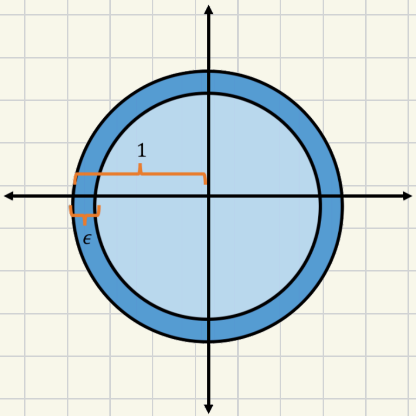
Takeaway: Random vectors tend to be far apart in high d.

$\therefore$ working with high d might <u>seem</u> hopeless, but our data is typically <u>not</u> random ie there is structure we can use

High d is weird part 2: Where Random Points Live

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

Q: What fraction of volume lies within ϵ of surface?



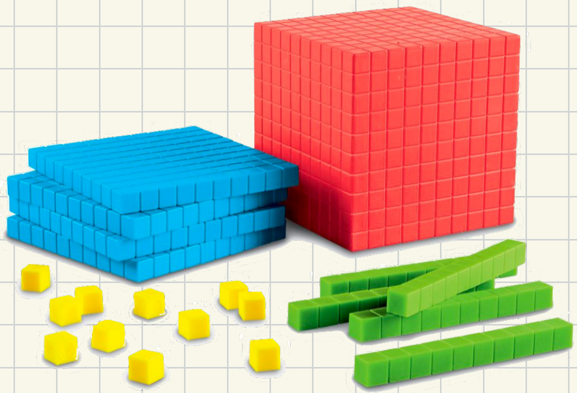
$$\mathrm{Vol}(B_d(R)) = \frac{\pi^{d/2}}{(d/2)!} R^d$$

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

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

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

All but $\frac{1}{2^{c\epsilon d}}$ fraction is ϵ-close to surface!

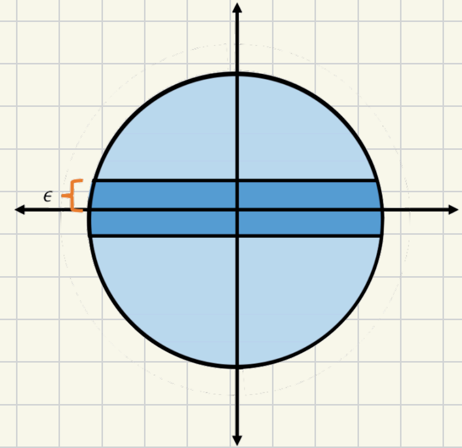


Q: What fraction of cubes are near surface?

d=1:

d=2:

d=3:



Q: What fraction of volume is ϵ close to equator?

A: All but $\frac{1}{2^{c\epsilon d}}$ fraction of volume

ϵ-close to any equator

Draw random points from unit ball

Goal: Show that $x \sim B_d$ has $|x_i| \leq \epsilon$ wp $1 - \frac{1}{2^{c\epsilon d}}$

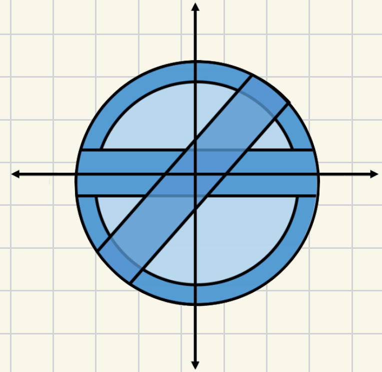
Given x from <u>interior</u> of unit ball, $w = \frac{x}{\|x\|_2}$ from surface

If $|w_1| \leq \epsilon \Rightarrow |x_1| < \epsilon$ since $\|x\|_2 \leq 1$

New Goal: Show that w from surface has $|x_i| \leq \epsilon$ wp $1 - \frac{1}{2^{c\epsilon d}}$

Let $g \sim N(0, I)$. $w = \frac{g}{\|g\|_2}$ from surface by rotational invariance

$\mathbb{E}\,\|g\|_2^2 = \ldots$



If:

① $\|g\|_2 \geq \sqrt{d/2}$

② $|g_1| \leq \sigma \sqrt{d/2}$

Then

$$|w_1| = \frac{|g_1|}{\|g\|_2} \leq \frac{\sigma\sqrt{d/2}}{\sqrt{d/2}} = \epsilon$$

$$\Pr(① \text{ or } ②) \leq \Pr(①) + \Pr(②)$$

$$\Pr(① \text{ and } ②) = \Pr(①) + \Pr(②) - \Pr(① \text{ or } ②)$$
$$\geq \Pr(①) + \Pr(②) - 1$$
$$= (1 - \Pr(①^c)) + (1 - \Pr(②^c)) - 1$$
$$= 1 - \Pr(①^c) + \Pr(②^c)$$

$$\Pr(|w_1| \leq \epsilon) \geq \Pr(① \text{ and } ②)$$
$$\geq 1 - \Pr(①^c) - \Pr(②^c)$$

$\Pr(①^c) = \Pr(\|g\|_2 < \sqrt{d/2}) \leq \frac{1}{2^{cd}}$ by Chi-squared concentration

$\Pr(②^c) = \Pr(|g_1| > \sigma\sqrt{d/2}) \leq \frac{1}{2^{(c\sigma\sqrt{d/2})^2}}$ by Gaussian tail bound

$$\geq 1 - \frac{1}{2^{c^2\sigma^2 d/2}} - \frac{1}{2^{cd}}$$

larger for small σ

$\boxed{\text{Unit Cube vs Sphere}}$

Stirling's Approximation

$$n! \approx \sqrt{2\pi n}\left(\frac{n}{e}\right)^n$$

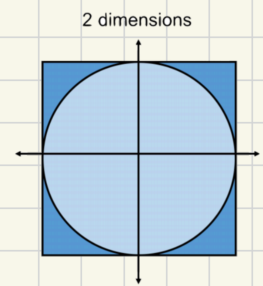
$$C_d = \{ x \in \mathbb{R}^d : |x_i| \le 1 \text{ for all } i \in [d] \}$$

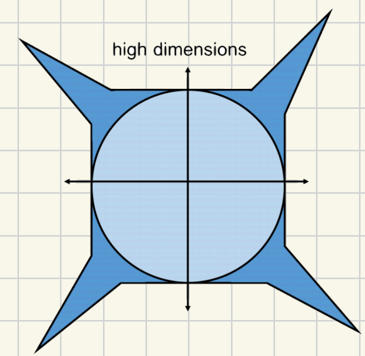
$$\frac{\mathrm{Vol}(C_d)}{\mathrm{Vol}(B_d)} = \frac{2^d (d/2)!}{\pi^{d/2}} \approx \frac{4^{d/2}\sqrt{2\pi d/2}\left(\frac{d/2}{e}\right)^{d/2}}{\pi^{d/2}}$$

$$= \sqrt{2\pi d/2}\left(\frac{2d}{\pi e}\right)^{d/2} \approx \sqrt{d}^{\;d}$$

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

$$\mathbb{E}_{x \sim B_d}\left[\|x\|_2^2\right] \le 1 \qquad \text{vs} \qquad \mathbb{E}_{x \sim C_d}\left[\|x\|_2^2\right] =$$





Next:

Play with high-dimensional data!

Primarily reduce dimension while retaining structure