

Tuesday, Feb 17

Plan

- Revisiting Load Balancing
- High dimensional geometry

Exponential Concentration Inequalities

x_1, \dots, x_n indep

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

↪ Make assumptions on x_i :

stronger assumption = stronger bound

↪ Proved using clever applications
of Markov's

Chernoff Bound

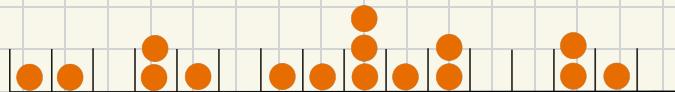
Binary $x_i \in \{0, 1\}$. For $0 < \epsilon < 1$,

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

For $\delta > 0$,

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

Revisiting Load Balancing



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

$$S = \max_i S_i$$

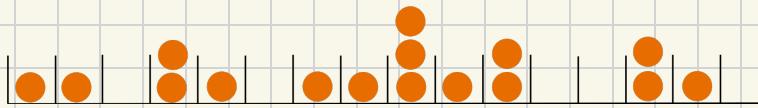
$$\Pr(S \geq c) \leq \frac{1}{10}$$

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

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

Q: Can we do better?

Power of Two Choices



use two hash functions,
choose least occupied

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

$$c = \log \log n$$

$$c = \log \log \log n$$

How about power of three choices?

High dimensional geometry

Unifying theme...

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

High-dimensional geometry is weird

Q: In d dimensions, how many orthogonal vectors are there?
 i.e., orthogonal if $\langle x, y \rangle = 0$

Setup

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

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

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

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

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,

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

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

$$\|x_i\|_2^2 =$$

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

$$\text{Var}(\langle x_i, x_j \rangle) =$$

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 were Gaussian,

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

and done by setting $\alpha = 6\sqrt{d}$.

But z is not Gaussian, so more work.

Chernoff? Write z as sum of binary.

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

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

$$\begin{aligned} z &= \sum_{k=1}^d c_k = \frac{z}{d} \sum_{k=1}^d \frac{d}{2} (B_k - 1/2) \\ &= \frac{z}{d} \sum_{k=1}^d (-1/2 + B_k) \\ &= \frac{z}{d} \left(-\frac{d}{2} + \sum_{k=1}^d B_k \right) \end{aligned}$$

$$z > 6 \iff$$

$$\sum_{k=1}^d B_k > \frac{d}{2} + \frac{d}{2} 6$$

$$z < 6 \iff$$

$$\sum_{k=1}^d B_k < \frac{d}{2} - \frac{d}{2} 6$$

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

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

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

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

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

Choose t so failure prob < t.

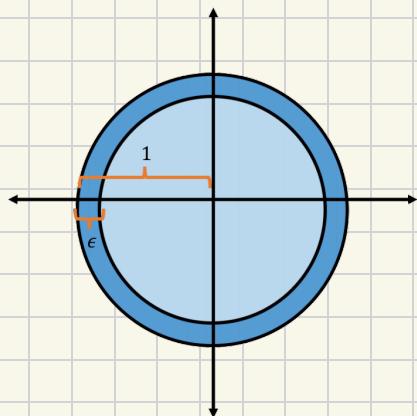
Takeaway: Random vectors
tend to be far apart in high d.

∴ working with high d
might seem hopeless, but
our data is typically not
random i.e. there is structure
we can use

High d is weird part 2: Where Random Points Lie

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

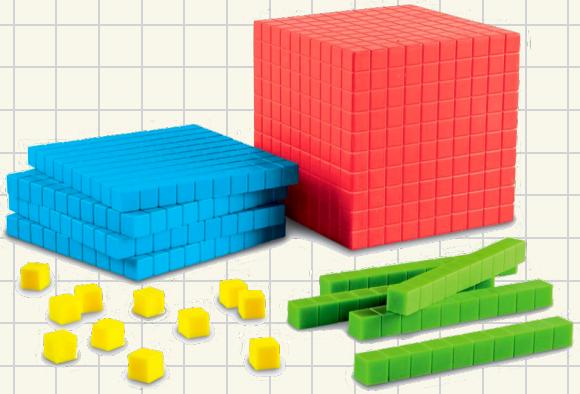
Q: What fraction of volume lies within ϵ of surface?



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

$$\begin{aligned}\frac{\text{Vol}(B_d(1)) - \text{Vol}(B_d(1-\epsilon))}{\text{Vol}(B_d(1))} &= 1 - (1-\epsilon)^d \\ &= 1 - ((1-\epsilon)/\epsilon)^{\epsilon d} \\ &\approx 1 - \frac{1}{e^{\epsilon d}}\end{aligned}$$

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

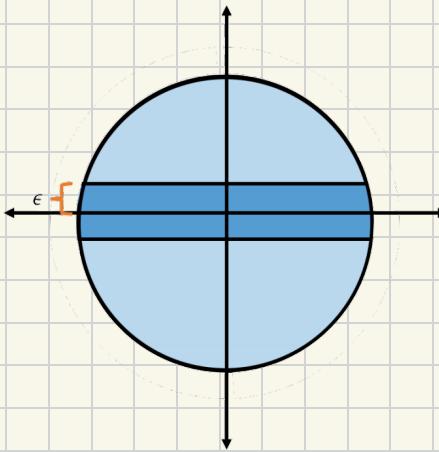


Q: What fraction of cubes
are near surface?

$d=1$:

$d=2$:

$d=3$:



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