2. Sampling via Gaussians

Friday, September 10, 2021 12:16 AM

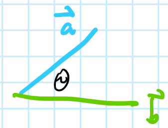
Last time: Most pts in hyperball are near the surface.

Today:
- Most pts in hyperball are near "equator", and therefore orthogonal.
- Sampling from a sphere (i.e. generating random directions) is hard.
- Gaussians to the rescue.

Next time:
- Gaussian annulus thm
- Johnson-Lindenstrauss Lemma and random projections

Most points in a unit ball are near equator

Recall dot product $\vec{a} \cdot \vec{b} = \sum_{i=1}^{d} a_i b_i = \|a\| \|b\| \cos\theta$

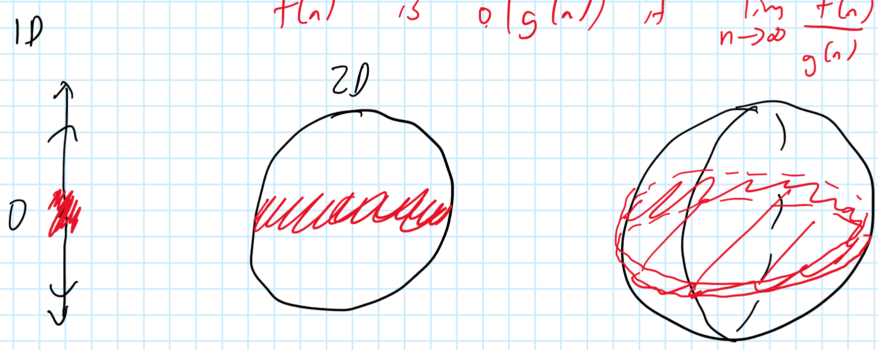


So $a \cdot b$ is small $\approx$ a + b are nearly orthogonal.
and $|a|$ and $|b|$ are not that small

WTS: most of the vol of the unit ball have $|x_1| \leq O\left(\frac{1}{\sqrt{d}}\right)$

Recall: big-O notation
$f(n)$ is $O(g(n))$ if $\exists c > 0$ s.t. $\forall n$, $f(n) \leq c\, g(n)$
$f(n)$ is $\Omega(g(n))$ if $\exists c > 0$ s.t. $\forall n$, $f(n) \geq c\, g(n)$
$f(n)$ is $\Theta(g(n))$ if both $f(n)$ is $O(g(n))$ and is $\Omega(g(n))$

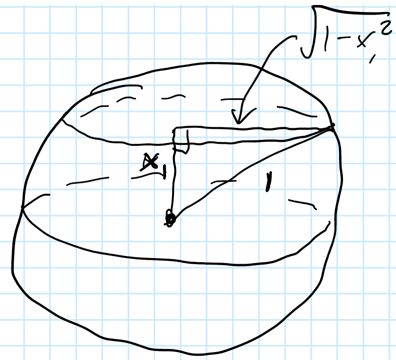
$f(n)$ is $o(g(n))$ if $\lim_{n\to\infty} \frac{f(n)}{g(n)} = 0$



Thm 2.7. For $c \geq 1$ and $d \geq 3$, at least $1 - \frac{2}{c} e^{-\frac{c^2}{2}}$ fraction of the vol of B_d has $|x_1| \leq \frac{c}{\sqrt{d-1}}$, $\vec{x} \in B_d$.

proof. By symmetry, consider just the top half.
Let $A = \{ \vec{x} \in B_d \mid x_1 \geq \frac{c}{\ldots} \}$

proof. By symmetry, consider just the top half.



Let $A = \left\{ \vec{x} \in B_d \mid x_1 \geq \frac{c}{\sqrt{d-1}} \right\}$

let $H = \left\{ \vec{x} \in B_d \mid x_1 \geq 0 \right\}$ upper hemisphere

Note
$$vol(A) = \int_{\frac{c}{\sqrt{d-1}}}^{1} \underbrace{\underbrace{(1 - x_1^2)^{\frac{d-1}{2}}}_{\text{scale down by } \sqrt{1-x_1^2}} \underbrace{Vol(B_{d-1})}_{\text{Vol of unit radius ball of dim } d-1}}_{\text{Vol of radius } \sqrt{1-x_1^2} \text{ ball of dim } d-1} dx_1$$

integrating over x_1 coord.

can integrate to ∞

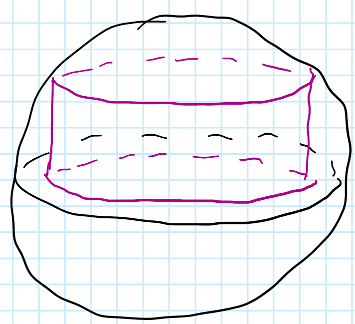
$$\leq \int_{\frac{c}{\sqrt{d-1}}}^{\infty} \underbrace{e^{-\frac{d-1}{2}x_1^2}}_{1-x \,\leq\, e^{-x}} Vol(B_{d-1}) dx_1$$

$$\leq \int_{\frac{c}{\sqrt{d-1}}}^{\infty} \frac{x_1\sqrt{d-1}}{c} e^{-\frac{d-1}{2}x_1^2} Vol(B_{d-1}) dx_1 \qquad \frac{x_1\sqrt{d-1}}{c} \geq 1 \text{ in } A$$

$$= Vol(B_{d-1}) \cdot \frac{\sqrt{d-1}}{c} \int_{\frac{c}{\sqrt{d-1}}}^{\infty} x_1 e^{-\frac{d-1}{2}x_1^2} dx_1$$

$$= Vol(B_{d-1}) \cdot \frac{\sqrt{d-1}}{c} \cdot \left[-\frac{1}{d-1} e^{-\frac{d-1}{2}x_1^2} \right] \Bigg|_{\frac{c}{\sqrt{d-1}}}^{\infty}$$

$$= \frac{Vol(B_{d-1})}{c\sqrt{d-1}} e^{-\frac{c^2}{2}} \quad \leftarrow \text{upper bound on } vol(A)$$



Also,
$$vol(H) \geq Vol\left(\left\{ \vec{x} \in H \mid x_1 \leq \frac{1}{\sqrt{d-1}} \right\} \right)$$
$$\geq Vol(B_{d-1}) \cdot \left(1 - \frac{1}{d-1}\right)^{\frac{d-1}{2}} \cdot \frac{1}{\sqrt{d-1}}$$
$$\geq \frac{Vol(B_{d-1})}{2\sqrt{d-1}} \qquad (1-x)^a \geq 1 - ax \text{ for } a \geq 1$$

Th... $vol(A)$ ... $\frac{Vol(B_{d-1})}{c\sqrt{d-1}} e^{-\frac{c^2}{2}}$ ... $2 \;\; e^{-\frac{c^2}{2}}$

Then $\quad \dfrac{vol(A)}{vol(H)} \le \dfrac{\dfrac{Vol(B_{d-1})}{c\sqrt{d-1}} e^{-\frac{c^2}{2}}}{\dfrac{Vol(B_{d-1})}{2\sqrt{d-1}}} = \dfrac{2}{c} e^{-\frac{c^2}{2}}$

$\square$

Thm 2.8 Consider drawing n points $\vec{x}_1, \ldots, \vec{x}_n \in B_d$ uniformly at random.
Then, with probability $1 - O\left(\frac{1}{n}\right)$

(1) $|\vec{x}_i| \ge 1 - \dfrac{2 \ln n}{d}$ for all i, and

(2) $|\vec{x}_i \cdot \vec{x}_j| \le \dfrac{\sqrt{6 \ln n}}{\sqrt{d-1}}$ for all $i \neq j$.

proof. (1) For any fixed i, $Prob(|\vec{x}_i| < 1 - \varepsilon) \le e^{-\varepsilon d}$ vol near surface

Then, $Prob\left(|\vec{x}_i| < 1 - \frac{2\ln n}{d}\right) \le e^{(-2\ln n)} = \frac{1}{n^2}$

By union bound, $Prob\left(\exists i \text{ s.t. } |\vec{x}_i| < 1 - \frac{2\ln n}{d}\right) \le \frac{1}{n}$

(2) By Thm 2.7 above,

$Prob\left(|\vec{x}_i \cdot \vec{e}_1| > \frac{c}{\sqrt{d-1}}\right) \le \frac{2}{c} e^{-\frac{c^2}{2}}$ for fixed i.

$\Rightarrow Prob\left(|\vec{x}_i \cdot \vec{e}_1| > \frac{\sqrt{6\ln n}}{\sqrt{d-1}}\right) \le \frac{2}{\sqrt{6 \ln n}} e^{-3\ln n} = \frac{2}{\sqrt{6\ln n}} \cdot \frac{1}{n^3} = O\left(\frac{1}{n^3}\right)$

true for any "north"!

There are $\binom{n}{2}$ pairs i and j, so for each pair, we define $\dfrac{\vec{x}_j}{|\vec{x}_j|}$ as "north"

By union bound for all $i \neq j$, the dot product condition fails with prob. at most $O\left(\binom{n}{2} n^{-3}\right) = O\left(\frac{1}{n}\right)$

$\square$

Corollary: $\lim_{d \to \infty} Vol(B_d) = 0$

Not obvious

Not obvious

$\text{vol} = 2 \quad < \quad \text{vol} = \pi \quad < \quad \text{vol} = \frac{4}{3} \cdot \pi$

proof. Let $c = 2\sqrt{\ln d}$. By Thm 2.7, at a $1 - \frac{1}{\sqrt{\ln d}} e^{-2\ln d} = 1 - \frac{1}{d^2\sqrt{\ln d}}$ fraction of the volume has $|x_1| \le \frac{2\sqrt{\ln d}}{\sqrt{d-1}}$.

$\Rightarrow$ For $\vec{x} \in B_d$ randomly drawn, $\text{Prob}\left(|x_1| > \frac{2\sqrt{\ln d}}{\sqrt{d-1}}\right) < \frac{1}{d^2\sqrt{\ln d}} < \frac{1}{d^2}$.

Let C be a box/hypercube with side-length $\frac{4\sqrt{\ln d}}{\sqrt{d-1}}$ centered at origin.

If $\vec{z} \in C \cap B_d$, then $|z_i| \le \frac{2\sqrt{\ln d}}{\sqrt{d-1}}$ for all $i \in \{1, \dots, d\}$.

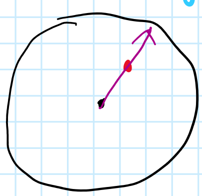
By union bound, $\text{Prob}\left(|x_i| > \frac{2\sqrt{\ln d}}{\sqrt{d-1}},\ \forall i\right) < \frac{d}{d^2} \le \frac{1}{2}$ for $d \ge 2$.

$\Rightarrow \text{Vol}(C \cap B_d) \ge \frac{1}{2} \text{Vol}(B_d)$

But $\text{Vol}(C) = \left(\frac{4\sqrt{\ln d}}{\sqrt{d-1}}\right)^d = \left(\frac{16 \ln d}{d-1}\right)^{\frac{d}{2}} \to 0$ as $d \to \infty$

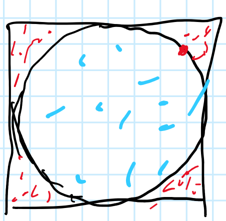
$\Rightarrow \text{Vol}(B_d) \to 0$ as $d \to \infty$. $\boxed{\diagup}$

How do we generate directions randomly?



One sol: sample ball and project out to sphere

How to sample from ball?



Naive solution: (1) sample from covering cube.

(rejection sampling) (2) reject points outside the ball

Fails in high dimensions because $\text{vol}(B_d) \to 0$ and $\text{vol}(\text{cube}) \to 2^d$.

Gaussian variable to sample directions

Gaussian variable to sample directions

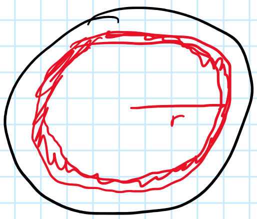
pdf of 1D Gaussian is $p(x) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{x^2}{2}\right)$

let $\vec{x} = \{x_1, \dots, x_d\}$ where $x_i \sim \mathcal{N}(0,1)$ i.i.d.

Then
$$\text{pdf}(\vec{x}) = \frac{1}{(2\pi)^{d/2}} \exp\underbrace{\left(\frac{-(x_1^2 + \dots + x_d^2)}{2}\right)}_{\text{spherically symmetric.}}$$

Then, for sampling hypersphere, normalize to unit length $\frac{\vec{x}}{|\vec{x}|}$.

To sample from hyperball:



Need to know volume in each shell

$\text{Vol}(r^d B_d) = r^d \, \text{Vol}(B_d)$ ← volume in ball up to radius r

$$\frac{d}{dr}\left[r^d \, \text{Vol}(B_d)\right] = \underbrace{d\,r^{d-1}}\,\text{Vol}(B_d)$$
└ use as pdf of radial coordinate.

Let ρ be a r.v. with p.d.f. $p(r) = d\,r^{d-1}$ for $r \in [0,1]$

Then take $\vec{y} = \rho \cdot \frac{\vec{x}}{|\vec{x}|}$ to get a random pt in B_d.