1 High dimensional space

Sunday, January 5, 2020 10:21 PM

High dimensional space & Johnson-Lindenstrauss Lemma

Key idea: often we can prove that a random variable only rarely deviates much from its expected behavior.

Let's recall a few standard examples from probability & statistics.

Theorem 2.1 (Markov's Inequality) Let x be a nonnegative random variable. Then for $a > 0$,

$$\text{Prob}(x \geq a) \leq \frac{\mathbb{E}(x)}{a}.$$

proof. Let x be continuous with probability density p. (similar for discrete)

$$\mathbb{E}x = \int_0^\infty x\,p(x)\,dx = \int_0^a x\,p(x)\,dx + \int_a^\infty x\,p(x)\,dx$$

$$\geq \int_a^\infty x\,p(x)\,dx \geq a\int_a^\infty p(x)\,dx = a \cdot \text{Prob}(x \geq a)$$

$$\Rightarrow \text{Prob}(x \geq a) \leq \frac{\mathbb{E}x}{a}.$$

$\boxtimes$

Corr. $\text{Prob}(x \geq b\,\mathbb{E}x) \leq \frac{1}{b}$.

Theorem 2.3 (Chebyshev's inequality): Let x be a r.v. with bounded variance. Then for $c > 0$,

$$\text{Prob}(|x - \mathbb{E}x| \geq c) \leq \frac{\text{Var}(x)}{c^2}.$$

Proof. $\text{Prob}(|x - \mathbb{E}x| \geq c) = \text{Prob}(|x - \mathbb{E}x|^2 \geq c^2)$.

Let $y = |x - \mathbb{E}x|^2$. Then y is a nonnegative r.v. and $\mathbb{E}y = \text{Var}(x)$.

Then by Markov's inequality,

$$\text{Prob}(|x - \mathbb{E}x| \geq c) = \text{Prob}(y \geq c^2) \leq \frac{\mathbb{E}y}{c^2} = \frac{\text{Var}(x)}{c^2}.$$

$\boxtimes$

Theorem 2.4 (Law of Large Numbers)

Let $x_1, \ldots, x_n$ be independent samples of a r.v. x. Then

$$\text{Prob}\left(\left|\frac{x_1 + \cdots + x_n}{n} - \mathbb{E}x\right| \geq \varepsilon\right) < \frac{\text{Var}(x)}{n\varepsilon^2}.$$

Proof. By Chebyshev's inequality,

$$\text{Prob}\left(\left|\frac{x_1+\cdots+x_n}{n}-\mathbb{E}x\right|\geq\varepsilon\right)\leq\frac{\text{Var}\left(\frac{x_1+\cdots+x_n}{n}\right)}{\varepsilon^2}$$

$$=\frac{1}{n^2\varepsilon^2}\text{Var}(x_1+\cdots+x_n)$$

$$=\frac{1}{n^2\varepsilon^2}\left(\text{Var}(x_1)+\cdots+\text{Var}(x_n)\right)$$

$$=\frac{\text{Var}(x)}{n\varepsilon^2}.$$

So why have we spent the last 10 minutes reviewing basic probability?

Because these tail bounds let us analyze "typical" random points in high dimensional space

High dimensional space different from ordinary 2D & 3D space.

Most volume of hypersphere concentrated near surface. (also near equator)

Volume of unit ball goes to 0 as $d\to\infty$

Let's analyze the behavior of random Gaussians, the unit hypersphere, and unit hypercube

(review some useful mathematical facts)

Consider random Gaussian pts in $\mathbb{R}^d$.

i.e. Let $\mathbf{y}=[y_1,\ldots,y_d]\in\mathbb{R}^d$, $\mathbf{z}=[z_1,\ldots,z_d]\in\mathbb{R}^d$, y_i's and z_i's are i.i.d. with p.d.f. $\varphi(x)=\frac{1}{\sqrt{2\pi}}e^{-\frac{1}{2}x^2}$ (i.e. $y_i\sim\mathcal{N}(0,1)$)

(i.e. unit-variance 1D Gaussians centered at 0)

What can we say about $|\mathbf{y}-\mathbf{z}|$? (the distance from $|y-z|$ is tightly bounded with high probability)

Note that $|y-z|^2=\sum_{i=1}^{d}(y_i-z_i)^2$.

Let $x_i=(y_i-z_i)^2$, which is a r.v. with bounded variance.

Then by the law of large numbers,

$$\text{Prob}\left(\left|\frac{x_1+\cdots+x_n}{n}-E(x)\right|\geq\varepsilon\right)\leq\frac{\text{Var}(x)}{n\varepsilon^2}.$$

So the distance is close to the expected difference

[illegible]

So the distance is close to the expected difference.

In the following, we will prove much stronger / rigorous facts about random Gaussians and hyperspheres in high-dimensional space.

Thm 2.5 (Master Tail Bounds Theorem)

Let $x = x_1 + x_2 + \cdots + x_n$, where $x_1, \ldots, x_n$ are mutually independent r.v. with zero mean and variance at most σ^2. Let $0 \le a \le \sqrt{2}\, n \sigma^2$.
Assume that $|\mathbb{E}\, x_i^s| \le \sigma^2 s!$ for $s = 3, 4, \ldots, \lfloor \frac{a^2}{4n\sigma^2} \rfloor$.
Then $\text{Prob}(|x| \ge a) \le 3e^{-a^2/(12 n \sigma^2)}$. (See proof in book)

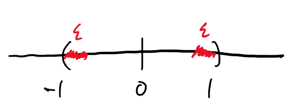
Useful tail bounds:

	Condition	Tail bound
Markov	$x \ge 0$	$\text{Prob}(x \ge a) \le \frac{\mathbb{E}\, x}{a}$
Chebyshev	Any x	$\text{Prob}(\lvert x - \mathbb{E}x \rvert \ge a) \le \frac{\text{Var}(x)}{a^2}$
Chernoff	$x = x_1 + \cdots + x_n$, $x_i \in [0,1]$ iid Bernoulli	$\text{Prob}(\lvert x - \mathbb{E}x \rvert \ge \varepsilon\, \mathbb{E}x) \le 3e^{-c\varepsilon^2 \mathbb{E}x}$ Many different forms based on proof.
Higher Moments	r positive even integer	$\text{Prob}(\lvert x \rvert \ge a) \le \frac{\mathbb{E}\, x^r}{a^r}$ (from Markov)
Gaussian Annulus	$x = \sqrt{x_1^2 + x_2^2 + \cdots + x_n^2}$, $x_i \sim N(0,1)$; $\beta \le \sqrt{n}$	$\text{Prob}(\lvert x - \sqrt{n} \rvert \ge \beta) \le 3e^{-c\beta^2}$

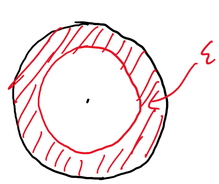
Used later to prove J-L.

Most volume of high-dimensional objects is near the surface

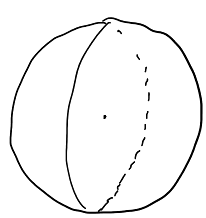
Ex.



Vol frac $\frac{2\varepsilon}{2} = \varepsilon$



Vol frac $\approx \frac{2\pi\varepsilon}{\pi} = 2\varepsilon$



Vol frac $\approx \frac{4\pi\varepsilon}{4\pi} = 3\varepsilon$

Vol frac $\approx \frac{2\pi\varepsilon}{\pi} = 2\varepsilon$ Vol frac $\approx \frac{4\pi\varepsilon}{\frac{4}{3}\pi} = 3\varepsilon$

As dimension increases, a larger fraction of the volume of a hypersphere is within distance ε of the surface.

More rigorously, consider any object $A \subset \mathbb{R}^d$. Shrink A by ε to produce $(1-\varepsilon)A = \{(1-\varepsilon)x \mid x \in A\}$. Then

$$\mathrm{vol}\big((1-\varepsilon)A\big) = (1-\varepsilon)^d \, \mathrm{vol}(A).$$

Proof sketch: Partition A into infinitesimal cubes. Then $(1-\varepsilon)A$ is the union of a set of cubes obtained by shrinking the cubes in A by a factor of $(1-\varepsilon)$.

A d-dimensional cube with side-length s has volume s^d. Shrinking by $(1-\varepsilon)$ gives side-lengths $(1-\varepsilon)s$, implying volume $(1-\varepsilon)^d s^d$. $\square$

Note that $1-x \le e^{-x}$, so for any object A in $\mathbb{R}^d$,

$$\frac{\mathrm{vol}\big((1-\varepsilon)A\big)}{\mathrm{vol}(A)} = (1-\varepsilon)^d \le e^{-\varepsilon d} \longrightarrow 0 \text{ as } d \to \infty.$$

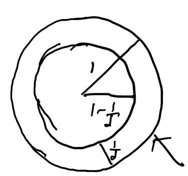
Thus, most of the volume does not belong to $(1-\varepsilon)A$.

Going back to the unit ball, let S be the d-dimensional unit ball.

Then $\frac{\mathrm{vol}\big((1-\varepsilon)S\big)}{\mathrm{vol}(S)} \le e^{-\varepsilon d}$, so $\mathrm{vol}\big(S \setminus (1-\varepsilon)S\big) \ge \big(1-e^{-\varepsilon d}\big)\,\mathrm{vol}(S)$.

Let $\varepsilon = \frac{1}{d}$. $\mathrm{vol}\big(S \setminus (1-\frac{1}{d})S\big) \ge (1-e^{-1})\,\mathrm{vol}(S) \approx 0.632\,\mathrm{vol}(S)$.

So most of the volume is contained in an annulus of width $\frac{1}{d}$ near the boundary.



Most points in a unit ball are nearly orthogonal (in high dimensions)

Recall that the dot product of two vectors $a \cdot b = \sum_{i=1}^{d} a_i b_i = \|a\| \|b\| \cos\theta$, where θ is the angle b/t the two vectors.

Thus, $a \cdot b$ is small when a & b are nearly orthogonal.

We can fix WLOG any unit vector as "north", corresponding to the first coordinate vector.

Then the dot product of a vector x with that coordinate vector is the first entry x_1.

We want to show that most of the volume is concentrated near the equator.

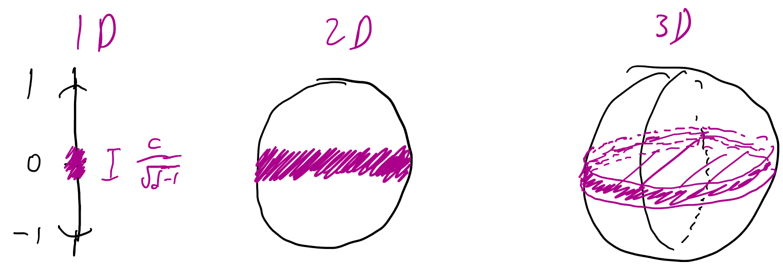
i.e. most of the volume of the unit ball has $|x_1| \le O\left(\frac{1}{\sqrt{d}}\right)$

Recall that big-O notation: $f(n)$ is $O(g(n))$ if $\exists c>0$ s.t. $\forall n$, $f(n) \le c g(n)$

$f(n)$ is $\Omega(g(n))$ if $\exists c>0$ s.t. $\forall n$, $f(n) \ge c g(n)$

$f(n)$ is $\Theta(g(n))$ if $f(n)$ is both $O(g(n))$ and $\Omega(g(n))$

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



As we increase dimensions, the equator gets "fatter", taking up more volume.

Thm. 2.7. For $c \ge 1$ and $d \ge 3$, at least a $1 - \frac{2}{c} e^{-c^2/2}$ fraction of the volume of the d-dimensional unit ball has $|x_1| \le \frac{c}{\sqrt{d-1}}$, where x_1 is the first coordinate.

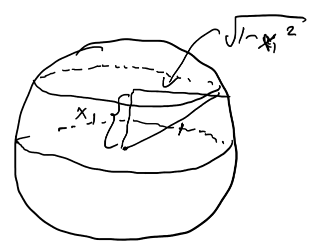
proof. By symmetry, need only consider top half of the ball.

i.e. show at most $\frac{2}{c} e^{-c^2/2}$ fraction of the half the ball with $x_1 \ge 0$ has $x_1 \ge \frac{c}{\sqrt{d-1}}$.

Let A be the part of the ball with $x_1 \ge \frac{c}{\sqrt{d-1}}$.

H be the upper hemisphere $(x_1 \ge 0)$.

Note that
$$\mathrm{vol}(A) = \underbrace{\int_{\frac{c}{\sqrt{d-1}}}^{1}}_{\text{integrating over } x_1 \text{ coordinate}} \underbrace{\underbrace{(1-x_1^2)^{\frac{d-1}{2}}}_{\text{scale down by } \sqrt{1-x_1^2}} \underbrace{V(d-1)}_{\text{volume of unit radius ball of dimension } d-1}}_{\text{volume of radius } \sqrt{1-x_1^2} \text{ ball of dim } d-1} dx_1$$



$$\le \int_{\frac{c}{\sqrt{d-1}}}^{\infty} (1-x_1^2)^{\frac{d-1}{2}} V(d-1)\, dx_1$$

integrate to infinity

$$\le \int_{\frac{c}{\cdots}}^{\infty} e^{-\frac{d-1}{2} x_1^2} V(d-1)\, dx_1$$

$1 - x \le e^{-x}$

$$\leq \int_{\frac{c}{\sqrt{d-1}}} e^{-\frac{d-1}{2} x_1^2} V(d-1)\, dx_1$$

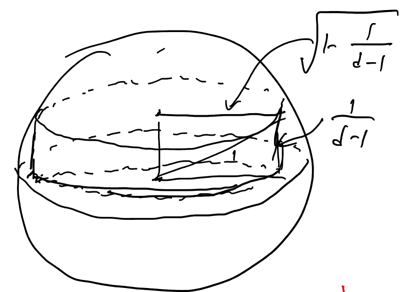
$$\leq \int_{\frac{c}{\sqrt{d-1}}}^{\infty} \frac{x_1\sqrt{d-1}}{c} e^{-\frac{d-1}{2}\cdot x_1^2} V(d-1)\, dx_1 \qquad \frac{x_1\sqrt{d-1}}{c} \geq 1 \text{ in integral range.}$$

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

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

$$= V(d-1)\cdot\frac{\sqrt{d-1}}{c}\cdot\left[\frac{1}{d-1} e^{-\frac{c^2}{2}}\right] = \frac{V(d-1)}{c\sqrt{d-1}}\cdot e^{-\frac{c^2}{2}}.$$

This gives us an upper bound on vol(A).
Now just need a good lower bound of vol(H).



$$Vol(H) \geq Vol\left(H \setminus \{x_1 > \tfrac{1}{\sqrt{d-1}}\}\right)$$
$$\geq V(d-1)\left(1-\frac{1}{d-1}\right)^{\frac{d-1}{2}}\cdot\frac{1}{\sqrt{d-1}}$$
$$\geq \frac{V(d-1)}{2\sqrt{d-1}} \qquad (1-x)^a > 1-ax \text{ for } a \geq 1$$

(Obviously not tight)

Thus,
$$\frac{vol(A)}{vol(H)} \leq \frac{\frac{V(d-1)}{c\sqrt{d-1}} e^{-\frac{c^2}{2}}}{\frac{V(d-1)}{2\sqrt{d-1}}} = \frac{2}{c} e^{-\frac{c^2}{2}}.$$

$\blacksquare$

Thus, most of the volume of the ball is near the equator.

But we also showed earlier that most of the volume of the ball is near the surface, and so has close to unit norm.

Given a pair of vectors in the ball, with high probability one has close to unit norm, and can be chosen as our north pole. But then we just showed that most other vectors are almost orthogonal to the north pole so most pairs of vectors are orthogonal.

Here we just showed that most other vectors are almost orthogonal to the north pole, so most pairs of vectors are orthogonal.

Thm 2.8 Consider drawing n points $x_1, \dots, x_n$ at random from the unit ball. With probability $1 - O(1/n)$

(1.) $|x_i| \geq 1 - \frac{2 \ln n}{d}$ for all i, and

(2.) $|x_i \cdot x_j| \leq \frac{\sqrt{6 \ln n}}{\sqrt{d-1}}$ for all $i \neq j$.

proof. For any fixed i, $\text{Prob}(|x_i| < 1 - \varepsilon) \leq e^{-\varepsilon d}$ (volume near surface)

Thus, $\text{Prob}\left(|x_i| < 1 - \frac{2 \ln n}{d}\right) \leq e^{-\left(\frac{2 \ln n}{d}\right) d} = \frac{1}{n^2}$

By union bound $\text{Prob}\left(\exists i \text{ s.t. } |x_i| < 1 - \frac{2 \ln n}{d}\right) \leq \frac{1}{n}$. (first part)

(second part) By Thm 2.7 above,

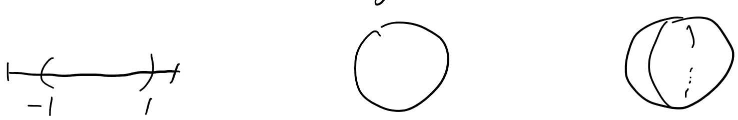
$$\text{Prob}\left(|x_i| > \frac{c}{\sqrt{d-1}}\right) \leq \frac{2}{c} e^{-\frac{c^2}{2}}.$$

$$\text{Prob}\left(|x_i| > \frac{\sqrt{6 \ln n}}{\sqrt{d-1}}\right) \leq \frac{2}{\sqrt{6 \ln n}} e^{-3 \ln n} = \frac{2}{\sqrt{6 \ln n}} \cdot \frac{1}{n^3} = O(n^{-3})$$

But this is true once a "north" is fixed.
There are $\binom{n}{2}$ pairs i and j and for each such pair we define x_j as "north".
By a union bound, the dot-product condition is violated with probability at most $O\left(\binom{n}{2} n^{-3}\right) = O\left(\frac{1}{n}\right)$. $\square$

Corollary: The volume of the unit ball goes to 0 as $d \to \infty$

Not obvious



vol = 2 vol = π vol = $\frac{4}{3} \cdot \pi$

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

Or alternately ... than $\frac{1}{\dots} < \frac{1}{d^2}$ of the volume has $|x_1| > \frac{2\sqrt{\ln d}}{\dots}$

Or alternately, no more than $\frac{1}{d^2\sqrt{\ln d}} < \frac{1}{d^2}$ of the volume has $|x_i| > \frac{2\sqrt{\ln d}}{\sqrt{d-1}}$.

Consider a box with side-length $\frac{4\sqrt{\ln d}}{\sqrt{d-1}}$, centered at the origin.

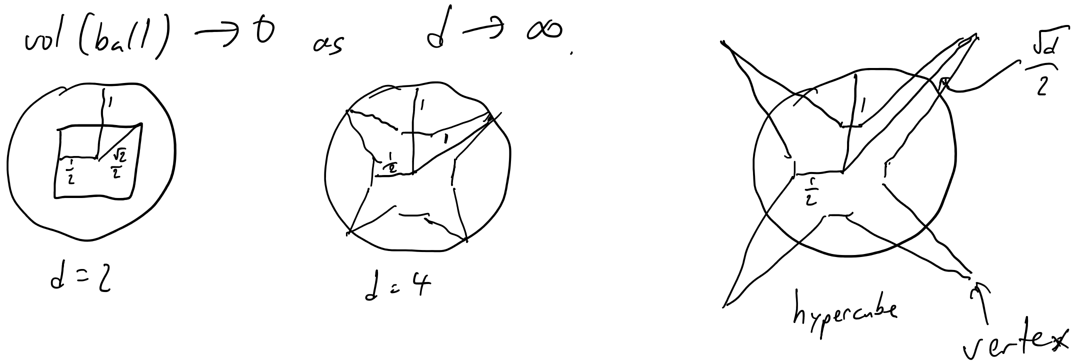
The intersection of this box with the unit ball requires $|x_i| \le \frac{2\sqrt{\ln d}}{\sqrt{d-1}}$, for all i in $1, \dots, d$.

By the union bound, at most a $O\left(\frac{1}{d}\right) \le \frac{1}{2}$ fraction of the volume of the ball lies outside this box.

So at least half of the ball is in the box.
Thus, the volume of the ball cannot be more than twice the volume of the box.

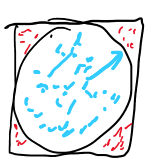
But $\text{vol}(\text{box}) = O\left(\left(\frac{\ln d}{d-1}\right)^{d/2}\right) \to 0$ as $d \to \infty$.

Thus $\text{vol}(\text{ball}) \to 0$ as $d \to \infty$.



Implication for generating points uniformly at random from a sphere (i.e. random directions)

Naive solution is to generate in a covering cube, and project onto the surface, discarding points that fall outside the sphere.



This fails in higher dimensions because the volume of the unit ball goes to 0 while the covering cube's vol grows.

Instead, we can generate points whose coordinates are ind. Gaussians.

Generate $x_1, \dots, x_d \sim N(0,1)$, i.e. with p.d.f. $\frac{1}{\sqrt{2\pi}} \exp(-x^2/2)$.

Then the p.d.f. of $x \in \mathbb{R}^d$ is $\frac{1}{\dots} \exp\left(-\frac{x_1^2 + x_2^2 + \dots + x_d^2}{\dots}\right)$

Then the p.d.f. of $x \in \mathbb{R}^d$ is $\frac{1}{(2\pi)^{d/2}} \exp\left(-\frac{x_1^2 + x_2^2 + \cdots + x_d^2}{2}\right)$,

which is spherically symmetric.

Then simply project onto the sphere by normalizing to unit length.

Next time:

- Gaussian annulus theorem
- Random projections and Johnson-Lindenstrauss
- Separating Gaussians in high dimensions