UC San Diego

Lecture 11:

High-Dimensional Geometry

Focus of Past High-D Geometry Research

From Low-D to High-D, Change of View

- From calculus view to statistical view
- From infinitesimal analysis to structural/topological analysis

- We will focus on the “odd” behaviors in high-D geometry

Some basic tools

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

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

Proof on board

Theorem 2.3 (Chebyshev's inequality) *Let x be a random variable with mean m and variance σ^2. Then*

$$Prob(|x - m| \geq a\sigma) \leq \frac{1}{a^2}.$$

Theorem 2.4 (Law of large numbers) *Let $x_1, x_2, \ldots, x_n$ be n samples of a random variable x. Then*

$$Prob\left(\left|\frac{x_1 + x_2 + \cdots + x_n}{n} - E(x)\right| > \epsilon\right) \leq \frac{\sigma^2}{n\epsilon^2}$$

Volume & Surface Distribution

Volume of Unit Sphere Goes to Zero

- Sphere volume:

$$V(d) = \int_{S^d} \int_{r=0}^{1} r^{d-1} dr d\Omega.$$

At radius r, the surface area of the top of the cone is $r^{d-1}d\Omega$ since the surface area is d – 1 dimensional and each dimension scales by r

$$V(d) = \int_{S^d} d\Omega \int_{r=0}^{1} r^{d-1} dr = \frac{1}{d} \int_{S^d} d\Omega = \frac{A(d)}{d}$$

Volume of Unit Sphere Goes to Zero

- Computation of A(d)
  - Trick: compare the integration of e^{-x^2} in Cartesian and Polar systems

Consider a different integral

$$I(d) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty} e^{-\left(x_1^2+x_2^2+\cdots x_d^2\right)} dx_d \cdots dx_2 dx_1 .$$

First, calculate $I(d)$ by integration in Cartesian coordinates.

$$I(d) = \left[\int_{-\infty}^{\infty} e^{-x^2} dx\right]^d = \left(\sqrt{\pi}\right)^d = \pi^{\frac{d}{2}} .$$

But

$$I(d) = \int_{S^d} d\Omega \int_0^{\infty} e^{-r^2} r^{d-1} dr .$$

Volume of Unit Sphere Goes to Zero

$$I\left(d\right)=\left[\int_{-\infty}^{\infty}e^{-x^2}dx\right]^d=\left(\sqrt{\pi}\right)^d=\pi^{\frac{d}{2}}.$$

$$I\left(d\right)=\int_{S^d}d\Omega\int_0^{\infty}e^{-r^2}r^{d-1}dr.$$

$$\int_0^{\infty}e^{-r^2}r^{d-1}dr=\frac{1}{2}\int_0^{\infty}e^{-t}t^{\frac{d}{2}}-1dt=\frac{1}{2}\Gamma\left(\frac{d}{2}\right)$$

and hence, $I(d)=A(d)\frac{1}{2}\Gamma\left(\frac{d}{2}\right)$

Lemma 2.5 *The surface area $A(d)$ and the volume $V(d)$ of a unit-radius sphere in d dimensions are given by*

$$A\left(d\right)=\frac{2\pi^{\frac{d}{2}}}{\Gamma\left(\frac{d}{2}\right)}\quad \text{and} \quad V\left(d\right)=\frac{2}{d}\frac{\pi^{\frac{d}{2}}}{\Gamma\left(\frac{d}{2}\right)}.$$

Volume of Unit Sphere Goes to Zero

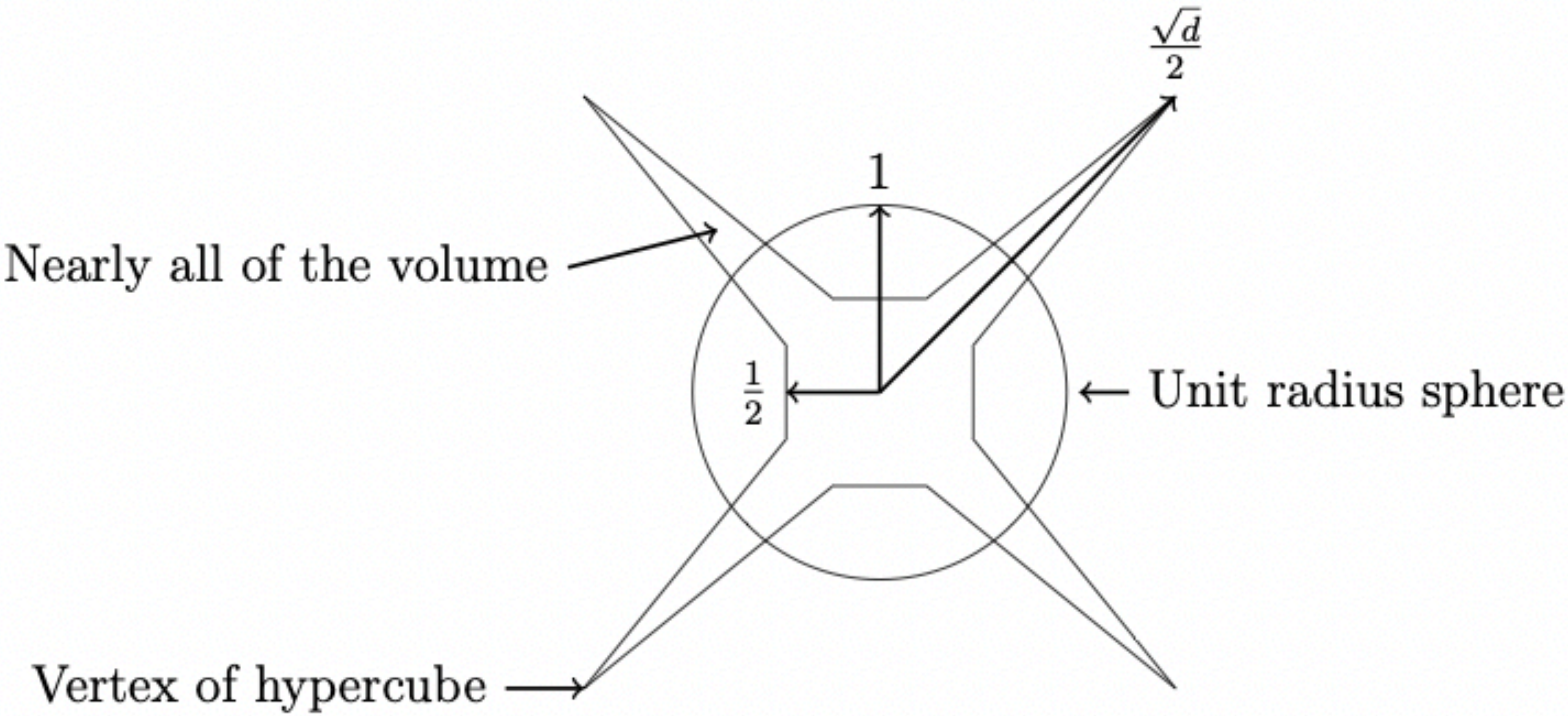


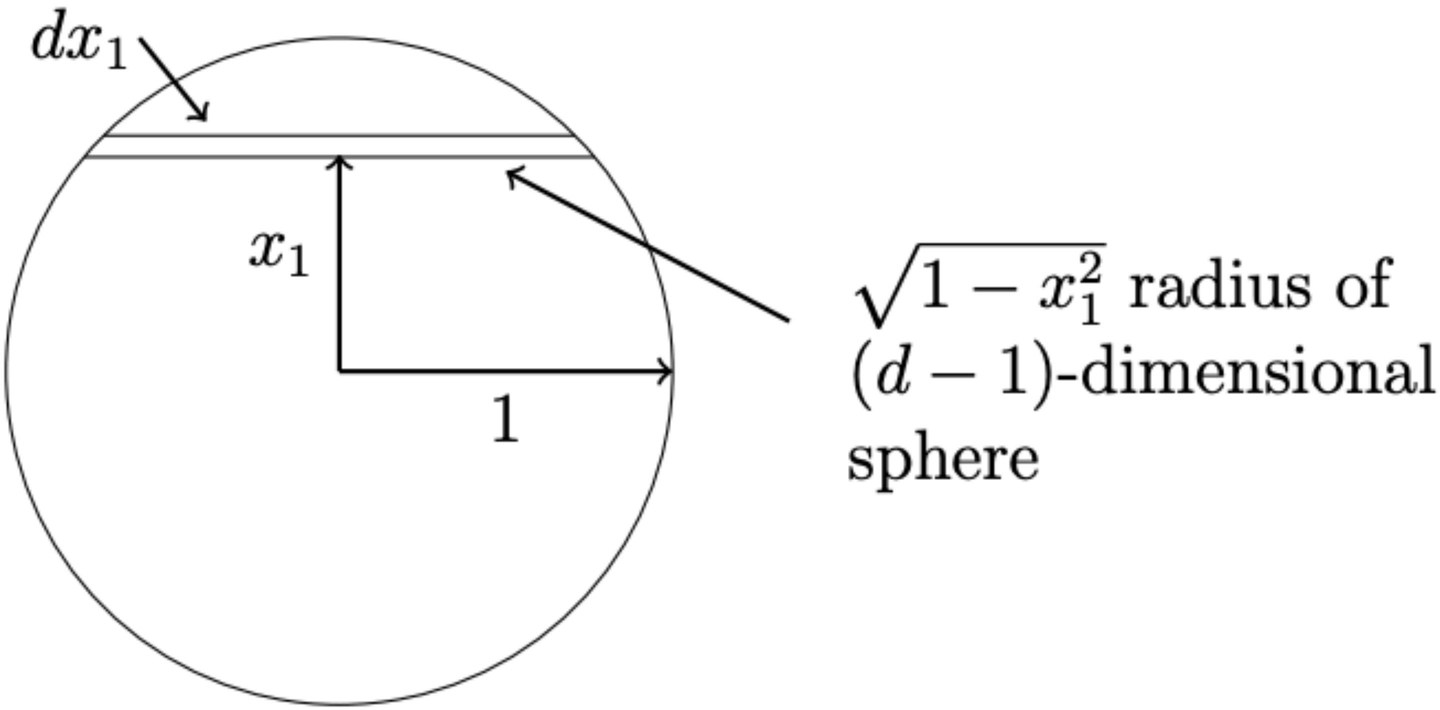
FIGURE 2.5: Conceptual drawing of a sphere and a cube.

The Volume is Near the Equator

- It turns out that essentially all of the volume of the upper hemisphere lies between the plane x1 = 0 and a parallel plane, x1 = ε, that is slightly higher.

- For what value of ε does essentially all the volume lie between x1 = 0 and x1 = ε ? The answer depends on the dimension. For dimension d, it is $O(\frac{1}{\sqrt{d-1}})$

The Volume is Near the Equator



The Volume is Near the Equator

$$\frac{\text{Volume above slice}}{\text{Volume upper hemisphere}} \leq \frac{\text{Upper bound volume above slice}}{\text{Lower bound volume upper hemisphere}}$$

Let $T = \left\{ \mathbf{x} \,\middle|\, |\mathbf{x}| \leq 1, x_1 \geq \varepsilon \right\}$ be the portion of the sphere above the slice.

$$\text{Volume}\,(T) = \int_{\varepsilon}^{1} \left(1 - x_1^2\right)^{\frac{d-1}{2}} V\,(d-1)\, dx_1 = V\,(d-1) \int_{\varepsilon}^{1} \left(1 - x_1^2\right)^{\frac{d-1}{2}} dx_1.$$

use the inequality $1 + x \leq e^x$ for all real x

$$\text{Volume}\,(T) \leq V\,(d-1) \int_{\varepsilon}^{\infty} e^{-\frac{d-1}{2} x_1^2} dx_1$$

$$\leq V(d-1) \int_{\varepsilon}^{\infty} \frac{x_1}{\varepsilon} e^{-\frac{d-1}{2} x_1^2} dx_1.$$

$$\text{Volume}\,(T) \leq \tfrac{1}{\varepsilon(d-1)} e^{-\frac{d-1}{2}\varepsilon^2} V\,(d-1)$$

The Volume is Near the Equator

Approximate Volume of upper hemisphere:

volume of the entire upper hemisphere. Clearly, the volume of the upper hemisphere is at least the volume between the slabs $x_1 = 0$ and $x_1 = \frac{1}{\sqrt{d-1}}$, which is at least the volume of the cylinder of radius $\sqrt{1 - \frac{1}{d-1}}$ and height $\frac{1}{\sqrt{d-1}}$. The volume of the cylinder is $1/\sqrt{d-1}$ times the $d-1$-dimensional volume of the disk $R = \left\{ \mathbf{x} \middle| \ |\mathbf{x}| \leq 1; x_1 = \frac{1}{\sqrt{d-1}} \right\}$. Now R is a $d-1$-dimensional sphere of radius $\sqrt{1 - \frac{1}{d-1}}$ and so its volume is

$$\text{Volume}(R) = V(d-1)\left(1 - \frac{1}{d-1}\right)^{(d-1)/2}.$$

Using $(1-x)^a \geq 1 - ax$

$$\text{Volume}(R) \geq V(d-1)\left(1 - \frac{1}{d-1}\frac{d-1}{2}\right) = \frac{1}{2}V(d-1).$$

Thus, the volume of the upper hemisphere is at least $\frac{1}{2\sqrt{d-1}}V(d-1)$.

The Volume is Near the Equator

Lemma 2.6 *For any $c > 0$, the fraction of the volume of the unit hemisphere above the plane $x_1 = \frac{c}{\sqrt{d-1}}$ is less than $\frac{2}{c}e^{-c^2/2}$.*

Proof: Substitute $\frac{c}{\sqrt{d-1}}$ for ε in the above. ∎

The Volume is in a Narrow Annulus

The ratio of the volume of a sphere of radius $1-\varepsilon$ to the volume of a unit sphere in d-dimensions is

$$\frac{(1-\varepsilon)^d V(d)}{V(d)} = (1-\varepsilon)^d,$$

and thus goes to zero as d goes to infinity when ε is a fixed constant. In high dimensions, all of the volume of the sphere is concentrated in a narrow annulus at the surface.

Since, $(1-\varepsilon)^d \leq e^{-\varepsilon d}$, if $\varepsilon = \frac{c}{d}$, for a large constant c, all but e^{-c} of the volume of the sphere is contained in a thin annulus of width c/d. The important item to remember is that most of the volume of the d-dimensional unit sphere is contained in an annulus of width $O(1/d)$ near the boundary. If the sphere is of radius r, then for sufficiently large d, the volume is contained in an annulus of width $O\left(\frac{r}{d}\right)$.

Gaussian in High Dimension

Gaussian Distribution

$$p(\mathbf{x}) = \frac{1}{(2\pi)^{d/2}\,\sigma^d} \exp\left(-\frac{|\mathbf{x}|^2}{2\sigma^2}\right)$$

Expected squared distance of a point from the center of a Gaussian

- A 1-dimensional Gaussian has its mass close to the origin.

Expected squared distance of a point from the center of a Gaussian

- However, as the dimension is increased something different happens
- When $\sigma^2 = 1$, integrating the probability density over a unit sphere centered at the origin yields nearly zero mass since the volume of a unit sphere is negligible.

Expected squared distance of a point from the center of a Gaussian

- However, as the dimension is increased something different happens
- When $\sigma^2 = 1$, integrating the probability density over a unit sphere centered at the origin yields nearly zero mass since the volume of a unit sphere is negligible.
- In fact, one needs to increase the radius of the sphere to $\sqrt{d}$ before there is a significant nonzero volume and hence a nonzero probability mass.

Expected squared distance of a point from the center of a Gaussian

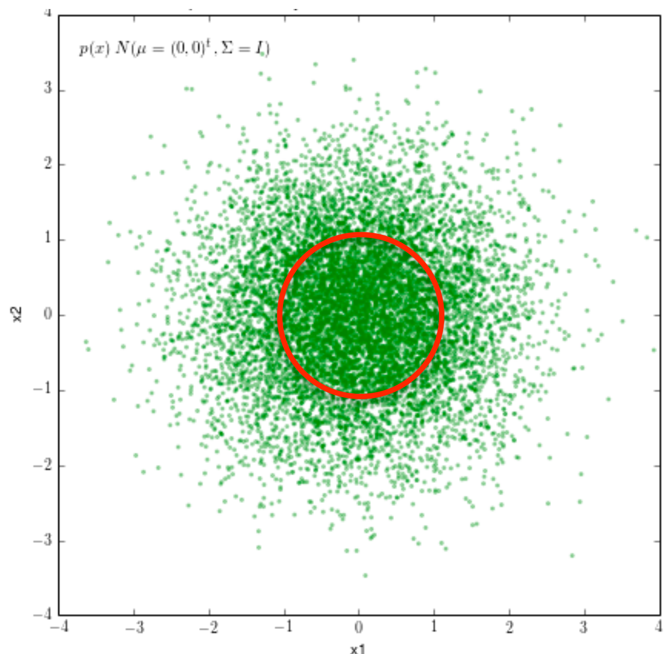
- However, as the dimension is increased something different happens
- When $\sigma^2 = 1$, integrating the probability density over a unit sphere centered at the origin yields nearly zero mass since the volume of a unit sphere is negligible.
- In fact, one needs to increase the radius of the sphere to $\sqrt{d}$ before there is a significant nonzero volume and hence a nonzero probability mass.
- If one increases the radius beyond $\sqrt{d}$, the integral ceases to increase, even though the volume increases, since the probability density is dropping off at a much higher rate. The natural scale for the Gaussian is in units of $\sigma\sqrt{d}$.

Some Facts

$$E\left(x_1^2 + x_2^2 + \cdots + x_d^2\right) = d\,E\left(x_1^2\right) = d\sigma^2.$$

The probability mass of a unit-variance Gaussian as a function of the distance from its center is given by $r^{d-1}e^{-r^2/2}$ times some constant normalization factor where r is the distance from the center and d is the dimension of the space. The probability mass function has its maximum at

$$r = \sqrt{d-1},$$



Concentration of Mass for Gaussian

Theorem 2.11 *For a d-dimensional, unit variance, spherical Gaussian, for any positive real number $\beta < \sqrt{d}$, all but $3e^{-\frac{\beta^2}{8}}$ of the mass lies within the annulus $\sqrt{d} - \beta \leq r \leq \sqrt{d} + \beta$.*