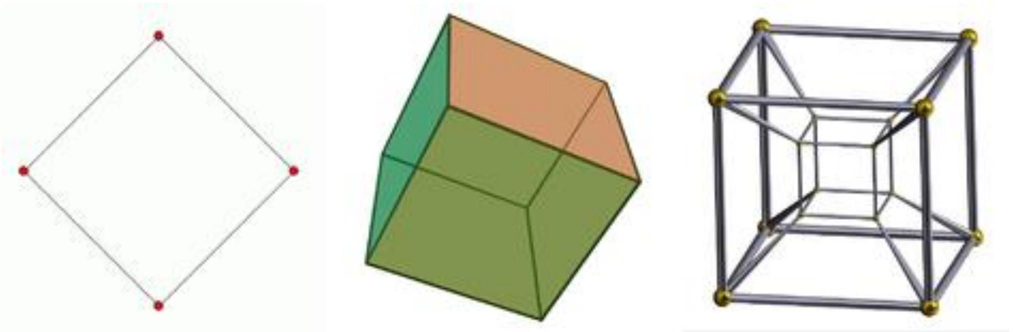
High Dimensional Space

longhuan@sjtu.edu.cn

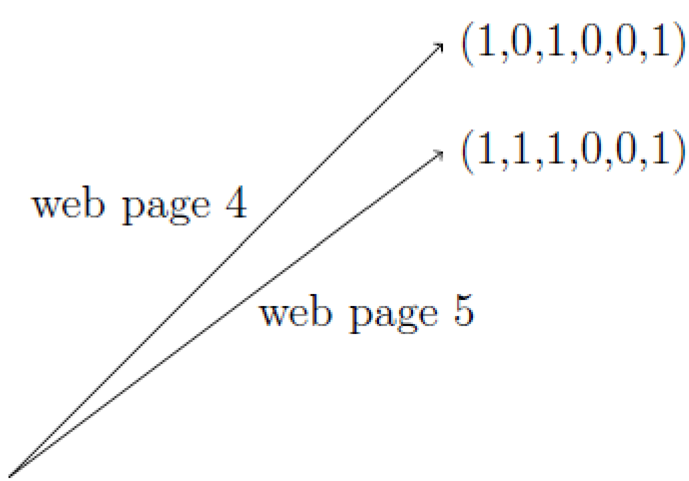


Word Vector Model

$$\begin{array}{l} \text{aardvark} \\ \text{abacus} \\ \ldots \\ \text{antitrust} \\ \ldots \\ \text{CEO} \\ \ldots \\ \text{Microsoft} \\ \ldots \\ \text{Windows} \end{array} \begin{bmatrix} 0 \\ 0 \\ \\ 42 \\ \\ 17 \\ \\ 61 \\ \\ 14 \end{bmatrix}$$

$$\begin{array}{l} \text{aardvark} \\ \text{abacus} \\ \ldots \\ \text{antitrust} \\ \ldots \\ \text{CEO} \\ \ldots \\ \text{Microsoft} \\ \ldots \\ \text{Windows} \end{array} \begin{bmatrix} 0 & 2 & 0 \\ 0 & 0 & 0 \\ & & \\ 42 & 0 & 0 \\ & & \\ 17 & 0 & 0 \\ & & \\ 61 & 0 & 0 \\ & & \\ 14 & 0 & 0 \end{bmatrix}$$

Web Page Model



- Nearest neighbor query
- Information retrieval
- Web page rank
- Online recommendation
- ……

- The law of Large numbers
- Properties of High-Dimensional space, unit ball
- Generating points uniformly at random from a ball
- Gaussians in High Dimension
- Random Projection and Johnson-Lindenstrauss Lemma
- Seperating Gaussians

Normal distribution (Gauss Distribution)

$X \sim N(\mu, \sigma^2)$, with density function:

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}}, \qquad -\infty < x < +\infty$$

Variance

$$\begin{aligned} Var(X) &= E((X - E[X])^2) \\ &= E(X^2 + E[X]^2 - 2XE[X]) \\ &= E(X^2 - E[X]^2) \\ &= E[X^2] - E[X]^2 \end{aligned}$$

Chebyshev's Inequality

$$\forall a > 0, \Pr(|X - E(X)| \geq a) \leq \frac{Var[X]}{a^2}$$

Law of Large Numbers

- In probability theory, the **law of large numbers (LLN)** is a theorem that describes the result of performing the same experiment a large number of times.
- According to the law, the average of the results obtained from a large number of trials should be close to the expected value, and will tend to become closer as more trials are performed.

Law of large numbers

Let $x_1, x_2, \ldots, x_n$ be n independent samples of a random variable x, then

$$\Pr\left(\left|\frac{x_1 + x_2 + \cdots x_n}{n} - E(x)\right| \geq \epsilon\right) \leq \frac{Var(x)}{n\epsilon^2}$$

Proof. (Chebychev's Inequality)

$$\Pr\left(\left|\frac{x_1 + x_2 + \cdots x_n}{n} - E(x)\right| \geq \epsilon\right) \leq \frac{Var(\frac{x_1 + x_2 + \cdots x_n}{n})}{\epsilon^2}$$

$$= \frac{Var(x_1 + x_2 + \cdots x_n)}{n^2\epsilon^2}$$

$$= \frac{Var(x)}{n\epsilon^2}$$

Application

- $\boldsymbol{x}$ be a d −dimensional random point whose coordinates are each selected from $N\left(0, \frac{1}{2\pi}\right)$,
- i.e. $\boldsymbol{x} = [x_1, x_2, \ldots, x_d]$ with $x_i \sim N\left(0, \frac{1}{2\pi}\right)$
- By LLN: $|\boldsymbol{x}|^2 = \sum_{i=1}^{d} x_i^2 = \frac{d}{2\pi} = \Theta(d)$ with high probability.
- The probability that point $\boldsymbol{x}$ lie in the unit ball is *vanishingly small*.

Application

- $\boldsymbol{x}, \boldsymbol{y}$ **:** $[z_1, z_2, \dots, z_d]$ with $z_i \sim N(0, 1)$
- $|\boldsymbol{x}|^2 \approx d, |\boldsymbol{y}|^2 \approx d,$
- $|\boldsymbol{x} - \boldsymbol{y}|^2 \approx ?$

Application

- $\boldsymbol{x}, \boldsymbol{y}$ **:** $[z_1, z_2, \ldots, z_d]$ with $z_i \sim N(0, 1)$
- $|\boldsymbol{x}|^2 \approx d, |\boldsymbol{y}|^2 \approx d,$
- $|\boldsymbol{x} - \boldsymbol{y}|^2 = \sum_{i=1}^{d} (x_i - y_i)^2$

$$\boldsymbol{E}(x_i - y_i)^2 = \boldsymbol{E}\left(x_i^2\right) + \boldsymbol{E}\left(y_i^2\right) - 2\boldsymbol{E}(x_i y_i)$$

$$= 1 + 1 - 2\boldsymbol{E}(x_i)\boldsymbol{E}(y_i) = 2.$$

Application

- $\boldsymbol{x}, \boldsymbol{y}$: $[z_1, z_2, \dots, z_d]$ with $z_i \sim N(0, 1)$
- $|\boldsymbol{x}|^2 \approx d, |\boldsymbol{y}|^2 \approx d,$
- $|\boldsymbol{x} - \boldsymbol{y}|^2 = \sum_{i=1}^{d} (x_i - y_i)^2 = 2d$

$$\boldsymbol{E}(x_i - y_i)^2 = \boldsymbol{E}\left(x_i^2\right) + \boldsymbol{E}\left(y_i^2\right) - 2\boldsymbol{E}(x_i y_i)$$

$$= 1 + 1 - 2\boldsymbol{E}(x_i)\boldsymbol{E}(y_i) = 2.$$

- $|\boldsymbol{x} - \boldsymbol{y}|^2 \approx |\boldsymbol{x}|^2 + |\boldsymbol{y}|^2$
- Pythagorean theorem $\Rightarrow$ random d –dimensional $\boldsymbol{x}, \boldsymbol{y}$ are approximately orthogonal.

Application

- $\boldsymbol{x}, \boldsymbol{y}$: $[z_1, z_2, \dots, z_d]$ with $z_i \sim N(0, 1)$
- Pythagorean theorem $\Rightarrow$ random d −dimensional $\boldsymbol{x}, \boldsymbol{y}$ are approximately orthogonal.

If we scale these random points to be unit length and call $\boldsymbol{x}$ the North Pole, *much of the surface area of the unit ball must lie near the equator*.

(to be formalized latter.)

Master Tail Bound Theorem

Theorem. Let $x = x_1 + x_2 + \cdots + x_n$, where $x_1, x_2, \ldots, x_n$ are mutually independent random variables with zero means and variance at most σ^2. Let $0 \le a \le \sqrt{2}n\sigma^2$.Assume that $|E(x_i^s)| \le \sigma^2 s!$ for $s = 3,4, \ldots, \lfloor(a^2/4n\sigma^2)\rfloor$ then

$$Prob(|x| \ge a) \le 3e^{-\frac{a^2}{12n\sigma^2}}.$$

Table of tail bounds

	Condition	Tail bound
Markov	$x \geq 0$	$\text{Prob}(x \geq a) \leq \frac{E(x)}{a}$
Chebychev	Any x	$\text{Prob}\big(\lvert x - E(x)\rvert \geq a\big) \leq \frac{\text{Var}(x)}{a^2}$
Chernoff	$x = x_1 + x_2 + \cdots + x_n$ $x_i \in [0,1]$ i.i.d. Bernoulli;	$\text{Prob}(\lvert x - E(x)\rvert \geq \varepsilon E(x))$ $\leq 3e^{-c\varepsilon^2 E(x)}$
Higher Moments	r positive even integer	$\text{Prob}(\lvert x\rvert \geq a) \leq E(x^r)/a^r$
Gaussian Annulus	$x = \sqrt{x_1^2 + x_2^2 + \cdots + x_n^2}$ $x_i \sim N(0,1); \beta \leq \sqrt{n}$ indep.	$\text{Prob}(\lvert x - \sqrt{n}\rvert \geq \beta) \leq 3e^{-c\beta^2}$
Power Law for x_i; order $k \geq 4$	$x = x_1 + x_2 + \ldots + x_n$ x_i i.i.d ; $\varepsilon \leq 1/k^2$	$\text{Prob}\big(\lvert x - E(x)\rvert \geq \varepsilon E(x)\big)$ $\leq (4/\varepsilon^2 kn)^{(k-3)/2}$

- The law of Large numbers
- Properties of High-Dimensional space, unit ball
- Generating points uniformly at random from a ball
- Gaussians in High Dimension
- Random Projection and Johnson-Lindenstrauss Lemma
- Seperating Gaussians

Geometry of High Dimensions

- Most of the volume of the high-dimensional objects is near the surface:

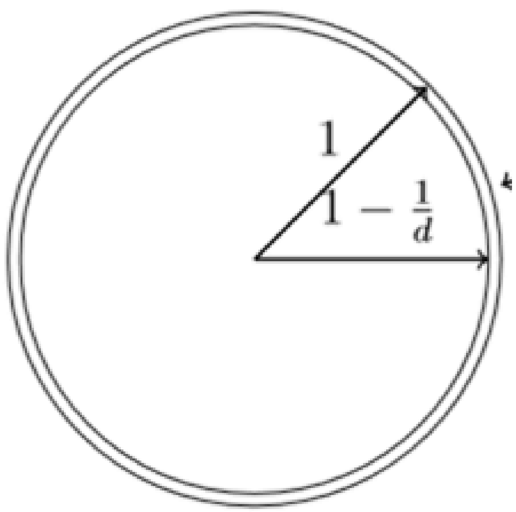
$$\frac{Volume\big((1-\epsilon)A\big)}{Volume(A)} = (1-\epsilon)^d \le e^{-\epsilon d}$$

Fix ϵ and letting $d \to \infty$, the above quantity rapidly approaches zero.

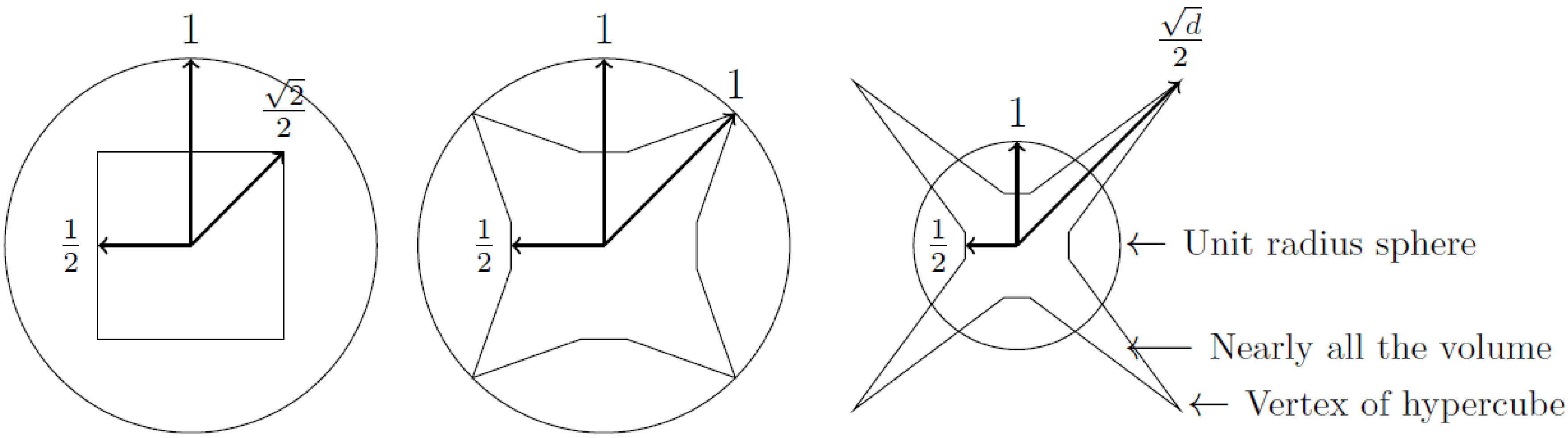
Application

S be the unit ball in d −dimensions (i.e., the set of points within distance 1 of the origin). Then $1 - e^{-\epsilon d}$ fraction of the volume is in $S\backslash(1-\epsilon)S$.

Especially, consider $\epsilon = \frac{1}{d}$.



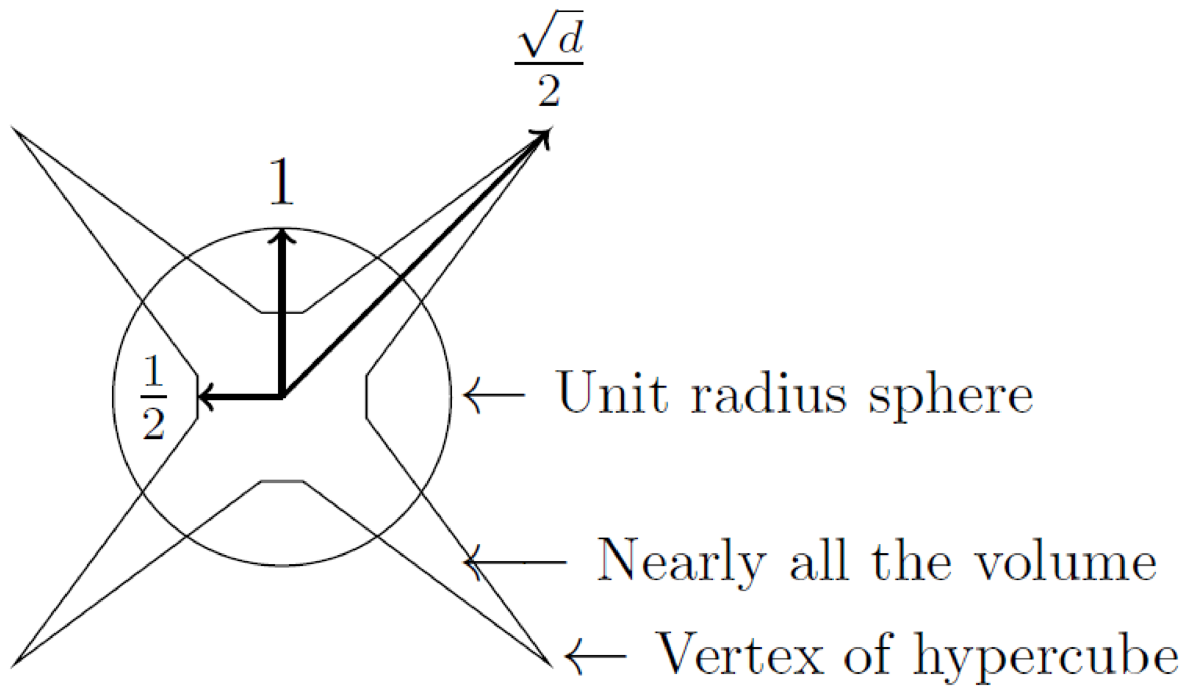
Relationship between the sphere and cube



John's Book Fig 2.4

The difference between the volume of a cube with unit-length sides and the volume of a unit-radius sphere at the dimensions: $2, 4$ and d.

Conceptual drawing of a sphere and a cube



For large d, almost all the volume of the cube is located outside the sphere.

Unit ball in d –dimensions

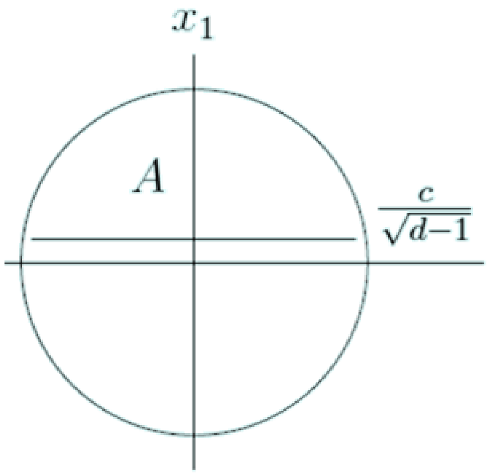
- Surface: $A(d) = \frac{2\pi^{\frac{d}{2}}}{\Gamma\left(\frac{d}{2}\right)}$, Volume: $\mathrm{V}(d) = \frac{2}{d}\frac{\pi^{\frac{d}{2}}}{\Gamma\left(\frac{d}{2}\right)}$.

$$\Gamma(n) = (n-1)!$$
$$\Gamma(x) = \int_0^{+\infty} t^{x-1}e^{-t}dt \ (x > 0)$$
$$\Gamma(1/2) = \sqrt{\pi}$$

Unit ball in d –dimensions

- Surface: $A(d) = \frac{2\pi^{\frac{d}{2}}}{\Gamma\left(\frac{d}{2}\right)}$, Volume: $\mathrm{V}(d) = \frac{2}{d} \cdot \frac{\pi^{\frac{d}{2}}}{\Gamma\left(\frac{d}{2}\right)}$.
- $V(2) = \pi, V(3) = \frac{4}{3}\pi$, $\lim_{n\to\infty} V(d) = 0.$ $n! \geq n^{n/2}$
- Most of the volume of a unit ball in high dimensions is concentrated near its equator no matter which direction is defined to be the North Pole.

Theorem: For $c \geq 1$ and $d \geq 3$, at least a $1 - \frac{2}{c} e^{-c^2/2}$ fraction of the volume of the d –dimensional unit ball has
$$|x_1| \leq \frac{c}{\sqrt{d-1}}.$$



Near orthogonality !

How it can be that nearly all the points in the unit ball are very close to the surface and yet at the same time nearly all points are in a box of side length $O(\frac{\ln d}{d-1})$?

A. Points on the surface of the ball satisfy

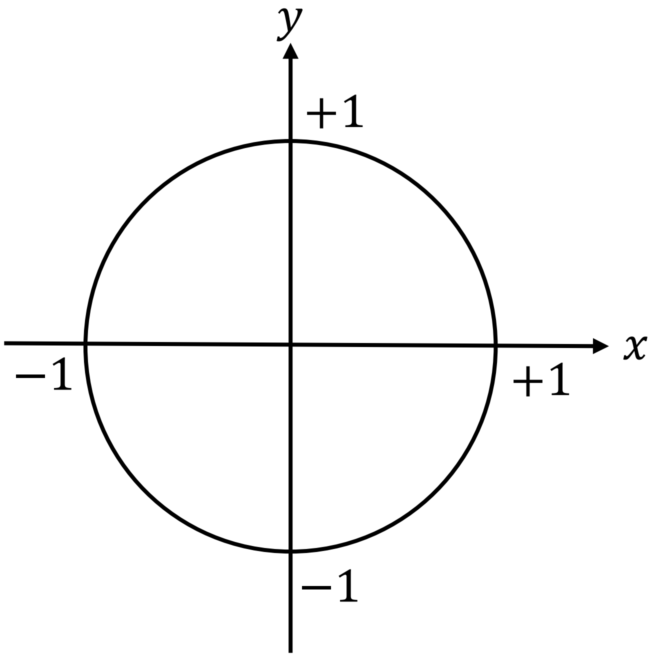
$$x_1^2 + x_2^2 + \cdots x_d^2 = 1,$$

so for each coordinate i, a typical value will be $\pm O\left(\frac{1}{\sqrt{d}}\right)$. In fact, it is helpful to think of picking a random point on the sphere as very similar to picking a random point of the form $\left(\pm\frac{1}{\sqrt{d}}, \pm\frac{1}{\sqrt{d}}, \cdots, \pm\frac{1}{\sqrt{d}}\right)$.

- The law of Large numbers
- Properties of High-Dimensional space, unit ball
- Generating points uniformly at random from a ball
- Gaussians in High Dimension
- Random Projection and Johnson-Lindenstrauss Lemma
- Seperating Gaussians

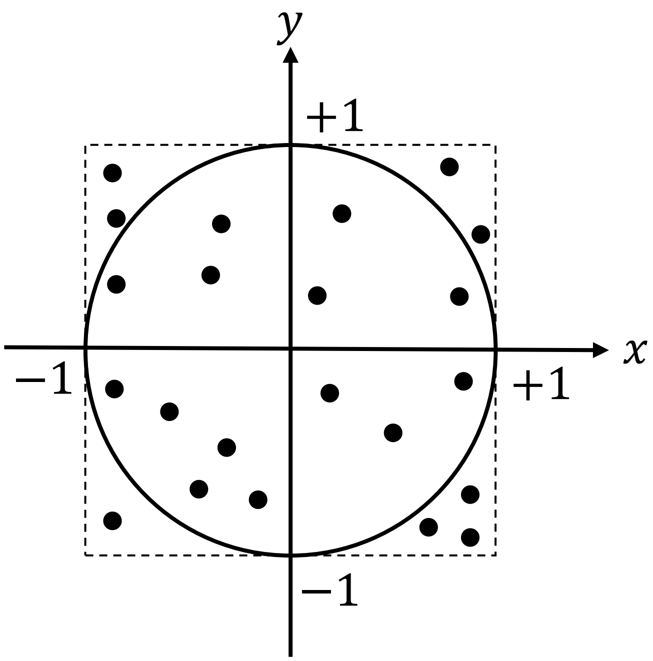
Generating points uniformly at random on the surface of the unit ball

- $d = 2$



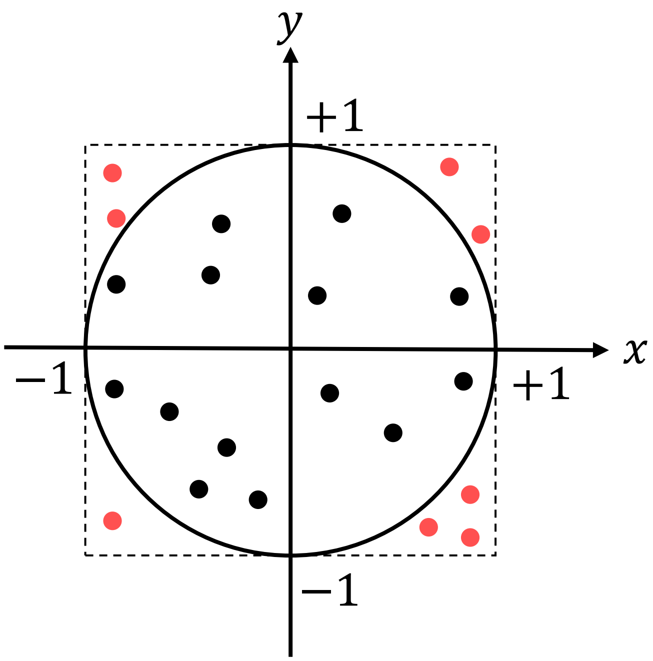
Generating points uniformly at random on the surface of the unit ball

- $d = 2$
  - Generate x_i, y_i u.a.r from the interval $[-1,1]$;



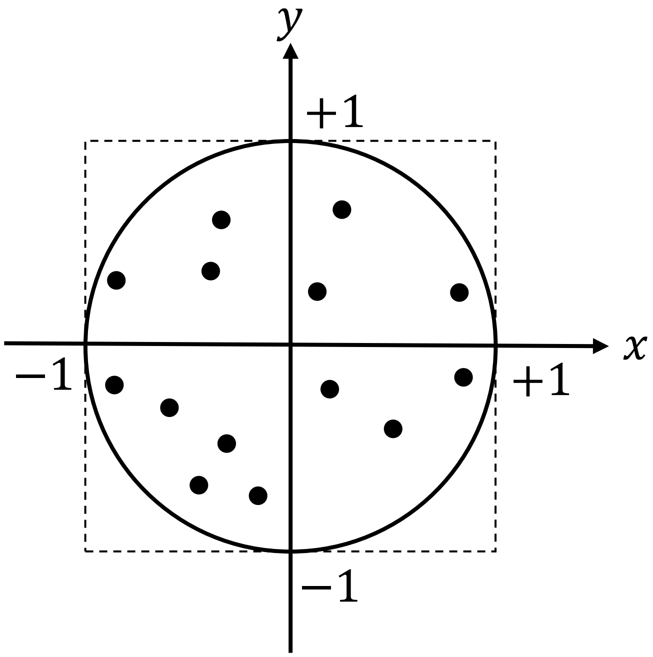
Generating points uniformly at random on the surface of the unit ball

- $d = 2$
  - Generate x_i, y_i u.a.r from the interval $[-1,1]$;



Generating points uniformly at random on the surface of the unit ball

- $d = 2$
  - Generate x_i, y_i u.a.r from the interval $[-1,1]$;
  - Discard the points outside the unit circle;

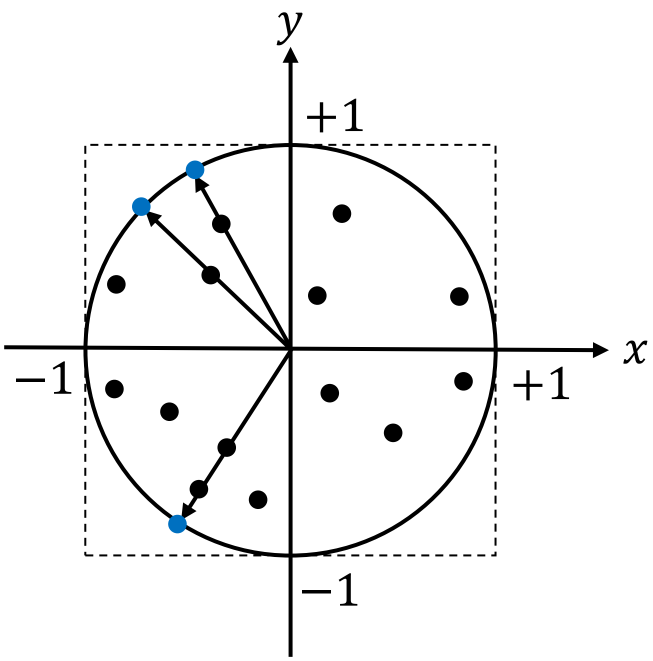


Generating points uniformly at random on the surface of the unit ball

- $d = 2$
  - Generate x_i, y_i u.a.r from the interval $[-1,1]$;
  - Discard the points outside the unit circle;
  - Project the remaining points onto the circle.
- How about d is large?
  - The above strategy would fail. (why?)

  ① Surface: Spherical normal distribution + Normalizing.

  ② Surface+interior: Scale the point on the surface.



Generating points uniformly at random on the surface of the unit ball

- When d is large, generate a point x:
  ① $r_i \sim N(0,1)$, i.e., $\frac{1}{\sqrt{2\pi}}\exp(-r^2/2)$ for all $i \in [d]$;
  ② Normalizing the vector to a unit vector $x = \frac{r}{|r|}$.

Generating points uniformly at random on the surface of the unit ball

- When d is large, generate a point x:

  ① $r_i \sim N(0,1)$, i.e., $\frac{1}{\sqrt{2\pi}}\exp(-r_i^2/2)$ for all $i \in [d]$;

  ② Normalizing the vector to a unit vector $x = \frac{r}{|r|}$.

- Proof. As every dimension is generated independently, then probability density of r is

$$P(r = \tilde{r}) = \frac{1}{(2\pi)^{d/2}} e^{-\frac{\tilde{r}_1^2 + \tilde{r}_2^2 + \cdots + \tilde{r}_d^2}{2}}$$

$$= \frac{1}{(2\pi)^{d/2}} e^{-\frac{|\tilde{r}|^2}{2}}$$

As the density only depends on the length of $\tilde{r}$ (i.e., $|\tilde{r}|^2$), the distribution is u.a.r..

Note that after step ②,coordinates are no longer statistically independent.

Generating points uniformly at random over the unit ball

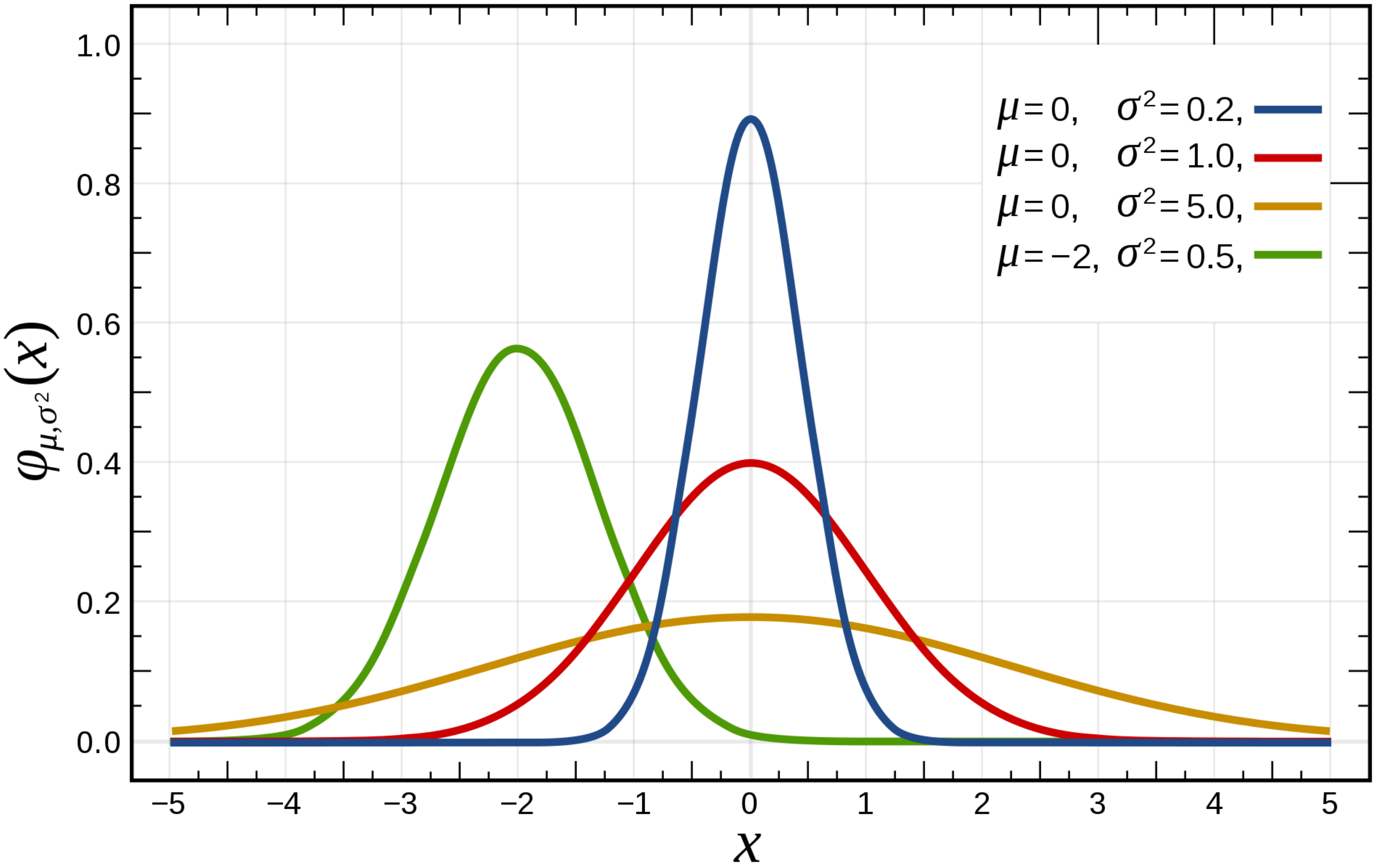
When d is large, generate a point y over the ball (surface and interior):

- Scale the point x generated on the surface by a scalar $\rho \in [0,1]$.
  - ✓ ρ should be a function of r,
  - ✓ As the volume of the radius r ball in d dimensions is $r^d V(d)$, the density of ρ at radius r is: $\frac{d}{dr}\left(r^d V(d)\right) = dr^{d-1}V(d)$.
- Thus, pick $\rho(r)$ with density for r over [0,1], i.e. $\rho(r) = dr^{d-1}$:

$$y = dr^{d-1} \cdot x$$

- The law of Large numbers
- Properties of High-Dimensional space, unit ball
- Generating points uniformly at random from a ball
- Gaussians in High Dimension
- Random Projection and Johnson-Lindenstrauss Lemma
- Seperating Gaussians

- 1-dimensional Gaussian



- d −dimensional spherical Gaussian with 0 means and variance σ^2 in each coordinate has density function:

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

- Integrate the PDF over a unit ball centered at the origin will cover almost 0 mass, for the volume of such a ball is negligible.
- The radius of the ball need to be nearly $\sqrt{d}$ before there is a significant volume and hence significant probability mass.

Gaussian Annulus Theorem

- For a d −dimensional spherical Gaussian <u>with unit variance in each direction</u> ,for any $\beta \leq \sqrt{d}$, all but at most $3e^{-c\beta^2}$ of the probability mass lies within the annulus

$$\sqrt{d} - \beta \leq |x| \leq \sqrt{d} + \beta$$

where c is a fixed positive constant.

- The law of Large numbers
- Properties of High-Dimensional space, unit ball
- Generating points uniformly at random from a ball
- Gaussians in High Dimension
- Random Projection and Johnson-Lindenstrauss Lemma
- Separating Gaussians

Database query: Nearest neighbor search

n points from R^d: $\begin{bmatrix} v_{11} & v_{21} & \vdots & v_{n1} \\ v_{12} & v_{22} & \vdots & v_{n2} \\ \vdots & \vdots & \vdots & \vdots \\ v_{1d} & v_{2d} & \vdots & v_{nd} \end{bmatrix}$

- Nearest neighbor search: find the nearest or approximately nearest database point to the query point.
- When d is large, it could cost more than expected.
- Dimension reduction : *Project* the database points to a k dimensional space with $k \ll d$. It will work so long as the relative distances between points are approximately preserved.

Projection function

- Pick k vectors $\boldsymbol{u_1}, \boldsymbol{u_2}, \ldots, \boldsymbol{u_k},$ in R^d with unit-variance coordinates independently , i.e., from the Gaussian distribution $\frac{1}{(2\pi)^{d/2}} exp\left(-\frac{|x|^2}{2}\right)$, for any vector $\boldsymbol{v}$, the projection $f: R^d \rightarrow R^k$ is:

$$f(\boldsymbol{v}) = (\boldsymbol{u_1} \cdot \boldsymbol{v}, \boldsymbol{u_2} \cdot \boldsymbol{v}, \ldots, \boldsymbol{u_k} \cdot \boldsymbol{v})$$

Projection function

Pick k vectors $\boldsymbol{u_1}, \boldsymbol{u_2}, \dots, \boldsymbol{u_k}$, independently from the Gaussian distribution $\frac{1}{(2\pi)^{d/2}} exp\left(-\frac{|x|^2}{2}\right)$, for any vector $\boldsymbol{v}$, the projection $f: R^d \to R^k$ is:

$$f(\boldsymbol{v}) = (\boldsymbol{u_1} \cdot \boldsymbol{v}, \boldsymbol{u_2} \cdot \boldsymbol{v}, \dots, \boldsymbol{u_k} \cdot \boldsymbol{v})$$

- $f(v_1 - v_2) = f(v_1) - f(v_2)$
- $|f(v)| \approx \sqrt{k}|v|$ w.h.p.
- To estimate $|v_1 - v_2|$, it suffices to compute $|f(v_1) - f(v_2)|$

Random Projection Theorem

- Let v be a fixed vector in R^d and let f be defined as above. Then there exists constant $c > 0$ such that for $\epsilon \in (0,1)$
$$\Pr\left(\left| |f(v)| - \sqrt{k}|v| \right| \geq \epsilon\sqrt{k}|v|\right) \leq 3e^{-ck\epsilon^2}$$

Johnson-Lindenstrass Lemma

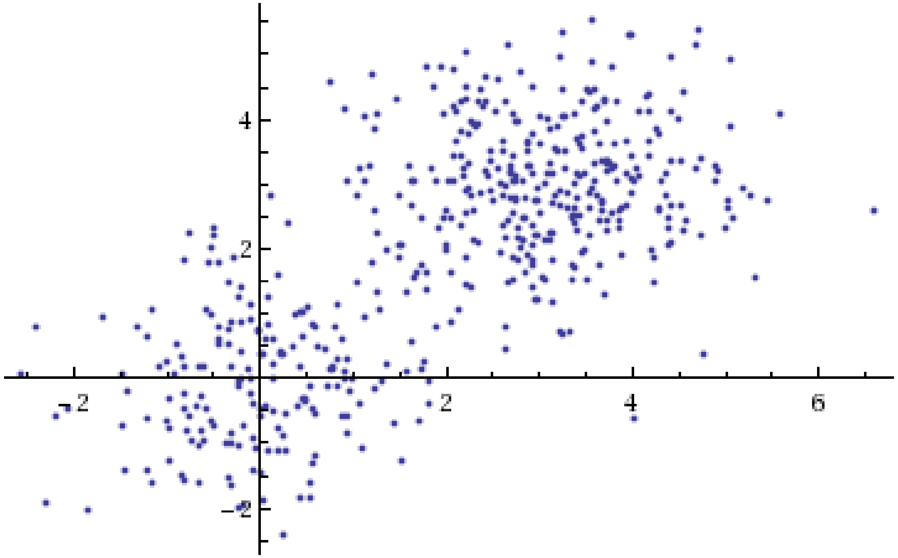
- For any $0 < \epsilon < 1$ and any integer n, let $k \geq \frac{3}{c\epsilon^2}\ln n$ for c as in the Gaussian Annulus theorem, for any set of n points in R^d, the random projection f defined above has the property that for all pairs of points v_i and v_j, with probability at least $1 - \frac{3}{2n}$.

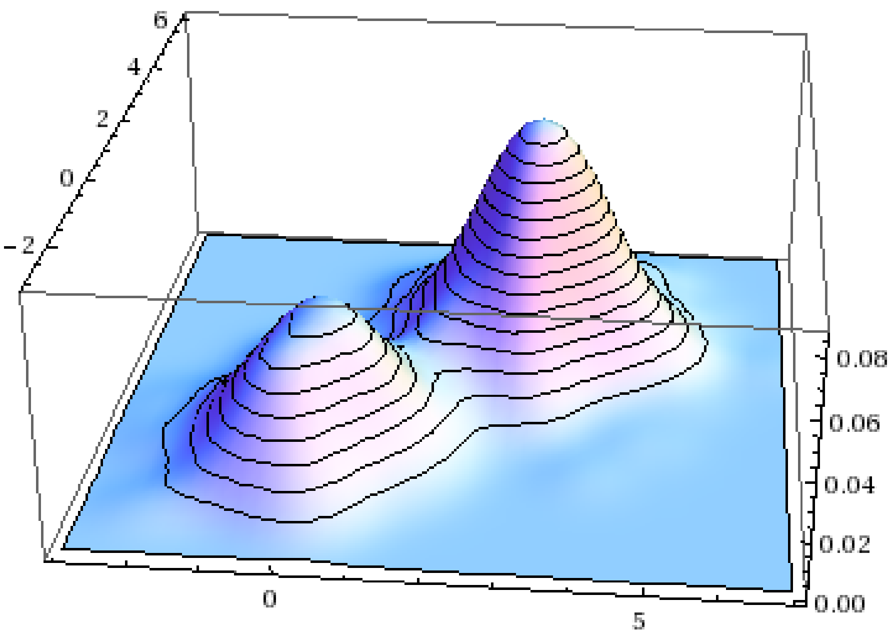
$$(1-\epsilon)\sqrt{k}|v_i - v_j| \leq |f(v_i) - f(v_j)| \leq (1+\epsilon)\sqrt{k}|v_i - v_j|.$$

Comments

- JL lemma works for all pairs of points,
- k depends on $\ln n$,
- To the database, JL Lemma says the algorithm will yield the right answer with high probability whatever the query is.

- The law of Large numbers
- Properties of High-Dimensional space, unit ball
- Generating points uniformly at random from a ball
- Gaussians in High Dimension
- Random Projection and Johnson-Lindenstrauss Lemma
- Separating Gaussians

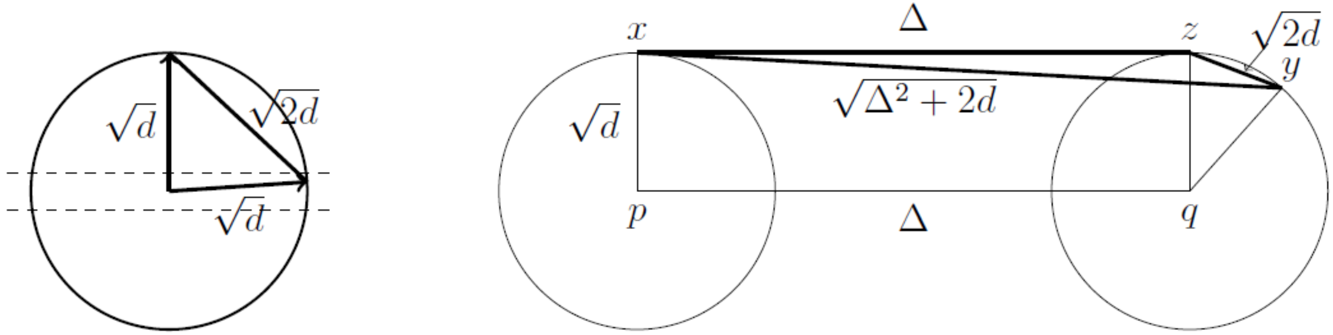




- Mixtures of Gaussians

- Parameter estimation problem

- When $\Delta \in \omega(d^{1/4})$



- Algorithm for separating points from two Gaussians: Calculate all pairwise distance between points. The cluster of smallest pairwise distances must come from a single Gaussian. Remove these points. The remaining points come from the second Gaussian.