CS 4850 Mathematical Foundations for the Information Age January 27, 2010

Lecture 2

Lecturer: John Hopcroft *Scribes: Gexu Zhang(gz54), Stanley Shen(ds676)*

1 Recap

Definition 1 *(1) Unit hypercube in d dimension:* $\{(x_1, x_2, \cdots, x_d) : -\frac{1}{2} \leq x_i \leq \frac{1}{2}, \forall i \in [d]\}$. *It is centered at the origin.*

(2) Unit hypersphere in d dimension: $\{(x_1, x_2, \cdots, x_d) : \sum_{i=1}^{d} x_i^2 \leq 1\}$. *It is also centered at the origin. We use* $V(d)$ *to denote the volume of a hypersphere in d dimension. And we use* $Vol(O)$ *to denote the volume of an object* O.

1.1 Cartesian coordinates

We can either integrate in Cartesian coordinates or polar coordinates. When we do it in Cartesian coordinates, it is hard:

$$\textbf{Cartesian coordinates: } V(d) = \int_{-1}^{1} \int_{-\sqrt{1-x_1^2}}^{\sqrt{1-x_1^2}} \cdots \int_{-\sqrt{1-x_1^2-\cdots-x_{d-1}^2}}^{\sqrt{1-x_1^2-\cdots-x_{d-1}^2}} \mathrm{d}x_d \mathrm{d}x_{d-1} \cdots \mathrm{d}x_1$$

1.2 Polar coordinates

Let's switch to polar coordinates:

$$\begin{aligned}
\textbf{Polar coordinates: } V(d) &= \int_{S^d} \int_{r=0}^{1} r^{d-1} \, \mathrm{d}r \mathrm{d}\Omega \\
&= (\int_{S^d} \mathrm{d}\Omega)(\int_{r=0}^{1} r^{d-1} \, \mathrm{d}r) \qquad \text{(Variables are independent.)} \\
&= \frac{1}{d} \underbrace{\int_{S^d} \mathrm{d}\Omega}_{A(d)} \\
&= \frac{1}{d} A(d)
\end{aligned}$$

Lemma 2

$$V(d) = \frac{1}{d} A(d) \qquad \blacksquare$$

So we need to calculate $A(d)$.

1.3 Consider a new problem

Let's forget about this problem and consider a new problem. The new problem is another integral

$$I(d) \stackrel{def}{=} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} \mathrm{e}^{-(x_1^2+x_2^2+\cdots+x_d^2)} \, \mathrm{d}x_d \cdots \mathrm{d}x_2 \mathrm{d}x_1$$

. We can calculate $I(d)$ in two ways:

Way I:

$$\begin{aligned} I(d) &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty} \mathrm{e}^{-(x_1^2+x_2^2+\cdots+x_d^2)}\,\mathrm{d}x_d\cdots\mathrm{d}x_2\mathrm{d}x_1 \\ &= [\int_{-\infty}^{\infty} \mathrm{e}^{-(x_1)^2}\,\mathrm{d}x_1]^d \\ &= (\sqrt{\pi})^d \\ &= \pi^{\frac{d}{2}} \end{aligned}$$

Way II:

$$\begin{aligned} I(d) &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty} \mathrm{e}^{-(x_1^2+x_2^2+\cdots+x_d^2)}\,\mathrm{d}x_d\cdots\mathrm{d}x_2\mathrm{d}x_1 \\ &= \int_{S^d}\int_{-\infty}^{\infty} \mathrm{e}^{-r^2} r^{d-1}\,\mathrm{d}r\mathrm{d}\Omega && \text{(In polar coordinates)} \\ &= (\int_{S^d} \mathrm{d}\Omega)(\int_{-\infty}^{\infty} \mathrm{e}^{-r^2} r^{d-1}\,\mathrm{d}r) \\ &= A(d)2\int_0^{\infty} e^{-t} t^{\frac{d-1}{2}} \frac{1}{2\sqrt{t}}\,\mathrm{d}t && \text{(Let } t=r^2\geq 0\text{, so } r=\sqrt{t},\ \mathrm{d}r=\tfrac{1}{2\sqrt{t}}\mathrm{d}t\text{)} \\ &= A(d)\underbrace{\int_0^{\infty} e^{-t} t^{\frac{d}{2}-1}\,\mathrm{d}t}_{\Gamma(\frac{d}{2})} \\ &= A(d)\Gamma(\frac{d}{2}) \end{aligned}$$

Definition 3 *For every $x\in\mathbb{R}$, its Gamma function is defined as follows:*

$$\Gamma(x) = \int_0^{+\infty} t^{-x}\mathrm{e}^{x-1}\,\mathrm{d}t$$

You can think of Gamma function as an extension of factorial function for non-integer values. It is $x-1$ "factorial". But the important thing is when x is not an integer. We can easily derive the following:

$$\begin{aligned} &\Gamma(x) = (x-1)\Gamma(x-1) \\ &\Gamma(2) = \Gamma(1) = 1 && \text{Similar to } 1! = 0! = 1 \\ &\Gamma(\frac{1}{2}) = \sqrt{\pi} \end{aligned}$$

Lemma 4

$$A(d) = \frac{2\pi^{\frac{d}{2}}}{\Gamma(\frac{d}{2})} \qquad \blacksquare$$

1.4 $V(d)$

Combining Lemma 2 and Lemma 4 we have the following:

Theorem 5 *The volume of the unit hypersphere in d dimensions is*

$$V(d) = \frac{1}{\frac{d}{2}} \frac{\pi^{\frac{d}{2}}}{\Gamma(\frac{d}{2})} = \frac{\pi^{\frac{d}{2}}}{\Gamma(\frac{d}{2}+1)}$$ ■

We can check $V(d)$: when $d = 2$, $V(2) = \pi = \pi r^2|_{r=1}$; when $d = 3$, $V(3) = \frac{4}{3}\pi = \frac{4\pi r^3}{3}|_{r=1}$.

So what happens to this volume as d increase to infinite? We can eliminate the effect of constants in asymptotic analysis. Note that in $V(d)$ the numerator grows exponentially, like 2^n. The denominator grows factorially, like $n!$. $n!$ grows much faster than 2^n. Therefore $\lim_{d\to+\infty} V(d) = 0$.Think about when volume is maximized.

2 Three counterintuitive facts about hyperspheres

Lemma 6 *For every real number x,*

$$1 + x \leq \mathrm{e}^x$$ ■

Lemma 7 (Bernoulli's inequality) *For every integer $r \geq 0$ and every real number $x \geq -1$,*

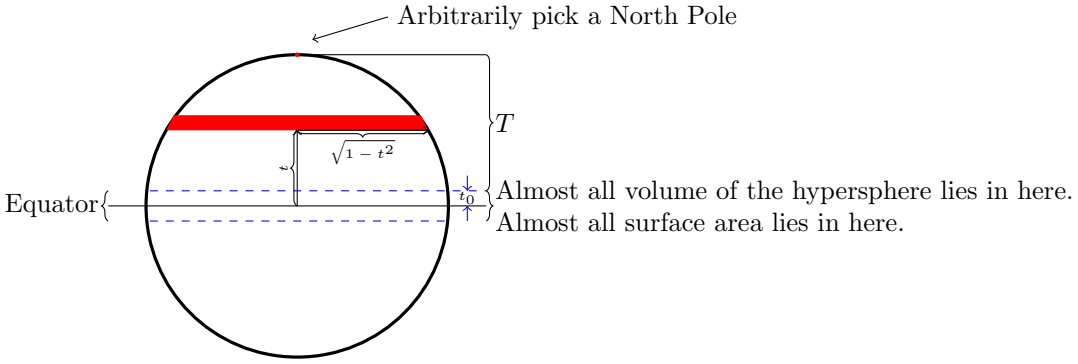
$$(1 + x)^r \geq 1 + rx$$ ■

2.1 Vast majority of the volume lies near the equator

We are interested in where is the volume of the unit sphere in high dimensions.

Definition 8 (Arbitrary "North Pole" and "equator") *Pick a vector on the hypersphere arbitrarily, call it the North Pole. Then the intersection of the hyperplane that is perpendicular to it and the hypersphere is called the equator.*

An equator of a d-dimensional hypersphere is a $(d-1)$-dimensional object. For example, a 3-dimensional sphere's equator is a (2-dimensional) disk; a 4-dimensional hypersphere's equator is a 3-dimensional sphere. However, the *surface area* of a d-dimensional hypersphere is also a $(d-1)$-dimensional object. The *circumference* at the equator is a $(d-2)$-dimensional object.



Theorem 9 *As d grows to infinity, almost all volume are near the equator.*

Proof We need to calculate a ratio. Firstly define an area T:

$$T = \{x : |x| \leq 1, x_1 \geq t_0\}$$

x_1 is in the direction of the North Pole. Because of symmetry, we only need to consider the upper hemisphere. We need to prove that the small slice above the equator contains almost all the volume of the upper hemisphere. It is hard to integrate from 0 to t_0, but easier to integrate from t_0 up to the North Pole. We will prove the upper bound on the percentage of the volume of T goes to zero when t_0 is fixed and d goes to infinity. Our roadmap is as follows:

$$\text{Upper bound on the percentage of volume above slice} = \frac{\text{Upper bound of } Vol(T)}{\text{Lower bound of } Vol(\text{hemisphere})}$$

That is, we want to find the upper bound of $\frac{Vol(T)}{\frac{1}{2}V(d)}$, and

$$\text{Upper bound of } \frac{Vol(T)}{\frac{1}{2}V(d)} = \frac{\text{Upper bound of } Vol(T)}{\text{Lower bound of } \frac{1}{2}V(d)} \tag{1}$$

$$\begin{aligned}
Vol(T) &= \int_{t_0}^{1} \Delta V \, \mathrm{d}t && (\Delta V \text{ is the hypersphere of one lower dimension}) \\
&= \int_{t_0}^{1} V(d-1) r^{d-1} \, \mathrm{d}t && (\Delta V = V(d-1)r^{d-1},\ r \text{ is } \Delta V\text{'s radius}) \\
&= V(d-1) \int_{t_0}^{1} (1-t)^{\frac{d-1}{2}} \, \mathrm{d}t && (r = \sqrt{1-t^2}) \\
&\leq V(d-1) \int_{t_0}^{1} \mathrm{e}^{-t^2 \frac{d-1}{2}} \, \mathrm{d}t && (\text{Lemma 6}) \\
&\leq V(d-1) \int_{t_0}^{\infty} \mathrm{e}^{-t^2 \frac{d-1}{2}} \, \mathrm{d}t && (\mathrm{e}^{-t^2 \frac{d-1}{2}} \geq 0 \ \forall t \in [t_0, \infty]) \\
&\leq V(d-1) \int_{t_0}^{\infty} \frac{t}{t_0} \mathrm{e}^{-t^2 \frac{d-1}{2}} \, \mathrm{d}t && (\frac{t}{t_0} \geq 1 \ \forall t \in [t_0, \infty]) \\
&= \frac{V(d-1)}{-2t_0 \frac{d-1}{2}} \int_{t_0}^{\infty} \mathrm{e}^{-t^2 \frac{d-1}{2}} \, \mathrm{d}(-t^2 \frac{d-1}{2}) \\
&= \frac{V(d-1)}{t_0(d-1)} \mathrm{e}^{-t^2 \frac{d-1}{2}} |_{t_0}^{\infty} \\
&= \frac{1}{(d-1)t_0} \mathrm{e}^{-\frac{d-1}{2} t_0^2} V(d-1)
\end{aligned}$$

$$
\begin{aligned}
\frac{1}{2}V(d) &= \int_0^1 V(d-1)r^{d-1}\,\mathrm{d}t \\
&= V(d-1)\int_0^1 (1-t^2)^{\frac{d-1}{2}}\,\mathrm{d}t \\
&\geq V(d-1)\int_0^{\frac{1}{\sqrt{d-1}}} (1-t^2)^{\frac{d-1}{2}}\,\mathrm{d}t \\
&\geq V(d-1)\int_0^{\frac{1}{\sqrt{d-1}}} (1-\frac{d-1}{2}t^2)\,\mathrm{d}t && \text{(Lemma 7)} \\
&\geq V(d-1)\int_0^{\frac{1}{\sqrt{d-1}}} (1-\frac{d-1}{2}\frac{1}{d-1})\,\mathrm{d}t && (t^2 \leq \frac{1}{d-1}\ \forall t \in [0, \frac{1}{\sqrt{d-1}}]) \\
&= \frac{1}{2\sqrt{d-1}}V(d-1)
\end{aligned}
$$

Remark We use integration to calculate $\frac{1}{2}V(d)$ instead of using the closed-form of $V(d)$ that we just got is because by integrating we can get a factor $V(d-1)$ in the denominator which will cancel the $V(d-1)$ in the numerator in ratio (1).

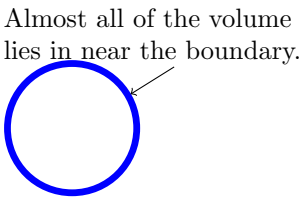
$$
\begin{aligned}
\frac{Vol(T)}{\frac{1}{2}V(d)} &\leq \text{Upper bound on the percentage of volume above slice} \\
&\leq \frac{\frac{1}{t_0(d-1)}\mathrm{e}^{-\frac{d-1}{2}t_0^2}V(d-1)}{\frac{1}{2\sqrt{d-1}}V(d-1)} \\
&= \frac{2}{t_0\sqrt{d-1}}\mathrm{e}^{-\frac{d-1}{2}t_0^2}
\end{aligned}
$$

This tells you how much of the volume is within $t_0 \leq x_1 \leq 1$. And the above shows that the amount of volume up there is dropping exponentially fast with the square of this distance (t_0) as d increases towards infinity. ■

Remark Theorem 9 holds when the North Pole is *arbitrarily* picked on the surface of this sphere. This is a little bit counterintuitive.

2.2 Vast majority of the volume lies near the boundary

Intuitively, what is the relationship between the volume of a sphere and the surface area of the sphere. Notice that the radius of this sphere is increased slightly, the volume increases by an incremental amount which is equal to the surface area times Δr. And that suggests the surface area is the derivative of the volume.



And so we can claim most of the volume is in a narrow annulus. We have the following theorem. Its proof was skipped in class.

Theorem 10 *The vast majority of the volume lies near the boundary of the hypersphere.*

Sketch of Proof Let $V_d(r)$ denote the volume of a sphere in d dimension with radius r. It is easy to prove that $V_d(r) = V(d)r^d$. Consider the volume that is not near the boundary: $A = \{(x_1, x_2, \cdots, x_d) : 0 \leq \sqrt{\sum_{i=1}^{d} x_i^2} \leq 1 - t_0\}$. Therefore,

$$\begin{aligned} \frac{Vol(A)}{V(d)} &= \frac{V_d(1-t_0)}{V_d(1)} \\ &= (1-t_0)^d \\ &\leq \mathrm{e}^{-t_0 d} \qquad \text{(Lemma 6)} \end{aligned}$$

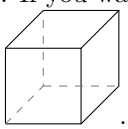
So the volume that is not near the boundary drops exponentially fast as d increases to infinity. ■

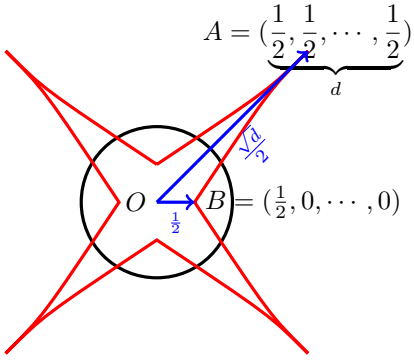
2.3 Vast majority of the surface area lies near the equator

The details of this part were skipped in class but are contained in the course notes.

3 Generating Points on the Surface of the Unit Hypersphere

3.1 Why generating points on the surface of a hypersphere?

Let's consider generating points uniformly at random on the surface of this sphere. If you want to generate some high dimensional data to test something, you cannot use a hypercube like this  . And generate points uniformly at random within the cube. Because vertices of this cube are a long way out of the origin. But the size of the sphere is very close to origin. Consider $A = (\frac{1}{2}, \frac{1}{2}, \cdots, \frac{1}{2})$ and $B = (\frac{1}{2}, 0, \cdots, 0)$ on the surface of the hypercube. Let O be the origin. Here is the picture.



The red part represents the boundary of the hypercube while the black circle represents the hypersphere. $|OA| = \frac{\sqrt{d}}{2}$ is greater than $|OB| = \frac{1}{2}$. And it turns out if this is the unit hypercube, $|OB|$ always comes the same as the dimension goes up. But $|OA|$ grows as $\sqrt{d}$. And so if you generate points in high dimensions like this, you have points out there around the vertices of the hypercube because *most of the volume of*

a hypercube lies near the vertices. We probably would not generate points in such a way. That's why people want to find out ways to generate points uniformly random at surface of the sphere in real world's applications.

3.2 First try: wrong

One way is generating points uniformly at random within a hypercube. Let's say we pick a hypercube which is bigger than a sphere, generate points in it and then project the points down to the surface of the sphere. So if a point $(x_1, x_2, \cdots, x_d)$ is generated, it is *normalized* (projected) by dividing every x_i by $\sqrt{x_1^2 + x_2^2 + \cdots + x_d^2}$, thus we get $(\frac{x_1}{\sqrt{\sum_{i=1}^d x_i^2}}, \frac{x_2}{\sqrt{\sum_{i=1}^d x_i^2}}, \cdots, \frac{x_d}{\sqrt{\sum_{i=1}^d x_i^2}})$. However, they are not equally distributed on the surface of the sphere. More points are in the direction of the vertices of the hypercube. So this does not work.

3.3 Second try: correct but inefficient

Next we might improve our former approach. Let's *generate the points at random within a hypercube bigger than the sphere, if the point is out of the sphere, we just throw it away; if it is inside the sphere, we keep it.* Because points will be symmetrically distributed, we can project (normalize) them on the surface of the sphere. Then they will be uniformly distributed over the surface of the sphere. That works in two dimensions. Slower in three dimensions. But it does not work in high dimensions because the volume of the sphere is almost zero, there will be hardly any points inside it.

3.4 Third try: correct and efficient

Let's consider an alternative method: generating points using an exponential distribution:

$$e^{-\frac{x_1^2+x_2^2+\cdots+x_d^2}{2}}$$

This means to generate each coordinate according to a Gaussian. Then they will be uniformly distributed spherely symmetrically and normalize them to unit length then they will be uniformly distributed over the unit hypersphere. This method works. In particular, each coordinate x_i ($i \in [d]$) is generated by a Gaussian $p(x_i) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x_i^2}{2}}$. We get a point $(x_1, x_2, \cdots, x_d)$. Note that the probability of getting it is $p(x_1, x_2, \cdots, x_d) = \frac{1}{(2\pi)^{\frac{d}{2}}} e^{-\frac{x_1^2+x_2^2\cdots+x_d^2}{2}} = \frac{1}{(2\pi)^{\frac{d}{2}}} e^{-\frac{r^2}{2}}$ where r is the distance from the origin to it. So the point are distributed spherely symmetrically. Then we normalize this point to make its norm 1 (i.e. project it on to the surface of the unit hypersphere). We have $(y_1, y_2, \cdots, y_d)$ with $\sum_{i=1}^d y_i^2 = 1$. Therefore, the points are uniformly distributed on the surface of the unit hypersphere. This method works.

4 Next Lecture

We will calculate the distance of two points on the unit sphere that were randomly generated, and show that they are $\sqrt{2}$ apart.