In this lecture, we analyze the shape and (somewhat surprising) properties of some basic geometric forms, which we understand very well in dimensions two and three, in high dimensions. For a comprehensive exposition of the topic see Chapter 2 in [BHK20].

1   The Geometry of High Dimensions

We show that "most" of the volume of the d-dimensional sphere is near its boundary. That is, for a d-dimensional sphere of radius r, most of the volume is contained in an annulus of width proportional to r/d.

Consider an object O in $\mathbb{R}^d$. If we shrink the object O by an ε factor, then

$$\text{volume}((1-\varepsilon)O) = (1-\varepsilon)^d \text{volume}(O),$$

where the shrunk object is defined as $(1-\varepsilon)O = \{(1-\varepsilon)x \mid x \in O\}$.

To see the relation between the volume of O and $(1-\varepsilon)O$, partition the object O into infinitesimal cubes. Then, shrinking O and getting the object $(1-\varepsilon)O$ is equivalent to shrinking the cubes and taking their union. By shrinking each side of a d-dimensional cube by a factor $(1-\varepsilon)$, it volumes shrinks by a factor of $(1-\varepsilon)^d$. By the fact that $1-x \le e^{-x}$, for any object O in $\mathbb{R}^d$,

$$\frac{\text{volume}((1-\varepsilon)O)}{\text{volume}(O)} = (1-\varepsilon)^d \le e^{-\varepsilon d}.$$

By fixing ε, as $d \to \infty$, the above quantity approaches to zero. This implies that nearly all volume of O must be in the portion of O that is not in $(1-\varepsilon)O$. Let S denote the unit ball in d dimensions. Then, at least a $(1-e^{-\varepsilon d})$-fraction of the volume of the unit ball is concentrated in $S \setminus (1-\varepsilon)S$. In particular, most of the volume of the d-dimensional unit ball is in the annulus of width $O(1/d)$ near its boundary. Similarly, for a general ball of radius r, then most of the volume is in the annulus of width $O(r/d)$ near its boundary.

2   Properties of the Unit Ball

Now, we focus more on the properties of the unit ball in d dimensions. For fixed dimension d, we observed that the volume of a sphere is a function of its radius and grows as r^d. However, for fixed radius, the volume of a sphere is a function of the dimension of the space. Interestingly (and somewhat surprisingly), the volume of a unit sphere goes to zero as the dimension of the sphere increases.

2.1   Volume of the Unit Ball

To calculate the volume of a sphere, one can integrate in either Cartesian or polar coordinates. In Cartesian coordinates the volume of a unit sphere is given by

$$V(d) = \int_{x_1=-1}^{x_1=1} \int_{x_2=-\sqrt{1-x_1^2}}^{x_2=\sqrt{1-x_1^2}} \cdots \int_{x_d=-\sqrt{1-x_1^2-x_2^2-\cdots x_{d-1}^2}}^{x_d=\sqrt{1-x_1^2-x_2^2-\cdots x_{d-1}^2}} dx_1 dx_2 \cdots dx_d$$

Since the limits of the integrals are complicated, it is easier to integrate using polar coordinates. In polar coordinates, $V(d)$ is given by

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

Here, $d\Omega$ is the surface area of the infinitesimal piece of the solid angle S^d of the unit sphere. We skip the detailed analysis of the above integral and directly state the final result here.

Lemma 2.1 *The volume $V(d)$ of a unit-radius ball in d-dimensions is given by $V(d) = \frac{2\pi^{\frac{d}{2}}}{d\Gamma(\frac{d}{2})}$.*

The gamma function $\Gamma(x)$ is a generalization of the factorial function for noninteger values of x. $\Gamma(x) = (x-1)\Gamma(x-1)$, $\Gamma(2) = \Gamma(1) = 1$, and $\Gamma(1/2) = \sqrt{\pi}$. For integer x, $\Gamma(x) = (x-1)!$. To check the formula for the volume of a unit ball, note that $V(2) = \pi$ and $V(3) = \frac{4}{3}\pi$, which are the correct volumes for the unit balls in two and three dimensions. Note that $\pi^{d/2}$ is an exponential in $d/2$ and $\Gamma(d/2)$ grows as the factorial of $d/2$. This implies that $\lim_{d\to\infty} V(d) = 0$, as claimed.

2.2   Volume near Equator

An interesting fact about the unit ball in high dimensions is that most of its volume is concentrated near its "equator". In particular, for any unit-length vector $\mathbf{v}$ defining "north", most of the volume of the unit ball lies in the thin slab of points whose inner product with $\mathbf{v}$ has magnitude $O(1/\sqrt{d})$. To show this fact, it suffices by symmetry to fix $\mathbf{v}$ to be the first coordinate vector. We show that most of the volume of the unit ball has $|x_1| = O(1/\sqrt{d})$.

Theorem 2.2 *For $c \geq 1$ and $d \geq 3$, at least $1 - \frac{2}{c}e^{-c^2/2}$ fraction of the volume of the d-dimensional unit ball has $|x_1| \leq c/\sqrt{d-1}$.*

Proof: By symmetry, we prove the statement for the upper hemisphere (i.e., the half of the ball with $x_1 > 0$). Let H denote the upper hemisphere and let A denote the portion of unit ball (or equivalently H) with $x_1 \geq c/\sqrt{d-1}$. To calculate volume of A, we write it as the integration of an incremental volume that is a disk of width dx_1 and has a $(d-1)$-dimensional ball of radius $\sqrt{1-x_1^2}$ as its face. Note that the volume of a $(d-1)$-dimensional unit ball of radius $\sqrt{1-x_1^2}$ is $(1-x_1^2)^{\frac{d-1}{2}}$ times the volume of the $(d-1)$-dimensional unit ball, $V(d-1)$.

$$\text{volume}(A) = \int_{c/\sqrt{d-1}}^{1} (1-x_1^2)^{\frac{d-1}{2}} V(d-1) dx_1$$

We use $1 - x \leq e^{-x}$ and integrate to infinity. Also, to compute the integral, we insert $\frac{x_1\sqrt{d-1}}{c}$ which is greater than one in the range of integration. Hence,

$$\text{volume}(A) \leq \int_{c/\sqrt{d-1}}^{\infty} \frac{x_1\sqrt{d-1}}{c} e^{-\frac{d-1}{2}x_1^2} V(d-1) dx_1 = V(d-1)\frac{\sqrt{d-1}}{c} \int_{c/\sqrt{d-1}}^{\infty} x_1 e^{-\frac{d-1}{2}x_1^2} dx_1$$

$$= \frac{V(d-1)}{c\sqrt{d-1}} e^{-\frac{c^2}{2}}$$

The volume of the hemisphere below the plane $x_1 = 1/\sqrt{d-1}$ is a lower bound on the entire volume of H. This volume is at least that of cylinder of height $a/\sqrt{d-1}$ and radius $\sqrt{1-1/(d-1)}$. The volume of the cylinder is $V(d-1)(1-\frac{1}{d-1})^{\frac{d-1}{2}}\frac{1}{\sqrt{d-1}}$. Using $(1-x)^a \geq 1 - ax$ for $a \geq 1$, the volume of cylinder is at least $\frac{V(d-1)}{2\sqrt{d-1}}$ for $d \geq 3$. So,

$$\frac{\text{volume}(A)}{\text{volume}(H)} \leq \frac{\frac{V(d-1)}{c\sqrt{d-1}} e^{-c^2/2}}{\frac{V(d-1)}{2\sqrt{d-1}}} = \frac{2}{c}e^{-c^2/2}$$

∎

Remark 2.3 *We computed a lower bound on the total hemisphere although we already know the volume of H equal to $V(d)/2$. We did so to get a formula with $V(d-1)$ in it to cancel the $V(d-1)$ in the numerator.*

Near orthogonality. One immediate implication of the above analysis is that if we draw two points at random from the unit ball, with high probability they will be nearly orthogonal. More precisely, with high probability both will be close to the surface and will

3

have length $1 - O(1/d)$. From our analysis, if we define the vector in the direction of the first point as "north", with high probability the second will have a projection of only $\pm O(1/\sqrt{d})$ in this direction, hence, their inner product is $\pm O(1/\sqrt{d})$. Moreover, we have the following theorem that states that if we draw n points at random in the unit ball, with high probability all points will be close to unit length and each pair of points will be almost orthogonal.

Theorem 2.4 *Consider drawing n points* $\mathbf{x}_1, \cdots, \mathbf{x}_n$ *at random from the unit ball. With probability* $1 - O(1/n)$,

- $|\mathbf{x}_i| \geq 1 - \frac{2\ln n}{d}$ *for all i, and*

- $|\mathbf{x}_i \cdot \mathbf{x}_j| \leq \frac{\sqrt{6\ln n}}{\sqrt{d-1}}$ *for all* $i \neq j$.

Discussion. One might wonder 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 d-dimensional box of side-length $O(\frac{\ln d}{d-1})$. The answer is to remember that points on the surface of the ball has norm one; so for each coordinate i, a typical value will be $\pm O(1/\sqrt{d})$. In fact, it is often helpful to think of picking a random point on the sphere as very similar to picking a random point of the form $(\pm\frac{1}{\sqrt{d}}, \cdots, \pm\frac{1}{\sqrt{d}})$.

3 Generating Points Uniformly at Random on the Surface of a Ball

Consider generating points uniformly at random on the surface of the unit ball. For the 2-dimensional version of generating points on the circumference of a unit-radius circle, independently generate each coordinate uniformly at random from the interval $[-1, 1]$. This produces points distributed over a square that completely contains the unit circle. Project each point onto the unit circle. The distribution is not uniform since more points fall on a line from the origin to a vertex of the square than fall on a line from the origin to the midpoint of an edge of the square due to the difference in length. To solve this issue in $\mathbb{R}^2$, we can instead discard all points outside the unit circle and project the remaining points onto the circle.

In higher dimensions, however, this method does not work since the fraction of points that fall inside the unit ball drops to zero and all of the points would be thrown away. The solution is to generate a point each of whose coordinates is an independent Gaussian variable. Generate $x_1, \cdots, x_d$, using a zero mean, unit variance Gaussian. Thus, the probability

4

density of **x** is

$$p(\mathbf{x}) = \frac{1}{(2\pi)^{d/2}} e^{-\frac{x_1^2 + \cdots + x_d^2}{2}}$$

and is spherically symmetric, i.e., *the function only depends on the magnitude of the input vector*. Normalizing the vector $\mathbf{x} = (x_1, x_2, \cdots, x_d)$ to a unit vector, namely $\frac{\mathbf{x}}{|\mathbf{x}|}$, gives a distribution that is uniform over the surface of the sphere.

References

[BHK20] Avrim Blum, John Hopcroft, and Ravindran Kannan, *Foundations of data science*, Cambridge University Press, 2020. 1