CHAPTER 6 High-dimensional Data

In data mining typically the data is very high dimensional, as the number of attributes can easily be in the hundreds or thousands. Understanding the nature of high-dimensional space, or *hyperspace*, is very important, especially because hyperspace does not behave like the more familiar geometry in two or three dimensions.

6.1 HIGH-DIMENSIONAL OBJECTS

Consider the $n \times d$ data matrix

$$\mathbf{D} = \left(\begin{array}{c|cccc} & X_1 & X_2 & \cdots & X_d \\ \hline \mathbf{x}_1 & x_{11} & x_{12} & \cdots & x_{1d} \\ \mathbf{x}_2 & x_{21} & x_{22} & \cdots & x_{2d} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \mathbf{x}_n & x_{n1} & x_{n2} & \cdots & x_{nd} \end{array}\right)$$

where each point $\mathbf{x}_i \in \mathbb{R}^d$ and each attribute $X_j \in \mathbb{R}^n$.

Hypercube

Let the minimum and maximum values for each attribute X_j be given as

$$\min(X_j) = \min_i \left\{x_{ij}\right\} \qquad \max(X_j) = \max_i \left\{x_{ij}\right\}$$

The data hyperspace can be considered as a d-dimensional *hyper-rectangle*, defined as

$$\begin{aligned} R_d &= \prod_{j=1}^{d} \Big[\min(X_j), \max(X_j)\Big] \\ &= \Big\{\mathbf{x} = (x_1, x_2, \ldots, x_d)^T \mid x_j \in [\min(X_j), \max(X_j)], \text{for } j = 1, \ldots, d\Big\} \end{aligned}$$

Assume the data is centered to have mean $\boldsymbol{\mu} = \mathbf{0}$. Let m denote the largest absolute value in $\mathbf{D}$, given as

$$m = \max_{j=1}^{d} \max_{i=1}^{n} \left\{ |x_{ij}| \right\}$$

The data hyperspace can be represented as a *hypercube*, centered at $\mathbf{0}$, with all sides of length $l = 2m$, given as

$$H_d(l) = \left\{ \mathbf{x} = (x_1, x_2, \dots, x_d)^T \mid \forall i,\ x_i \in [-l/2, l/2] \right\}$$

The hypercube in one dimension, $H_1(l)$, represents an interval, which in two dimensions, $H_2(l)$, represents a square, and which in three dimensions, $H_3(l)$, represents a cube, and so on. The *unit hypercube* has all sides of length $l = 1$, and is denoted as $H_d(1)$.

Hypersphere

Assume that the data has been centered, so that $\boldsymbol{\mu} = \mathbf{0}$. Let r denote the largest magnitude among all points:

$$r = \max_i \left\{ \|\mathbf{x}_i\| \right\}$$

The data hyperspace can also be represented as a d-dimensional *hyperball* centered at $\mathbf{0}$ with radius r, defined as

$$B_d(r) = \left\{ \mathbf{x} \mid \ \|\mathbf{x}\| \le r \right\}$$

$$\text{or } B_d(r) = \left\{ \mathbf{x} = (x_1, x_2, \dots, x_d) \mid \sum_{j=1}^{d} x_j^2 \le r^2 \right\}$$

The surface of the hyperball is called a *hypersphere*, and it consists of all the points exactly at distance r from the center of the hyperball, defined as

$$S_d(r) = \left\{ \mathbf{x} \mid \ \|\mathbf{x}\| = r \right\}$$

$$\text{or } S_d(r) = \left\{ \mathbf{x} = (x_1, x_2, \dots, x_d) \mid \sum_{j=1}^{d} (x_j)^2 = r^2 \right\}$$

Because the hyperball consists of all the surface and interior points, it is also called a *closed hypersphere*.

Example 6.1. Consider the 2-dimensional, centered, Iris dataset, plotted in Figure 6.1. The largest absolute value along any dimension is $m = 2.06$, and the point with the largest magnitude is $(2.06, 0.75)$, with $r = 2.19$. In two dimensions, the hypercube representing the data space is a square with sides of length $l = 2m = 4.12$. The hypersphere marking the extent of the space is a circle (shown dashed) with radius $r = 2.19$.

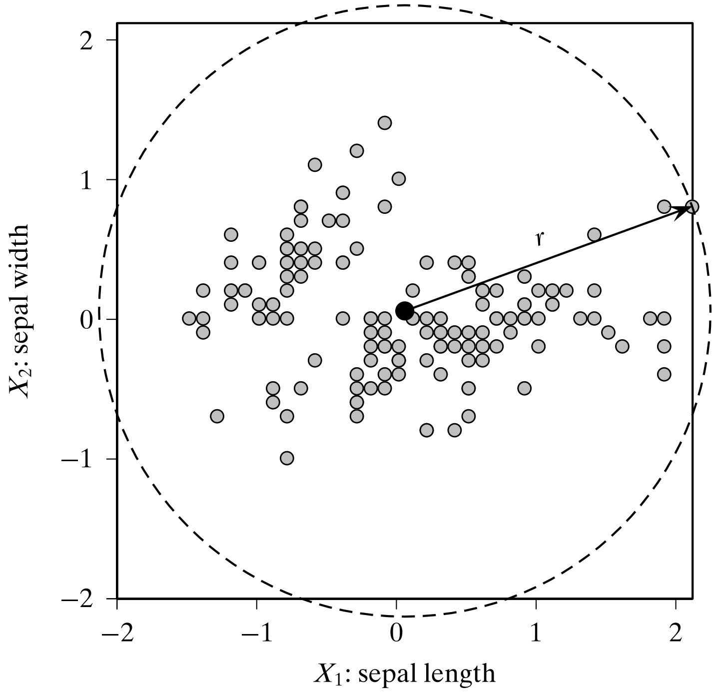

Figure 6.1. Iris data hyperspace: hypercube (solid; with $l = 4.12$) and hypersphere (dashed; with $r = 2.19$).

6.2 HIGH-DIMENSIONAL VOLUMES

Hypercube

The volume of a hypercube with edge length l is given as

$$\mathrm{vol}(H_d(l)) = l^d$$

Hypersphere

The volume of a hyperball and its corresponding hypersphere is identical because the volume measures the total content of the object, including all internal space. Consider the well known equations for the volume of a hypersphere in lower dimensions

$$\mathrm{vol}(S_1(r)) = 2r \tag{6.1}$$

$$\mathrm{vol}(S_2(r)) = \pi r^2 \tag{6.2}$$

$$\mathrm{vol}(S_3(r)) = \frac{4}{3}\pi r^3 \tag{6.3}$$

As per the derivation in Appendix 6.7, the general equation for the volume of a d-dimensional hypersphere is given as

$$\mathrm{vol}(S_d(r)) = K_d r^d = \left(\frac{\pi^{\frac{d}{2}}}{\Gamma\left(\frac{d}{2}+1\right)} \right) r^d \tag{6.4}$$

where

$$K_d = \frac{\pi^{d/2}}{\Gamma(\frac{d}{2}+1)} \tag{6.5}$$

is a scalar that depends on the dimensionality d, and Γ is the gamma function [Eq. (3.17)], defined as (for $\alpha > 0$)

$$\Gamma(\alpha) = \int_0^\infty x^{\alpha-1} \mathrm{e}^{-x} dx \tag{6.6}$$

By direct integration of Eq. (6.6), we have

$$\Gamma(1) = 1 \qquad \text{and} \qquad \Gamma\left(\frac{1}{2}\right) = \sqrt{\pi} \tag{6.7}$$

The gamma function also has the following property for any $\alpha > 1$:

$$\Gamma(\alpha) = (\alpha-1)\Gamma(\alpha-1) \tag{6.8}$$

For any integer $n \geq 1$, we immediately have

$$\Gamma(n) = (n-1)! \tag{6.9}$$

Turning our attention back to Eq. (6.4), when d is even, then $\frac{d}{2}+1$ is an integer, and by Eq. (6.9) we have

$$\Gamma\left(\frac{d}{2}+1\right) = \left(\frac{d}{2}\right)!$$

and when d is odd, then by Eqs. (6.8) and (6.7), we have

$$\Gamma\left(\frac{d}{2}+1\right) = \left(\frac{d}{2}\right)\left(\frac{d-2}{2}\right)\left(\frac{d-4}{2}\right)\cdots\left(\frac{d-(d-1)}{2}\right)\Gamma\left(\frac{1}{2}\right) = \left(\frac{d!!}{2^{(d+1)/2}}\right)\sqrt{\pi}$$

where $d!!$ denotes the double factorial (or multifactorial), given as

$$d!! = \begin{cases} 1 & \text{if } d=0 \text{ or } d=1 \\ d \cdot (d-2)!! & \text{if } d \geq 2 \end{cases}$$

Putting it all together we have

$$\Gamma\left(\frac{d}{2}+1\right) = \begin{cases} \left(\frac{d}{2}\right)! & \text{if } d \text{ is even} \\ \sqrt{\pi}\left(\frac{d!!}{2^{(d+1)/2}}\right) & \text{if } d \text{ is odd} \end{cases} \tag{6.10}$$

Plugging in values of $\Gamma(d/2+1)$ in Eq. (6.4) gives us the equations for the volume of the hypersphere in different dimensions.

Example 6.2. By Eq. (6.10), we have for $d = 1$, $d = 2$ and $d = 3$:

$$\Gamma(1/2+1) = \frac{1}{2}\sqrt{\pi}$$
$$\Gamma(2/2+1) = 1! = 1$$
$$\Gamma(3/2+1) = \frac{3}{4}\sqrt{\pi}$$

Thus, we can verify that the volume of a hypersphere in one, two, and three dimensions is given as

$$\text{vol}(S_1(r)) = \frac{\sqrt{\pi}}{\frac{1}{2}\sqrt{\pi}} r = 2r$$
$$\text{vol}(S_2(r)) = \frac{\pi}{1} r^2 = \pi r^2$$
$$\text{vol}(S_3(r)) = \frac{\pi^{3/2}}{\frac{3}{4}\sqrt{\pi}} r^3 = \frac{4}{3}\pi r^3$$

which match the expressions in Eqs. (6.1), (6.2), and (6.3), respectively.

Surface Area The *surface area* of the hypersphere can be obtained by differentiating its volume with respect to r, given as

$$\text{area}(S_d(r)) = \frac{d}{dr}\text{vol}(S_d(r)) = \left(\frac{\pi^{\frac{d}{2}}}{\Gamma\left(\frac{d}{2}+1\right)}\right) d r^{d-1} = \left(\frac{2\pi^{\frac{d}{2}}}{\Gamma\left(\frac{d}{2}\right)}\right) r^{d-1}$$

We can quickly verify that for two dimensions the surface area of a circle is given as $2\pi r$, and for three dimensions the surface area of sphere is given as $4\pi r^2$.

Asymptotic Volume An interesting observation about the hypersphere volume is that as dimensionality increases, the volume first increases up to a point, and then starts to decrease, and ultimately vanishes. In particular, for the unit hypersphere with $r = 1$,

$$\lim_{d\to\infty} \text{vol}(S_d(1)) = \lim_{d\to\infty} \frac{\pi^{\frac{d}{2}}}{\Gamma(\frac{d}{2}+1)} \to 0$$

Example 6.3. Figure 6.2 plots the volume of the unit hypersphere in Eq. (6.4) with increasing dimensionality. We see that initially the volume increases, and achieves the highest volume for $d = 5$ with $\text{vol}(S_5(1)) = 5.263$. Thereafter, the volume drops rapidly and essentially becomes zero by $d = 30$.

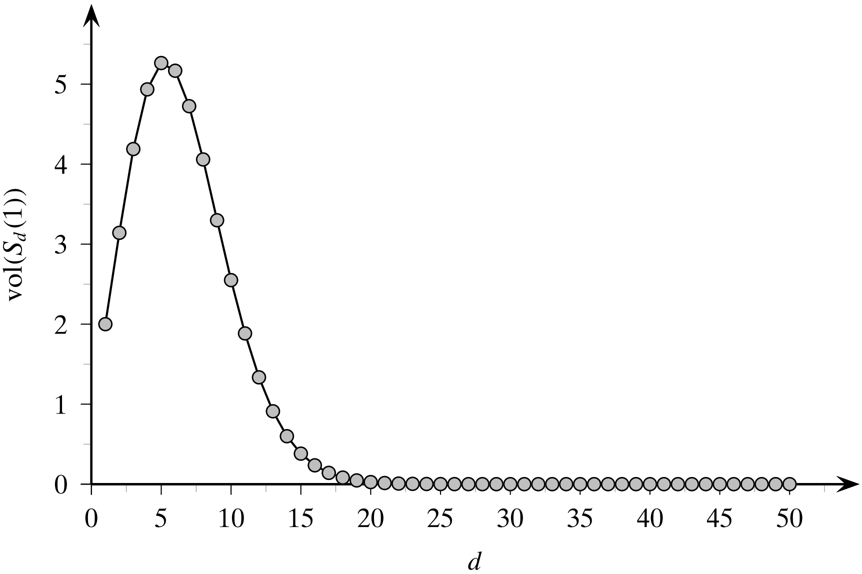

Figure 6.2. Volume of a unit hypersphere.

6.3 HYPERSPHERE INSCRIBED WITHIN HYPERCUBE

We next look at the space enclosed within the largest hypersphere that can be accommodated within a hypercube (which represents the dataspace). Consider a hypersphere of radius r inscribed in a hypercube with sides of length $2r$. When we take the ratio of the volume of the hypersphere of radius r to the hypercube with side length $l = 2r$, we observe the following trends.

In two dimensions, we have

$$\frac{\text{vol}(S_2(r))}{\text{vol}(H_2(2r))} = \frac{\pi r^2}{4r^2} = \frac{\pi}{4} = 78.5\%$$

Thus, an inscribed circle occupies $\frac{\pi}{4}$ of the volume of its enclosing square, as illustrated in Figure 6.3a.

In three dimensions, the ratio is given as

$$\frac{\text{vol}(S_3(r))}{\text{vol}(H_3(2r))} = \frac{\frac{4}{3}\pi r^3}{8r^3} = \frac{\pi}{6} = 52.4\%$$

An inscribed sphere takes up only $\frac{\pi}{6}$ of the volume of its enclosing cube, as shown in Figure 6.3b, which is quite a sharp decrease over the 2-dimensional case.

For the general case, as the dimensionality d increases asymptotically, we get

$$\lim_{d\to\infty} \frac{\text{vol}(S_d(r))}{\text{vol}(H_d(2r))} = \lim_{d\to\infty} \frac{\pi^{d/2}}{2^d\Gamma(\frac{d}{2}+1)} \to 0$$

This means that as the dimensionality increases, most of the volume of the hypercube is in the "corners," whereas the center is essentially empty. The mental picture that

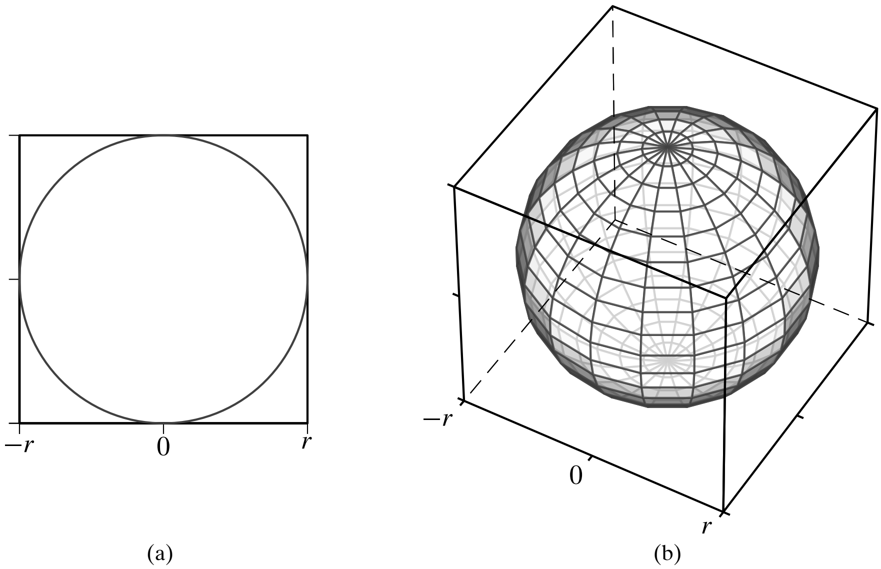

Figure 6.3. Hypersphere inscribed inside a hypercube: in (a) two and (b) three dimensions.

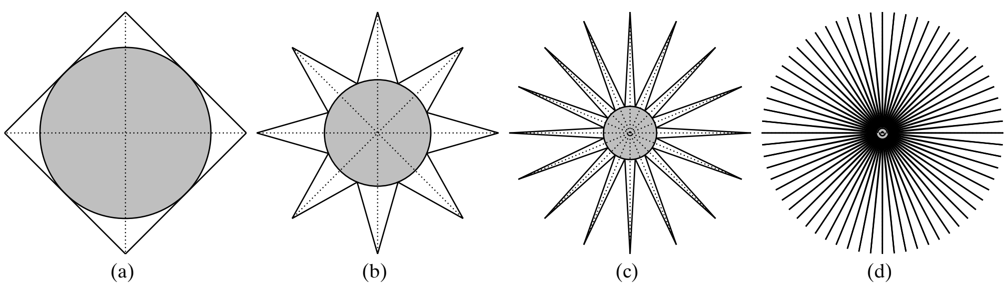

Figure 6.4. Conceptual view of high-dimensional space: (a) two, (b) three, (c) four, and (d) higher dimensions. In d dimensions there are 2^d "corners" and 2^{d-1} diagonals. The radius of the inscribed circle accurately reflects the difference between the volume of the hypercube and the inscribed hypersphere in d dimensions.

emerges is that high-dimensional space looks like a rolled-up porcupine, as illustrated in Figure 6.4.

6.4 VOLUME OF THIN HYPERSPHERE SHELL

Let us now consider the volume of a thin hypersphere shell of width ϵ bounded by an outer hypersphere of radius r, and an inner hypersphere of radius $r-\epsilon$. The volume of the thin shell is given as the difference between the volumes of the two bounding hyperspheres, as illustrated in Figure 6.5.

Let $S_d(r,\epsilon)$ denote the thin hypershell of width ϵ. Its volume is given as

$$\mathrm{vol}(S_d(r,\epsilon)) = \mathrm{vol}(S_d(r)) - \mathrm{vol}(S_d(r-\epsilon)) = K_d r^d - K_d(r-\epsilon)^d.$$

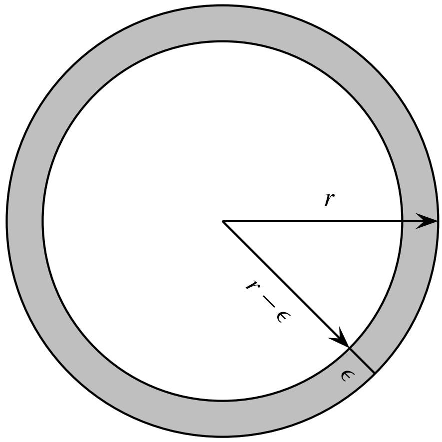

Figure 6.5. Volume of a thin shell (for $\epsilon > 0$).

Let us consider the ratio of the volume of the thin shell to the volume of the outer sphere:

$$\frac{\text{vol}(S_d(r,\epsilon))}{\text{vol}(S_d(r))} = \frac{K_d r^d - K_d (r-\epsilon)^d}{K_d r^d} = 1 - \left(1 - \frac{\epsilon}{r}\right)^d$$

Example 6.4. For example, for a circle in two dimensions, with $r = 1$ and $\epsilon = 0.01$ the volume of the thin shell is $1 - (0.99)^2 = 0.0199 \simeq 2\%$. As expected, in two-dimensions, the thin shell encloses only a small fraction of the volume of the original hypersphere. For three dimensions this fraction becomes $1 - (0.99)^3 = 0.0297 \simeq 3\%$, which is still a relatively small fraction.

Asymptotic Volume

As d increases, in the limit we obtain

$$\lim_{d\to\infty} \frac{\text{vol}(S_d(r,\epsilon))}{\text{vol}(S_d(r))} = \lim_{d\to\infty} 1 - \left(1 - \frac{\epsilon}{r}\right)^d \to 1$$

That is, almost all of the volume of the hypersphere is contained in the thin shell as $d \to \infty$. This means that in high-dimensional spaces, unlike in lower dimensions, most of the volume is concentrated around the surface (within ϵ) of the hypersphere, and the center is essentially void. In other words, if the data is distributed uniformly in the d-dimensional space, then all of the points essentially lie on the boundary of the space (which is a $d - 1$ dimensional object). Combined with the fact that most of the hypercube volume is in the corners, we can observe that in high dimensions, data tends to get scattered on the boundary and corners of the space.

6.5 DIAGONALS IN HYPERSPACE

Another counterintuitive behavior of high-dimensional spaces deals with the diagonals. Let us assume that we have a d-dimensional hypercube, with origin $\mathbf{0}_d = (0_1, 0_2, \ldots, 0_d)$, and bounded in each dimension in the range $[-1, 1]$. Then each "corner" of the hyperspace is a d-dimensional vector of the form $(\pm 1_1, \pm 1_2, \ldots, \pm 1_d)^T$. Let $\mathbf{e}_i = (0_1, \ldots, 1_i, \ldots, 0_d)^T$ denote the d-dimensional canonical unit vector in dimension i, and let $\mathbf{1}$ denote the d-dimensional diagonal vector $(1_1, 1_2, \ldots, 1_d)^T$.

Consider the angle θ_d between the diagonal vector $\mathbf{1}$ and the first axis $\mathbf{e}_1$, in d dimensions:

$$\cos\theta_d = \frac{\mathbf{e}_1^T \mathbf{1}}{\|\mathbf{e}_1\| \, \|\mathbf{1}\|} = \frac{\mathbf{e}_1^T \mathbf{1}}{\sqrt{\mathbf{e}_1^T \mathbf{e}_1}\sqrt{\mathbf{1}^T \mathbf{1}}} = \frac{1}{\sqrt{1}\sqrt{d}} = \frac{1}{\sqrt{d}}$$

Example 6.5. Figure 6.6 illustrates the angle between the diagonal vector $\mathbf{1}$ and $\mathbf{e}_1$, for $d = 2$ and $d = 3$. In two dimensions, we have $\cos\theta_2 = \frac{1}{\sqrt{2}}$ whereas in three dimensions, we have $\cos\theta_3 = \frac{1}{\sqrt{3}}$.

Asymptotic Angle

As d increases, the angle between the d-dimensional diagonal vector $\mathbf{1}$ and the first axis vector $\mathbf{e}_1$ is given as

$$\lim_{d\to\infty} \cos\theta_d = \lim_{d\to\infty} \frac{1}{\sqrt{d}} \to 0$$

which implies that

$$\lim_{d\to\infty} \theta_d \to \frac{\pi}{2} = 90^\circ$$

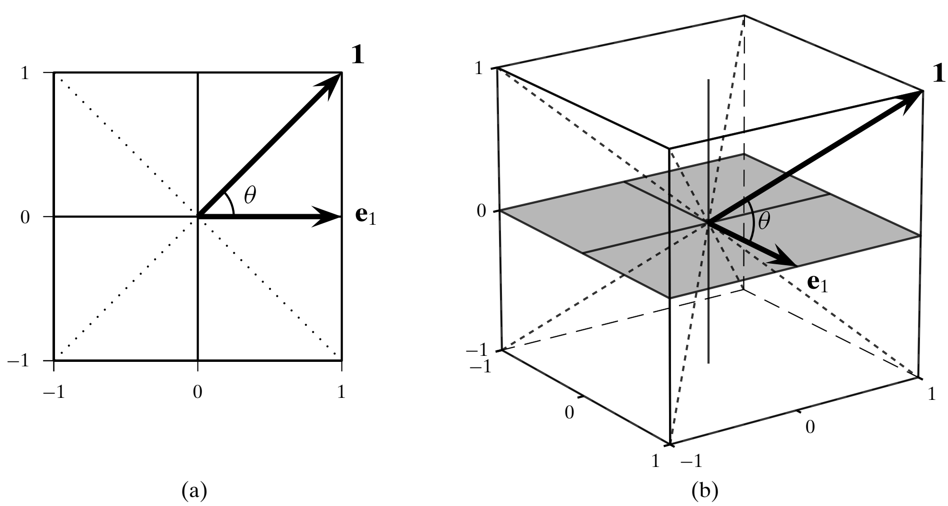

Figure 6.6. Angle between diagonal vector $\mathbf{1}$ and $\mathbf{e}_1$: in (a) two and (b) three dimensions.

This analysis holds for the angle between the diagonal vector $\mathbf{1}_d$ and any of the d principal axis vectors $\mathbf{e}_i$ (i.e., for all $i \in [1, d]$). In fact, the same result holds for any diagonal vector and any principal axis vector (in both directions). This implies that in high dimensions all of the diagonal vectors are perpendicular (or orthogonal) to all the coordinates axes! Because there are 2^d corners in a d-dimensional hyperspace, there are 2^d diagonal vectors from the origin to each of the corners. Because the diagonal vectors in opposite directions define a new axis, we obtain 2^{d-1} new axes, each of which is essentially orthogonal to all of the d principal coordinate axes! Thus, in effect, high-dimensional space has an exponential number of orthogonal "axes." A consequence of this strange property of high-dimensional space is that if there is a point or a group of points, say a cluster of interest, near a diagonal, these points will get projected into the origin and will not be visible in lower dimensional projections.

6.6 DENSITY OF THE MULTIVARIATE NORMAL

Let us consider how, for the standard multivariate normal distribution, the density of points around the mean changes in d dimensions. In particular, consider the probability of a point being within a fraction $\alpha > 0$, of the peak density at the mean.

For a multivariate normal distribution [Eq. (2.33)], with $\boldsymbol{\mu} = \mathbf{0}_d$ (the d-dimensional zero vector), and $\boldsymbol{\Sigma} = \mathbf{I}_d$ (the $d \times d$ identity matrix), we have

$$f(\mathbf{x}) = \frac{1}{(\sqrt{2\pi})^d} \exp\left\{-\frac{\mathbf{x}^T\mathbf{x}}{2}\right\} \tag{6.11}$$

At the mean $\mu = \mathbf{0}_d$, the peak density is $f(\mathbf{0}_d) = \frac{1}{(\sqrt{2\pi})^d}$. Thus, the set of points $\mathbf{x}$ with density at least α fraction of the density at the mean, with $0 < \alpha < 1$, is given as

$$\frac{f(\mathbf{x})}{f(\mathbf{0})} \geq \alpha$$

which implies that

$$\begin{aligned}
\exp\left\{-\frac{\mathbf{x}^T\mathbf{x}}{2}\right\} &\geq \alpha \\
\text{or } \mathbf{x}^T\mathbf{x} &\leq -2\ln(\alpha) \\
\text{and thus} \sum_{i=1}^{d} (x_i)^2 &\leq -2\ln(\alpha)
\end{aligned} \tag{6.12}$$

It is known that if the random variables $X_1, X_2, \ldots, X_k$ are independent and identically distributed, and if each variable has a standard normal distribution, then their squared sum $X^2 + X_2^2 + \cdots + X_k^2$ follows a χ^2 distribution with k degrees of freedom, denoted as χ_k^2. Because the projection of the standard multivariate normal onto any attribute X_j is a standard univariate normal, we conclude that $\mathbf{x}^T\mathbf{x} = \sum_{i=1}^{d}(x_i)^2$ has a χ^2 distribution with d degrees of freedom. The probability that a point $\mathbf{x}$ is within α times the density at the mean can be computed from the χ_d^2 density function using Eq. (6.12), as follows:

$$P\left(\frac{f(\mathbf{x})}{f(\mathbf{0})} \geq \alpha\right) = P\left(\mathbf{x}^T\mathbf{x} \leq -2\ln(\alpha)\right)$$
$$= \int_0^{-2\ln(\alpha)} f_{\chi_d^2}(\mathbf{x}^T\mathbf{x})$$
$$= F_{\chi_d^2}(-2\ln(\alpha)) \tag{6.13}$$

where $f_{\chi_q^2}(x)$ is the chi-squared probability density function [Eq. (3.16)] with q degrees of freedom:

$$f_{\chi_q^2}(x) = \frac{1}{2^{q/2}\Gamma(q/2)} x^{\frac{q}{2}-1} e^{-\frac{x}{2}}$$

and $F_{\chi_q^2}(x)$ is its cumulative distribution function.

As dimensionality increases, this probability decreases sharply, and eventually tends to zero, that is,

$$\lim_{d\to\infty} P\left(\mathbf{x}^T\mathbf{x} \leq -2\ln(\alpha)\right) \to 0 \tag{6.14}$$

Thus, in higher dimensions the probability density around the mean decreases very rapidly as one moves away from the mean. In essence the entire probability mass migrates to the tail regions.

Example 6.6. Consider the probability of a point being within 50% of the density at the mean, that is, $\alpha = 0.5$. From Eq. (6.13) we have

$$P\left(\mathbf{x}^T\mathbf{x} \leq -2\ln(0.5)\right) = F_{\chi_d^2}(1.386)$$

We can compute the probability of a point being within 50% of the peak density by evaluating the cumulative χ^2 distribution for different degrees of freedom (the number of dimensions). For $d = 1$, we find that the probability is $F_{\chi_1^2}(1.386) = 76.1\%$. For $d = 2$ the probability decreases to $F_{\chi_2^2}(1.386) = 50\%$, and for $d = 3$ it reduces to 29.12%. Looking at Figure 6.7, we can see that only about 24% of the density is in the tail regions for one dimension, but for two dimensions more than 50% of the density is in the tail regions.

Figure 6.8 plots the χ_d^2 distribution and shows the probability $P\left(\mathbf{x}^T\mathbf{x} \leq 1.386\right)$ for two and three dimensions. This probability decreases rapidly with dimensionality; by $d = 10$, it decreases to 0.075%, that is, 99.925% of the points lie in the extreme or tail regions.

Distance of Points from the Mean

Let us consider the average distance of a point $\mathbf{x}$ from the center of the standard multivariate normal. Let r^2 denote the square of the distance of a point $\mathbf{x}$ to the center $\boldsymbol{\mu} = \mathbf{0}$, given as

$$r^2 = \|\mathbf{x} - \mathbf{0}\|^2 = \mathbf{x}^T\mathbf{x} = \sum_{i=1}^d x_i^2$$

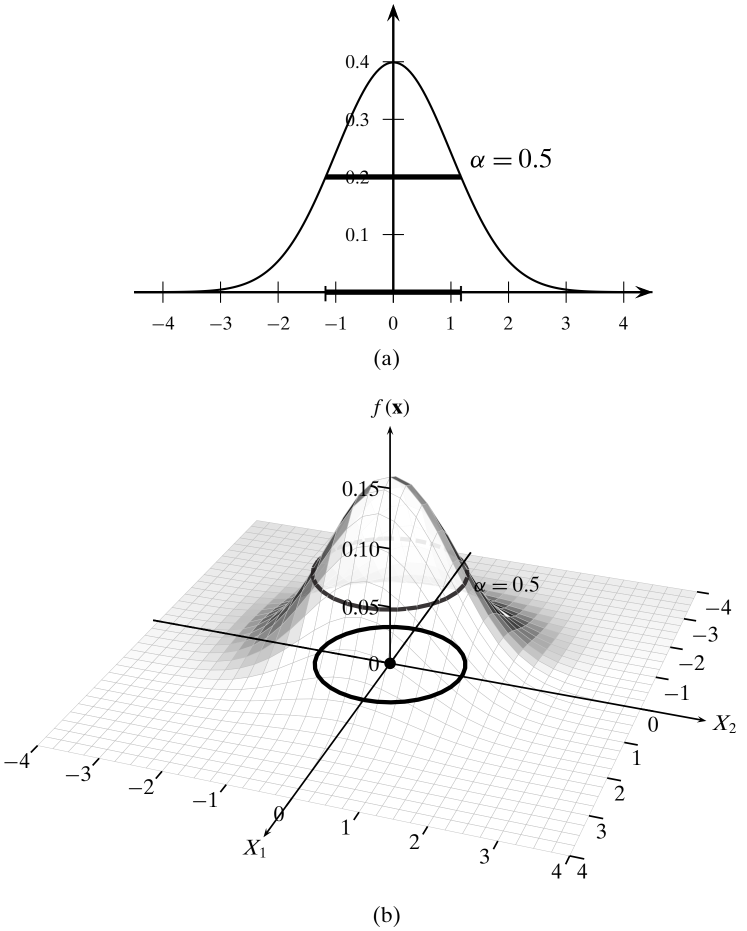

Figure 6.7. Density contour for α fraction of the density at the mean: in (a) one and (b) two dimensions.

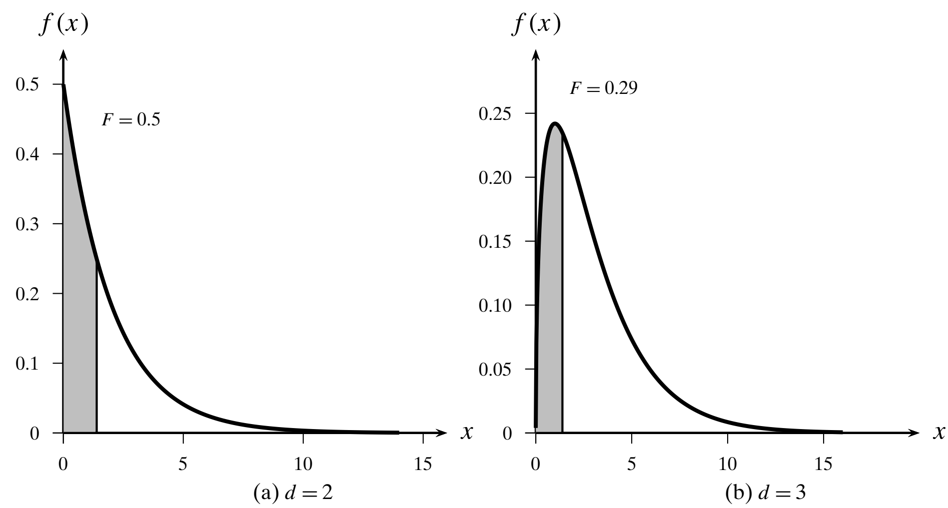

Figure 6.8. Probability $P(\mathbf{x}^T\mathbf{x} \leq -2\ln(\alpha))$, with $\alpha = 0.5$.

$\mathbf{x}^T\mathbf{x}$ follows a χ^2 distribution with d degrees of freedom, which has mean d and variance $2d$. It follows that the mean and variance of the random variable r^2 is

$$\mu_{r^2} = d \qquad \sigma_{r^2}^2 = 2d$$

By the central limit theorem, as $d \to \infty$, r^2 is approximately normal with mean d and variance $2d$, which implies that r^2 is concentrated about its mean value of d. As a consequence, the distance r of a point $\mathbf{x}$ to the center of the standard multivariate normal is likewise approximately concentrated around its mean $\sqrt{d}$.

Next, to estimate the spread of the distance r around its mean value, we need to derive the standard deviation of r from that of r^2. Assuming that σ_r is much smaller compared to r, then using the fact that $\frac{d\log r}{dr} = \frac{1}{r}$, after rearranging the terms, we have

$$\begin{aligned}\frac{dr}{r} &= d\log r\\ &= \frac{1}{2}d\log r^2\end{aligned}$$

Using the fact that $\frac{d\log r^2}{dr^2} = \frac{1}{r^2}$, and rearranging the terms, we obtain

$$\frac{dr}{r} = \frac{1}{2}\frac{dr^2}{r^2}$$

which implies that $dr = \frac{1}{2r}dr^2$. Setting the change in r^2 equal to the standard deviation of r^2, we have $dr^2 = \sigma_{r^2} = \sqrt{2d}$, and setting the mean radius $r = \sqrt{d}$, we have

$$\sigma_r = dr = \frac{1}{2\sqrt{d}}\sqrt{2d} = \frac{1}{\sqrt{2}}$$

We conclude that for large d, the radius r (or the distance of a point $\mathbf{x}$ from the origin $\mathbf{0}$) follows a normal distribution with mean $\sqrt{d}$ and standard deviation $1/\sqrt{2}$. Nevertheless, the density at the mean distance $\sqrt{d}$, is exponentially smaller than that at the peak density because

$$\frac{f(\mathbf{x})}{f(\mathbf{0})} = \exp\left\{-\mathbf{x}^T\mathbf{x}/2\right\} = \exp\{-d/2\}$$

Combined with the fact that the probability mass migrates away from the mean in high dimensions, we have another interesting observation, namely that, whereas the density of the standard multivariate normal is maximized at the center $\mathbf{0}$, most of the probability mass (the points) is concentrated in a small band around the mean distance of $\sqrt{d}$ from the center.

6.7 APPENDIX: DERIVATION OF HYPERSPHERE VOLUME

The volume of the hypersphere can be derived via integration using spherical polar coordinates. We consider the derivation in two and three dimensions, and then for a general d.

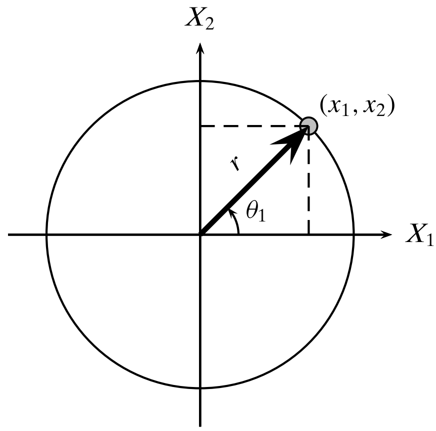

Figure 6.9. Polar coordinates in two dimensions.

Volume in Two Dimensions

As illustrated in Figure 6.9, in $d = 2$ dimensions, the point $\mathbf{x} = (x_1, x_2) \in \mathbb{R}^2$ can be expressed in polar coordinates as follows:

$$x_1 = r\cos\theta_1 = rc_1$$
$$x_2 = r\sin\theta_1 = rs_1$$

where $r = \|\mathbf{x}\|$, and we use the notation $\cos\theta_1 = c_1$ and $\sin\theta_1 = s_1$ for convenience.

The *Jacobian matrix* for this transformation is given as

$$J(\theta_1) = \begin{pmatrix} \frac{\partial x_1}{\partial r} & \frac{\partial x_1}{\partial \theta_1} \\ \frac{\partial x_2}{\partial r} & \frac{\partial x_2}{\partial \theta_1} \end{pmatrix} = \begin{pmatrix} c_1 & -rs_1 \\ s_1 & rc_1 \end{pmatrix}$$

The determinant of the Jacobian matrix is called the *Jacobian*. For $J(\theta_1)$, the Jacobian is given as

$$\det(J(\theta_1)) = rc_1^2 + rs_1^2 = r(c_1^2 + s_1^2) = r \tag{6.15}$$

Using the Jacobian in Eq. (6.15), the volume of the hypersphere in two dimensions can be obtained by integration over r and θ_1 (with $r > 0$, and $0 \le \theta_1 \le 2\pi$)

$$\begin{aligned}
\text{vol}(S_2(r)) &= \int_r \int_{\theta_1} \left|\det(J(\theta_1))\right| \, dr \, d\theta_1 \\
&= \int_0^r \int_0^{2\pi} r \, dr \, d\theta_1 = \int_0^r r \, dr \int_0^{2\pi} d\theta_1 \\
&= \frac{r^2}{2}\bigg|_0^r \cdot \theta_1\bigg|_0^{2\pi} = \pi r^2
\end{aligned}$$

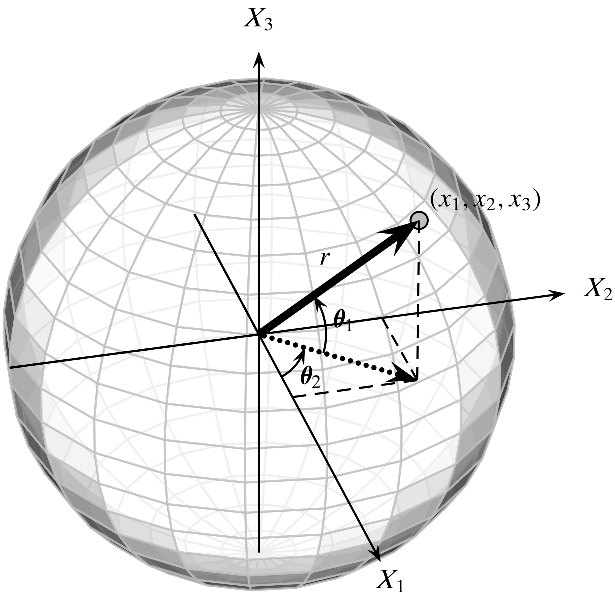

Figure 6.10. Polar coordinates in three dimensions.

Volume in Three Dimensions

As illustrated in Figure 6.10, in $d = 3$ dimensions, the point $\mathbf{x} = (x_1, x_2, x_3) \in \mathbb{R}^3$ can be expressed in polar coordinates as follows:

$$x_1 = r\cos\theta_1 \cos\theta_2 = rc_1c_2$$
$$x_2 = r\cos\theta_1 \sin\theta_2 = rc_1s_2$$
$$x_3 = r\sin\theta_1 = rs_1$$

where $r = \|\mathbf{x}\|$, and we used the fact that the dotted vector that lies in the X_1–X_2 plane in Figure 6.10 has magnitude $r\cos\theta_1$.

The Jacobian matrix is given as

$$J(\theta_1,\theta_2) = \begin{pmatrix} \frac{\partial x_1}{\partial r} & \frac{\partial x_1}{\partial \theta_1} & \frac{\partial x_1}{\partial \theta_2} \\ \frac{\partial x_2}{\partial r} & \frac{\partial x_2}{\partial \theta_1} & \frac{\partial x_2}{\partial \theta_2} \\ \frac{\partial x_3}{\partial r} & \frac{\partial x_3}{\partial \theta_1} & \frac{\partial x_3}{\partial \theta_2} \end{pmatrix} = \begin{pmatrix} c_1c_2 & -rs_1c_2 & -rc_1s_2 \\ c_1s_2 & -rs_1s_2 & rc_1c_2 \\ s_1 & rc_1 & 0 \end{pmatrix}$$

The Jacobian is then given as

$$\begin{aligned} \det(J(\theta_1,\theta_2)) &= s_1(-rs_1)(c_1)\det(J(\theta_2)) - rc_1c_1c_1\det(J(\theta_2)) \\ &= -r^2c_1(s_1^2 + c_2^2) = -r^2c_1 \end{aligned} \tag{6.16}$$

In computing this determinant we made use of the fact that if a column of a matrix $\mathbf{A}$ is multiplied by a scalar s, then the resulting determinant is $s\det(\mathbf{A})$. We also relied on the fact that the $(3,1)$-*minor* of $J(\theta_1,\theta_2)$, obtained by deleting row 3 and column 1 is actually $J(\theta_2)$ with the first column multiplied by $-rs_1$ and the second column multiplied by c_1. Likewise, the $(3,2)$-minor of $J(\theta_1, \theta_2))$ is $J(\theta_2)$ with both the columns multiplied by c_1.

The volume of the hypersphere for $d = 3$ is obtained via a triple integral with $r > 0$, $-\pi/2 \leq \theta_1 \leq \pi/2$, and $0 \leq \theta_2 \leq 2\pi$

$$\begin{aligned}
\text{vol}(S_3(r)) &= \int_r \int_{\theta_1} \int_{\theta_2} \left|\det(J(\theta_1, \theta_2))\right| dr\, d\theta_1\, d\theta_2 \\
&= \int_0^r \int_{-\pi/2}^{\pi/2} \int_0^{2\pi} r^2 \cos\theta_1\, dr\, d\theta_1\, d\theta_2 = \int_0^r r^2\, dr \int_{-\pi/2}^{\pi/2} \cos\theta_1 d\theta_1 \int_0^{2\pi} d\theta_2 \\
&= \frac{r^3}{3}\bigg|_0^r \cdot \sin\theta_1 \bigg|_{-\pi/2}^{\pi/2} \cdot \theta_2 \bigg|_0^{2\pi} = \frac{r^3}{3} \cdot 2 \cdot 2\pi = \frac{4}{3}\pi r^3 \qquad (6.17)
\end{aligned}$$

Volume in d Dimensions

Before deriving a general expression for the hypersphere volume in d dimensions, let us consider the Jacobian in four dimensions. Generalizing the polar coordinates from three dimensions in Figure 6.10 to four dimensions, we obtain

$$\begin{aligned}
x_1 &= r\cos\theta_1 \cos\theta_2 \cos\theta_3 = rc_2c_2c_3 \\
x_2 &= r\cos\theta_1 \cos\theta_2 \sin\theta_3 = rc_1c_2s_3 \\
x_3 &= r\cos\theta_1 \sin\theta_2 = rc_1s_1 \\
x_4 &= r\sin\theta_1 = rs_1
\end{aligned}$$

The Jacobian matrix is given as

$$J(\theta_1, \theta_2, \theta_3) = \begin{pmatrix} \frac{\partial x_1}{\partial r} & \frac{\partial x_1}{\partial \theta_1} & \frac{\partial x_1}{\partial \theta_2} & \frac{\partial x_1}{\partial \theta_3} \\ \frac{\partial x_2}{\partial r} & \frac{\partial x_2}{\partial \theta_1} & \frac{\partial x_2}{\partial \theta_2} & \frac{\partial x_2}{\partial \theta_3} \\ \frac{\partial x_3}{\partial r} & \frac{\partial x_3}{\partial \theta_1} & \frac{\partial x_3}{\partial \theta_2} & \frac{\partial x_3}{\partial \theta_3} \\ \frac{\partial x_4}{\partial r} & \frac{\partial x_4}{\partial \theta_1} & \frac{\partial x_4}{\partial \theta_2} & \frac{\partial x_4}{\partial \theta_3} \end{pmatrix} = \begin{pmatrix} c_1c_2c_3 & -rs_1c_2c_3 & -rc_1s_2c_3 & rc_1c_2s_3 \\ c_1c_2s_3 & -rs_1c_2s_3 & -rc_1s_2s_3 & rc_1c_2c_3 \\ c_1s_2 & -rs_1s_2 & rc_1c_2 & 0 \\ s_1 & rc_1 & 0 & 0 \end{pmatrix}$$

Utilizing the Jacobian in three dimensions [Eq. (6.16)], the Jacobian in four dimensions is given as

$$\begin{aligned}
\det(J(\theta_1, \theta_2, \theta_3)) &= s_1(-rs_1)(c_1)(c_1)\det(J(\theta_2, \theta_3)) - rc_1(c_1)(c_1)(c_1)\det(J(\theta_2, \theta_3)) \\
&= r^3s_1^2c_1^2c_2 + r^3c_1^4c_2 = r^3c_1^2c_2(s_1^2 + c_1^2) = r^3c_1^2c_2
\end{aligned}$$

Jacobian in d Dimensions By induction, we can obtain the d-dimensional Jacobian as follows:

$$\det(J(\theta_1, \theta_2, \ldots, \theta_{d-1})) = (-1)^d r^{d-1} c_1^{d-2} c_2^{d-3} \ldots c_{d-2}$$

The volume of the hypersphere is given by the d-dimensional integral with $r > 0$, $-\pi/2 \le \theta_i \le \pi/2$ for all $i = 1, \ldots, d-2$, and $0 \le \theta_{d-1} \le 2\pi$:

$$\begin{aligned}\text{vol}(S_d(r)) &= \int_r \int_{\theta_1} \int_{\theta_2} \ldots \int_{\theta_{d-1}} \left|\det(J(\theta_1, \theta_2, \ldots, \theta_{d-1}))\right| \, dr \, d\theta_1 \, d\theta_2 \ldots d\theta_{d-1} \\ &= \int_0^r r^{d-1} dr \int_{-\pi/2}^{\pi/2} c_1^{d-2} d\theta_1 \ldots \int_{-\pi/2}^{\pi/2} c_{d-2} d\theta_{d-2} \int_0^{2\pi} d\theta_{d-1} \end{aligned} \tag{6.18}$$

Consider one of the intermediate integrals:

$$\int_{-\pi/2}^{\pi/2} (\cos\theta)^k d\theta = 2 \int_0^{\pi/2} \cos^k \theta d\theta \tag{6.19}$$

Let us substitute $u = \cos^2\theta$, then we have $\theta = \cos^{-1}(u^{1/2})$, and the Jacobian is

$$J = \frac{\partial \theta}{\partial u} = -\frac{1}{2} u^{-1/2} (1-u)^{-1/2} \tag{6.20}$$

Substituting Eq. (6.20) in Eq. (6.19), we get the new integral:

$$\begin{aligned} 2\int_0^{\pi/2} \cos^k \theta d\theta &= \int_0^1 u^{(k-1)/2} (1-u)^{-1/2} du \\ &= B\left(\frac{k+1}{2}, \frac{1}{2}\right) = \frac{\Gamma\left(\frac{k+1}{2}\right)\Gamma\left(\frac{1}{2}\right)}{\Gamma\left(\frac{k}{2}+1\right)} \end{aligned} \tag{6.21}$$

where $B(\alpha, \beta)$ is the *beta function*, given as

$$B(\alpha, \beta) = \int_0^1 u^{\alpha-1} (1-u)^{\beta-1} du$$

and it can be expressed in terms of the gamma function [Eq. (6.6)] via the identity

$$B(\alpha, \beta) = \frac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha+\beta)}$$

Using the fact that $\Gamma(1/2) = \sqrt{\pi}$, and $\Gamma(1) = 1$, plugging Eq. (6.21) into Eq. (6.18), we get

$$\begin{aligned} \text{vol}(S_d(r)) &= \frac{r^d}{d} \frac{\Gamma\left(\frac{d-1}{2}\right)\Gamma\left(\frac{1}{2}\right)}{\Gamma\left(\frac{d}{2}\right)} \frac{\Gamma\left(\frac{d-2}{2}\right)\Gamma\left(\frac{1}{2}\right)}{\Gamma\left(\frac{d-1}{2}\right)} \ldots \frac{\Gamma(1)\Gamma\left(\frac{1}{2}\right)}{\Gamma\left(\frac{3}{2}\right)} 2\pi \\ &= \frac{\pi \Gamma\left(\frac{1}{2}\right)^{d/2-1} r^d}{\frac{d}{2}\Gamma\left(\frac{d}{2}\right)} \\ &= \left(\frac{\pi^{d/2}}{\Gamma\left(\frac{d}{2}+1\right)}\right) r^d \end{aligned}$$

which matches the expression in Eq. (6.4).

6.8 FURTHER READING

For an introduction to the geometry of d-dimensional spaces see Kendall (1961) and also Scott (1992, Section 1.5). The derivation of the mean distance for the multivariate normal is from MacKay (2003, p. 130).

Kendall, M. G. (1961). *A Course in the Geometry of n Dimensions*. New York: Hafner.

MacKay, D. J. (2003). *Information Theory, Inference and Learning Algorithms*. New York: Cambridge University Press.

Scott, D. W. (1992). *Multivariate Density Estimation: Theory, Practice, and Visualization*. New York: John Wiley & Sons.

6.9 EXERCISES

Q1. Given the gamma function in Eq. (6.6), show the following:

(a) $\Gamma(1) = 1$

(b) $\Gamma\left(\frac{1}{2}\right) = \sqrt{\pi}$

(c) $\Gamma(\alpha) = (\alpha - 1)\Gamma(\alpha - 1)$

Q2. Show that the asymptotic volume of the hypersphere $S_d(r)$ for any value of radius r eventually tends to zero as d increases.

Q3. The ball with center $\mathbf{c} \in \mathbb{R}^d$ and radius r is defined as

$$B_d(\mathbf{c}, r) = \left\{\mathbf{x} \in \mathbb{R}^d \mid \delta(\mathbf{x}, \mathbf{c}) \leq r\right\}$$

where $\delta(\mathbf{x}, \mathbf{c})$ is the distance between $\mathbf{x}$ and $\mathbf{c}$, which can be specified using the L_p-norm:

$$L_p(\mathbf{x}, \mathbf{c}) = \left(\sum_{i=1}^{d} |x_i - c_i|^p\right)^{\frac{1}{p}}$$

where $p \neq 0$ is any real number. The distance can also be specified using the L_∞-norm:

$$L_\infty(\mathbf{x}, \mathbf{c}) = \max_i \left\{|x_i - c_i|\right\}$$

Answer the following questions:

(a) For $d = 2$, sketch the shape of the hyperball inscribed inside the unit square, using the L_p-distance with $p = 0.5$ and with center $\mathbf{c} = (0.5, 0.5)^T$.

(b) With $d = 2$ and $\mathbf{c} = (0.5, 0.5)^T$, using the L_∞-norm, sketch the shape of the ball of radius $r = 0.25$ inside a unit square.

(c) Compute the formula for the maximum distance between any two points in the unit hypercube in d dimensions, when using the L_p-norm. What is the maximum distance for $p = 0.5$ when $d = 2$? What is the maximum distance for the L_∞-norm?

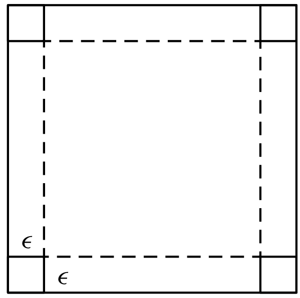

Figure 6.11. For Q4.

Q4. Consider the corner hypercubes of length $\epsilon \leq 1$ inside a unit hypercube. The 2-dimensional case is shown in Figure 6.11. Answer the following questions:

(a) Let $\epsilon = 0.1$. What is the fraction of the total volume occupied by the corner cubes in two dimensions?

(b) Derive an expression for the volume occupied by all of the corner hypercubes of length $\epsilon < 1$ as a function of the dimension d. What happens to the fraction of the volume in the corners as $d \to \infty$?

(c) What is the fraction of volume occupied by the thin hypercube shell of width $\epsilon < 1$ as a fraction of the total volume of the outer (unit) hypercube, as $d \to \infty$? For example, in two dimensions the thin shell is the space between the outer square (solid) and inner square (dashed).

Q5. Prove Eq. (6.14), that is, $\lim_{d\to\infty} P\left(\mathbf{x}^T\mathbf{x} \leq -2\ln(\alpha)\right) \to 0$, for any $\alpha \in (0,1)$ and $\mathbf{x} \in \mathbb{R}^d$.

Q6. Consider the conceptual view of high-dimensional space shown in Figure 6.4. Derive an expression for the radius of the inscribed circle, so that the area in the spokes accurately reflects the difference between the volume of the hypercube and the inscribed hypersphere in d dimensions. For instance, if the length of a half-diagonal is fixed at 1, then the radius of the inscribed circle is $\frac{1}{\sqrt{2}}$ in Figure 6.4a.

Q7. Consider the unit hypersphere (with radius $r = 1$). Inside the hypersphere inscribe a hypercube (i.e., the largest hypercube you can fit inside the hypersphere). An example in two dimensions is shown in Figure 6.12. Answer the following questions:

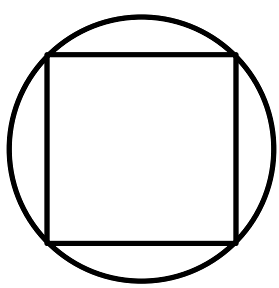

Figure 6.12. For Q7.

(a) Derive an expression for the volume of the inscribed hypercube for any given dimensionality d. Derive the expression for one, two, and three dimensions, and then generalize to higher dimensions.

(b) What happens to the ratio of the volume of the inscribed hypercube to the volume of the enclosing hypersphere as $d \to \infty$? Again, give the ratio in one, two and three dimensions, and then generalize.

Q8. Assume that a unit hypercube is given as $[0,1]^d$, that is, the range is $[0,1]$ in each dimension. The main diagonal in the hypercube is defined as the vector from $(\mathbf{0}, 0) = (\overbrace{0, \ldots, 0}^{d-1}, 0)$ to $(\mathbf{1}, 1) = (\overbrace{1, \ldots, 1}^{d-1}, 1)$. For example, when $d = 2$, the main diagonal goes from $(0,0)$ to $(1,1)$. On the other hand, the main anti-diagonal is defined as the vector from $(\mathbf{1}, 0) = (\overbrace{1, \ldots, 1}^{d-1}, 0)$ to $(\mathbf{0}, 1) = (\overbrace{0, \ldots, 0}^{d-1}, 1)$ For example, for $d = 2$, the anti-diagonal is from $(1,0)$ to $(0,1)$.

(a) Sketch the diagonal and anti-diagonal in $d = 3$ dimensions, and compute the angle between them.

(b) What happens to the angle between the main diagonal and anti-diagonal as $d \to \infty$. First compute a general expression for the d dimensions, and then take the limit as $d \to \infty$.

Q9. Draw a sketch of a hypersphere in four dimensions.