Chapter 6

High Dimensional Data

In data mining typically the data is very high dimensional, since 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 since 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

$$\min(X_j) = \min_{i=1}^{n} \{x_{ij}\}$$
$$\max(X_j) = \max_{i=1}^{n} \{x_{ij}\}$$

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

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

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

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

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

$$H_d(l) = \Big\{ \mathbf{x} = (x_1, x_2, \cdots, x_d)^T \;:\; \forall i,\; x_i \in [-l/2, l/2] \Big\} \tag{6.1}$$

Thus, $H_1(l)$ represents an interval in 1D, $H_2(l)$ represents a square in 2D space, $H_3(l)$ represents a cube in 3D space, and so on. The *unit hypercube* has all sides of length $l = 1$, 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

$$r = \max_{i=1}^{n} \Big\{ \|\mathbf{x}_i\| \Big\}$$

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

$$B_d(r) = \Big\{ \mathbf{x} = (x_1, x_2, \cdots, x_d) \;:\; \sum_{j=1}^{d} x_j^2 \le r^2 \Big\} \tag{6.2}$$

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

$$S_d(r) = \Big\{ \mathbf{x} = (x_1, x_2, \cdots, x_d) \;:\; \sum_{j=1}^{d} (x_j)^2 = r^2 \Big\} \tag{6.3}$$

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

Example 6.1: Consider the 2-dimensional, centered, Iris dataset, plotted in Figure 6.1. The largest absolute value among all the points is $m = 2.06$, and the point with the largest magnitude is $(2.06, 0.75)$, with $r = 2.19$. In 2D, 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 $r = 2.19$.

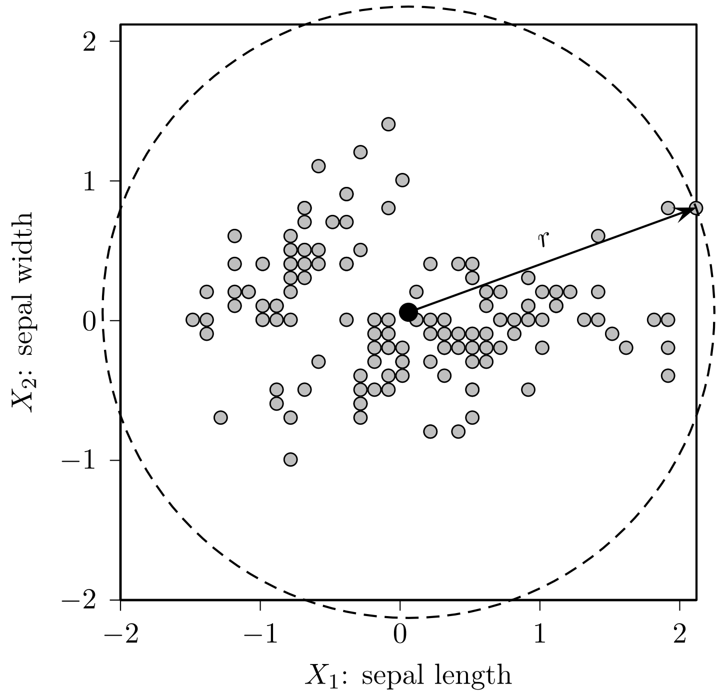


Figure 6.1: Iris Data Hyperspace: Hypercube (solid) and Hypersphere (dashed)

6.2 High Dimensional Volumes

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

$$\mathrm{vol}(H_d(l)) = l^d \tag{6.4}$$

The volume of a hyperball and its corresponding hypersphere is identical, since 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

$$\mathrm{vol}(S_1(r)) = 2r \tag{6.5}$$

$$\mathrm{vol}(S_2(r)) = \pi r^2 \tag{6.6}$$

$$\mathrm{vol}(S_3(r)) = \frac{4}{3}\pi r^3 \tag{6.7}$$

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

$$\mathrm{vol}(S_d(r)) = K_d r^d = \left(\frac{\pi^{\frac{d}{2}}}{\Gamma\left(\frac{d}{2}+1\right)}\right) r^d \tag{6.8}$$

where

$$K_d = \frac{\pi^{d/2}}{\Gamma(\frac{d}{2}+1)} \tag{6.9}$$

is a scalar that depends on the dimensionality d, and Γ is the Gamma function (3.32), defined as (for $\alpha > 0$)

$$\Gamma(\alpha) = \int_0^\infty x^{\alpha-1} e^{-x} dx \tag{6.10}$$

By direct integration of (6.10), we have

$$\Gamma(1) = 1 \qquad \text{and} \qquad \Gamma\left(\frac{1}{2}\right) = \sqrt{\pi} \tag{6.11}$$

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

$$\Gamma(\alpha) = (\alpha - 1)\Gamma(\alpha - 1) \tag{6.12}$$

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

$$\Gamma(n) = (n-1)! \tag{6.13}$$

Turning our attention back to (6.8), when d is even, then $\frac{d}{2}+1$ is an integer, and by (6.13) we have

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

and when d is odd, then by (6.12) and (6.11), 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

$$\Gamma\left(\frac{d}{2}+1\right) = \begin{cases} \left(\frac{d}{2}\right)! & \text{if } d \text{ is even} \\ \sqrt{\pi}\left(\frac{d!!}{2^{(d+1)/2}}\right) & \text{if } d \text{ is odd} \end{cases} \tag{6.14}$$

Plugging in values of $\Gamma(d/2+1)$ in (6.8) give us the equations for the volume of the hypersphere in different dimensions.

Example 6.2: By (6.14), 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 the expressions for the volume of the hypersphere in 1D (6.5), 2D (6.6), and 3D (6.7),

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

Surface Area We can also compute the *surface area* of the hypersphere, which is given as

$$\mathrm{area}(S_d(r)) = \frac{d}{dr}\,\mathrm{vol}(S_d(r)) = \left(\frac{\pi^{\frac{d}{2}}}{\Gamma\left(\frac{d}{2}+1\right)}\right) d r^{d-1} = \left(\frac{2\pi^{\frac{d}{2}}}{\Gamma\left(\frac{d}{2}\right)}\right) r^{d-1} \tag{6.15}$$

We can quickly verify that for 2D, the surface area of a circle is given as $2\pi r$, and in 3D 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 $r = 1$,

$$\lim_{d\to\infty} \mathrm{vol}(S_d(1)) = \lim_{d\to\infty} \frac{\pi^{\frac{d}{2}}}{\Gamma(\frac{d}{2}+1)} \to 0 \tag{6.16}$$

Example 6.3: Figure 6.2 plots the volume of the unit hypersphere with increasing dimensionality. We see that initially the volume increases, and (6.8) achieves the highest volume for $d = 5$ with $\mathrm{vol}(S_5(1)) = 5.263$. Thereafter the volume drops rapidly, and essentially becomes zero by $d = 30$.

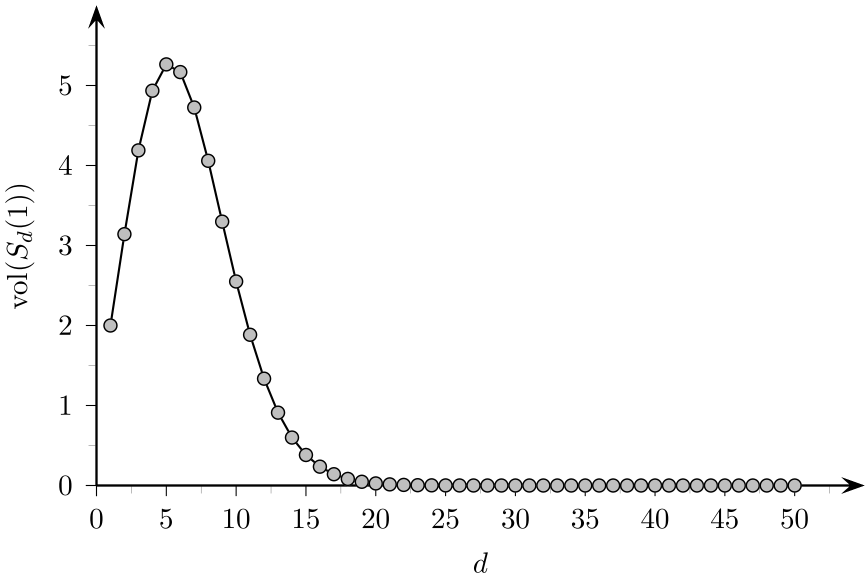


Figure 6.2: Volume of the unit hypersphere

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 data-space). 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 2 Dimensions

$$\frac{\text{vol}(S_2(r))}{\text{vol}(H_2(2r))} = \frac{\pi r^2}{4r^2} = \frac{\pi}{4} = 78.5\% \tag{6.17}$$

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

In 3 Dimensions

$$\frac{\text{vol}(S_3(r))}{\text{vol}(H_3(2r))} = \frac{\frac{4}{3}\pi r^3}{8r^3} = \frac{\pi}{6} = 52.4\% \tag{6.18}$$

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 2D case.

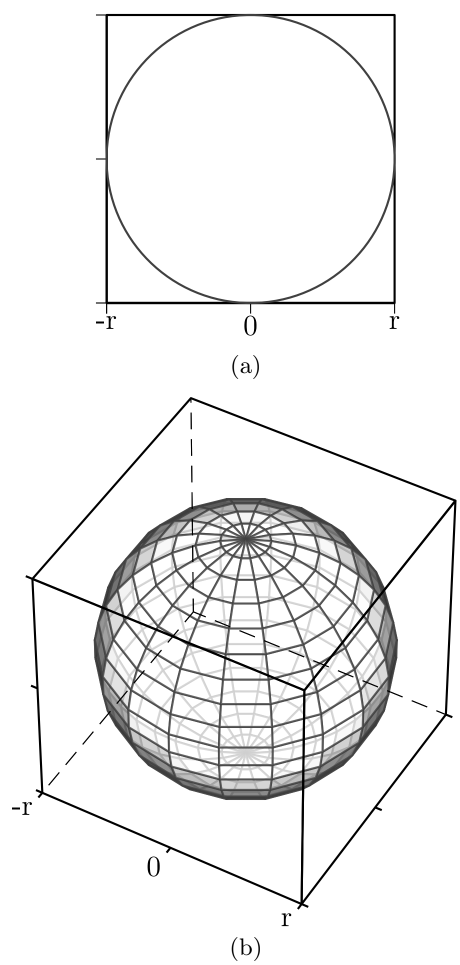


Figure 6.3: Hypersphere inscribed inside a hypercube in (a) 2D, (b) 3D

In d Dimensions As the dimensionality d increases asymptotically, we get

$$\lim_{d\to\infty} \frac{\text{vol}(S_d(r))}{\text{vol}(H_d(2r))} = \lim_{d\to\infty} \frac{\pi^{d/2}}{2^d\Gamma(\frac{d}{2}+1)} \to 0 \tag{6.19}$$

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

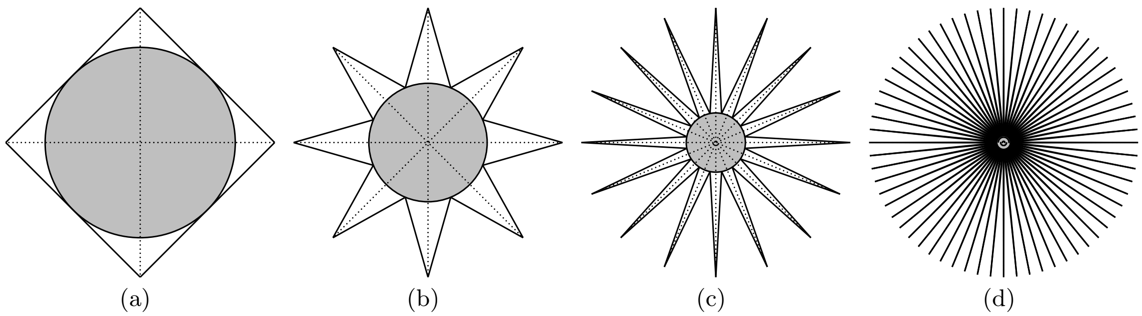


Figure 6.4: Conceptual View of High Dimensional Space in (a) 2D, (b) 3D, (c) 4D, and (d) Higher Dimensions. In d dimensions there are 2^d "corners" and 2^{d-1} diagonals.
The radius of the inscribed circle accurately reflects the difference between the volume of the hypercube and the inscribed hypersphere in d dimensions.

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.

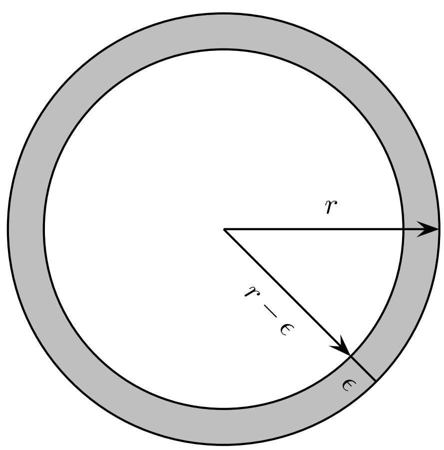


Figure 6.5: Volume of the Thin Shell: Comparing $\text{vol}(S_d(r-\epsilon))$ to $\text{vol}(S_d(r))$ for small $\epsilon > 0$

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

$$\text{vol}(S_d(r,\epsilon)) = \text{vol}(S_d(r)) - \text{vol}(S_d(r-\epsilon)) = K_d r^d - K_d(r-\epsilon)^d. \tag{6.20}$$

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

$$\frac{\mathrm{vol}(S_d(r,\epsilon))}{\mathrm{vol}(S_d(r))} = \frac{K_d r^d - K_d(r-\epsilon)^d}{K_d r^d} = 1 - \left(1 - \frac{\epsilon}{r}\right)^d \tag{6.21}$$

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

$$\lim_{d\to\infty} \frac{\mathrm{vol}(S_d(r,\epsilon))}{\mathrm{vol}(S_d(r))} = \lim_{d\to\infty} 1 - \left(1 - \frac{\epsilon}{r}\right)^d \to 1 \tag{6.22}$$

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

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

$$\cos(\theta_d) = \frac{\mathbf{e}_1^T \mathbf{1}}{\|\mathbf{e}_1\|\,\|\mathbf{1}\|} = \frac{\mathbf{e}_1^T \mathbf{1}}{\sqrt{\mathbf{e}_1^T \mathbf{e}_1}\sqrt{\mathbf{1}^T \mathbf{1}}} = \frac{1}{\sqrt{1}\sqrt{d}} = \frac{1}{\sqrt{d}} \tag{6.23}$$

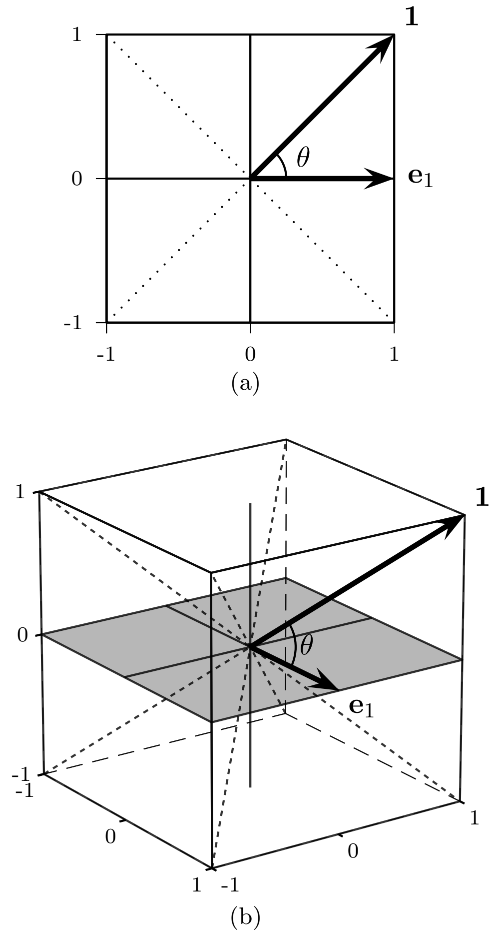


Figure 6.6: Angle between diagonal $\mathbf{1}$ vector and $\mathbf{e}_1$ in (a) 2D, (b) 3D

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

$$\lim_{d\to\infty} \cos(\theta_d) = \lim_{d\to\infty} \frac{1}{\sqrt{d}} \to 0 \implies \lim_{d\to\infty} \theta_d \to \frac{\pi}{2} = 90^\circ \tag{6.24}$$

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 axes! Since there are 2^d corners in a d-dimensional hyperspace, there are 2^d diagonal vectors from the origin to each of the corners. Since 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.

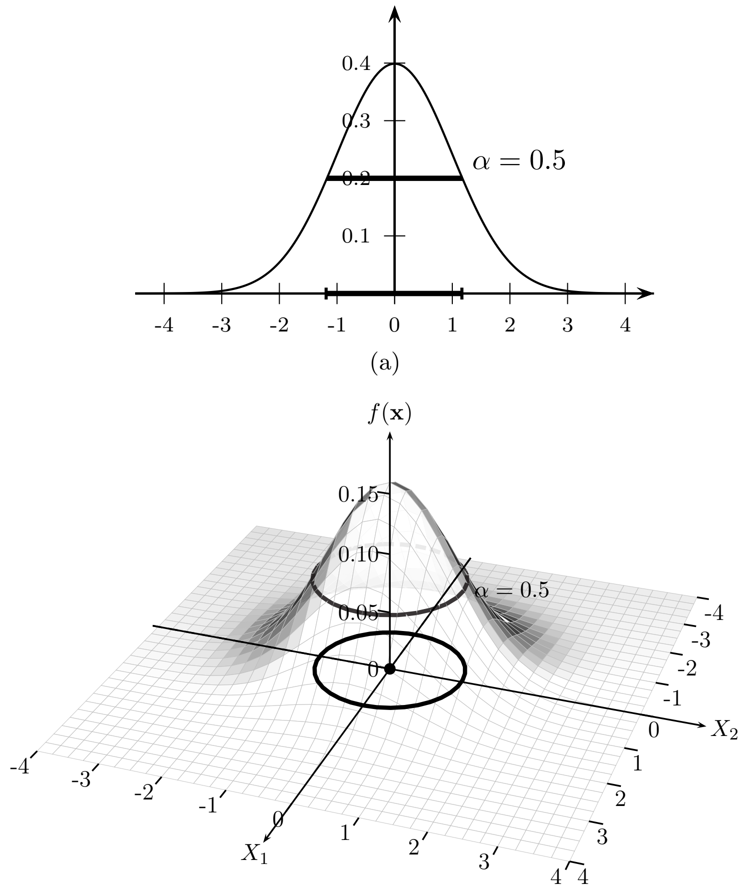
For a multivariate normal distribution (2.54), with $\boldsymbol{\mu} = \mathbf{0}_d$ (the d dimensional zero vector), and $\boldsymbol{\Sigma} = \mathbf{I}_d$ (the $d \times d$ identity matrix), we have

$$f(\mathbf{x}) = \frac{1}{(\sqrt{2\pi})^d} \exp\left\{-\frac{\mathbf{x}^T\mathbf{x}}{2}\right\} \tag{6.25}$$

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 times the density at the mean, is given as

$$\begin{aligned}
\frac{f(\mathbf{x})}{f(\mathbf{0})} &\geq \alpha \\
\exp\left\{-\frac{\mathbf{x}^T\mathbf{x}}{2}\right\} &\geq \alpha \\
\mathbf{x}^T\mathbf{x} \leq -2\ln(\alpha) \implies \sum_{i=1}^{d}(x_i)^2 &\leq -2\ln(\alpha)
\end{aligned} \tag{6.26}$$

It is known that if the random variables X_1, X_2, $\cdots$, 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. Since 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 the α times the density at the mean, given as $P\left(\frac{f(\mathbf{x})}{f(\mathbf{0})} \geq \alpha\right)$, is thus



(b)

Figure 6.7: Density contour for α times the density at the mean:(a) 1D, (b) 2D

the same as the probability $P\left(\mathbf{x}^T\mathbf{x} \le -2\ln(\alpha)\right)$, which can be computed from the χ^2_d density function, i.e.,

$$
\begin{aligned}
P\left(\frac{f(\mathbf{x})}{f(\mathbf{0})} \ge \alpha\right) &= P\left(\mathbf{x}^T\mathbf{x} \le -2\ln(\alpha)\right) \\
&= \int_0^{-2\ln(\alpha)} f_{\chi^2_d}(\mathbf{x}^T\mathbf{x}) = F_{\chi^2_d}(-2\ln(\alpha))
\end{aligned}
\tag{6.27}
$$

where $f_{\chi^2_q}(x)$ is the chi-squared density function with q degrees of freedom (3.31)

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

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

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

$$\lim_{d\to\infty} P\left(\mathbf{x}^T\mathbf{x} \le -2\ln(\alpha)\right) \to 0 \tag{6.28}$$

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, i.e., $\alpha = 0.5$. From (6.27) we have

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

We can compute the probability of a point being within 50% of the mean 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 $F_{\chi^2_1}(1.386) = 76.1\%$. For $d = 2$ the probability decreases to 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 1D, but by 2D over 50% of the density is in the tail regions. Figure 6.8 plots the χ^2_d distribution for different dimensions, and shows the probability $P\left(\mathbf{x}^T\mathbf{x} \le 1.386\right)$ for 2D and 3D. 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 tail regions.

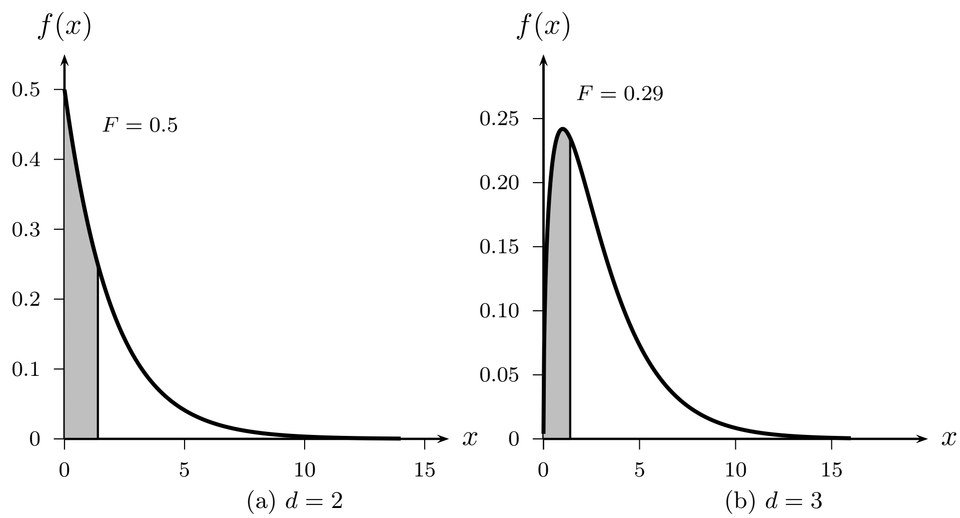


Figure 6.8: Probability $P(\mathbf{x}^T\mathbf{x} \le -2\ln(\alpha))$, with $\alpha = 0.5$

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 $\boldsymbol{\mu} = \mathbf{0}$ of a standard multivariate normal, given as

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

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

$$\mu_{r^2} = d \qquad\qquad \sigma^2_{r^2} = 2d$$

By the central limit theorem, as $d \to \infty$, the r^2 is approximately normal with mean d (and variance $2d$), which implies that r^2 is concentrated about its mean value (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 the mean $\sqrt{d}$. Next, we need to derive the standard deviation of r from that of r^2. Assuming that σ_r is much smaller compared to r, we have

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

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

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

We conclude for large d, the radius r follows a normal distribution with mean $\sqrt{d}$ and standard deviation $1/\sqrt{2}$. Furthermore, the density at the mean distance $\sqrt{d}$, is exponentially smaller than that at the peak density, since

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

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 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 2D, 3D, and then for a general d.

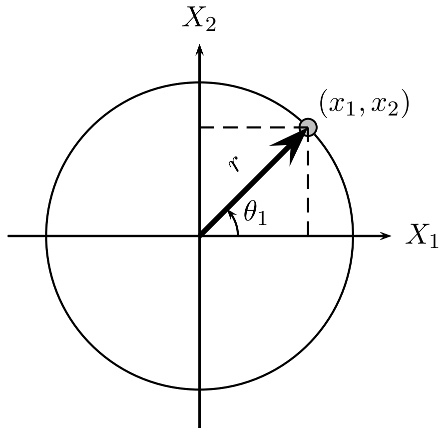


Figure 6.9: Polar coordinates in 2D

Volume in 2D 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 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_1, the Jacobian is given as

$$\det(J(\theta_1)) = rc_1^2 + rs_1^2 = r(c_1^2 + s_1^2) = r \tag{6.29}$$

Using the Jacobian in (6.29), the volume of the hypersphere in 2D can be obtained by integration over r and θ_1 (with $r > 0$, and $0 \le \theta_1 \le 2\pi$)

$$\begin{aligned} \mathrm{vol}(S_2(r)) &= \int_r \int_{\theta_1} \left|\det(J(\theta_1))\right| \, dr \, d\theta_1 \\ &= \int_0^r \int_0^{2\pi} r \, dr \, d\theta_1 = \int_0^r r \, dr \int_0^{2\pi} d\theta_1 \\ &= \frac{r^2}{2}\bigg|_0^r \cdot \theta_1\bigg|_0^{2\pi} = \pi r^2 \end{aligned} \tag{6.30}$$

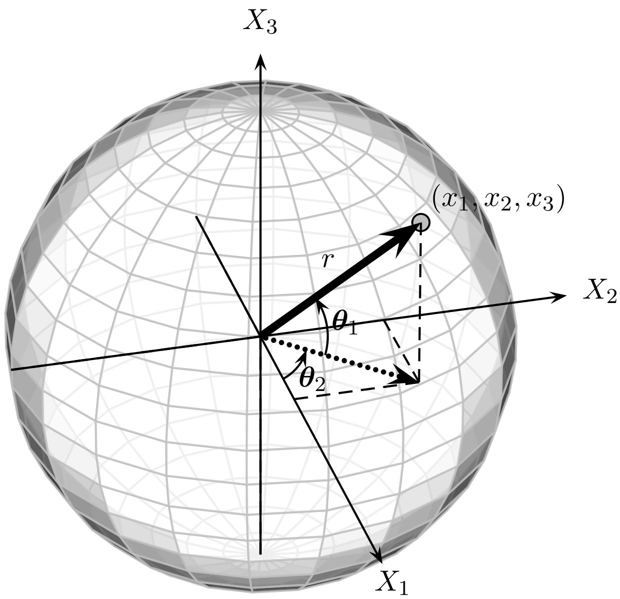


Figure 6.10: Polar coordinates in 3D

Volume in 3D 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

$$\begin{aligned}
x_1 &= r\cos\theta_1\cos\theta_2 = rc_1c_2\\
x_2 &= r\cos\theta_1\sin\theta_2 = rc_1s_2\\
x_3 &= r\sin\theta_1 = rs_1
\end{aligned}$$

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

$$\begin{aligned}
\det(J(\theta_1,\theta_2)) &= s_1(-rs_1)(c_1)\det(J(\theta_2)) - rc_1c_1c_1\det(J(\theta_2))\\
&= -r^2c_1(s_1^2+c_2^2) = -r^2c_1
\end{aligned} \tag{6.31}$$

In computing this determinant we made use of the fact that if a column of a matrix A is multiplied by a scalar s, then the resulting determinant is $s\det(A)$. 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 in $d = 3$ is obtained via a triple integral with $r > 0$, $-\pi/2 \leq \theta_1 \leq \pi/2$, and $0 \leq \theta_2 \leq 2\pi$

$$\begin{aligned}
\mathrm{vol}(S_3(r)) &= \int_r \int_{\theta_1} \int_{\theta_2} \Big|\det(J(\theta_1, \theta_2))\Big| \; dr \; d\theta_1 \; d\theta_2 \\
&= \int_0^r \int_{-\pi/2}^{\pi/2} \int_0^{2\pi} r^2 \cos\theta_1 \; dr \; d\theta_1 \; d\theta_2 = \int_0^r r^2 \; dr \int_{-\pi/2}^{\pi/2} \cos\theta_1 d\theta_1 \int_0^{2\pi} d\theta_2 \\
&= \frac{r^3}{3}\Big|_0^r \cdot \sin\theta_1\Big|_{-\pi/2}^{\pi/2} \cdot \theta_2\Big|_0^{2\pi} = \frac{r^3}{3} \cdot 2 \cdot 2\pi = \frac{4}{3}\pi r^3
\end{aligned} \tag{6.32}$$

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 3D in Figure 6.10 to 4D, 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 3D Jacobian (6.31), the Jacobian in 4D is given as

$$\begin{aligned}
\det(J(\theta_1, \theta_2, \theta_3)) &= s_1(-rs_1)(c_1)(c_1)\det(J(\theta_2, \theta_3)) - rc_1(c_1)(c_1)(c_1)\det(J(\theta_2, \theta_3)) \\
&= r^3 s_1^2 c_1^2 c_2 + r^3 c_1^4 c_2 = r^3 c_1^2 c_2 (s_1^2 + c_1^2) = r^3 c_1^2 c_2
\end{aligned} \tag{6.33}$$

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

$$\det(J(\theta_1, \theta_2, \ldots, \theta_{d-1})) = (-1)^d r^{d-1} c_1^{d-2} c_2^{d-3} \ldots c_{d-2} \tag{6.34}$$

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

$$\begin{aligned}
\text{vol}(S_d(r)) &= \int\limits_r \int\limits_{\theta_1} \int\limits_{\theta_2} \dots \int\limits_{\theta_{d-1}} \Big|\det(J(\theta_1, \theta_2, \dots, \theta_{d-1}))\Big| \, dr \, d\theta_1 \, d\theta_2 \dots d\theta_{d-1} \\
&= \int\limits_0^r r^{d-1} dr \int\limits_{-\pi/2}^{\pi/2} c_1^{d-2} d\theta_1 \, \dots \int\limits_{-\pi/2}^{\pi/2} c_{d-2} d\theta_{d-2} \int\limits_0^{2\pi} d\theta_{d-1} \qquad (6.35)
\end{aligned}$$

Consider one of the intermediate integrals

$$\int\limits_{-\pi/2}^{\pi/2} (\cos\theta)^k d\theta = 2 \int\limits_0^{\pi/2} \cos^k \theta d\theta \qquad (6.36)$$

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

$$J = \frac{\partial \theta}{\partial u} = -\frac{1}{2} u^{-1/2} (1-u)^{-1/2} \qquad (6.37)$$

Substituting (6.37) in (6.36), we get the new integral

$$\begin{aligned}
2 \int\limits_0^{\pi/2} \cos^k \theta d\theta &= \int\limits_0^1 u^{(k-1)/2} (1-u)^{-1/2} du \\
&= B\left(\frac{k+1}{2}, \frac{1}{2}\right) = \frac{\Gamma\left(\frac{k+1}{2}\right) \Gamma\left(\frac{1}{2}\right)}{\Gamma\left(\frac{k}{2}+1\right)} \qquad (6.38)
\end{aligned}$$

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

$$B(\alpha, \beta) = \int\limits_0^1 u^{\alpha-1} (1-u)^{\beta-1} du$$

and it can be expressed in terms of the Gamma function (6.10) 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 (6.38) into (6.35), we get

$$\begin{aligned}
\text{vol}(S_d(r)) &= \frac{r^d}{d} \frac{\Gamma\left(\frac{d-1}{2}\right)\Gamma\left(\frac{1}{2}\right)}{\Gamma\left(\frac{d}{2}\right)} \frac{\Gamma\left(\frac{d-2}{2}\right)\Gamma\left(\frac{1}{2}\right)}{\Gamma\left(\frac{d-1}{2}\right)} \cdots \frac{\Gamma(1)\Gamma\left(\frac{1}{2}\right)}{\Gamma\left(\frac{3}{2}\right)} 2\pi \\
&= \frac{\pi \Gamma(1/2)^{d/2-1} r^d}{\frac{d}{2}\Gamma\left(\frac{d}{2}\right)} \\
&= \left(\frac{\pi^{d/2}}{\Gamma\left(\frac{d}{2}+1\right)}\right) r^d \qquad (6.39)
\end{aligned}$$