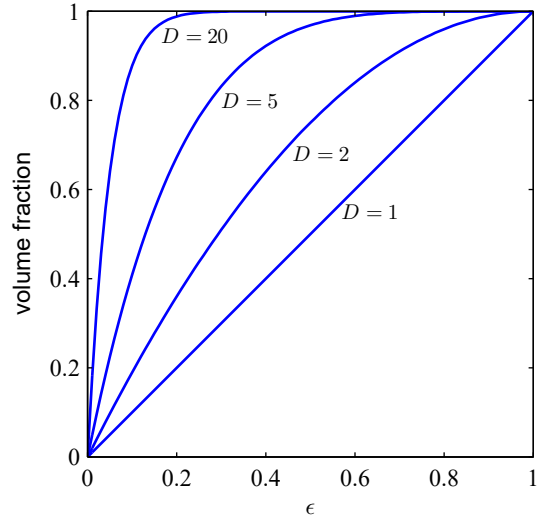**Figure 6.4** Plot of the fraction of the volume of a hypersphere of radius $r = 1$ lying in the range $r = 1 - \epsilon$ to $r = 1$ for various values of the dimensionality D.

data to ensure that the cells are not empty. We have already seen in Figure 6.2 that some cells contain no training points. Hence, a test point in such cells cannot be classified. Clearly, we have no hope of applying such a technique in a space of more than a few variables. The difficulties with both the polynomial regression example and the Iris data classification example arise because the basis functions were chosen independently of the problem being solved. We will need to be more sophisticated *Section 6.1.4* in our choice of basis functions if we are to circumvent the curse of dimensionality.

6.1.2 High-dimensional spaces

First, however, we will look more closely at the properties of spaces with higher dimensionality where our geometrical intuitions, formed through a life spent in a space of three dimensions, can fail badly. As a simple example, consider a hypersphere of radius $r = 1$ in a space of D dimensions, and ask what is the fraction of the volume of the hypersphere that lies between radius $r = 1 - \epsilon$ and $r = 1$. We can evaluate this fraction by noting that the volume $V_D(r)$ of a hypersphere of radius r in D dimensions must scale as r^D, and so we write

$$V_D(r) = K_D r^D \tag{6.4}$$

Exercise 6.1 where the constant K_D depends only on D. Thus, the required fraction is given by

$$\frac{V_D(1) - V_D(1 - \epsilon)}{V_D(1)} = 1 - (1 - \epsilon)^D, \tag{6.5}$$

which is plotted as a function of ϵ for various values of D in Figure 6.4. We see that, for large D, this fraction tends to 1 even for small values of ϵ. Thus, we arrive at the remarkable result that, in spaces of high dimensionality, most of the volume of a hypersphere is concentrated in a thin shell near the surface!