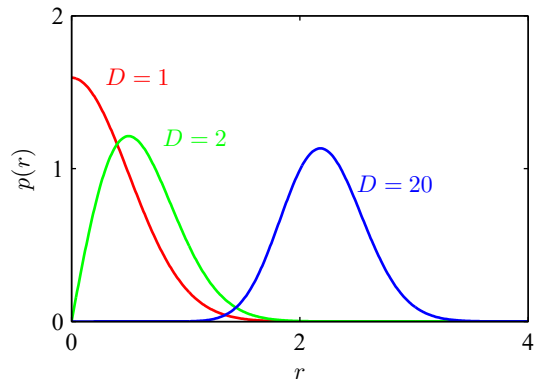


**Figure 6.5** Plot of the probability density with respect to radius  $r$  of a Gaussian distribution for various values of the dimensionality  $D$ . In a high-dimensional space, most of the probability mass of a Gaussian is located within a thin shell at a specific radius.



As a further example of direct relevance to machine learning, consider the behaviour of a Gaussian distribution in a high-dimensional space. If we transform from Cartesian to polar coordinates and then integrate out the directional variables, we obtain an expression for the density  $p(r)$  as a function of radius  $r$  from the origin. Thus,  $p(r)\delta r$  is the probability mass inside a thin shell of thickness  $\delta r$  located at radius  $r$ . This distribution is plotted, for various values of  $D$ , in Figure 6.5, and we see that for large  $D$ , the probability mass of the Gaussian is concentrated in a thin shell at a specific radius.

### Exercise 6.3

In this book, we make extensive use of illustrative examples involving one or two variables, because this makes it particularly easy to visualize these spaces graphically. The reader should be warned, however, that not all intuitions developed in spaces of low dimensionality will generalize to situations involving many dimensions.

Finally, although we have talked about the curse of dimensionality, there can also be advantages to working in high-dimensional spaces. Consider the situation shown in Figure 6.6. We see that this data set, in which each data point consists of a pair of values  $(x_1, x_2)$ , is linearly separable, but when only the value of  $x_1$  is observed, the classes have a strong overlap. The classification problem is therefore much easier in the higher-dimensional space.

### 6.1.3 Data manifolds

With both the polynomial regression model and the grid-based classifier in Figure 6.2, we saw that the number of basis functions grows rapidly with dimensionality, making such methods impractical for applications involving even a few dozen variables, never mind the millions of inputs that often arise with, say, image processing. The problem is that the basis functions are fixed ahead of time and do not depend on the data, or indeed even on the specific problem being solved. We need to find a way to create basis functions that are tuned to the particular application.

Although the curse of dimensionality certainly raises important issues for machine learning applications, it does not prevent us from finding effective techniques applicable to high-dimensional spaces. One reason for this is that real data will generally be confined to a region of the data space having lower effective dimen-