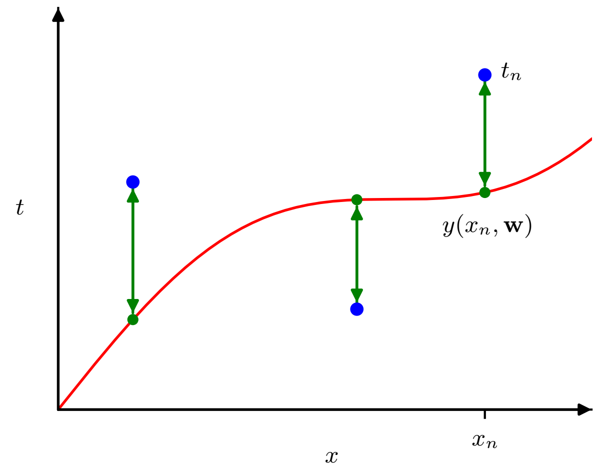

Figure 1.5 The error function (1.2) corresponds to (one half of) the sum of the squares of the displacements (shown by the vertical green arrows) of each data point from the function $y(x, \mathbf{w})$.

the squares of the differences between the predictions $y(x_n, \mathbf{w})$ for each data point x_n and the corresponding target value t_n, given by

$$E(\mathbf{w}) = \frac{1}{2} \sum_{n=1}^{N} \{y(x_n, \mathbf{w}) - t_n\}^2 \tag{1.2}$$

Section 2.3.4

where the factor of $1/2$ is included for later convenience. We will later derive this error function starting from probability theory. Here we simply note that it is a non-negative quantity that would be zero if, and only if, the function $y(x, \mathbf{w})$ were to pass exactly through each training data point. The geometrical interpretation of the sum-of-squares error function is illustrated in Figure 1.5.

Exercise 4.1

We can solve the curve fitting problem by choosing the value of $\mathbf{w}$ for which $E(\mathbf{w})$ is as small as possible. Because the error function is a quadratic function of the coefficients $\mathbf{w}$, its derivatives with respect to the coefficients will be linear in the elements of $\mathbf{w}$, and so the minimization of the error function has a unique solution, denoted by $\mathbf{w}^{\star}$, which can be found in closed form. The resulting polynomial is given by the function $y(x, \mathbf{w}^{\star})$.

1.2.4 Model complexity

There remains the problem of choosing the order M of the polynomial, and as we will see this will turn out to be an example of an important concept called *model comparison* or *model selection*. In Figure 1.6, we show four examples of the results of fitting polynomials having orders $M = 0, 1, 3$, and 9 to the data set shown in Figure 1.4.

Notice that the constant ($M = 0$) and first-order ($M = 1$) polynomials give poor fits to the data and consequently poor representations of the function $\sin(2\pi x)$. The third-order ($M = 3$) polynomial seems to give the best fit to the function $\sin(2\pi x)$ of the examples shown in Figure 1.6. When we go to a much higher order polynomial ($M = 9$), we obtain an excellent fit to the training data. In fact, the polynomial