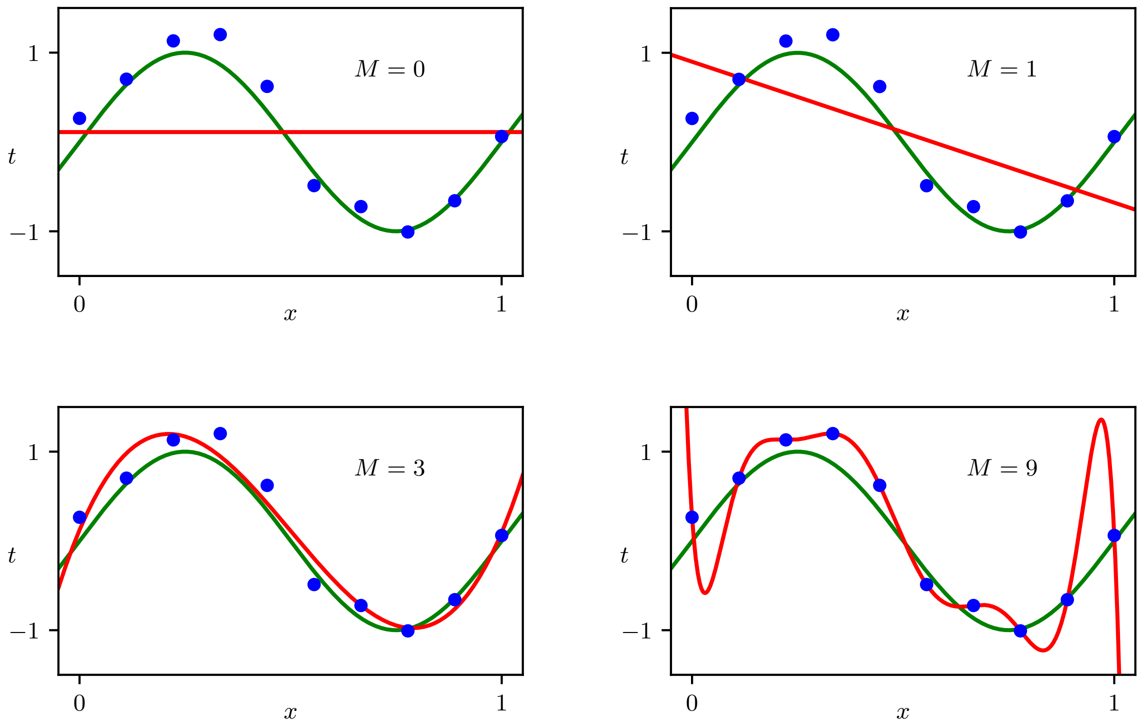

Figure 1.6 Plots of polynomials having various orders M, shown as red curves, fitted to the data set shown in Figure 1.4 by minimizing the error function (1.2).

passes exactly through each data point and $E(\mathbf{w}^\star) = 0$. However, the fitted curve oscillates wildly and gives a very poor representation of the function $\sin(2\pi x)$. This latter behaviour is known as *over-fitting*.

Our goal is to achieve good generalization by making accurate predictions for new data. We can obtain some quantitative insight into the dependence of the generalization performance on M by considering a separate set of data known as a *test set*, comprising 100 data points generated using the same procedure as used to generate the training set points. For each value of M, we can evaluate the residual value of $E(\mathbf{w}^\star)$ given by (1.2) for the training data, and we can also evaluate $E(\mathbf{w}^\star)$ for the test data set. Instead of evaluating the error function $E(\mathbf{w})$, it is sometimes more convenient to use the root-mean-square (RMS) error defined by

$$E_{\mathrm{RMS}} = \sqrt{\frac{1}{N}\sum_{n=1}^{N}\{y(x_n, \mathbf{w}) - t_n\}^2} \tag{1.3}$$

in which the division by N allows us to compare different sizes of data sets on an equal footing, and the square root ensures that E_{RMS} is measured on the same scale (and in the same units) as the target variable t. Graphs of the training-set and test-set