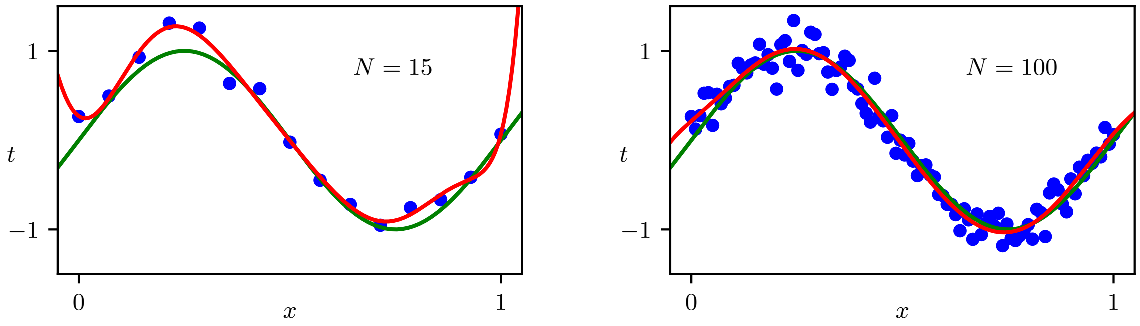

Figure 1.8 Plots of the solutions obtained by minimizing the sum-of-squares error function (1.2) using the $M = 9$ polynomial for $N = 15$ data points (left plot) and $N = 100$ data points (right plot). We see that increasing the size of the data set reduces the over-fitting problem.

as the size of the data set increases. Another way to say this is that with a larger data set, we can afford to fit a more complex (in other words more flexible) model to the data. One rough heuristic that is sometimes advocated in classical statistics is that the number of data points should be no less than some multiple (say 5 or 10) of the number of learnable parameters in the model. However, when we discuss deep learning later in this book, we will see that excellent results can be obtained using models that have significantly more parameters than the number of training data points.

Section 9.3.2

1.2.5 Regularization

There is something rather unsatisfying about having to limit the number of parameters in a model according to the size of the available training set. It would seem more reasonable to choose the complexity of the model according to the complexity of the problem being solved. One technique that is often used to control the over-fitting phenomenon, as an alternative to limiting the number of parameters, is that of *regularization*, which involves adding a penalty term to the error function (1.2) to discourage the coefficients from having large magnitudes. The simplest such penalty

Table 1.1 Table of the coefficients $\mathbf{w}^{\star}$ for polynomials of various order. Observe how the typical magnitude of the coefficients increases dramatically as the order of the polynomial increases.

	$M = 0$	$M = 1$	$M = 3$	$M = 9$
$w_0^{\star}$	0.11	0.90	0.12	0.26
$w_1^{\star}$		−1.58	11.20	−66.13
$w_2^{\star}$			−33.67	1,665.69
$w_3^{\star}$			22.43	−15,566.61
$w_4^{\star}$				76,321.23
$w_5^{\star}$				−217,389.15
$w_6^{\star}$				370,626.48
$w_7^{\star}$				−372,051.47
$w_8^{\star}$				202,540.70
$w_9^{\star}$				−46,080.94