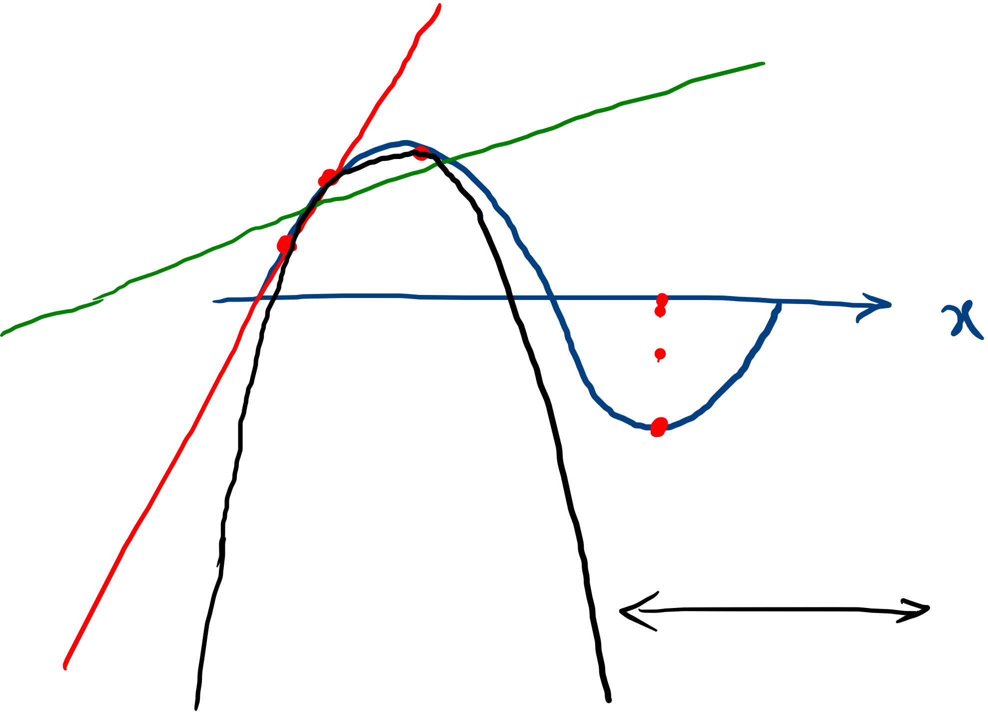
$$D = \{(x_1, y_1), \ldots, (x_n, y_n)\}$$

$$g(x) = w^T x + w_0 \qquad J(w) = SSE = \sum_{i=1}^{n} (\underbrace{w^T x_i + w_0}_{g(x_i)} - y_i)^2$$

$$\boxed{Err = Bias + \underline{Variance}}$$

$$g(x) = \underline{\underline{w_1}}\, x + \underline{\underline{w_0}}$$

$$g(x) = \underline{\underline{w_2}}\, x^2 + \underline{\underline{w_1}}\, x + \underline{\underline{w_0}}$$

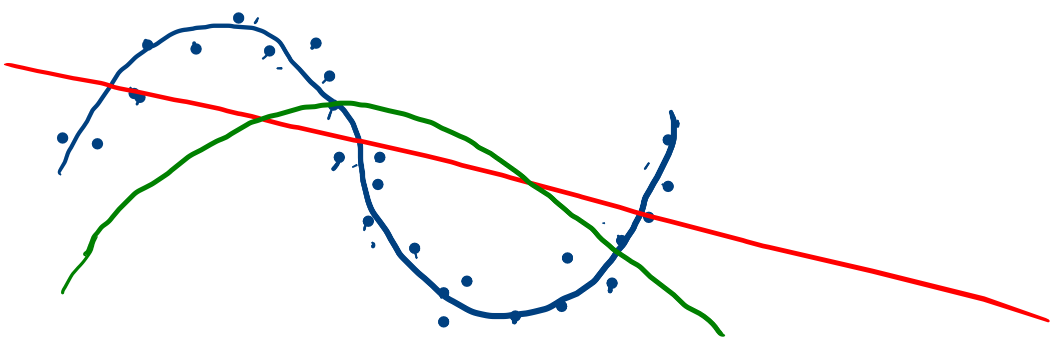


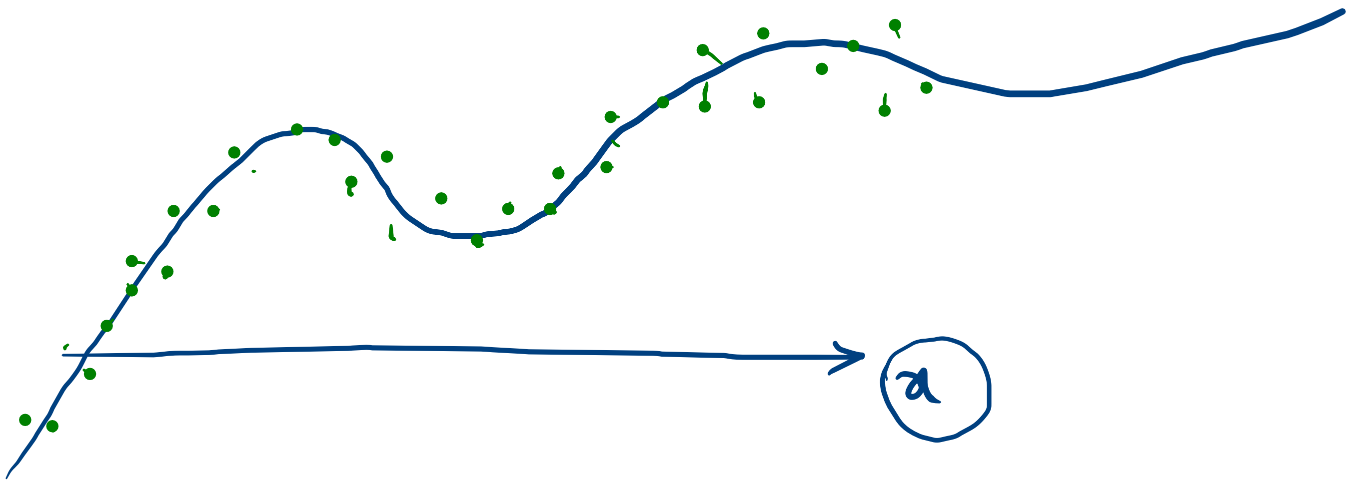
$$f(x) = \sin \pi x$$

$$g(x) \doteq w_1 x + w_0$$

$$|w_1| < \theta$$

$$g(x) \doteq w_2 x^2 + w_1 x + w_0$$





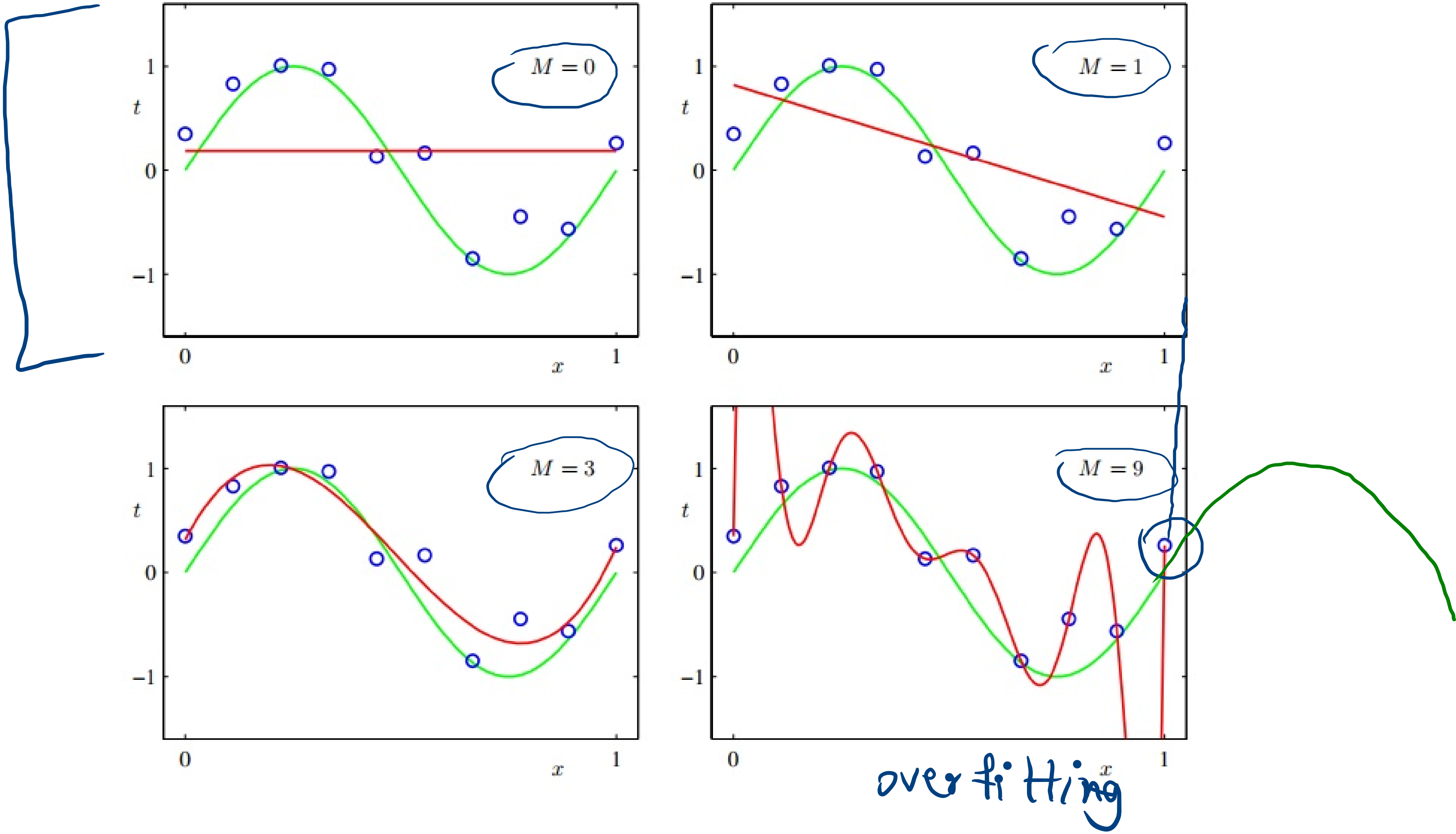
$$g(x) = w_4 x^4 + w_3 x^3 + w_2 x^2 + w_1 x + w_0$$

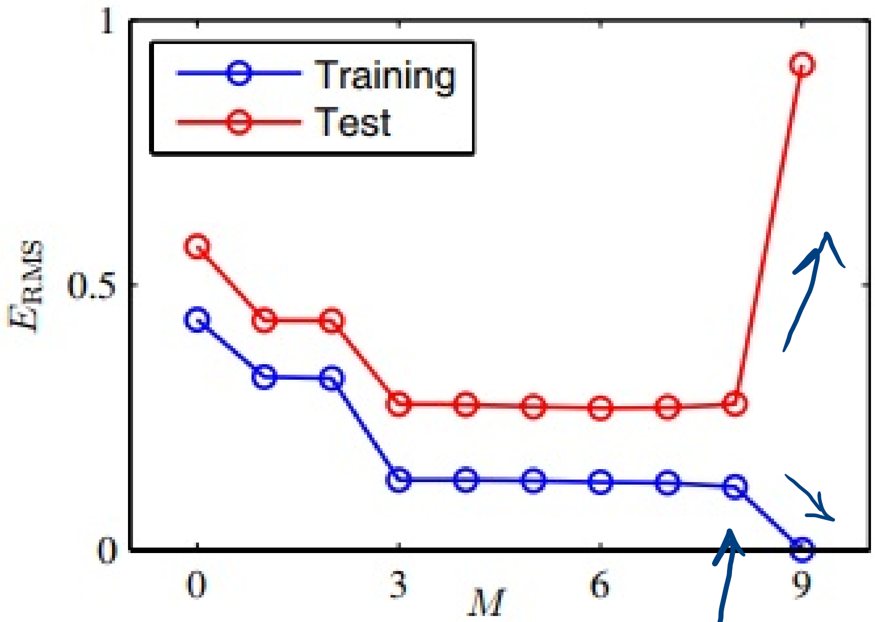
$$\phi(x) = \begin{bmatrix} 1 \\ x \\ x^2 \\ x^3 \\ x^4 \end{bmatrix}$$

$$\boxed{g(x) = w^T \phi(x) \qquad w = \begin{bmatrix} w_1 \\ \vdots \\ w_4 \end{bmatrix}}$$

$$y(x, \mathbf{w}) = w_0 + w_1 x + w_2 x^2 + \ldots + w_M x^M = \sum_{j=0}^{M} w_j x^j$$

M





	$M = 0$	$M = 1$	$M = 3$	$M = 9$
$w_0^\star$	0.19	0.82	0.31	0.35
$w_1^\star$		-1.27	7.99	232.37
$w_2^\star$			-25.43	-5321.83
$w_3^\star$			17.37	48568.31
$w_4^\star$				-231639.30
$w_5^\star$				640042.26
$w_6^\star$				-1061800.52
$w_7^\star$				1042400.18
$w_8^\star$				-557682.99
$w_9^\star$				125201.43

$$|w_i| < \theta_i$$

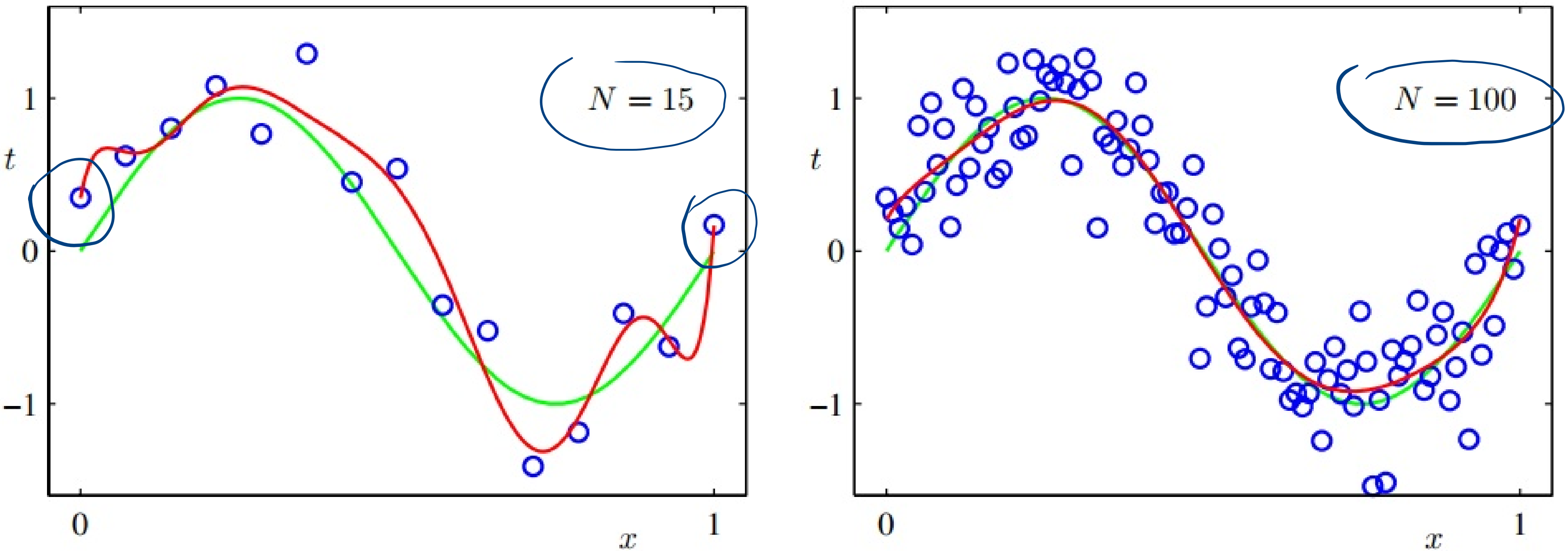


Figure 1.6 Plots of the solutions obtained by minimizing the sum-of-squares error function using the $M = 9$ polynomial for $N = 15$ data points (left plot) and $N = 100$ data points (right plot). We see that increasing the size of the data set reduces the over-fitting problem.

$$g(x) = \sum_{i=1}^{M} w_i x^i$$

$$J(w) = \sum_{j=1}^{n} \left(g(x_j) - y_j \right)^2$$

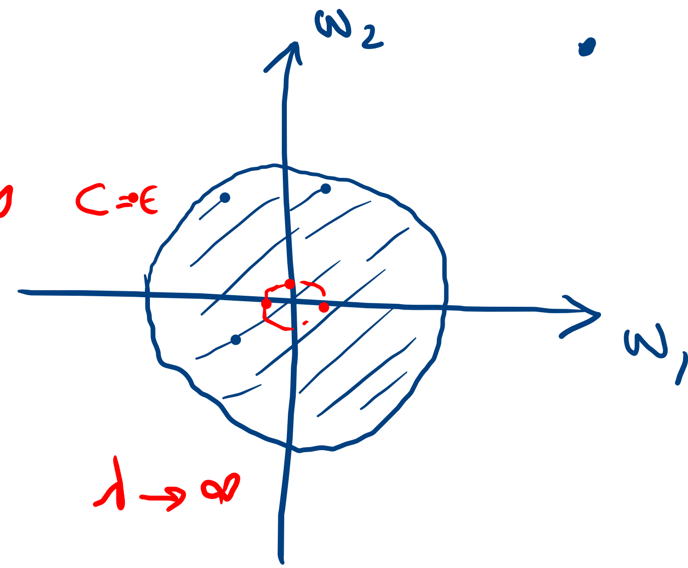
$$\|w\|_2^2 \leq C$$

$$g(x) = w_1 x_1 + w_2 x_2$$

$$\boxed{\min_{w} J(w)}$$

subject to: $\|w\|_2^2 \leq C$

$C = +\infty \qquad C = \epsilon$



$$\boxed{\min_{w} \underline{J(w)} + \underline{\lambda \|w\|_2^2}}$$

$\lambda = 0 \qquad \lambda \to \infty$

$$w^* = \arg\min_w \underbrace{J(w)} + \underbrace{\lambda \|w\|_2^2}_{\text{regularization term}}$$

ridge regression

$$\|w\|_2^2 = \sum_{i=1}^{d} w_i^2$$

$w_i = 0$

$$\boxed{w^* = \arg\min_w J(w) \quad \text{s.t.} \quad \boxed{\|w\|_0 \leq \underline{\underline{C}}}}$$

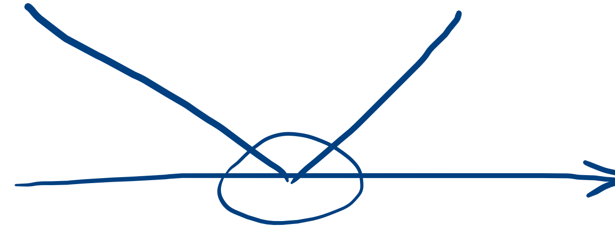
5

$$\Longrightarrow \quad w^* = \arg\min_w \underline{J(w) + \lambda \|w\|_0}$$

$$w^* = \arg\min_{w} J(w) + \lambda \|w\|_1$$

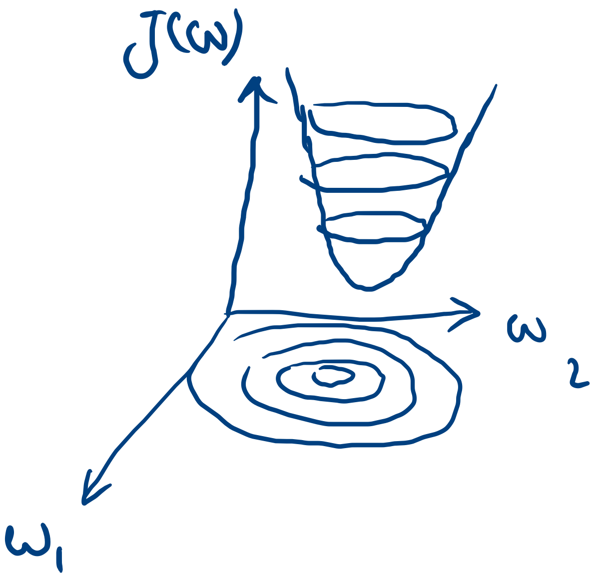
$$aw^2 + bw + c$$

$$a > 0$$



$$\|w\|_1 = \sum_{i=1}^{d} |w_i|$$

$$g(x) = w_1 x_1 + w_2 x_2$$



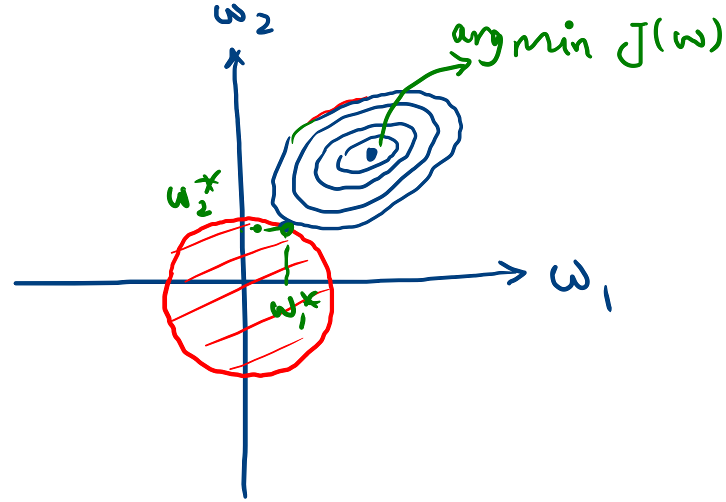
$$J(w) = \sum_{i=1}^{n} (w^T x_i - y_i)^2$$

$$\underbrace{J(w) = w^T A w + b^T w + c}_{\text{quadratic form}}$$

$$A > 0$$

$$\min_{w} J(w)$$

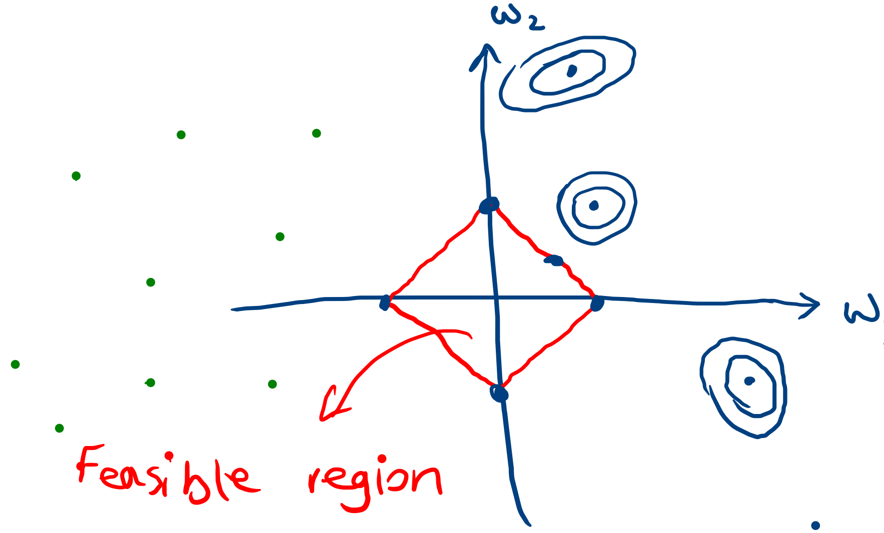
$$\text{s.t.} \quad \|w\|_2^2 \leq C$$



$$\min_{w} J(w)$$

$$\text{s.t.} \quad \|w\|_1 \leq C$$

$$|w_1| + |w_2| < C$$



$$\min_{w} \underbrace{\sum_{i=1}^{n} (w^T x_i - y_i)^2}_{J(w)} + \underline{\underline{\lambda}} \|w\|_1$$

LASSO

$J(w) + \lambda \|w\|_2^2$

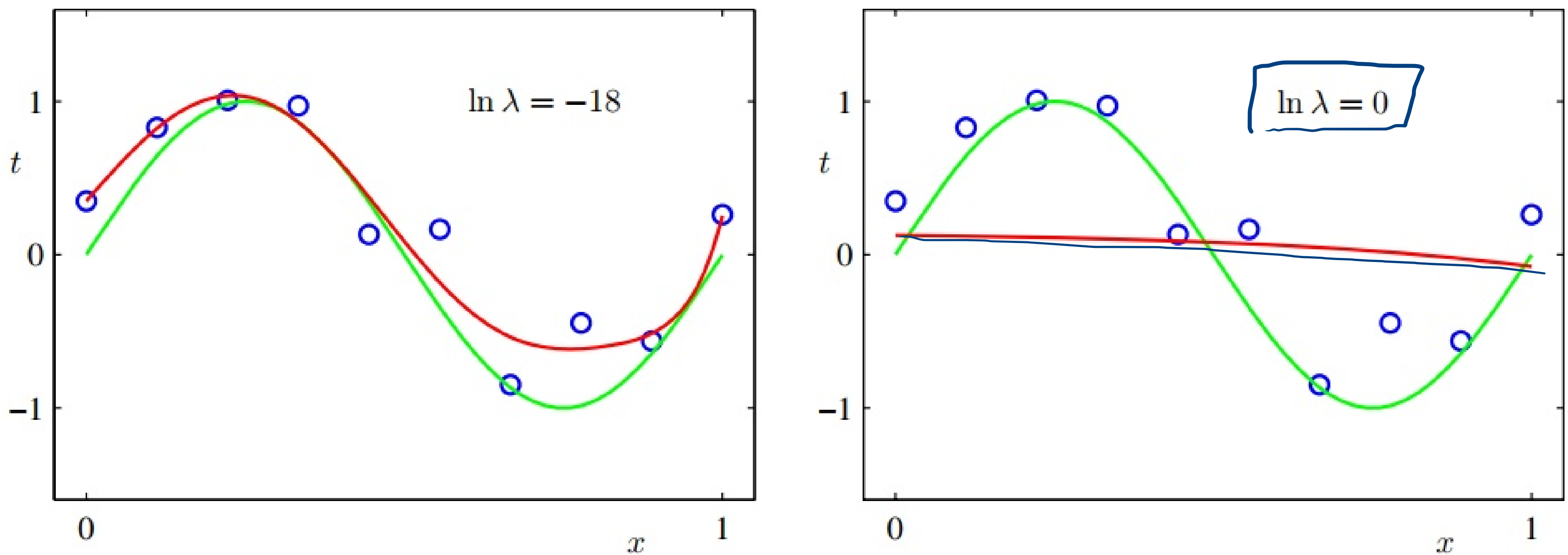


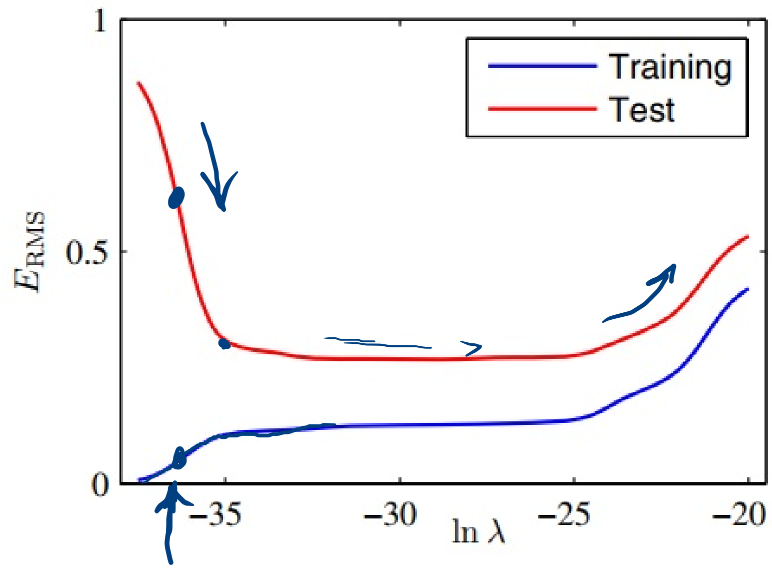
Figure 1.7 Plots of $M = 9$ polynomials fitted to the data set shown in Figure 1.2 using the regularized error function (1.4) for two values of the regularization parameter λ corresponding to $\ln \lambda = -18$ and $\ln \lambda = 0$. The case of no regularizer, i.e., $\lambda = 0$, corresponding to $\ln \lambda = -\infty$, is shown at the bottom right of Figure 1.4.

Table 1.2 Table of the coefficients $\mathbf{w}^\star$ for $M = 9$ polynomials with various values for the regularization parameter λ. Note that $\ln \lambda = -\infty$ corresponds to a model with no regularization, i.e., to the graph at the bottom right in Figure 1.4. We see that, as the value of λ increases, the typical magnitude of the coefficients gets smaller.

	$\ln \lambda = -\infty$	$\ln \lambda = -18$	$\ln \lambda = 0$
$w_0^\star$	0.35	0.35	0.13
$w_1^\star$	232.37	4.74	-0.05
$w_2^\star$	-5321.83	-0.77	-0.06
$w_3^\star$	48568.31	-31.97	-0.05
$w_4^\star$	-231639.30	-3.89	-0.03
$w_5^\star$	640042.26	55.28	-0.02
$w_6^\star$	-1061800.52	41.32	-0.01
$w_7^\star$	1042400.18	-45.95	-0.00
$w_8^\star$	-557682.99	-91.53	0.00
$w_9^\star$	125201.43	72.68	0.01

بدون نرم منظم ساز

Figure 1.8 Graph of the root-mean-square error (1.3) versus $\ln \lambda$ for the $M = 9$ polynomial.



train val test

Data