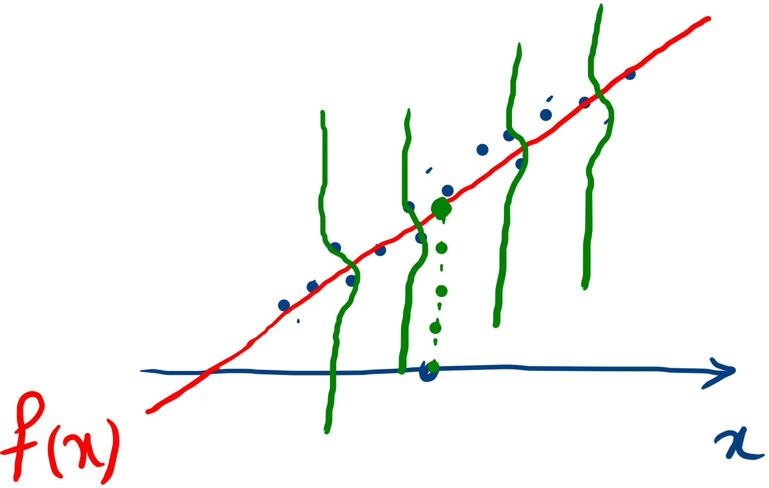
$$g(x) = \underset{?}{\underline{\underline{w}}}^T x$$

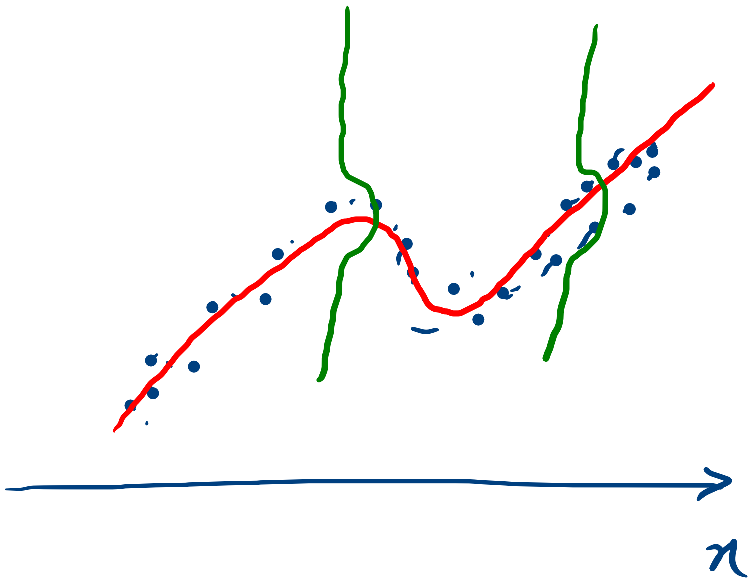
$$w^* = \arg\min_{w} \underbrace{\sum_{i=1}^{n} (w^T x_i - y_i)^2}_{SSE}$$

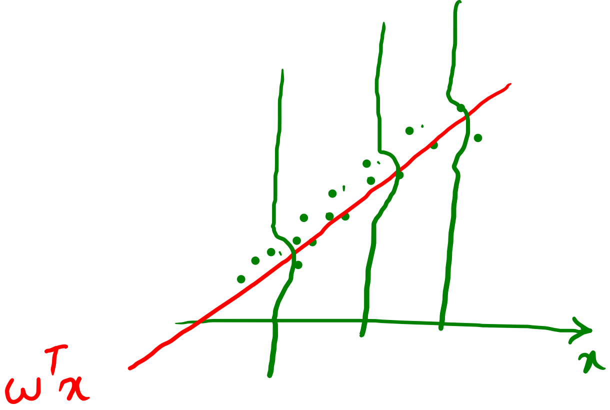
Least Squares



$$y = f(x) + n$$

$$\boxed{n \sim N(0, \sigma^2)}$$





$$\begin{cases} y = w^T x + \epsilon \\ \epsilon \sim N(0, \sigma^2) \end{cases}$$

$$\boxed{P(y|x) = N(y; w^T x, \sigma^2)}$$

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

$$\underset{P(y|x)}{E[y]} = E[y|x] = E[w^T x + \epsilon | x] = w^T x$$

$$w^*_{ML} = \arg\max_w \ln P(D|w) = \arg\max_w \ln P(y_1, \dots, y_n | x_1, \dots, x_n, w)$$

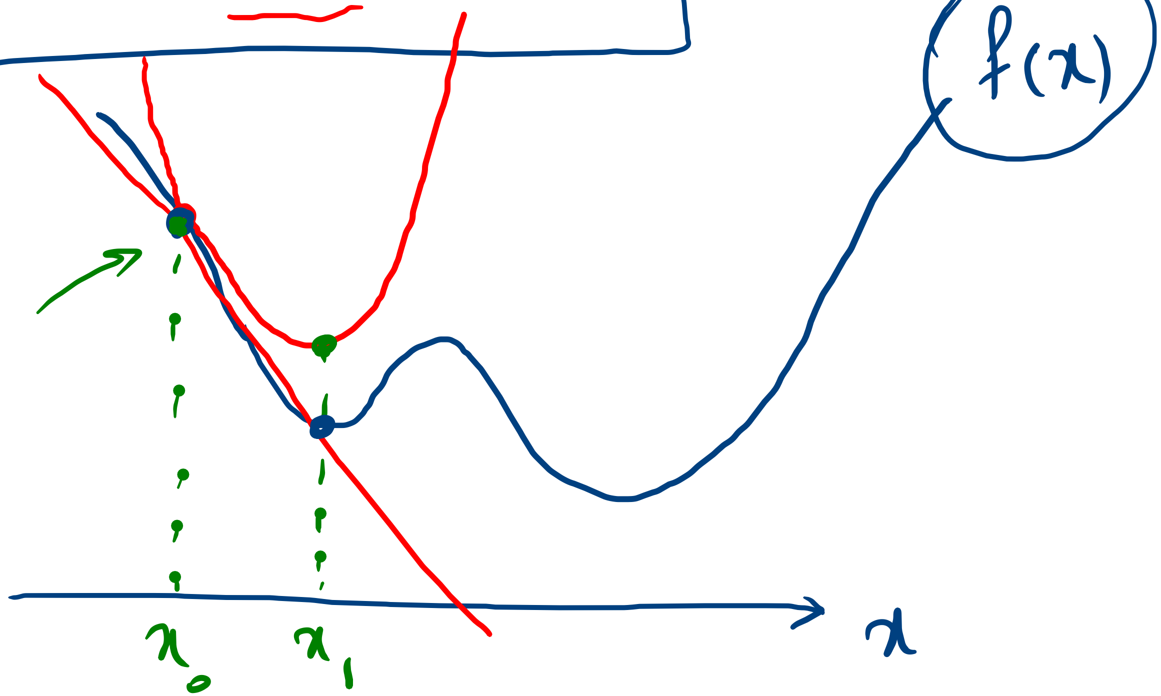
$L(w)$

$$\overset{i.i.d.}{=} \arg\max_w \sum_{i=1}^n \ln P(y_i | x_i, w) = \arg\max_w \sum_{i=1}^n \ln N(y_i; w^T x_i, \sigma^2)$$

$$= \arg\max_w \sum_{i=1}^n \ln \frac{1}{\sqrt{2\pi}\,\sigma} \exp\left(\frac{-1}{2} \frac{(w^T x_i - y_i)^2}{\sigma^2}\right)$$

$$= \arg\max_w \sum_i -\ln\sqrt{2\pi}\,\sigma - \frac{1}{2}\frac{(w^T x_i - y_i)^2}{\sigma^2} = \arg\min_w \overbrace{\sum_{i=1}^n (w^T x_i - y_i)^2}^{SSE \quad f(w)}$$

$$\boxed{w^* = \underline{(X^T X)}^{-1} X^T y}$$



$ax + b$

$a = f'(x_0)$

$ax^2 + bx + c$

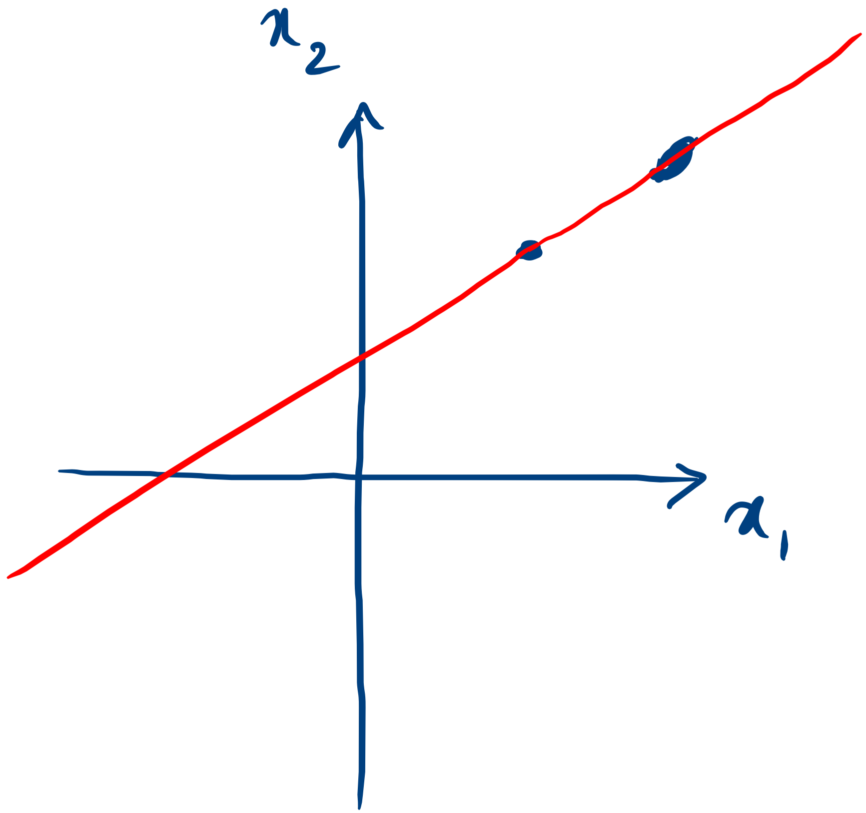
$X_{n \times d}$

$(X^T X)_{d \times d}$

$O(d^3) = O(10^{12})$

$2ax + b = f'(x_0)$

$2a = f''(x_0)$

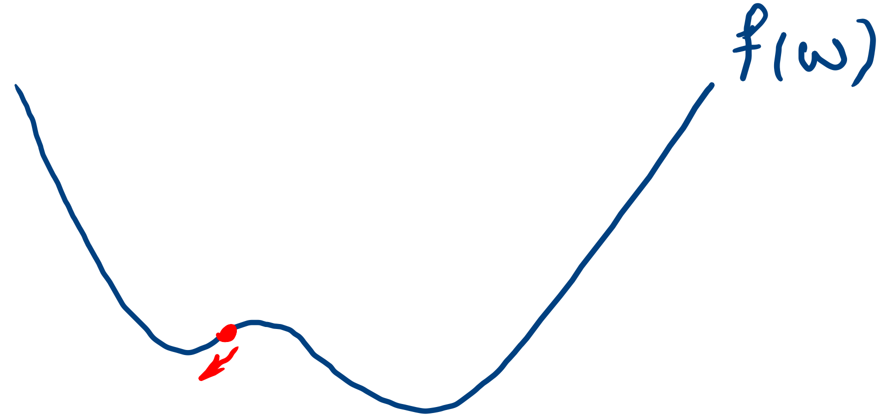


$d = 2$

$n = 1$

$\max_{w} \log L(w)$

$\min_{w} Loss(w)$



Batch

→ stochastic Gradient Descent (SGD)

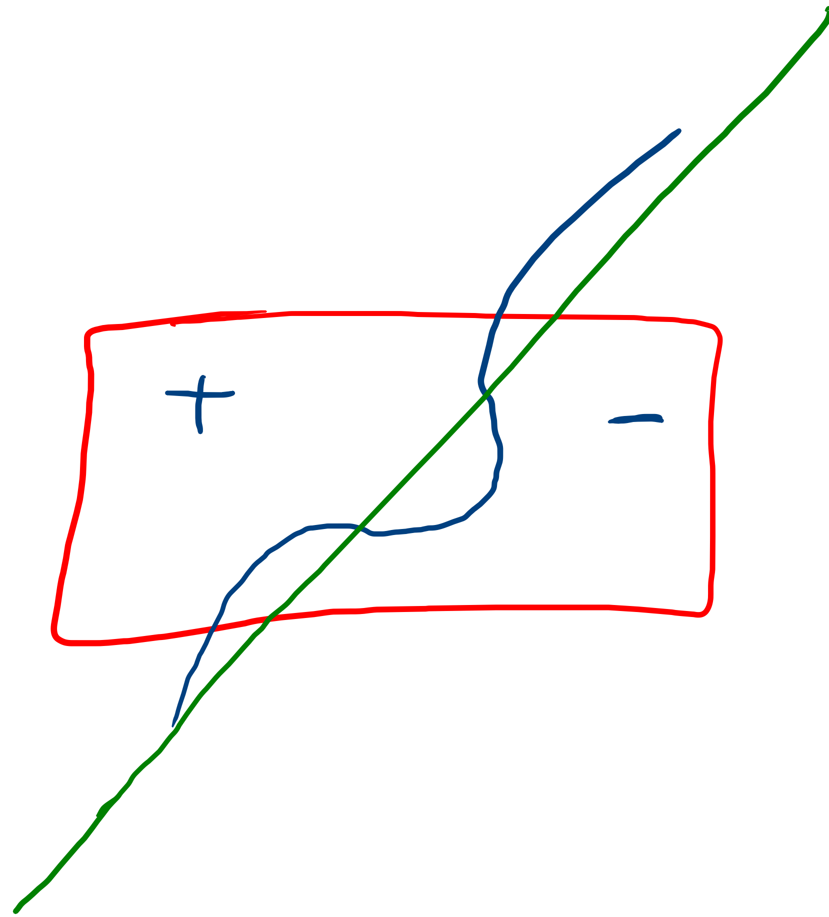
Mini-batch

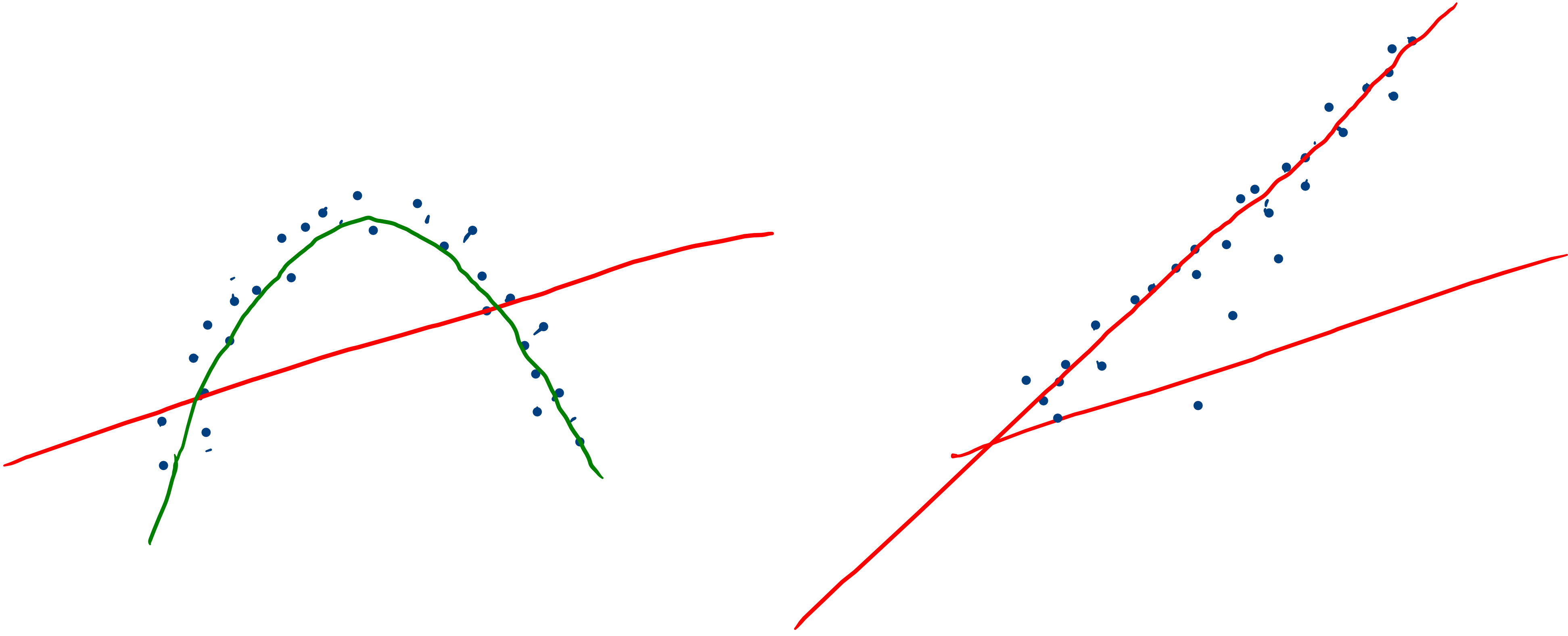
ناشی از مدل

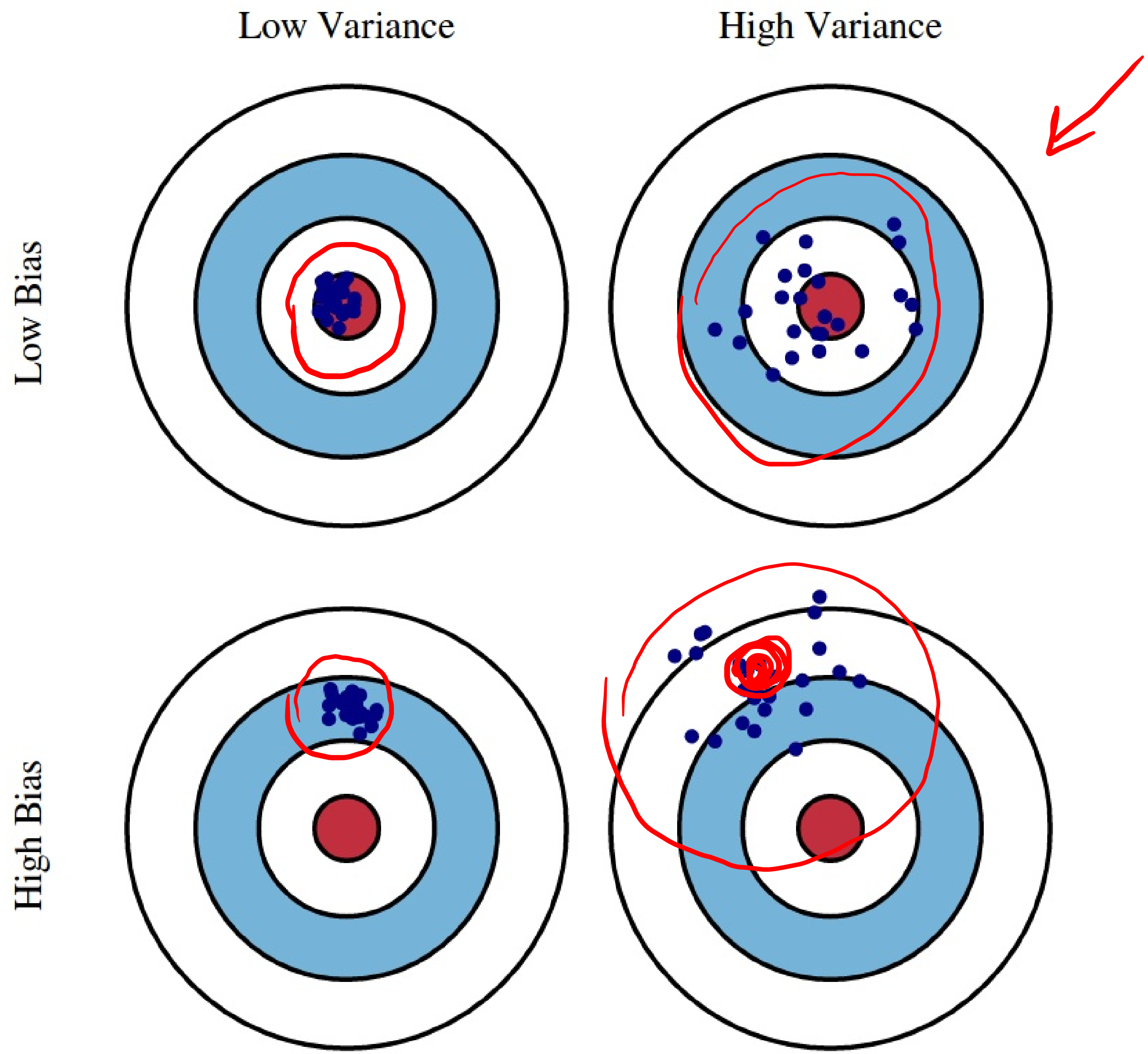
Error = Bias + Variance

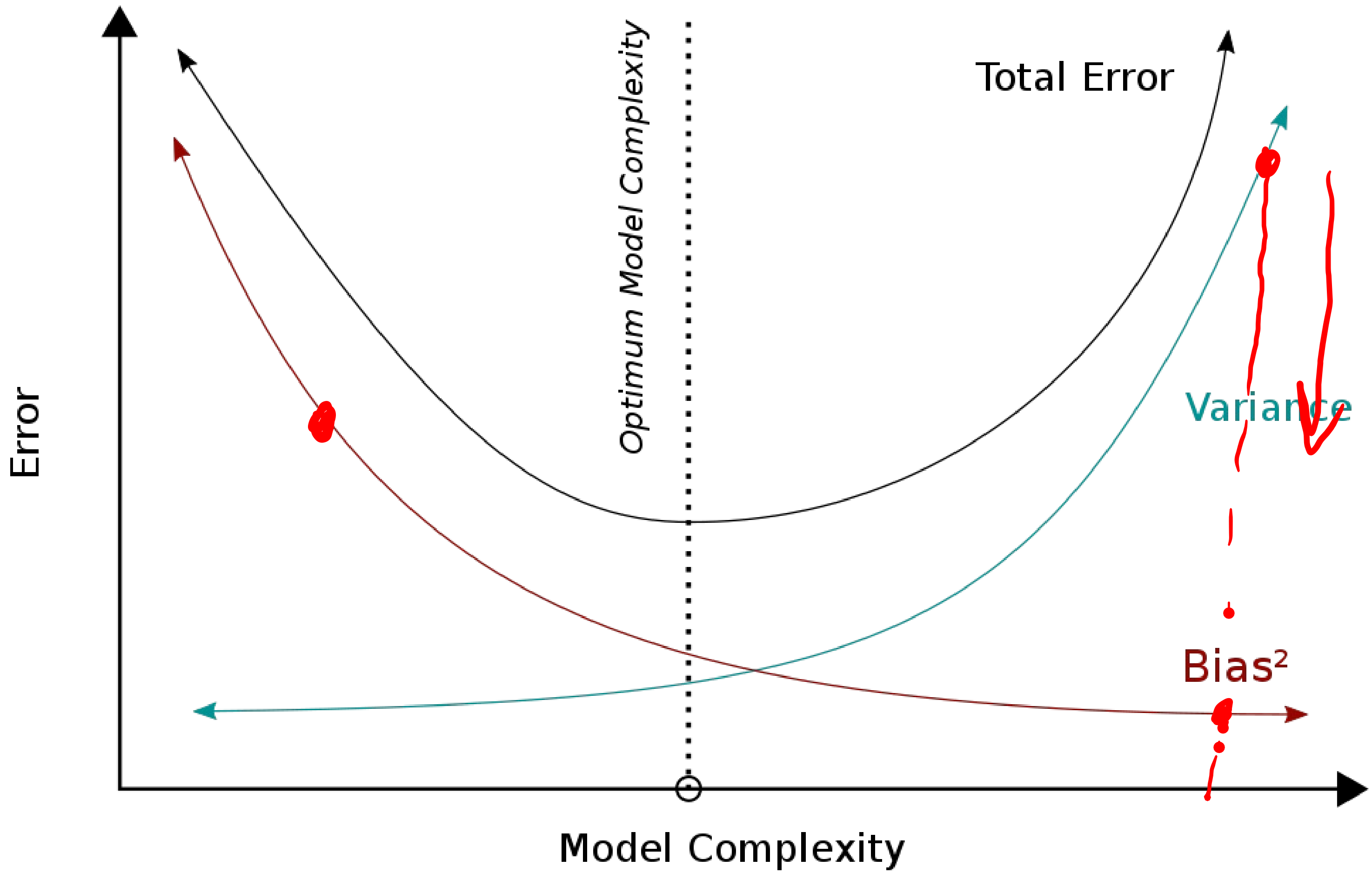
Bias_ Variance trade off

$n \uparrow \implies$ variance $\downarrow$

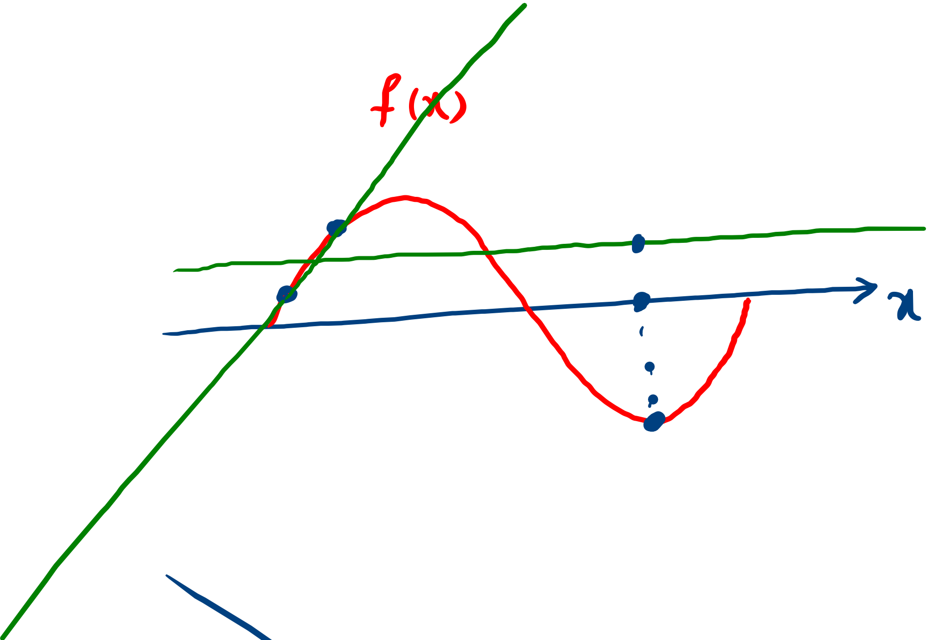





Low Variance
High Variance
Low Bias
High Bias

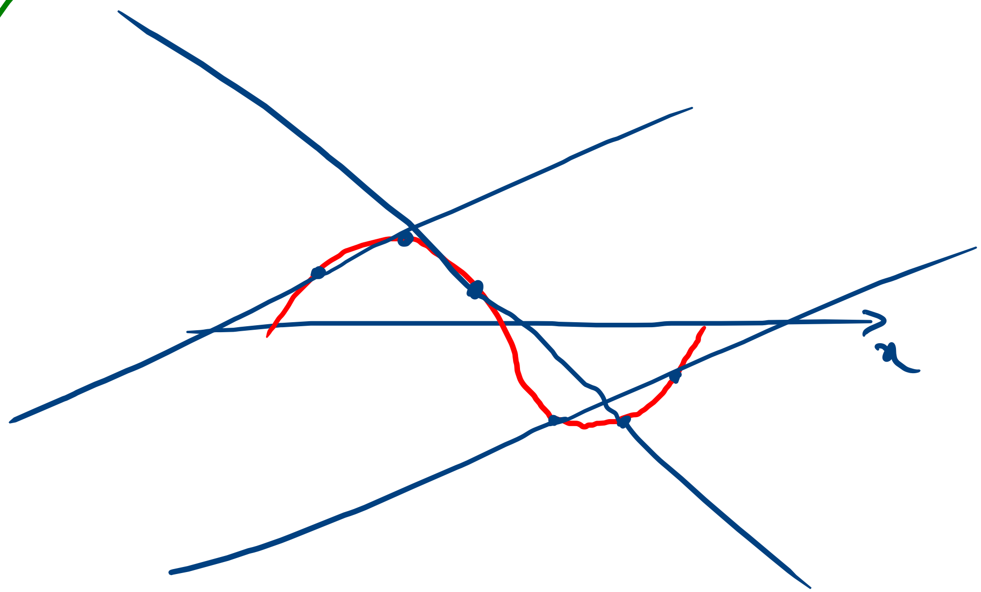


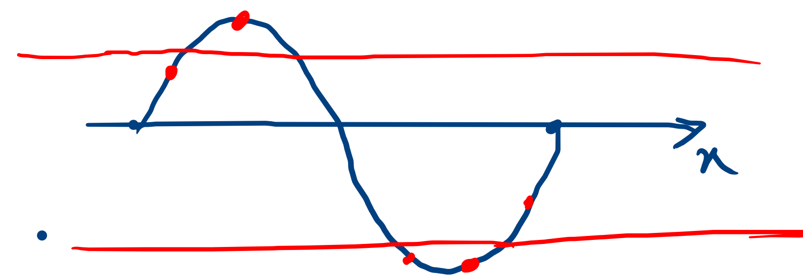
Error
Optimum Model Complexity
Total Error
Variance
Bias²
Model Complexity



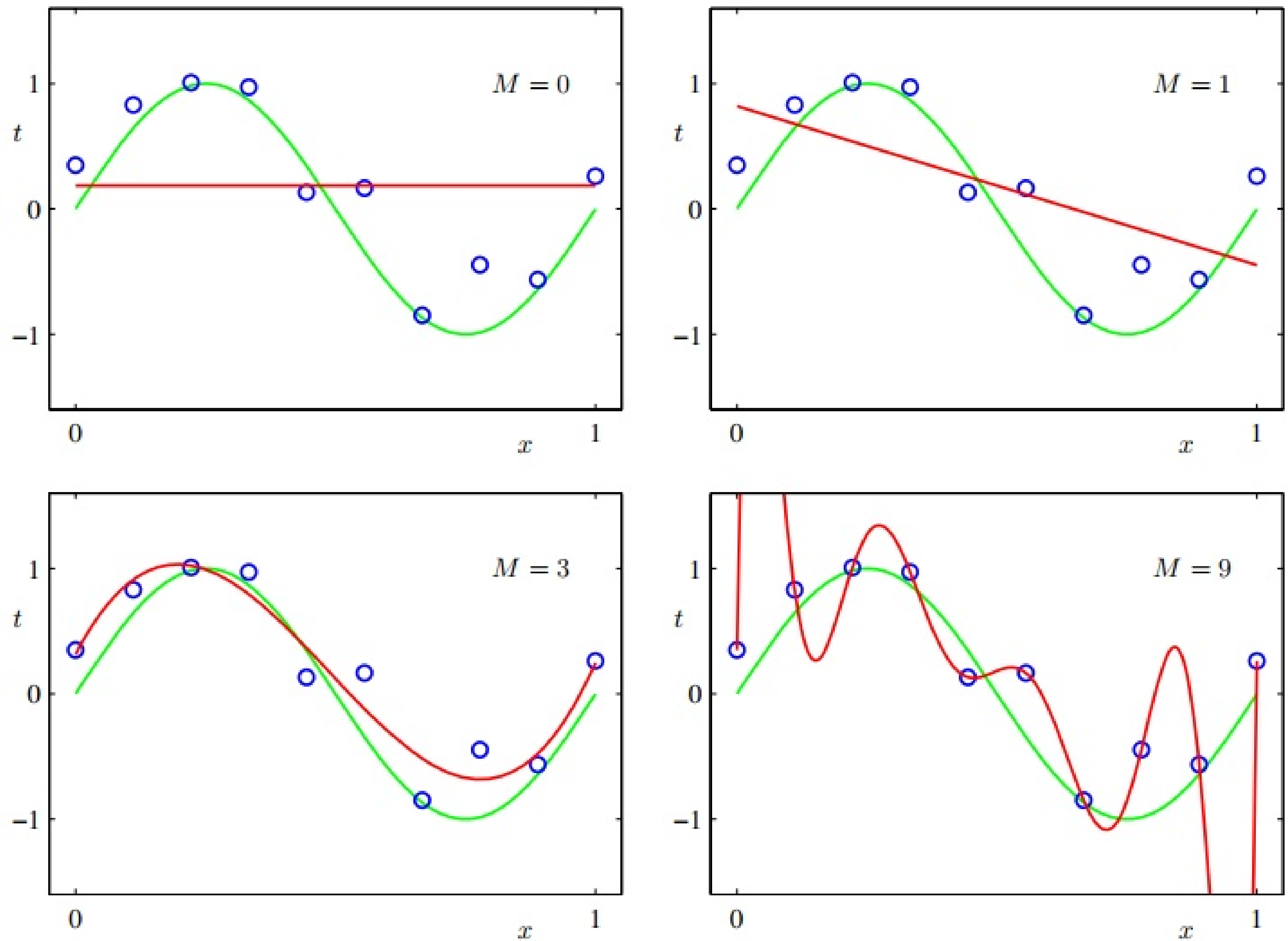
$g(x) = b$

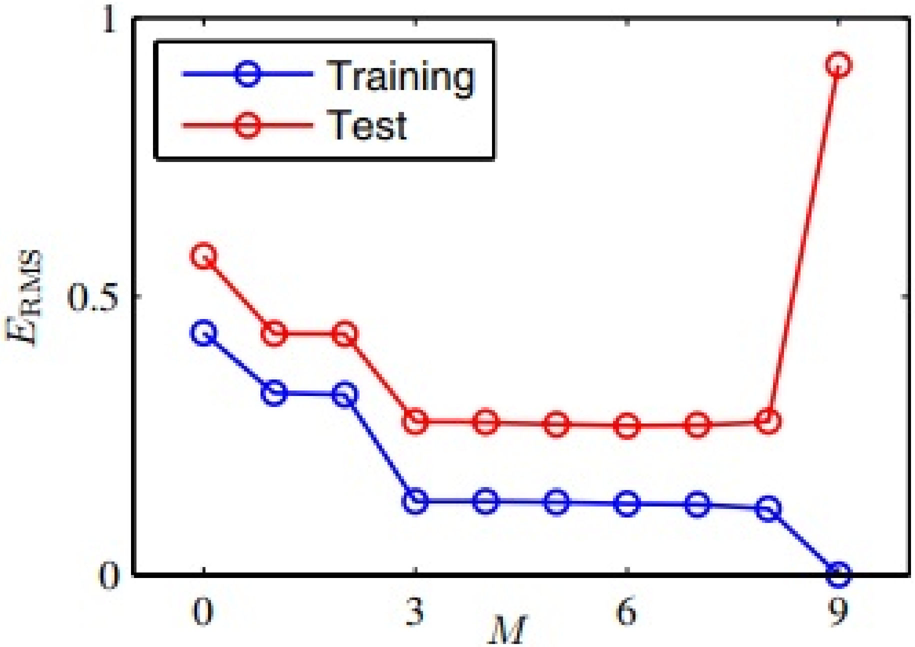
$g(x) = ax + b$





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





	$M=0$	$M=1$	$M=6$	$M=9$
$w_0^\star$	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

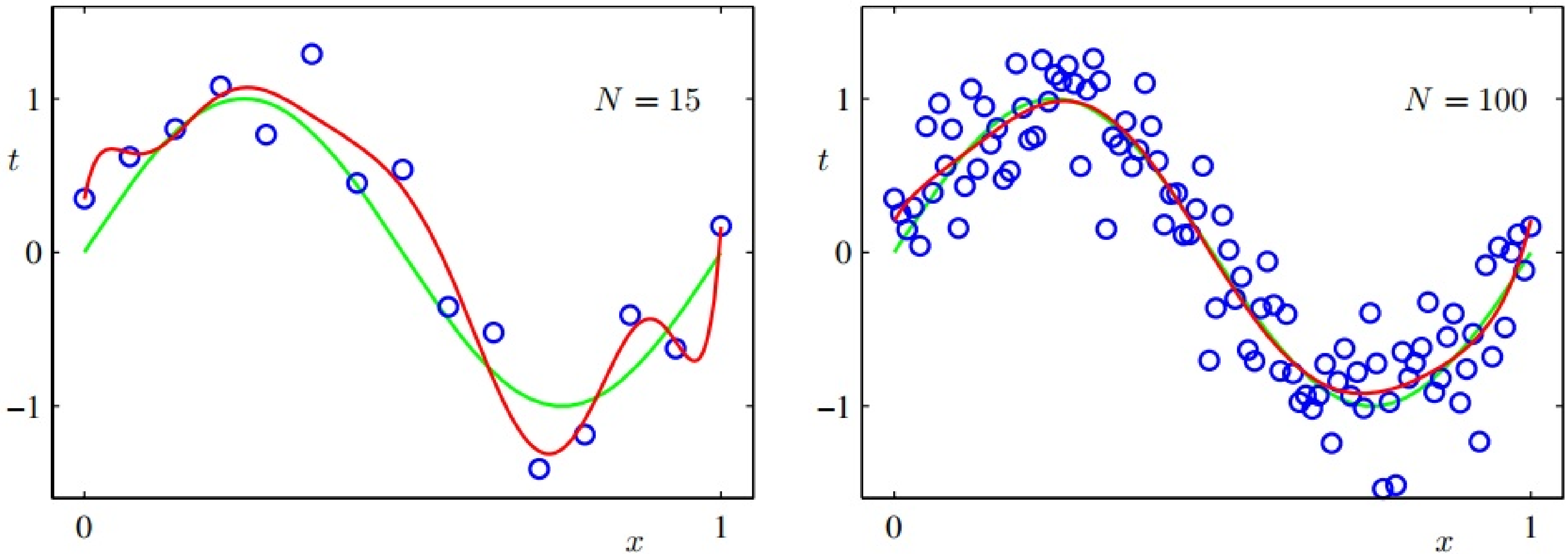


Figure 1.6 Plots of the solutions obtained by minimizing the sum-of-squares error function using the $M = 9$ polynomial for $N = 15$ data points (left plot) and $N = 100$ data points (right plot). We see that increasing the size of the data set reduces the over-fitting problem.

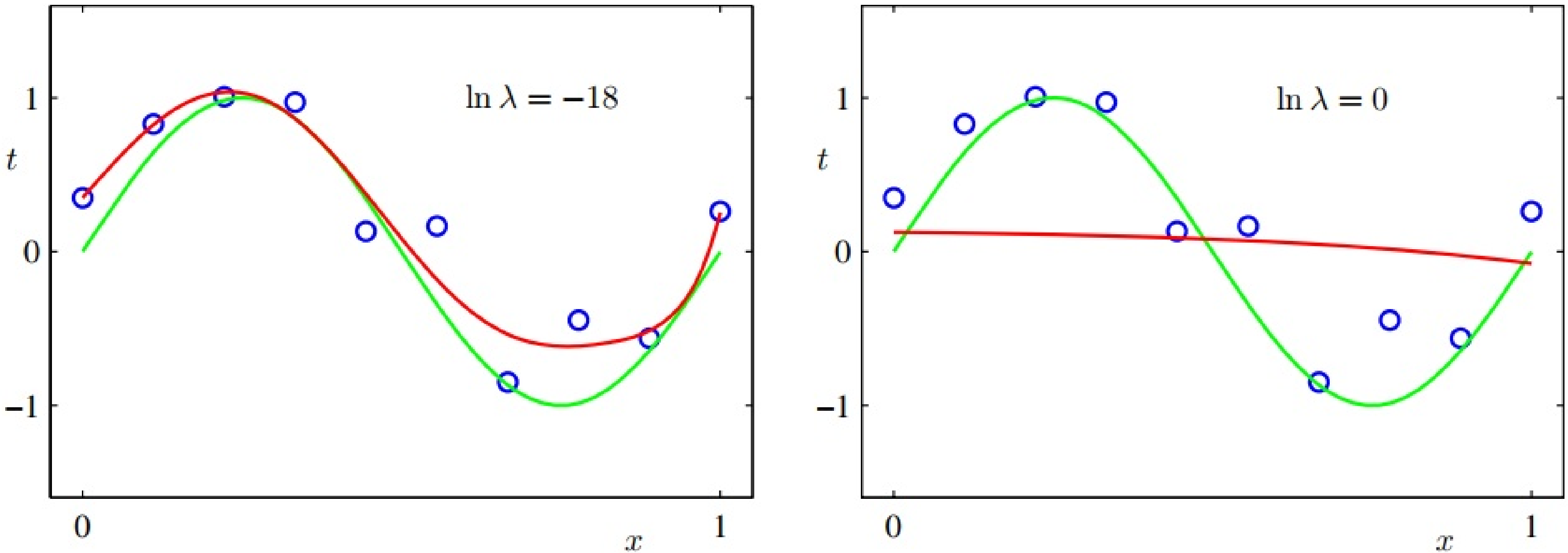


Figure 1.7 Plots of $M = 9$ polynomials fitted to the data set shown in Figure 1.2 using the regularized error function (1.4) for two values of the regularization parameter λ corresponding to $\ln \lambda = -18$ and $\ln \lambda = 0$. The case of no regularizer, i.e., $\lambda = 0$, corresponding to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

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