CSE 575
Statistical Machine Learning

Lecture 2

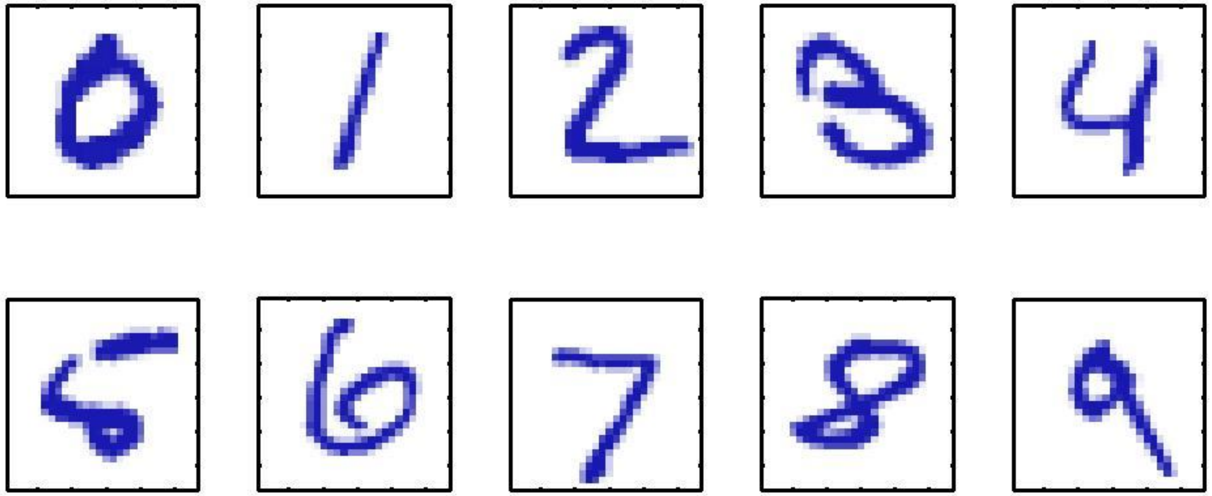
YooJung Choi

Fall 2022

Announcements

- Additional references for prerequisites
- Slack workspace—through Canvas

Example: Hand-written digits



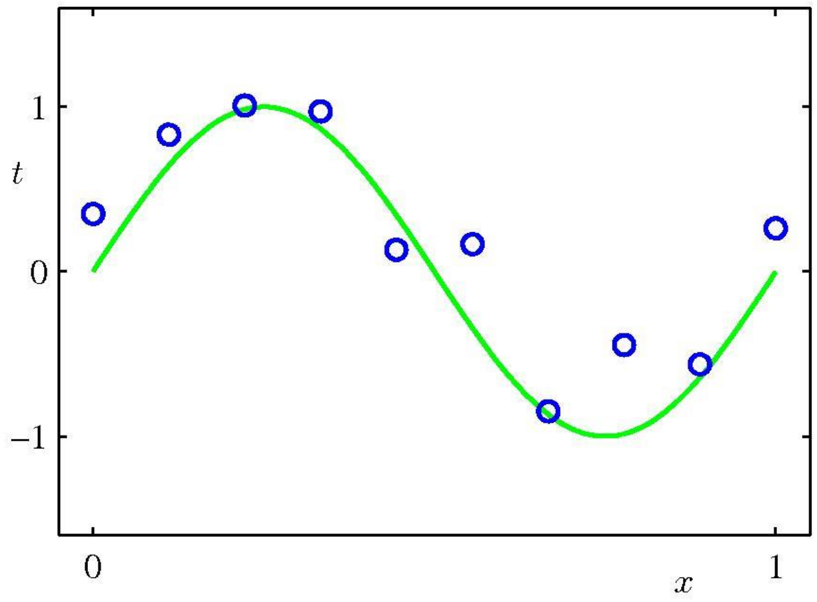
- 28x28 pixel images (784-dim vector)
- Each image corresponds to a digit

Basic terminology

- **Training set** $\{\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_N\}$ of N digits
- Each associated with a **target vector** $\mathbf{t}$ (identity of the digit)
- **Training (learning) phase**: determining a function $\mathbf{y}(\mathbf{x})$ that generates an output vector for a digit image $\mathbf{x}$
- An instance of **supervised learning**
- "How well can the trained model categorize new images (i.e. **test set**)?"

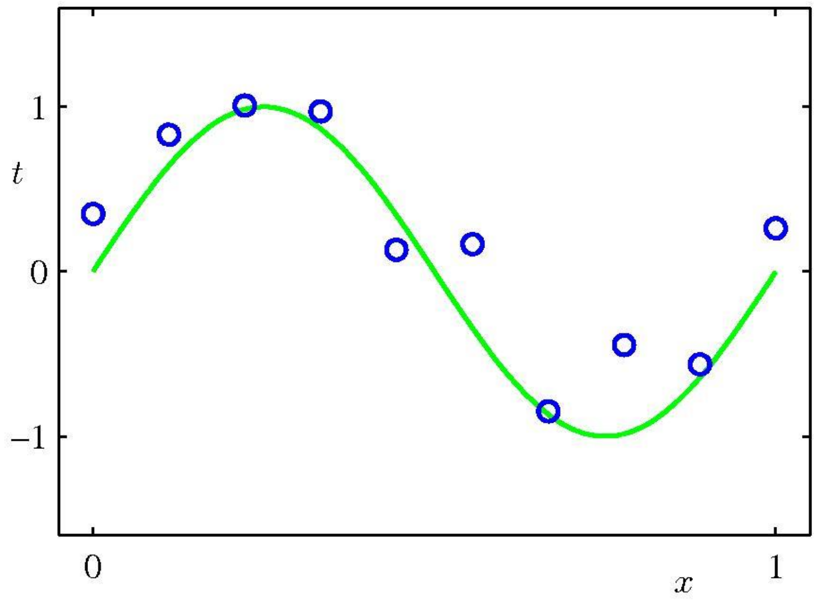
  Equivalently… "how well does it **generalize**?"

Example: Polynomial curve fitting



- Training set of 10 points
- Can we learn the green function that maps x to t?
- Another example of supervised learning (regression)

Example: Polynomial curve fitting



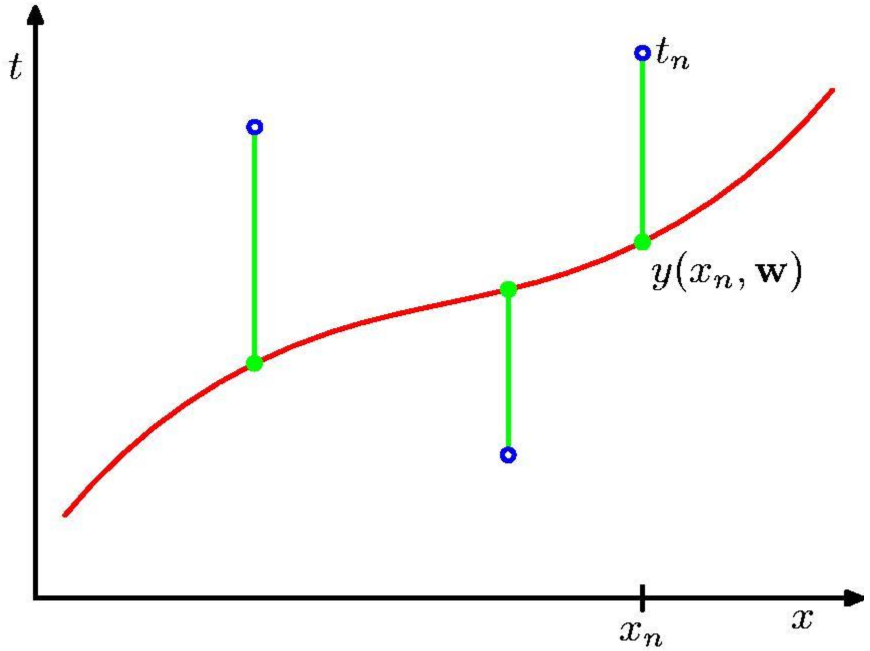
- Use a polynomial to approximate:

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

Example: Polynomial curve fitting

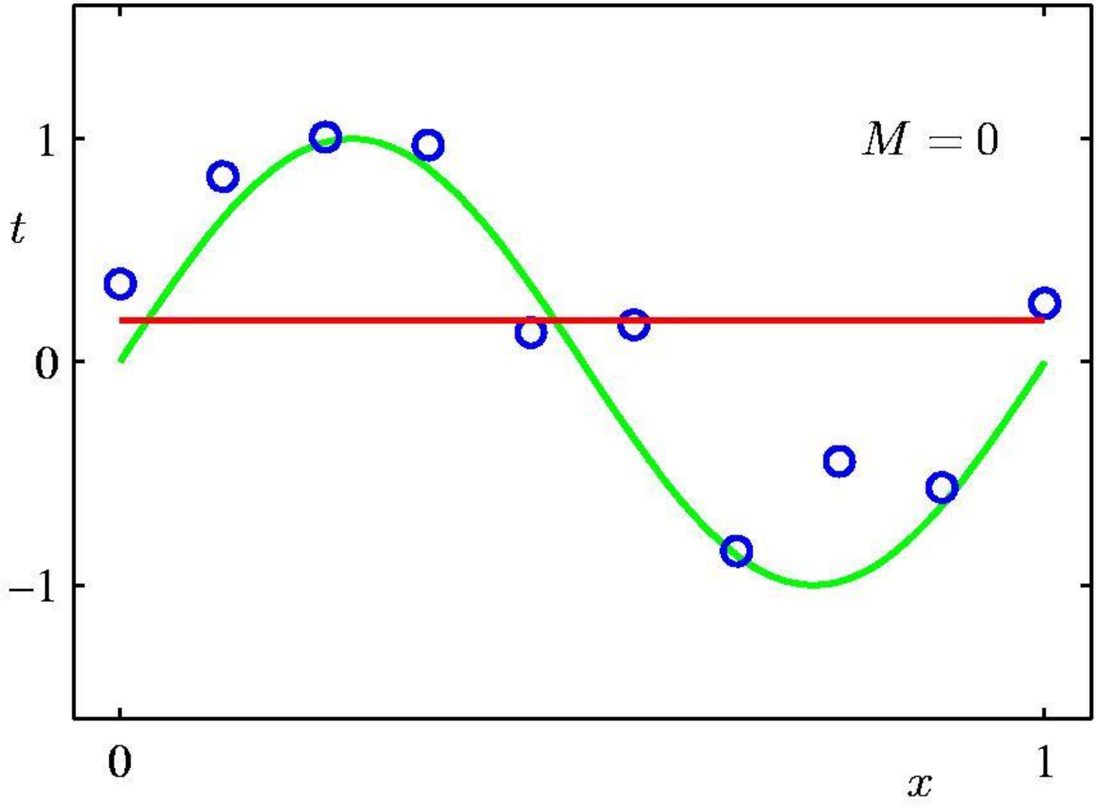
- How to determine the weights **w**?
- Sum-of-squares error function:

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

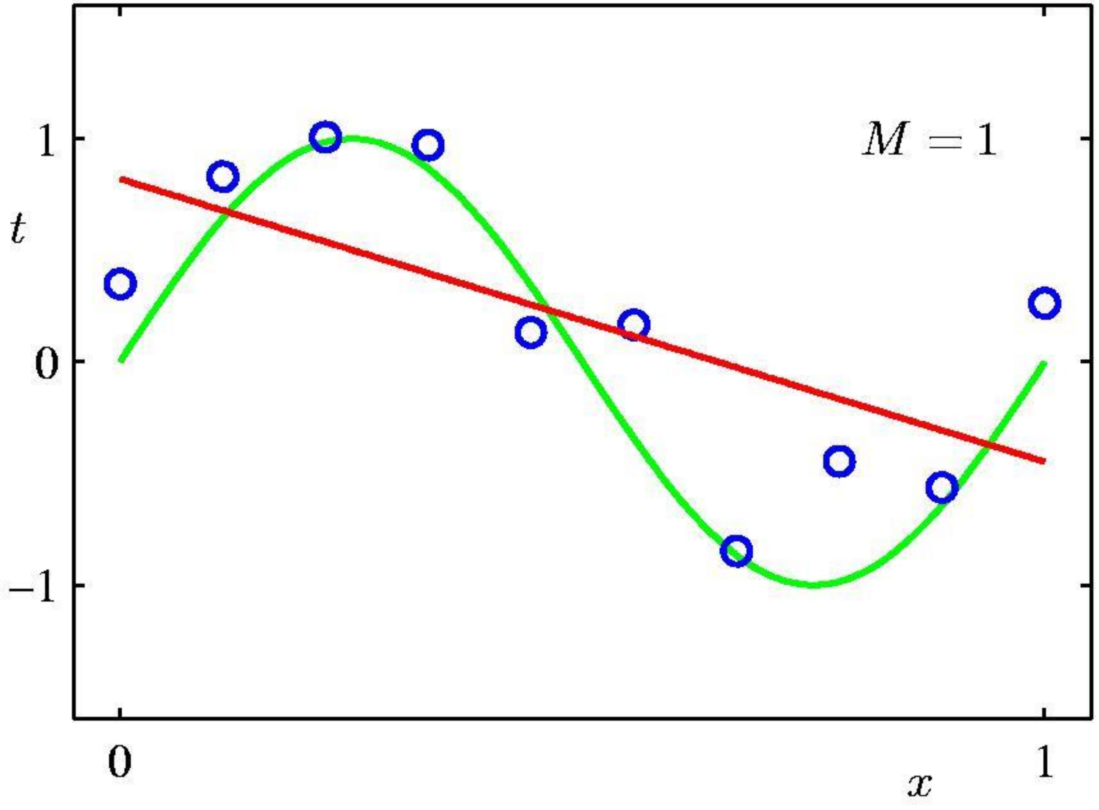


- Minimize E(**w**) to determine the optimal parameters
- How to minimize E(**w**)? “E(**w**) is convex quadratic”

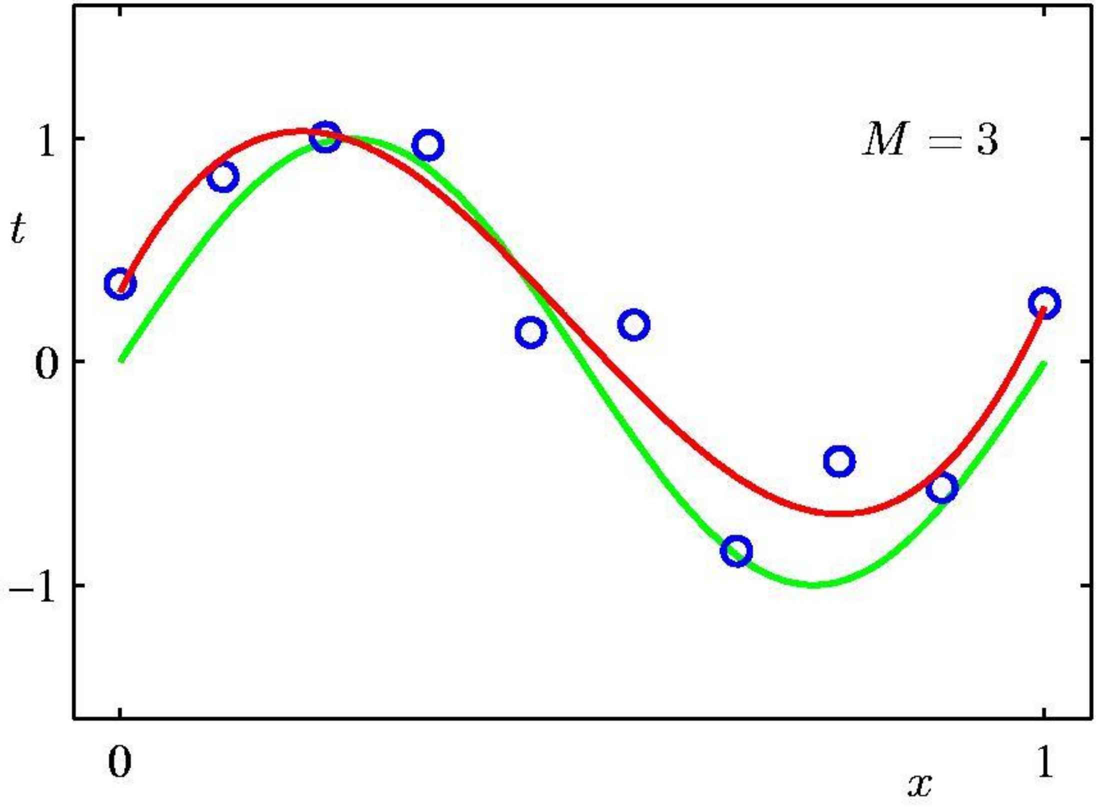
0^{th}-order polynomial



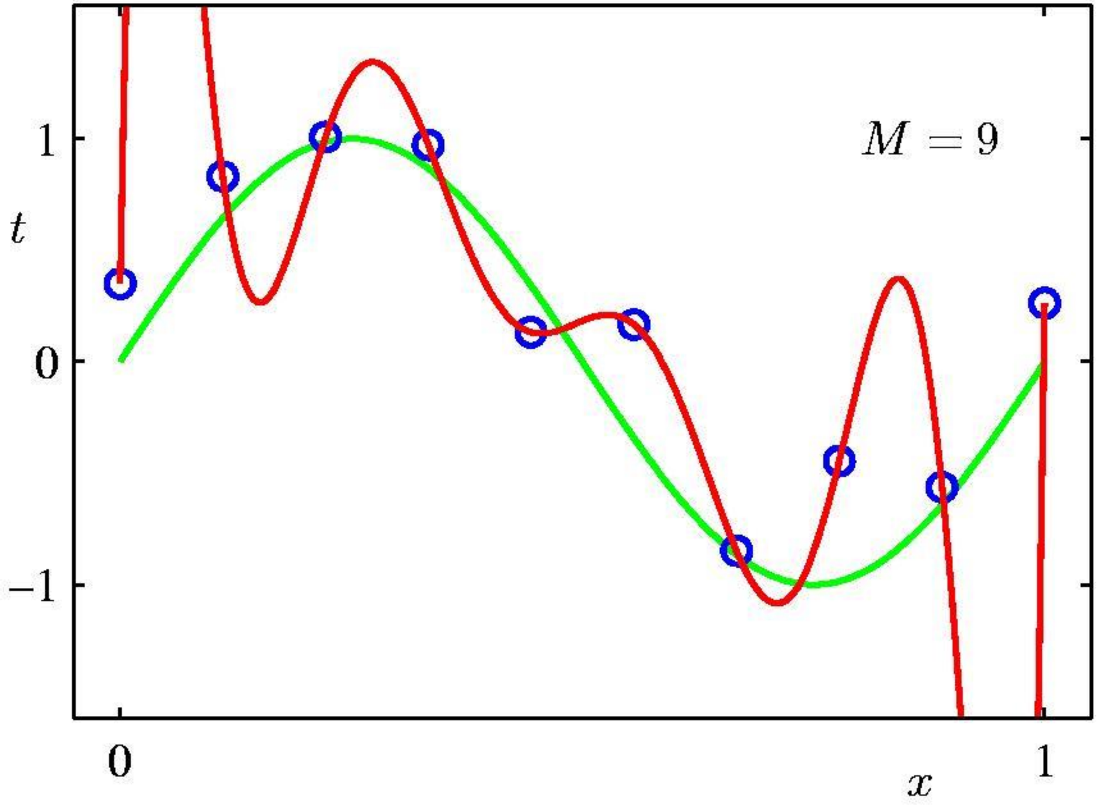
1^{st}-order polynomial



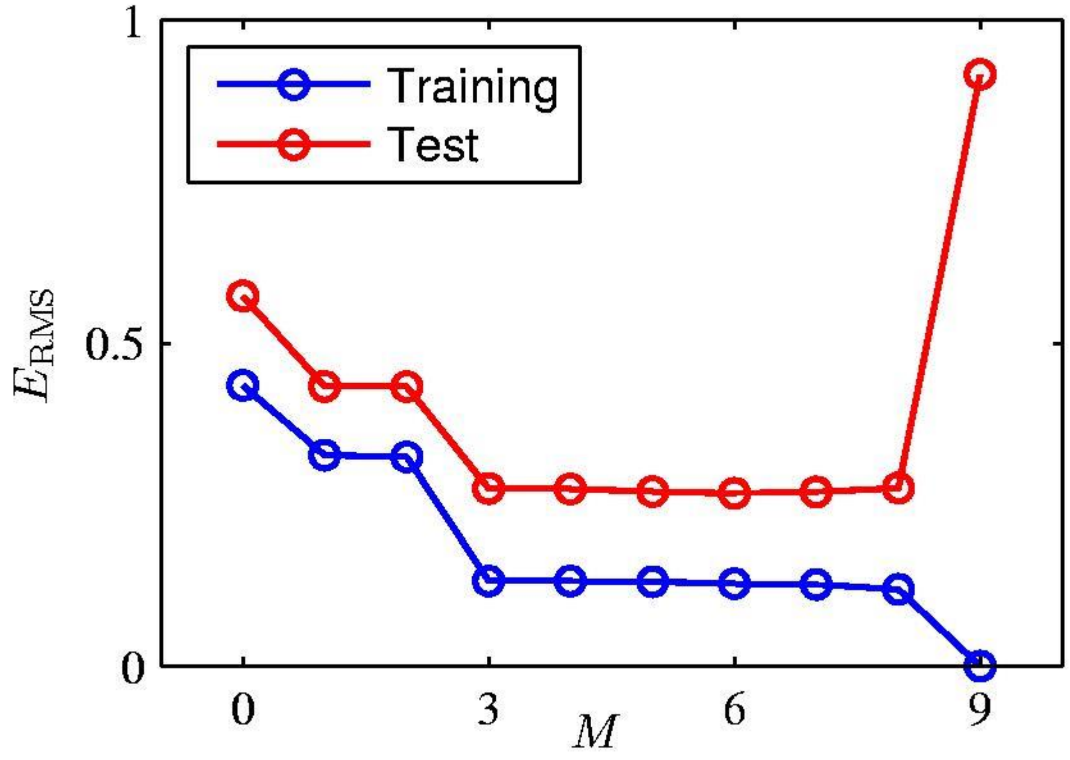
3rd-order polynomial



9^{th}-order polynomial



Overfitting

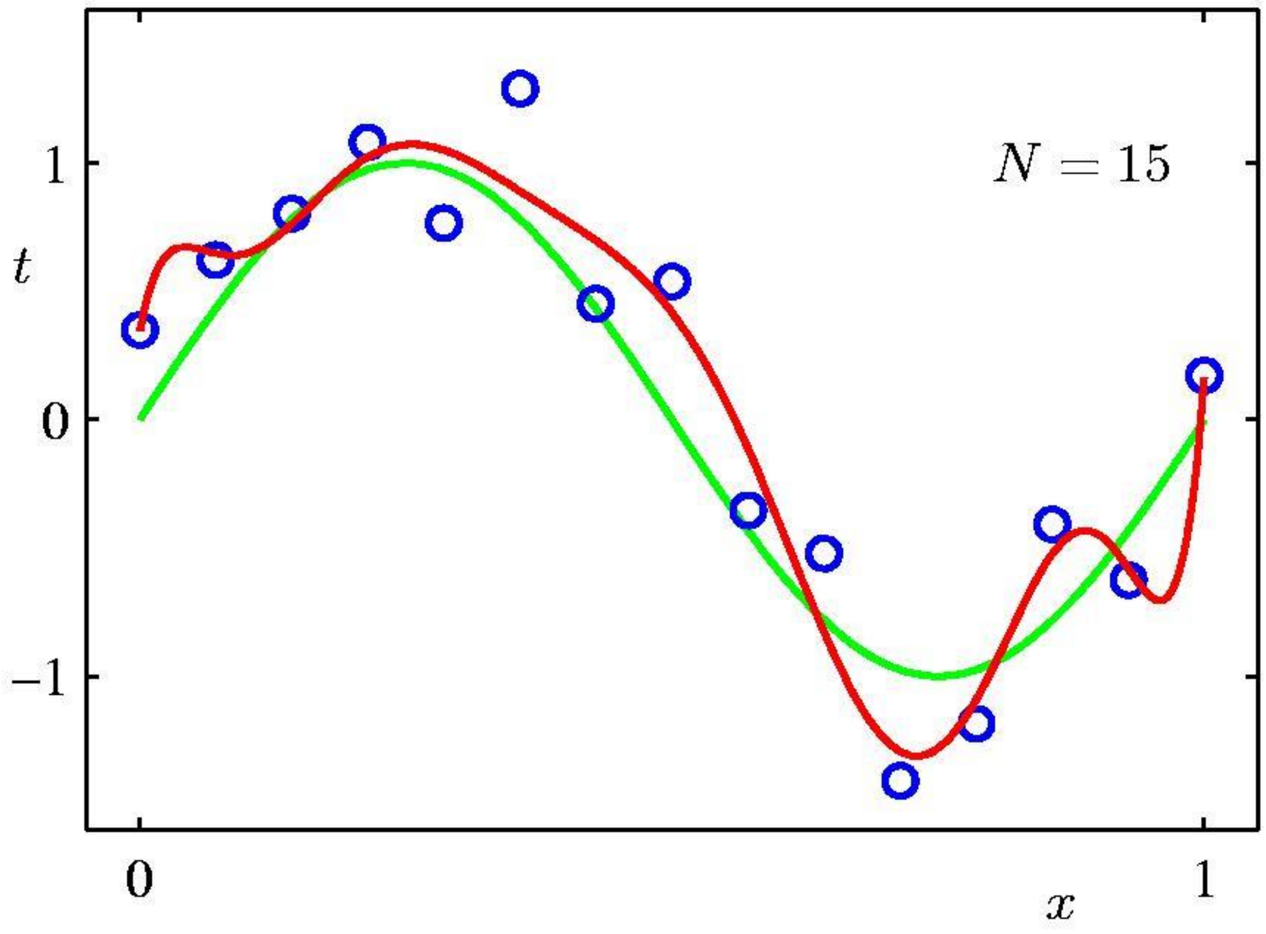


Root-mean-square error: $E_{RMS} = \sqrt{2E(\mathbf{w}^*)/N}$

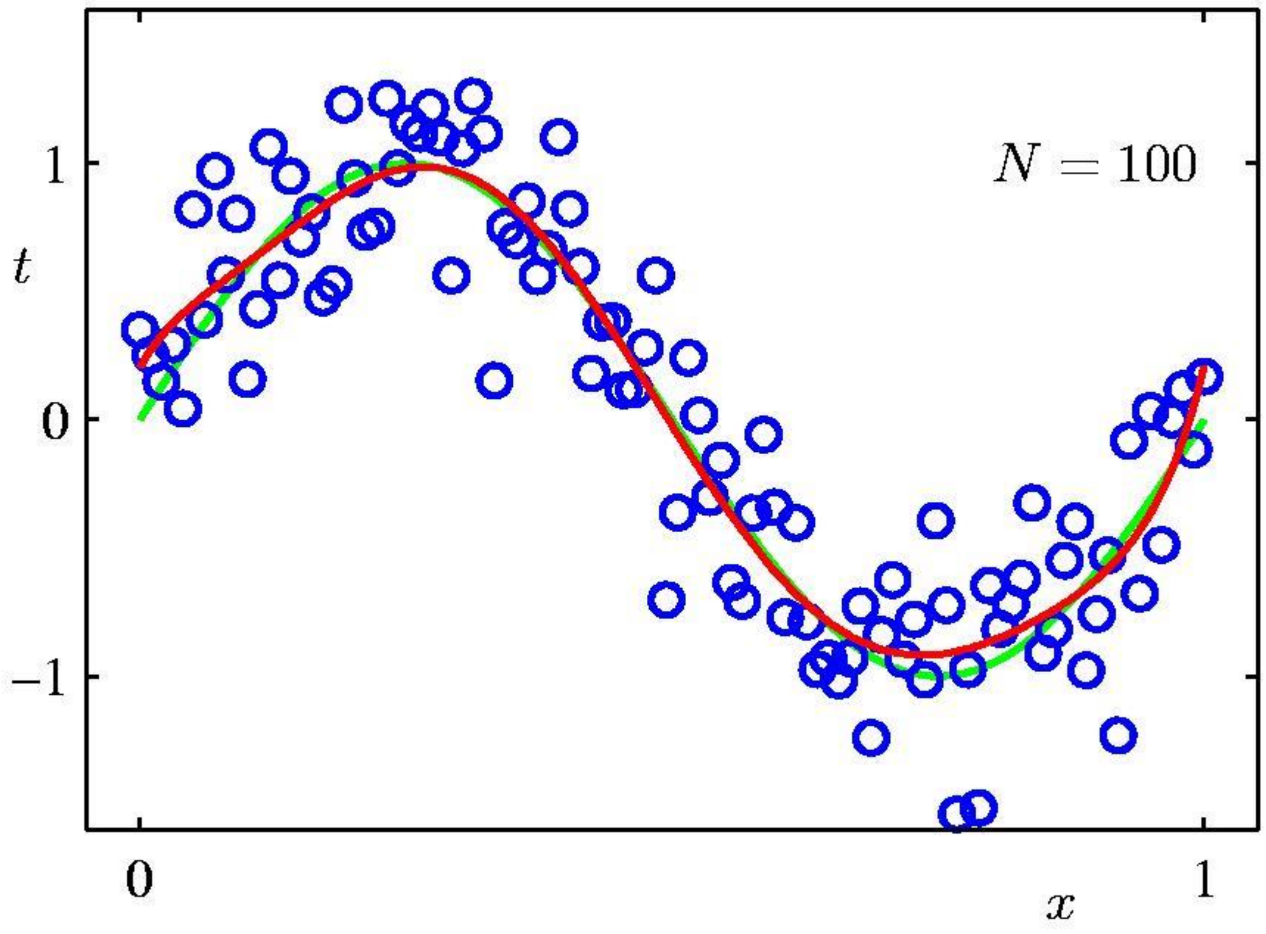
Learned polynomial coefficients

	$M=0$	$M=1$	$M=3$	$M=9$
$w_0^\star$	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

Increasing the size of datasets



Increasing the size of datasets



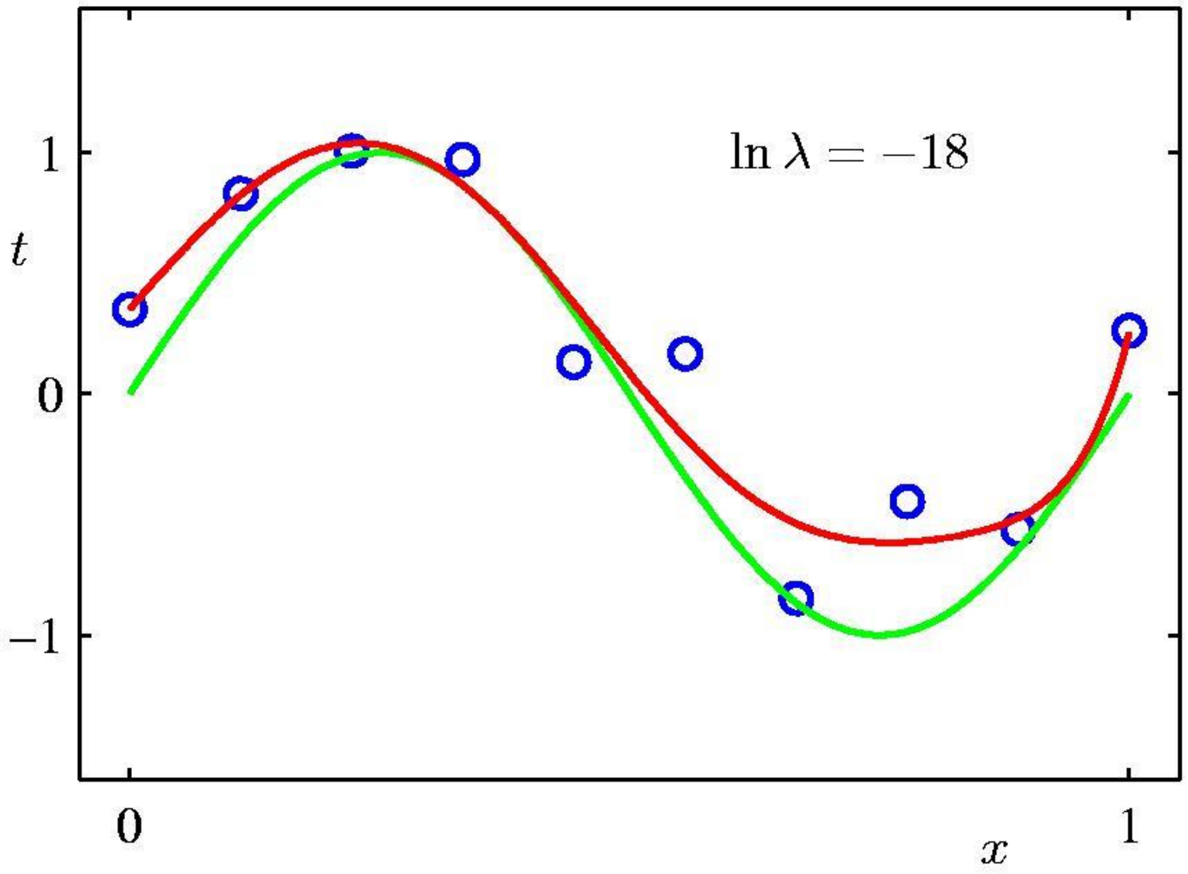
Regularization

- Penalize large coefficient values

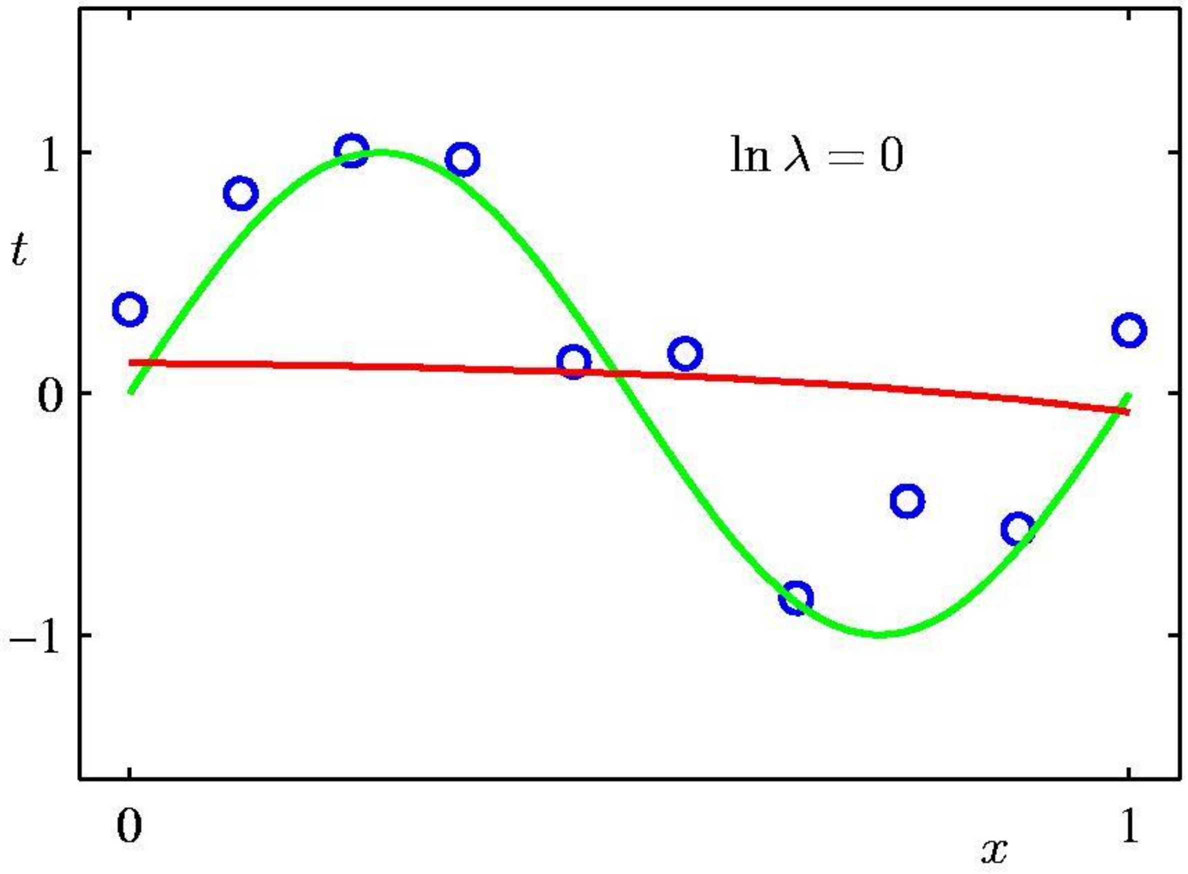
$$\widetilde{E}(\mathbf{w}) = \frac{1}{2}\sum_{n=1}^{N}\{y(x_n, \mathbf{w}) - t_n\}^2 + \frac{\lambda}{2}\|\mathbf{w}\|^2$$

- What is the best value of λ?

Regularization



Regularization



Polynomial coefficients

	$\ln\lambda = -\infty$	$\ln\lambda = -18$	$\ln\lambda = 0$
$w_0^\star$	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

RMSE vs. λ