



---

# **MACHINE LEARNING**

## **CHAPTER 1: INTRODUCTION**

---

# What is Machine Learning?

---

**Machine learning** is a subfield of computer science that evolved from the study of pattern recognition and computational learning theory in artificial intelligence.

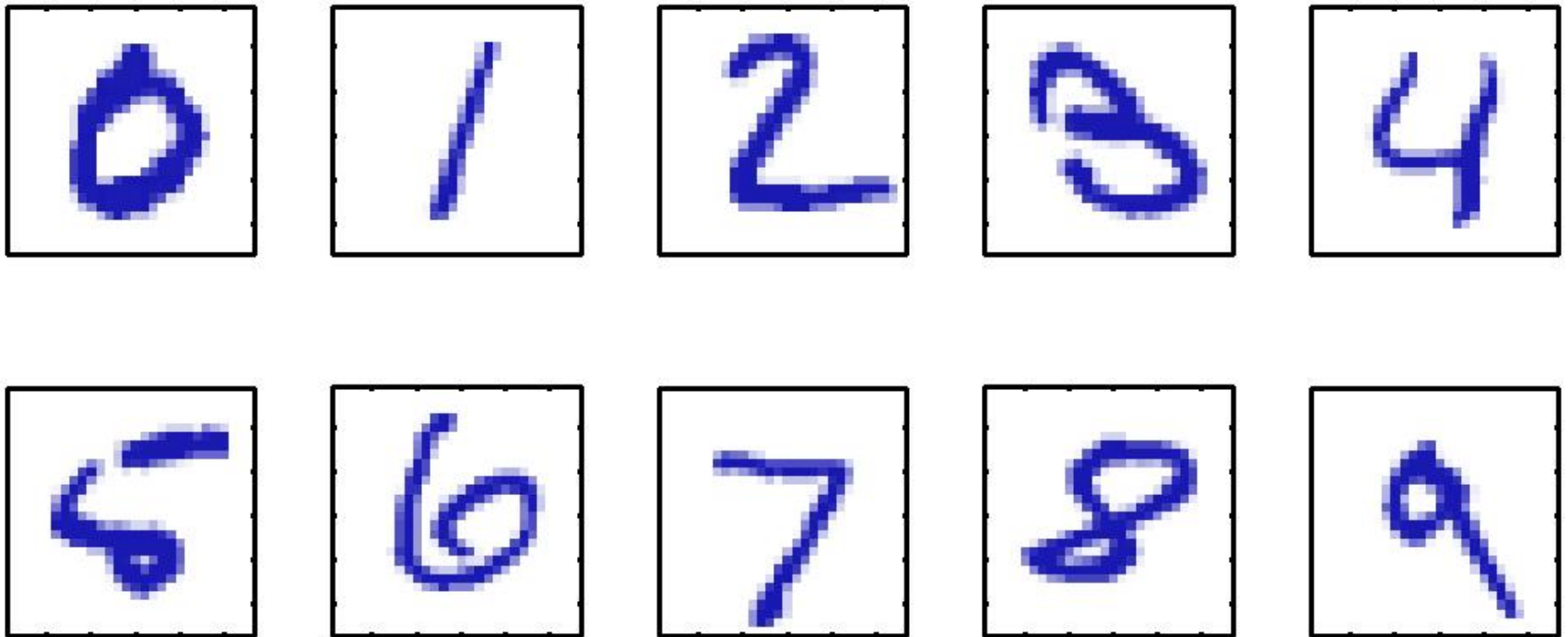
The emphasis of machine learning is on *automatic* methods. In other words, the goal is to devise learning algorithms that do the learning automatically without human intervention or assistance.

---

# Example

---

## Handwritten Digit Recognition



# Example

---

**Training set:** a set of  $N$  digits  $\{\mathbf{x}_1, \dots, \mathbf{x}_N\}$ , which is used to tune the parameters of an adaptive model.

**Target vector  $\mathbf{t}$ :** represents the identity of the corresponding digit.

---

# Example

---

**Training phase or learning phase:** the process of learning a prediction function  $y(\mathbf{x})$ .

**Test set:** new digital images whose identity can be determined by  $y(\mathbf{x})$ .

**Feature extraction:** the images in the digitals are typically translated and scaled so that each digit contained within a box of a fixed size.

---

# Example

---

**Supervised learning:** the training data comprises examples of the input vectors along with their corresponding target are known as supervised learning.

---

# Example

---

**Classification:** the aim is to assign each input vector to one of a finite number of discrete categories.

**Regression:** the desired output for each input consist of one or more continuous variables.

---

# Example

---

**Unsupervised learning:** the training data consists of a set of input vectors without any corresponding target values.

**Clustering:** discover groups of similar examples within the data (without any target values).

---



# Example

---

## Learning to detect objects in images

(Prof. H. Schneiderman)



Example training images  
for each orientation



# Example

---

Learning to classify text documents



→ Company home page

vs

Personal home page

vs

University home page

vs

...

# Example

## Email spam

data

prediction

☆ **Osman Khan** to Carlos [show details](#) Jan 7 (6 days ago) [Reply](#)

sounds good  
+ok

Carlos Guestrin wrote:  
Let's try to chat on Friday a little to coordinate and more on Sunday in person?  
Carlos

**Welcome to New Media Installation: Art that Learns**

☆ **Carlos Guestrin** to 10615-announce, Osman, Miche [show details](#) 3:15 PM (8 hours ago) [Reply](#)

Hi everyone,

Welcome to New Media Installation:Art that Learns

The class will start tomorrow.  
\*\*\*Make sure you attend the first class, even if you are on the Wait List.\*\*\*  
The classes are held in Doherty Hall C316, and will be Tue, Thu 01:30-4:20 PM.

By now, you should be subscribed to our course mailing list: [10615-announce@cs.cmu.edu](mailto:10615-announce@cs.cmu.edu).  
You can contact the instructors by emailing: [10615-instructors@cs.cmu.edu](mailto:10615-instructors@cs.cmu.edu)

**Natural \_LoseWeight SuperFood Endorsed by Oprah Winfrey, Free Trial 1 bottle, pay only \$5.95 for shipping mfw rik** [Spam](#) [X](#)

☆ **Jaquelyn Halley** to nherlein, bcc: thehorney, bcc: ang [show details](#) 9:52 PM (1 hour ago) [Reply](#)

=== Natural WeightLOSS Solution ===

Vital Acai is a natural WeightLOSS product that Enables people to lose wieght and cleansing their bodies faster than most other products on the market.

Here are some of the benefits of Vital Acai that You might not be aware of. These benefits have helped people who have been using Vital Acai daily to Achieve goals and reach new heights in there dieting that they never thought they could.

- \* Rapid WeightLOSS
- \* Increased metabolism - BurnFat & calories easily!
- \* Better Mood and Attitude
- \* More Self Confidence
- \* Cleanse and Detoxify Your Body
- \* Much More Energy
- \* BetterSexLife

# Machine Learning - Practice

Data:

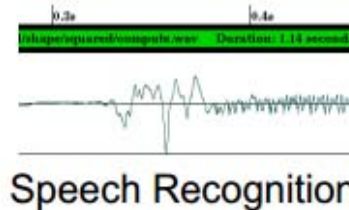
Actual	Model	Actual	Model	Actual	Model
Apr 23	Apr 23	Apr 23	Apr 23	Apr 23	Apr 23
Emergency: no	Emergency: no	Emergency: no	Emergency: no	Emergency: no	Emergency: no
Admission: no	Admission: no	Admission: no	Admission: no	Admission: no	Admission: no
Parous/Paraenacted: no	Parous/Paraenacted: no	Parous/Paraenacted: no	Parous/Paraenacted: no	Parous/Paraenacted: no	Parous/Paraenacted: no
Obstructed: ?	Obstructed: no	Obstructed: no	Obstructed: no	Obstructed: no	Obstructed: no
Delivery C-Section: ?	Delivery C-Section: no	Delivery C-Section: no	Delivery C-Section: no	Delivery C-Section: no	Delivery C-Section: no
Emergency C-Section: ?	Emergency C-Section: ?	Emergency C-Section: ?	Emergency C-Section: ?	Emergency C-Section: ?	Emergency C-Section: ?

One of 18 learned rules:

If No previous vaginal delivery, and  
Abnormal 2nd Trimester Ultrasound, and  
Malpresentation at admission  
Then Probability of Emergency C-Section is 0.6

Over training data: 26/41 = .63,  
Over test data: 12/20 = .60

Mining Databases



Control learning

Text analysis

Peter H. van Oppen, Chairman of the Board & Chief Executive Officer. Mr. van Oppen has served as Chairman of the Board and Chief Executive Officer of ADIQ since its acquisition by Interpoint in 1994 and a director of ADIQ since 1996. Until its acquisition by Crane Co. in October 1996, Mr. van Oppen served as Chairman of the Board of ADIQ and Chief Executive Officer of ADIQ. Prior to 1996, Mr. van Oppen worked as a consulting engineer at Price Waterhouse LLP and at Bain & Company in Boston and London. He has additional experience in medical electronics and venture capital. Mr. van Oppen also serves as a Director of Health Partners, Inc. and Spacelabs Medical, Inc. He holds a B.A. from Whitman College and an M.B.A. from Harvard Business School, where he was a Baker Scholar.



Object recognition

- Supervised learning
- Bayesian networks
- Hidden Markov models
- Unsupervised clustering
- Reinforcement learning
- ....

# Example: Polynomial Curve Fitting

---

$x$ : real-valued input variable

$t$ : real-valued target variable

$\mathbf{x} = (x_1, \dots, x_N)^T$ :  $N$  observations of  $x$

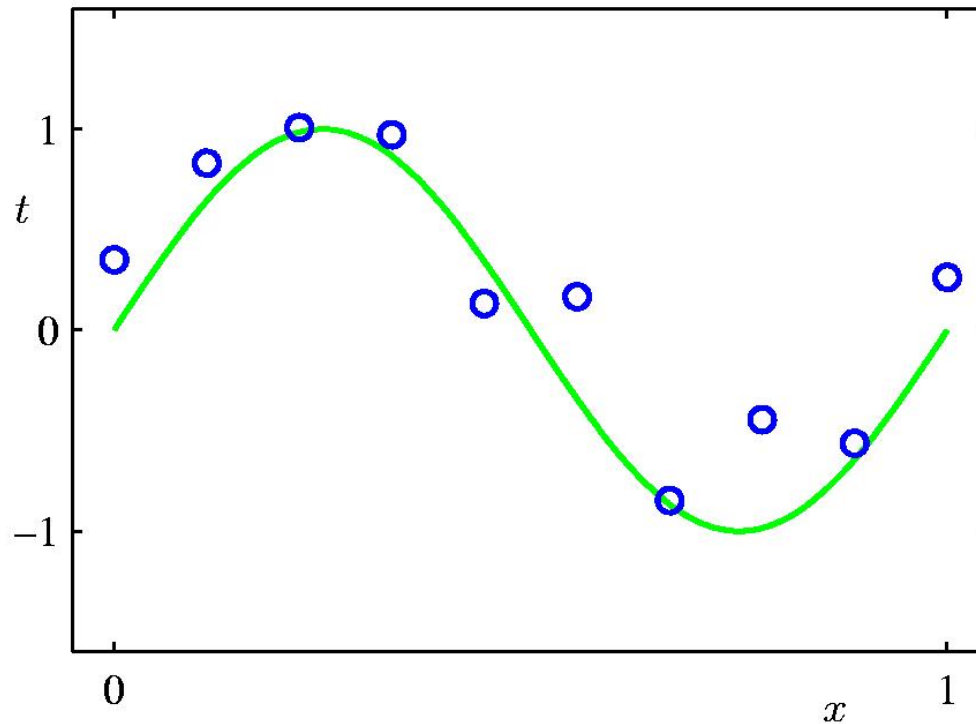
$\mathbf{t} = (t_1, \dots, t_N)^T$ : observations of target variable  $t$

$x_n$  is spaced uniformly in range  $[0, 1]$ , and the target data set  $\mathbf{t}$  was obtained by first computing the corresponding values of the function  $\sin(2\pi x)$  and then adding small level of random noise having a Gaussian distribution.

---

# Example: Polynomial Curve Fitting

---

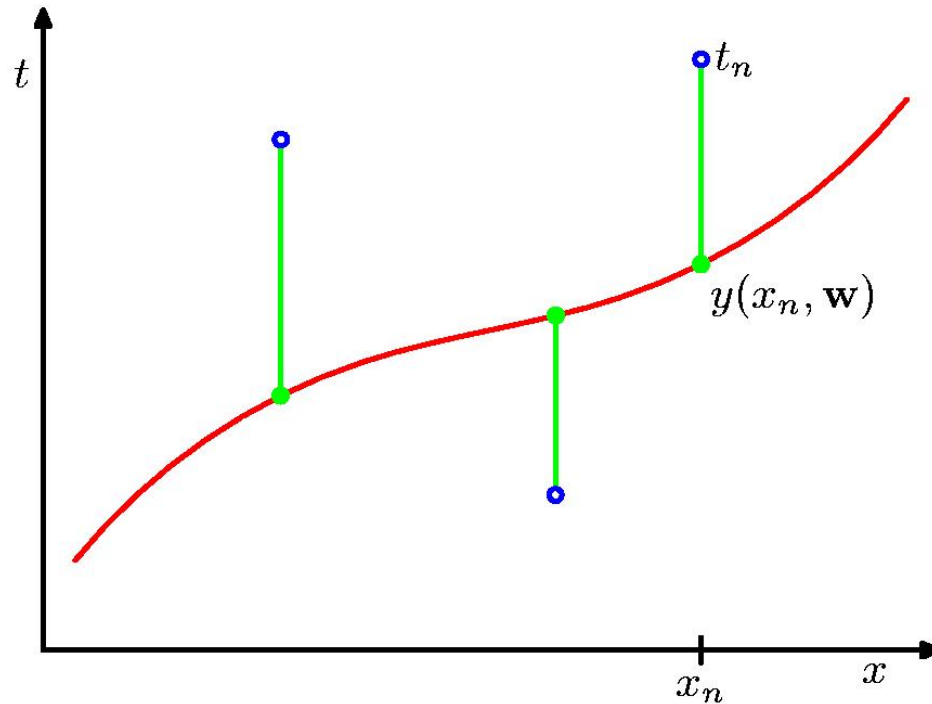


$$y(x, \mathbf{w}) = w_0 + w_1x + w_2x^2 + \dots + w_Mx^M = \sum_{j=0}^M w_jx^j$$

---

# Sum-of-Squares Error Function

---

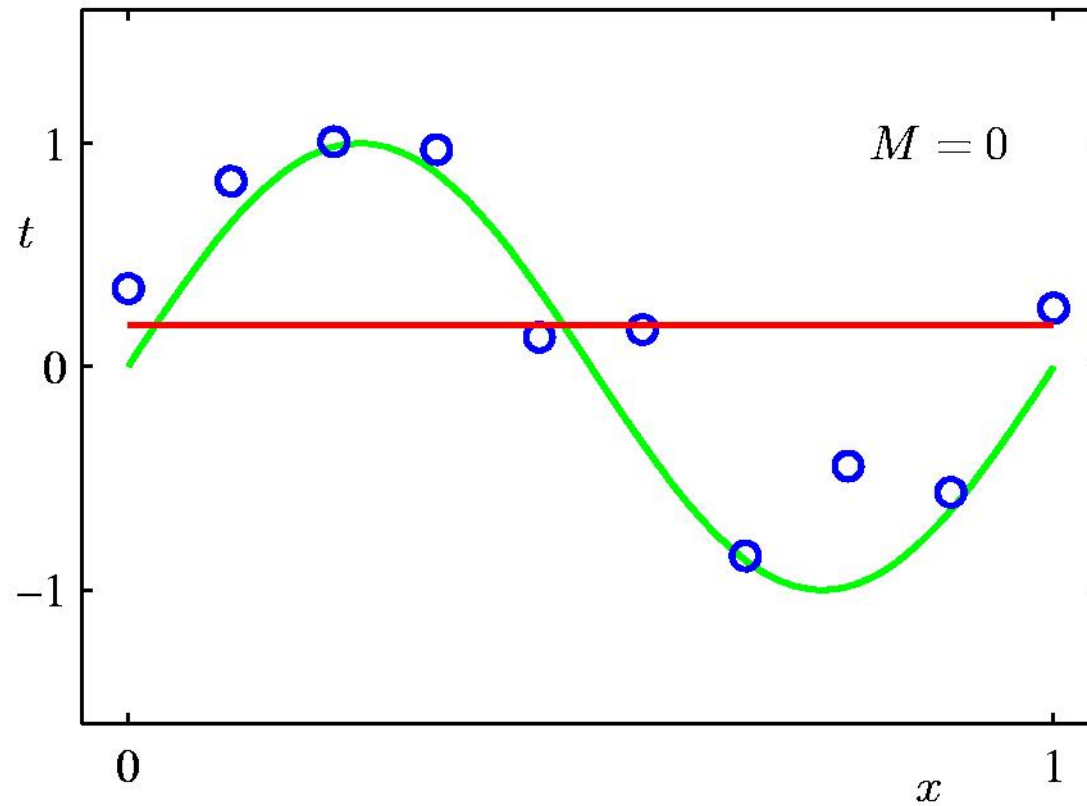


$$E(\mathbf{w}) = \frac{1}{2} \sum_{n=1}^N \{y(x_n, \mathbf{w}) - t_n\}^2 \quad \mathbf{w}^* = \operatorname{argmin} E(\mathbf{w})$$

---

# 0<sup>th</sup> Order Polynomial

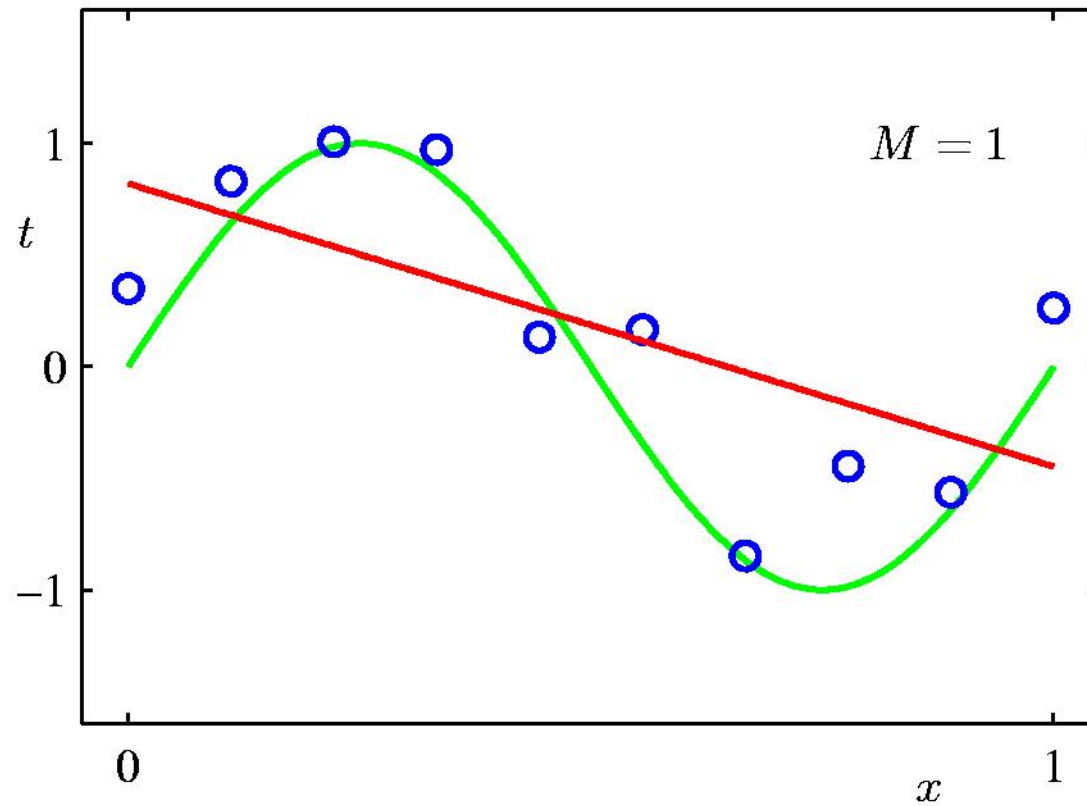
---





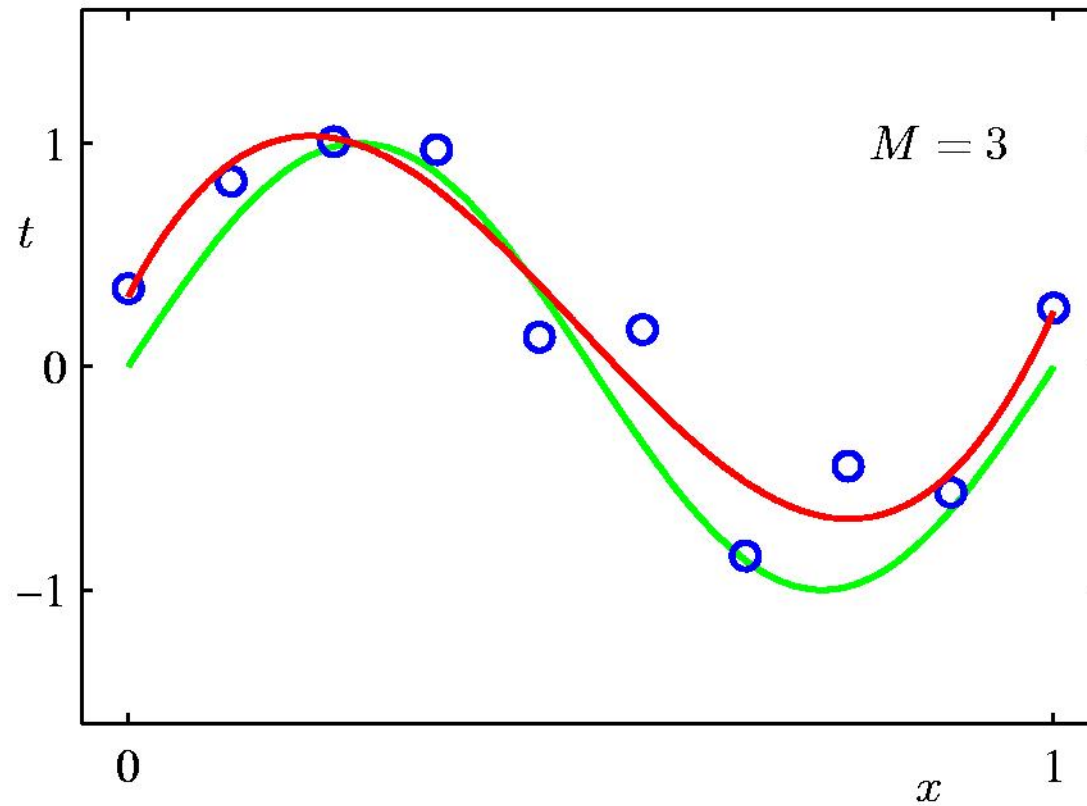
# 1<sup>st</sup> Order Polynomial

---



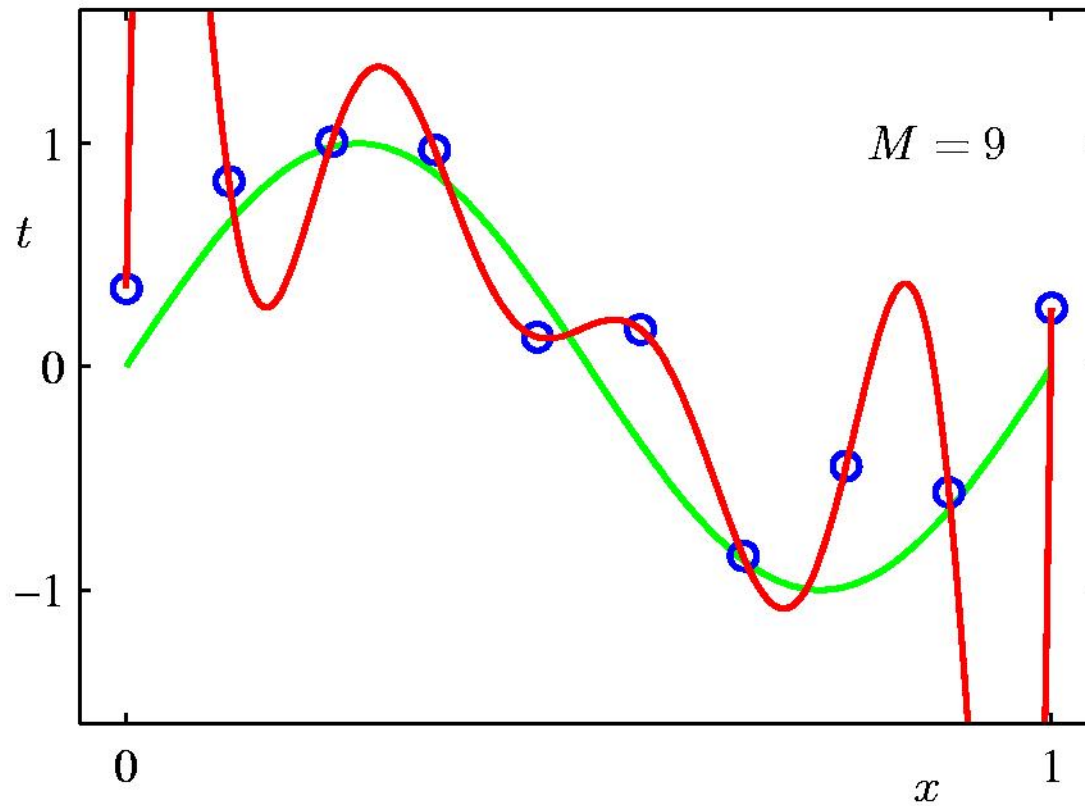
# 3<sup>rd</sup> Order Polynomial

---



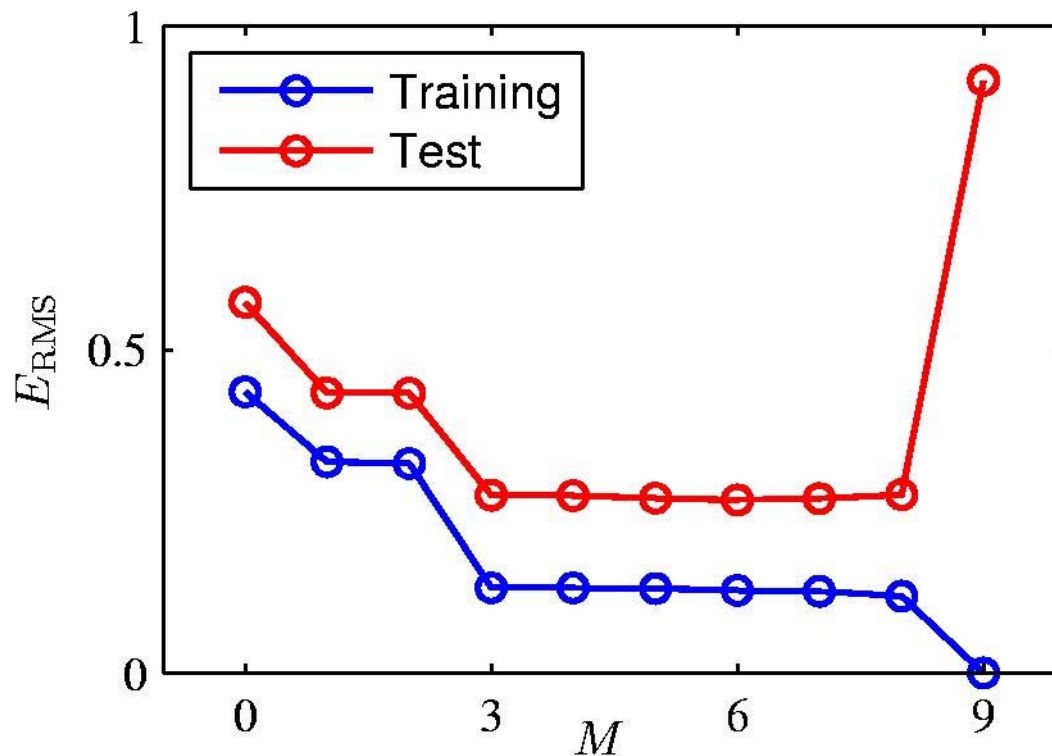
# 9<sup>th</sup> Order Polynomial

---



# Over-fitting

---



Root-Mean-Square (RMS) Error:  $E_{\text{RMS}} = \sqrt{2E(\mathbf{w}^*)/N}$

---

# Polynomial Coefficients

---

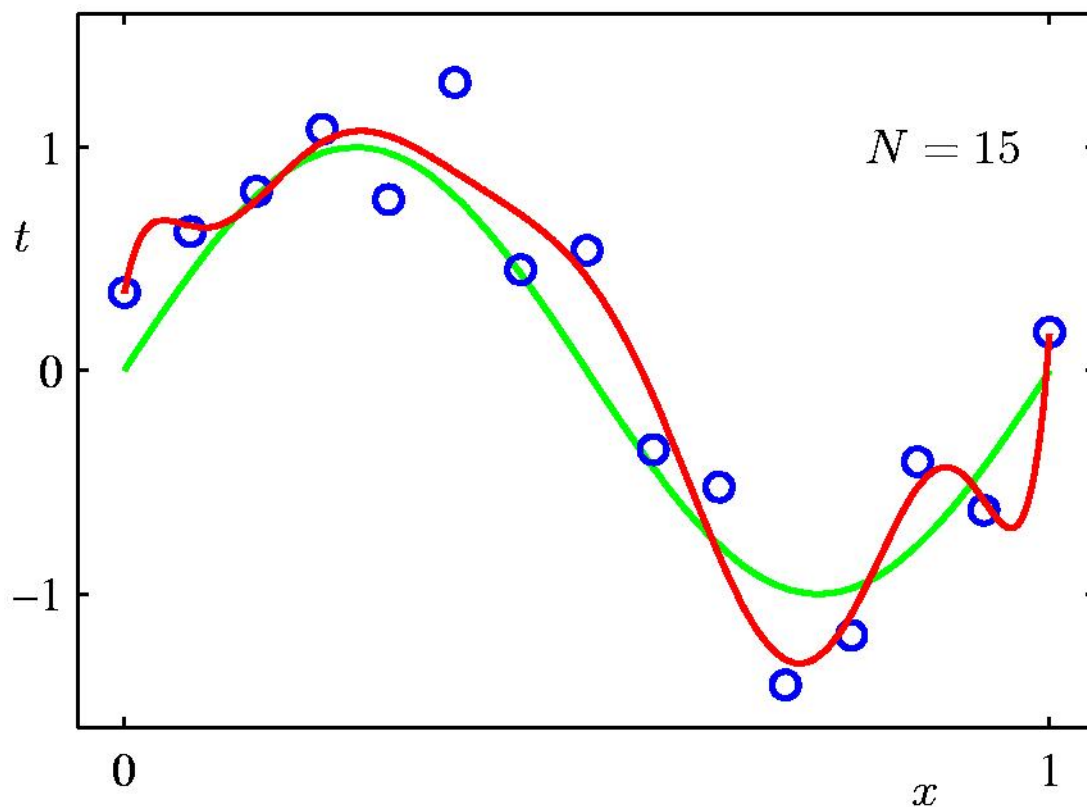
	$M = 0$	$M = 1$	$M = 3$	$M = 9$
$w_0^*$	0.19	0.82	0.31	0.35
$w_1^*$		-1.27	7.99	232.37
$w_2^*$			-25.43	-5321.83
$w_3^*$			17.37	48568.31
$w_4^*$				-231639.30
$w_5^*$				640042.26
$w_6^*$				-1061800.52
$w_7^*$				1042400.18
$w_8^*$				-557682.99
$w_9^*$				125201.43

---

# Data Set Size: $N = 15$

---

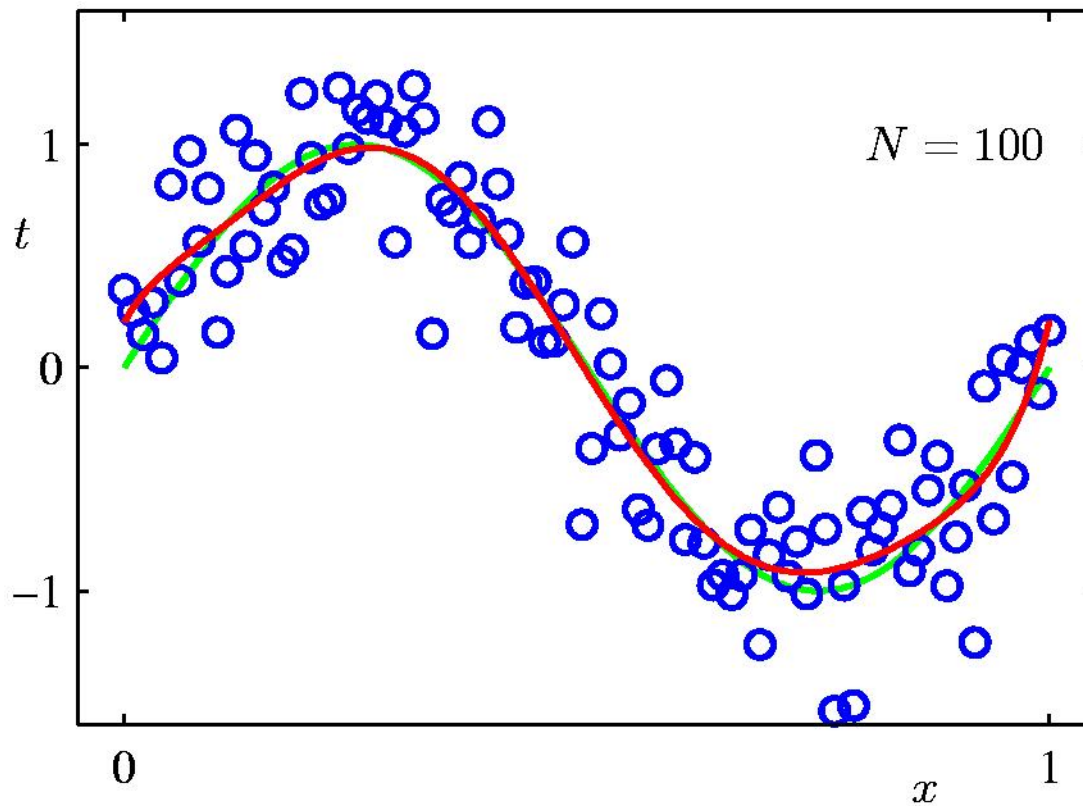
9<sup>th</sup> Order Polynomial



# Data Set Size: $N = 100$

---

9<sup>th</sup> Order Polynomial



# Regularization

---

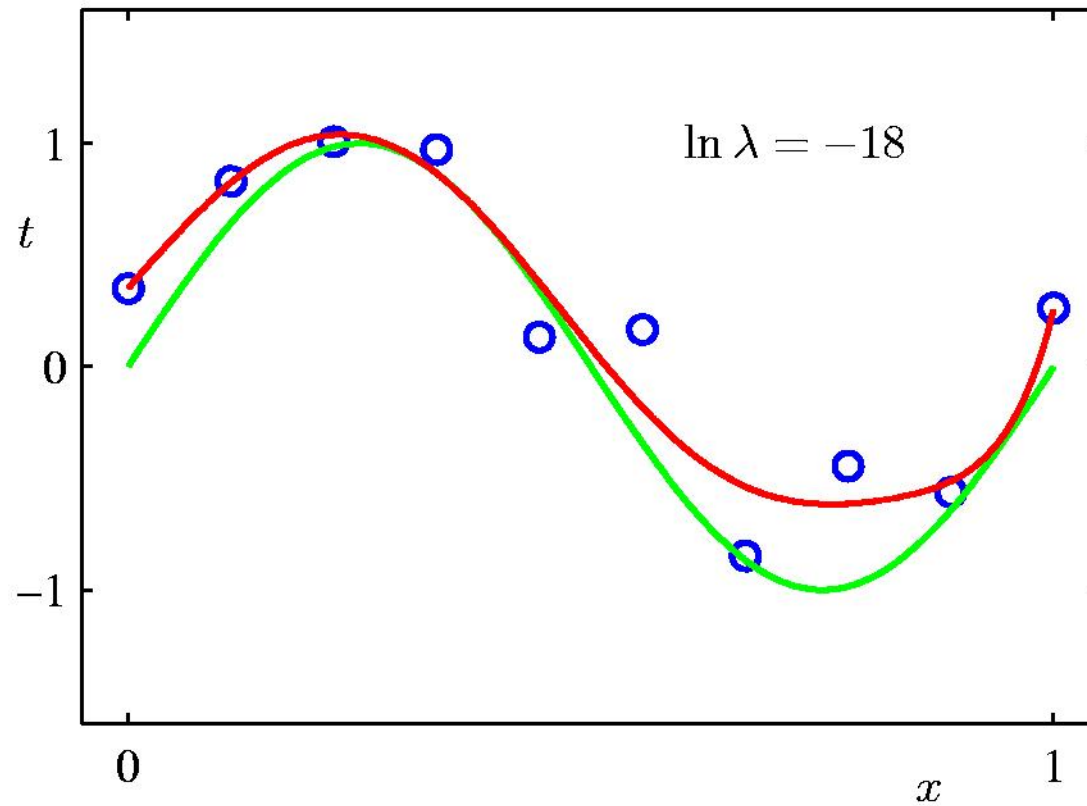
Penalize large coefficient values

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



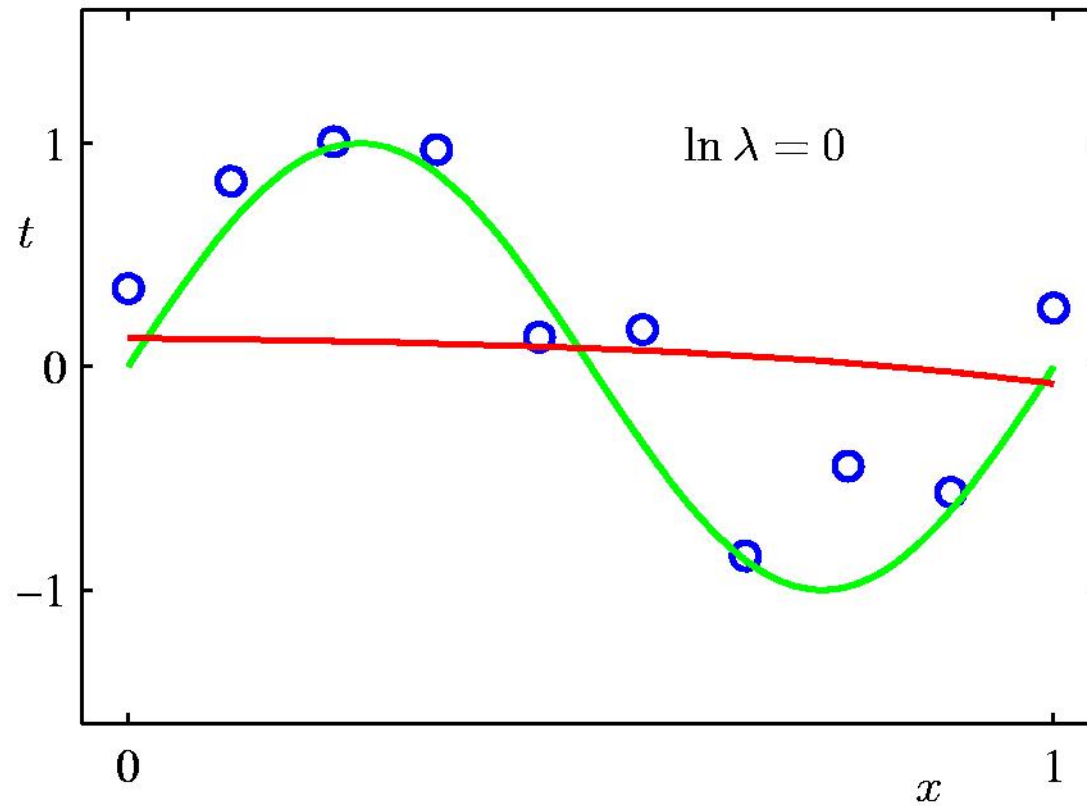
# Regularization: $\ln \lambda = -18$

---



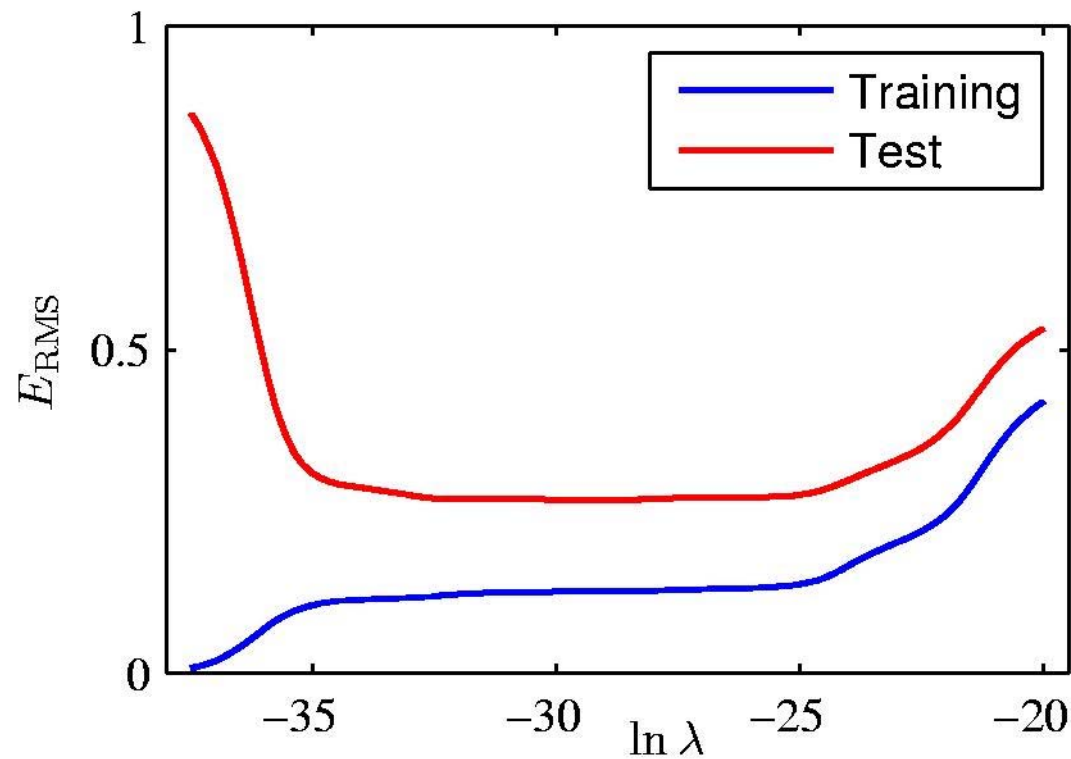
# Regularization: $\ln \lambda = 0$

---



# Regularization: $E_{\text{RMS}}$ vs. $\ln \lambda$

---



# Polynomial Coefficients

---

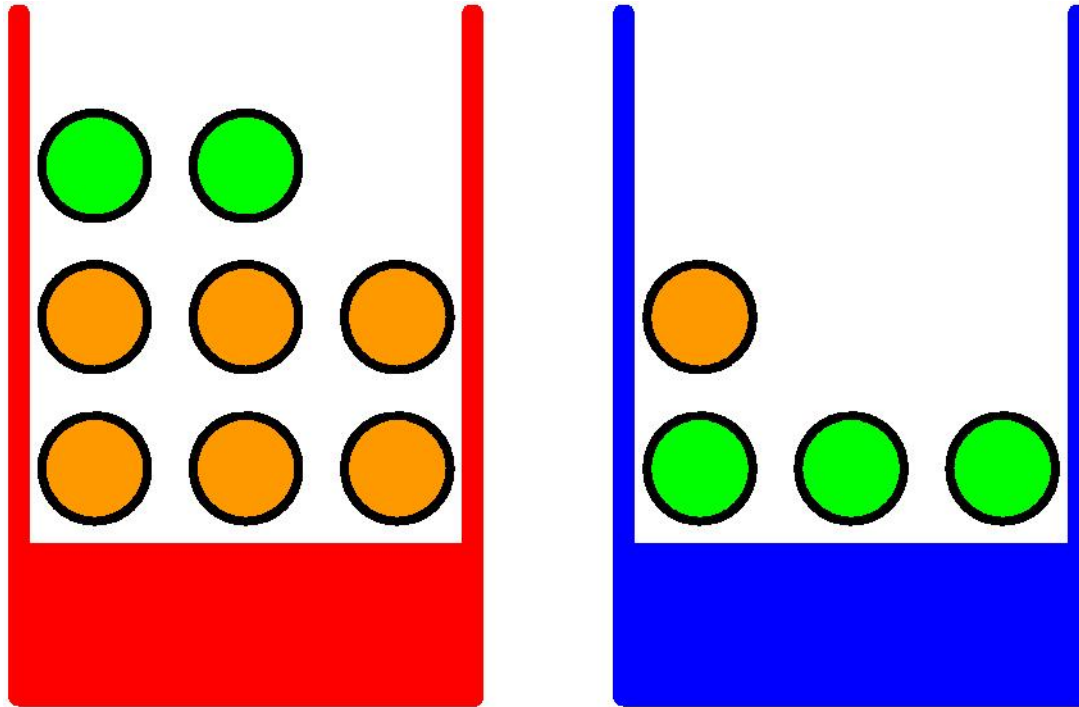
	$\ln \lambda = -\infty$	$\ln \lambda = -18$	$\ln \lambda = 0$
$w_0^*$	0.35	0.35	0.13
$w_1^*$	232.37	4.74	-0.05
$w_2^*$	-5321.83	-0.77	-0.06
$w_3^*$	48568.31	-31.97	-0.05
$w_4^*$	-231639.30	-3.89	-0.03
$w_5^*$	640042.26	55.28	-0.02
$w_6^*$	-1061800.52	41.32	-0.01
$w_7^*$	1042400.18	-45.95	-0.00
$w_8^*$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

---

# Probability Theory

---

Apples and Oranges



# Probability Theory

---

B: a random variable denotes the identity of the box that will be chosen.

F: a random variable denotes the fruit

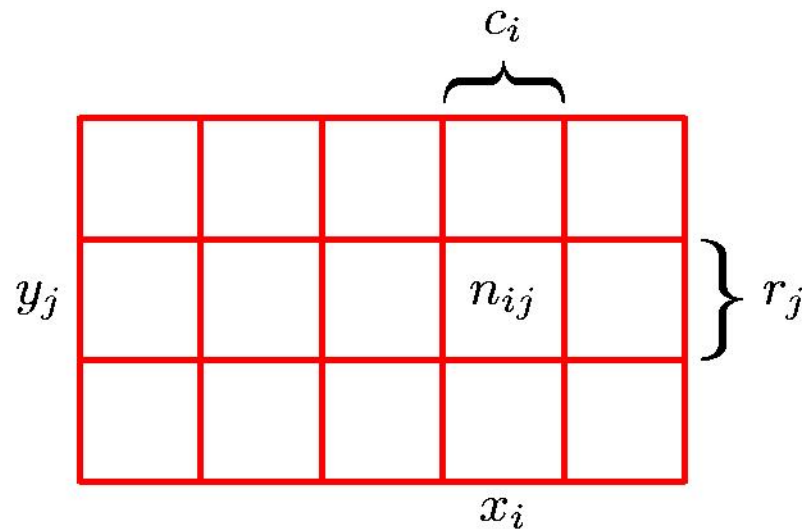
**Prior probability:**  $p(B)$  is the probability available before we observe the identity of the fruit.

**Posterior probability:**  $p(B | F)$  is the probability obtained after we have observed F.

---

# Probability Theory

---



Marginal Probability

$$p(X = x_i) = \frac{c_i}{N}.$$

Joint Probability

$$p(X = x_i, Y = y_j) = \frac{n_{ij}}{N}$$

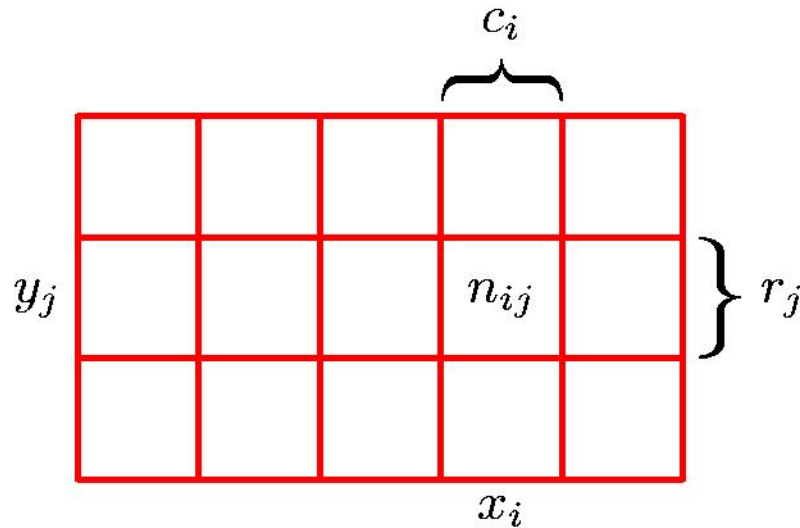
Conditional Probability

$$p(Y = y_j | X = x_i) = \frac{n_{ij}}{c_i}$$

---

# Probability Theory

---



Sum Rule

$$\begin{aligned} p(X = x_i) &= \frac{c_i}{N} = \frac{1}{N} \sum_{j=1}^L n_{ij} \\ &= \sum_{j=1}^L p(X = x_i, Y = y_j) \end{aligned}$$

Product Rule

$$\begin{aligned} p(X = x_i, Y = y_j) &= \frac{n_{ij}}{N} = \frac{n_{ij}}{c_i} \cdot \frac{c_i}{N} \\ &= p(Y = y_j | X = x_i) p(X = x_i) \end{aligned}$$

---



# The Rules of Probability

---

Sum Rule

$$p(X) = \sum_Y p(X, Y)$$

Product Rule

$$p(X, Y) = p(Y|X)p(X)$$

---

# Bayes' Theorem

---

$$p(Y|X) = \frac{p(X|Y)p(Y)}{p(X)}$$

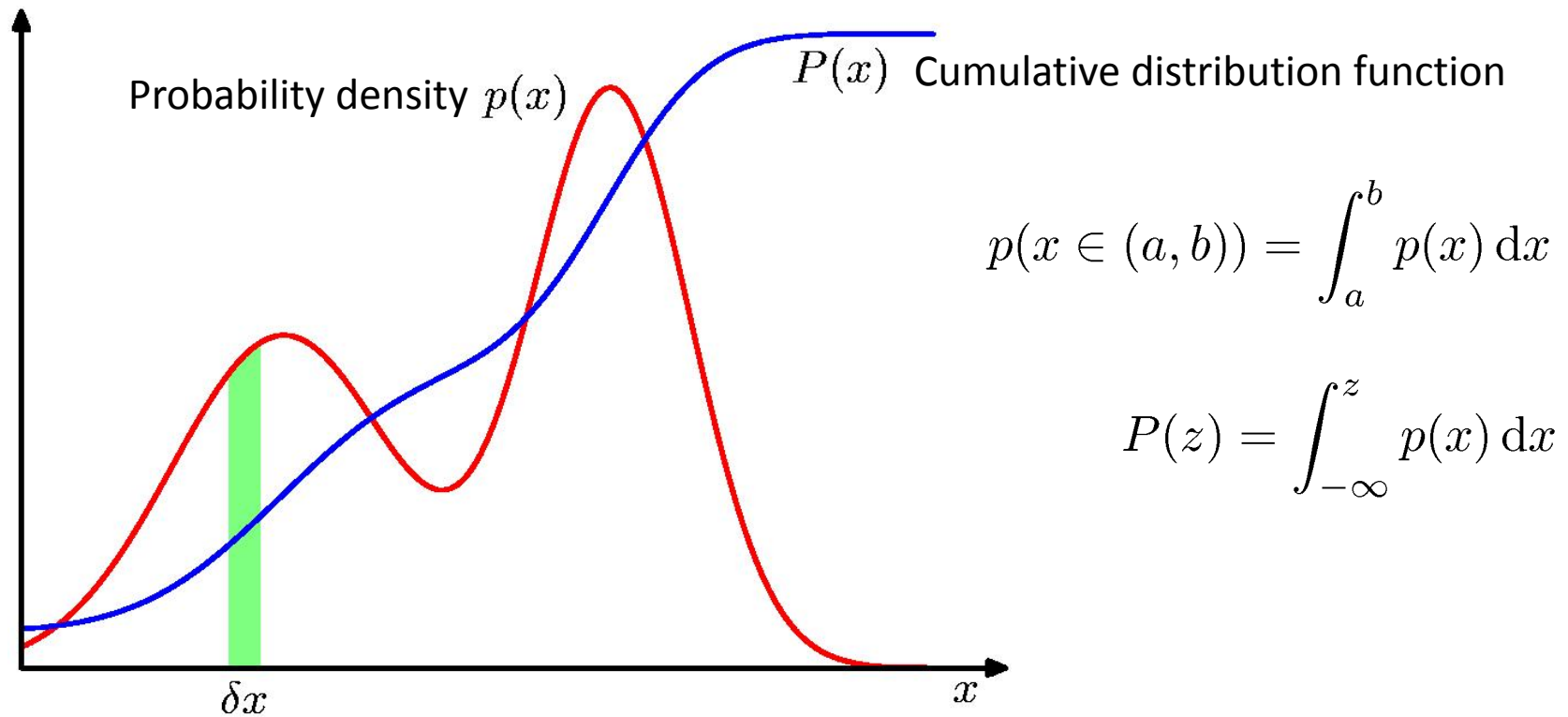
$$p(X) = \sum_Y p(X|Y)p(Y)$$

posterior  $\propto$  likelihood  $\times$  prior

---

# Probability Densities

---



$$p(x \in (a, b)) = \int_a^b p(x) dx$$

$$P(z) = \int_{-\infty}^z p(x) dx$$

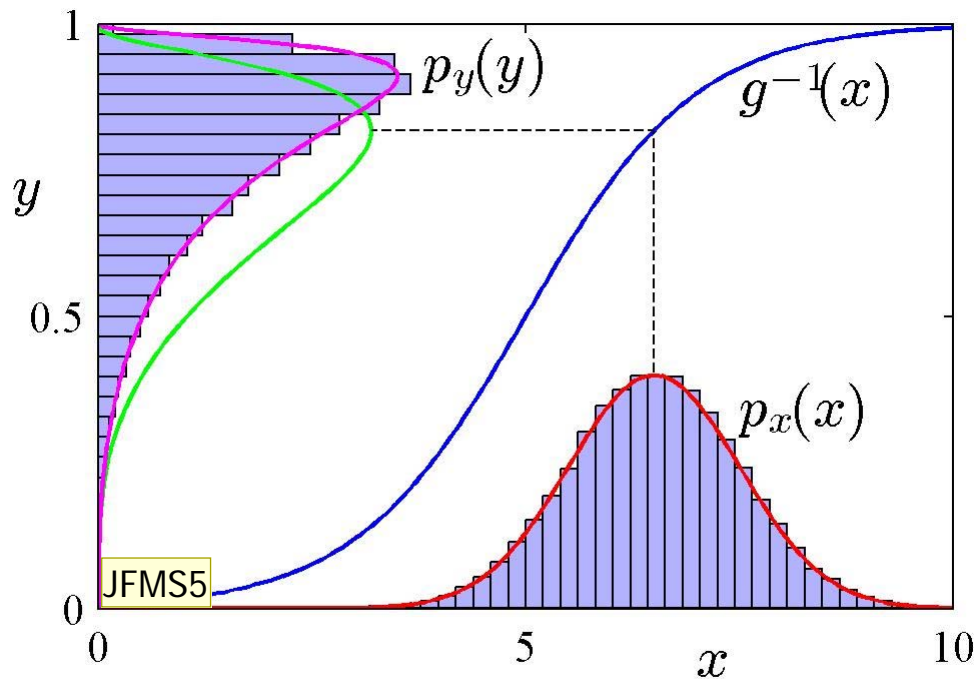
$$p(x) \geq 0$$

$$\int_{-\infty}^{\infty} p(x) dx = 1$$

---

# Transformed Densities

---




$$\begin{aligned} p_y(y) &= p_x(x) \left| \frac{dx}{dy} \right| \\ &= p_x(g(y)) |g'(y)| \end{aligned}$$

# Expectations

---

$$\mathbb{E}[f] = \sum_x p(x)f(x)$$

$$\mathbb{E}[f] = \int p(x)f(x) \, dx$$

$$\mathbb{E}_x[f|y] = \sum_x p(x|y)f(x)$$


Conditional Expectation  
(discrete)

$$\mathbb{E}[f] \simeq \frac{1}{N} \sum_{n=1}^N f(x_n)$$

Approximate Expectation  
(discrete and continuous)

---

# Variances and Covariances

---

$$\text{var}[f] = \mathbb{E} \left[ (f(x) - \mathbb{E}[f(x)])^2 \right] = \mathbb{E}[f(x)^2] - \mathbb{E}[f(x)]^2$$

$$\begin{aligned} \text{cov}[x, y] &= \mathbb{E}_{x,y} [\{x - \mathbb{E}[x]\} \{y - \mathbb{E}[y]\}] \\ &= \mathbb{E}_{x,y} [xy] - \mathbb{E}[x]\mathbb{E}[y] \end{aligned}$$

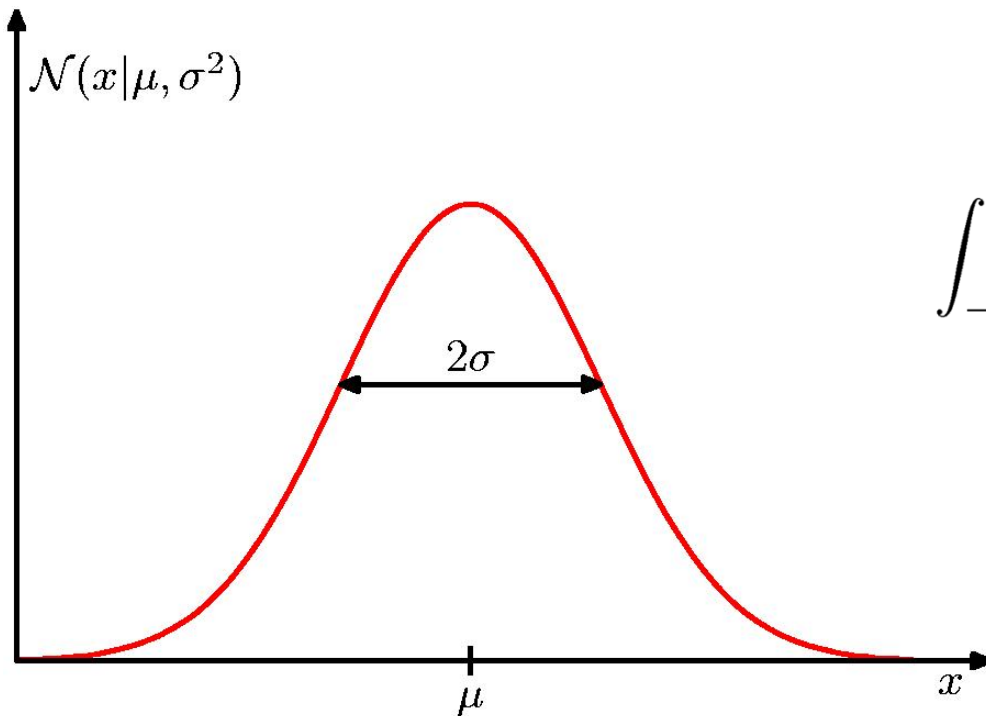
$$\begin{aligned} \text{cov}[\mathbf{x}, \mathbf{y}] &= \mathbb{E}_{\mathbf{x}, \mathbf{y}} [\{\mathbf{x} - \mathbb{E}[\mathbf{x}]\} \{\mathbf{y}^T - \mathbb{E}[\mathbf{y}^T]\}] \\ &= \mathbb{E}_{\mathbf{x}, \mathbf{y}} [\mathbf{x} \mathbf{y}^T] - \mathbb{E}[\mathbf{x}] \mathbb{E}[\mathbf{y}^T] \end{aligned}$$

---

# The Gaussian Distribution

---

$$\mathcal{N}(x|\mu, \sigma^2) = \frac{1}{(2\pi\sigma^2)^{1/2}} \exp \left\{ -\frac{1}{2\sigma^2}(x - \mu)^2 \right\}$$



$$\mathcal{N}(x|\mu, \sigma^2) > 0$$

$$\int_{-\infty}^{\infty} \mathcal{N}(x|\mu, \sigma^2) dx = 1$$

# Gaussian Mean and Variance

---

$$\mathbb{E}[x] = \int_{-\infty}^{\infty} \mathcal{N}(x|\mu, \sigma^2) x \, dx = \mu$$

$$\mathbb{E}[x^2] = \int_{-\infty}^{\infty} \mathcal{N}(x|\mu, \sigma^2) x^2 \, dx = \mu^2 + \sigma^2$$

$$\text{var}[x] = \mathbb{E}[x^2] - \mathbb{E}[x]^2 = \sigma^2$$

---



# The Multivariate Gaussian

---

$$\mathcal{N}(\mathbf{x}|\boldsymbol{\mu}, \boldsymbol{\Sigma}) = \frac{1}{(2\pi)^{D/2}} \frac{1}{|\boldsymbol{\Sigma}|^{1/2}} \exp \left\{ -\frac{1}{2}(\mathbf{x} - \boldsymbol{\mu})^T \boldsymbol{\Sigma}^{-1}(\mathbf{x} - \boldsymbol{\mu}) \right\}$$

