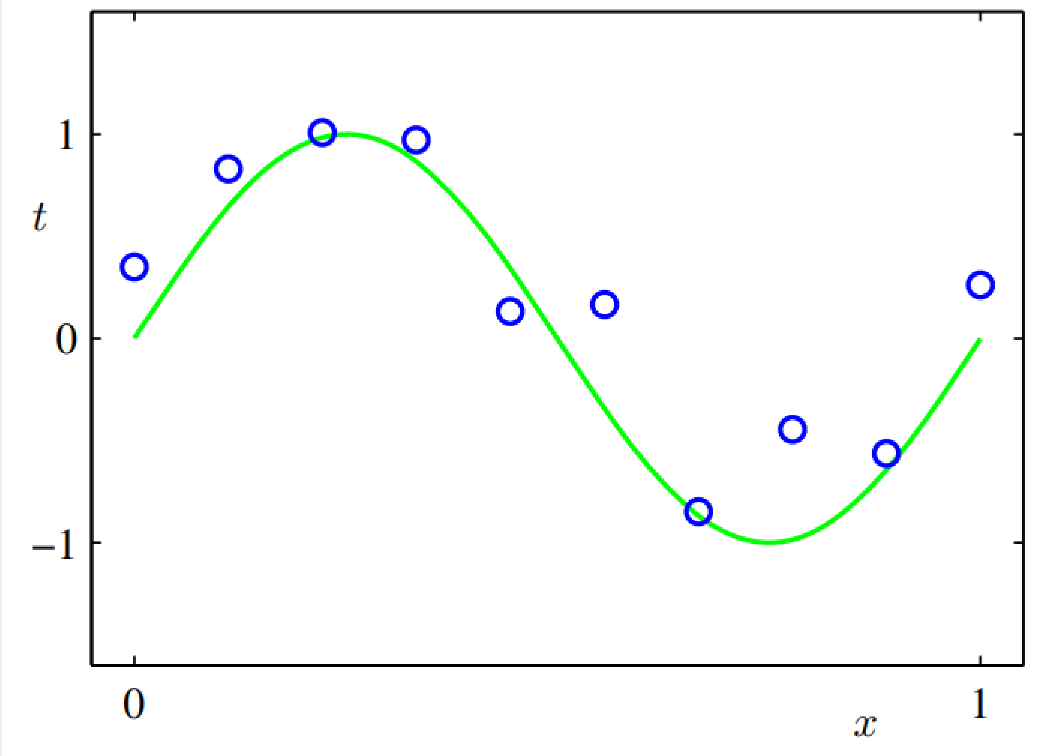
Component of learning

By Mustafa Shiple

2

REGRESSION EXAMPLE:

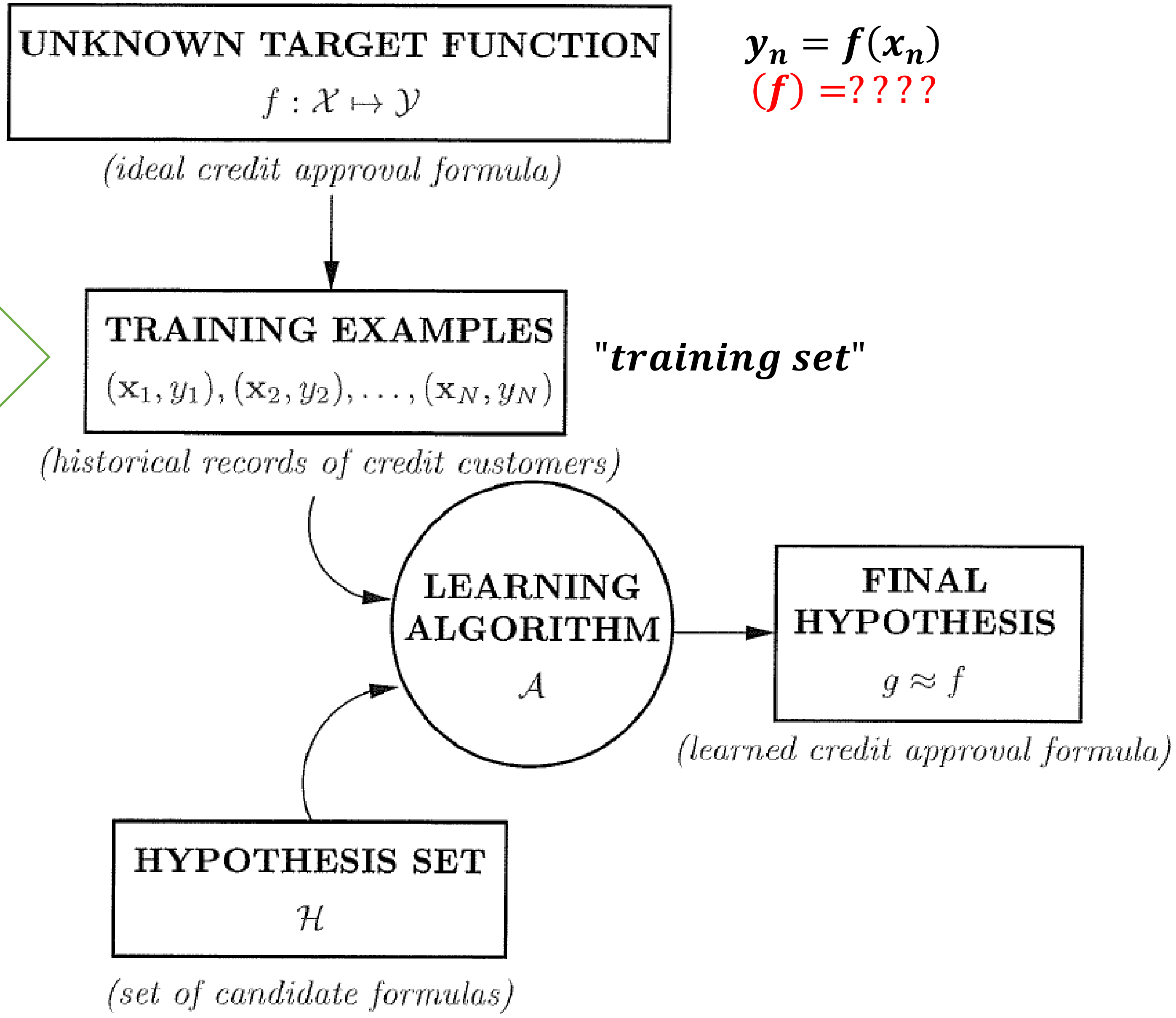
- - The green curve shows the function $\sin(2\pi x)$ used to generate date with gaussian noise distribution



Components of learning

(χ): ***input space***, (y): ***output space***

n	item	Value (χ)	Decision (y)
1	Age	23	Yes
	Annual salary	$ (30 000)	
	Residence (years)	1	
	Current debt	$ (15 000)	
2	Age	30	No
	Annual salary	$ (38 000)	
	Residence (years)	2	
	Current debt	$ (30 000)	
3	Age	35	Yes
	Annual salary	$ (38 000)	
	Residence (years)	10	
	Current debt	$ (10 000)	
⋮	⋮	⋮	⋮
N	Age	24	Yes
	Annual salary	$ (25 000)	
	Residence (years)	2	
	Current debt	$ (5 000)	



$y_n = f(x_n)$

(f) =????

"***training set***"

Component of learning

Problem Setting:

- Set of possible instances $(\boldsymbol{\chi})$
- Unknown target function $\boldsymbol{y_n} = \boldsymbol{\mathcal{F}}\,(\boldsymbol{x_n})$, $(\mathcal{X} \rightarrow \mathcal{Y})$
- Set of function hypotheses $\mathcal{H}$: $\left(\hbar | \hbar\text{:}\ \mathcal{X} \rightarrow \hat{\mathcal{Y}}\right)$

Input:

- Training examples $\{\ (\mathcal{X}_i \rightarrow \mathcal{Y}_i)\ \}$ of unknown target function $\mathcal{F}$

Output:

- Hypothesis $\hbar \in\ \mathcal{H}$ that best approximates target function

Exercise 1.1

Express each of the following tasks in the framework of learning from data by specifying :
the input space X, output space Y, target function ƒ: X → Y, and the specifics of the data set that we will learn from.

(a) Medical diagnosis: A patient walks in with a medical history and some symptoms, and you want to identify the problem.

- ❑ **Input space (χ):**
- ❑ **Output space (y):**
- ❑ **Target function $f : \chi \rightarrow y$:**
- ❑ **Data set:**

Exercise 1.1

Express each of the following tasks in the framework of learning from data by specifying :
the input space X, output space Y, target function ƒ: X → Y, and the specifics of the data set that we will learn from.

(a) Medical diagnosis: A patient walks in with a medical history and some symptoms, and you want to identify the problem.

- ❑ **Input space (χ):** patient's medical history, symptoms, personal health information etc.
- ❑ **Output space (y):** all possible diseases
- ❑ **Target function** $f : \chi \rightarrow y$: ideal formula to identify a patient's problem
- ❑ **Data set:** All available patients' information and their corresponding correct problem diagnostic.

Exercise 1.1

Express each of the following tasks in the framework of learning from data by specifying :
the input space X, output space Y, target function ƒ: X → Y, and the specifics of the data set that we will learn from.

(B) Handwritten digit recognition (for example postal code recognition for mail sorting).

- ❑ **Input space (χ):**
- ❑ **Output space (y):**
- ❑ **Target function $f : \chi \rightarrow y$:**
- ❑ **Data set:**

Exercise 1.1

Express each of the following tasks in the framework of learning from data by specifying :
the input space X, output space Y, target function ƒ: X → Y, and the specifics of the data set that we will learn from.

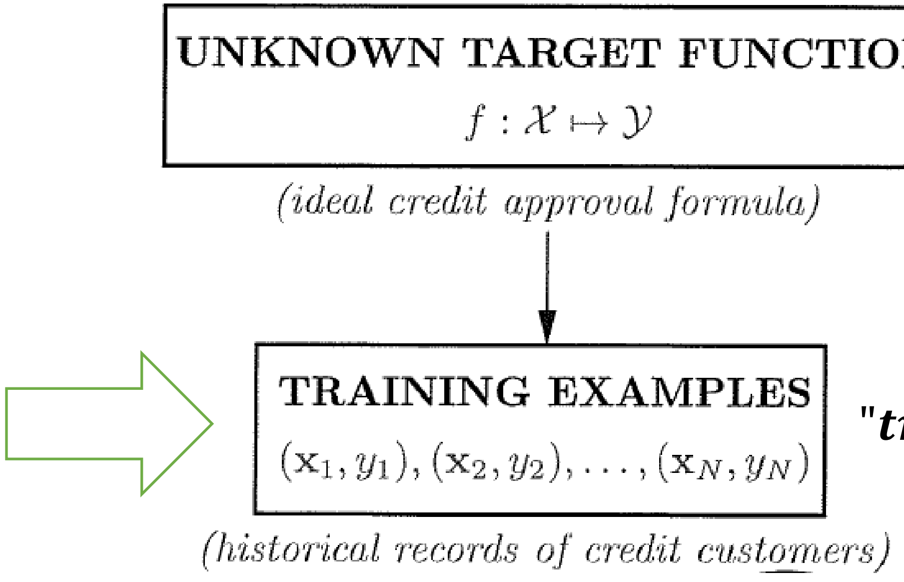
(B) Handwritten digit recognition (for example postal code recognition for mail sorting).

- **Input space (χ):** handwritten digits (digitalized) .
- **Output space (y):** 0-9 digits
- **Target function** $f : \chi \rightarrow y$: ideal formula match a handwritten digit to a correct digit
- **Data set:** handwritten digits and their corresponding correct matches

Components of learning

(χ): ***input space***, (y): ***output space***

n	item	Value (χ)	Decision (y)
1	Age	23	Yes
	Annual salary	$ (30 000)	
	Residence (years)	1	
	Current debt	$ (15 000)	
2	Age	30	No
	Annual salary	$ (38 000)	
	Residence (years)	2	
	Current debt	$ (30 000)	
3	Age	35	Yes
	Annual salary	$ (38 000)	
	Residence (years)	10	
	Current debt	$ (10 000)	
⋮	⋮	⋮	⋮
N	Age	24	Yes
	Annual salary	$ (25 000)	
	Residence (years)	2	
	Current debt	$ (5 000)	



$y_n = f(x_n)$

$(f) = ????$

"***training set***"

$$y = \{-1, +1\}$$

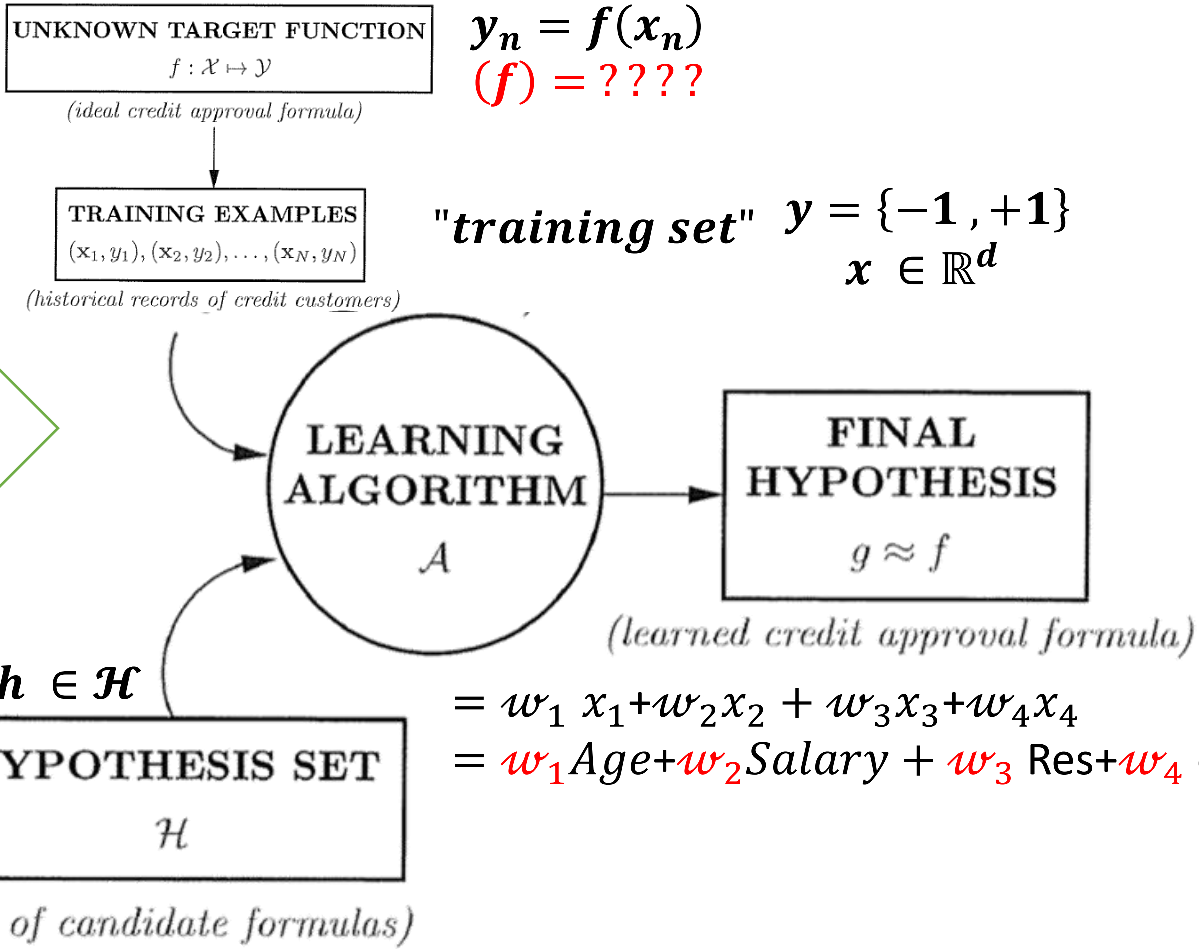
$$x \in \mathbb{R}^d$$

Where d : multiple dimensions represents age, salary …etc.

Components of learning

(χ): ***input space***, (y): ***output space***

n	item	Value (χ)	Decision (y)
1	Age	23	Yes
	Annual salary	$ (30 000)	
	Residence (years)	1	
	Current debt	$ (15 000)	
2	Age	30	No
	Annual salary	$ (38 000)	
	Residence (years)	2	
	Current debt	$ (30 000)	
3	Age	35	Yes
	Annual salary	$ (38 000)	
	Residence (years)	10	
	Current debt	$ (10 000)	
⋮	⋮	⋮	⋮
N	Age	24	Yes
	Annual salary	$ (25 000)	
	Residence (years)	2	
	Current debt	$ (5 000)	



$y_n = f(x_n)$

$(f) = ????$

"***training set***" $y = \{-1, +1\}$

$x \in \mathbb{R}^d$

$h \in \mathcal{H}$

$= w_1 x_1 + w_2 x_2 + w_3 x_3 + w_4 x_4$

$= w_1 Age + w_2 Salary + w_3$ Res $+ w_4$ debt

$$\sum_i^d w_i x_i > thereshold \iff y = +1$$

$$\sum_i^d w_i x_i < thereshold \iff y = -1$$

Components of learning

HYPOTHESIS SET
$\mathcal{H}$

$$\sum_{i}^{d} w_i x_i > thershold \quad \Leftrightarrow y = +1$$

$$\sum_{i}^{d} w_i x_i < thershold \quad \Leftrightarrow y = -1$$

$$h(x) = sign\left(\left(\sum_{i=1}^{d} w_i x_i\right) + b\right) \qquad \boldsymbol{h} \in \boldsymbol{\mathcal{H}}$$

Where b : threshold

$$h(x) = sign\left(\left(\sum_{i=1}^{d} w_i x_i\right) + (w_0 \times 1)\right)$$

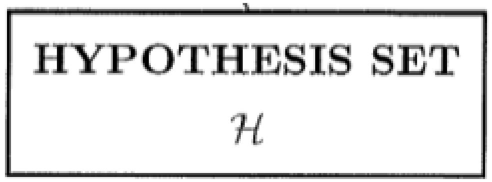
$$h(x) = sign\left(\left(\sum_{i=1}^{d} w_i x_i\right) + (w_0 \times x_0)\right)$$

$$h(x) = sign\left(\left(\sum_{i=0}^{d} w_i x_i\right)\right) \qquad where\ x_0 = 1$$

Bias → weights

$$= w_0 + w_1 x_1 + w_2 x_2 + w_3 x_3 + w_4 x_4$$

Components of learning

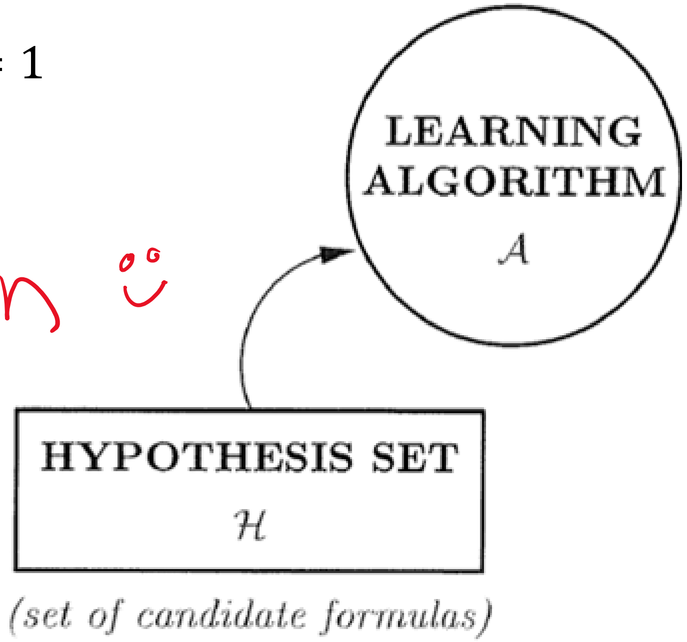


$$h(x) = sign\left(\left(\sum_{i=0}^{d} w_i x_i\right)\right) \quad where\ x_0 = 1$$

$$h(x) = sign(w^T x)$$

← Perceptron ☺

Perceptron
Learning Algorithm
(PLA)



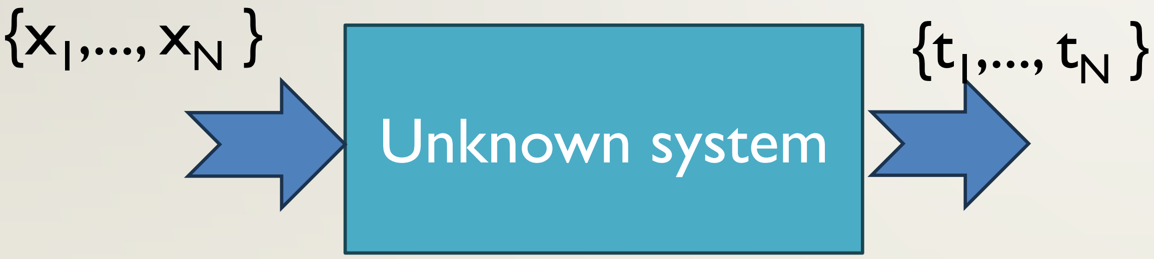
REGRESSION EXAMPLE: CURVE FITTING (LINEAR MODEL)

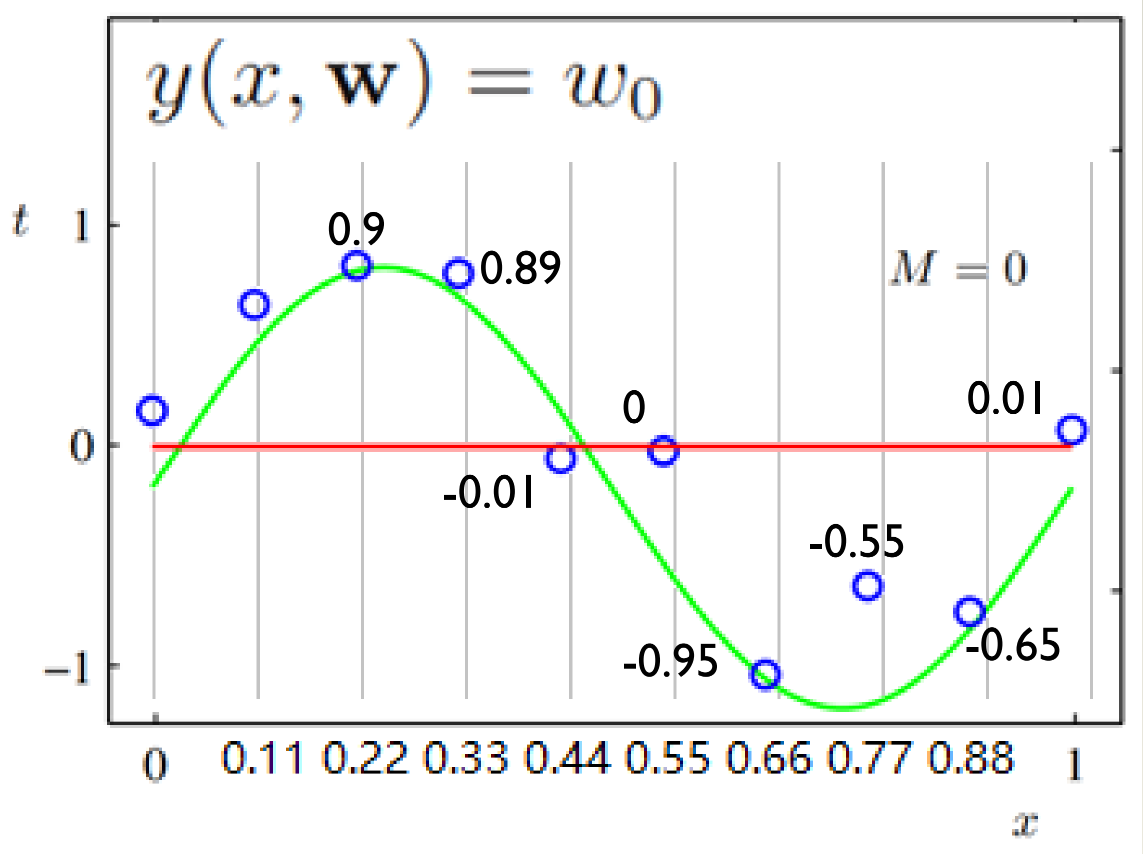
Training set (X) :

$\{x_1,\ldots, x_N\}$ → {0.11,0.22,0.33,0.44,0.55,0.66,0.77,0.88}

Target vector (t) :

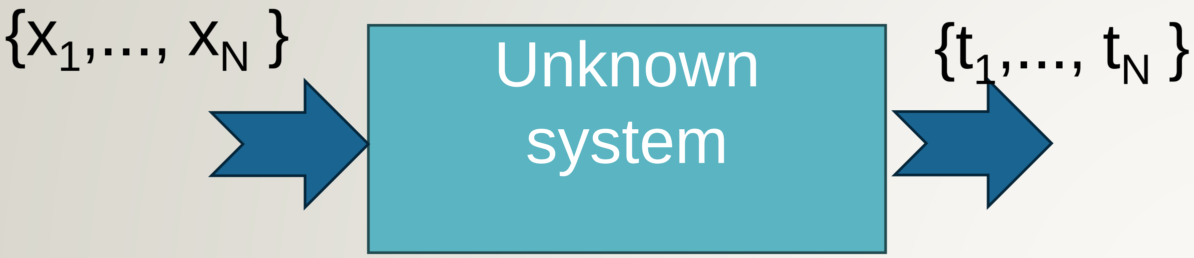
$\{t_1,\ldots, t_N\}$ → {0.1,0.6,0.9,0.89,-0.01,0,-0.95,-0.55…etc}





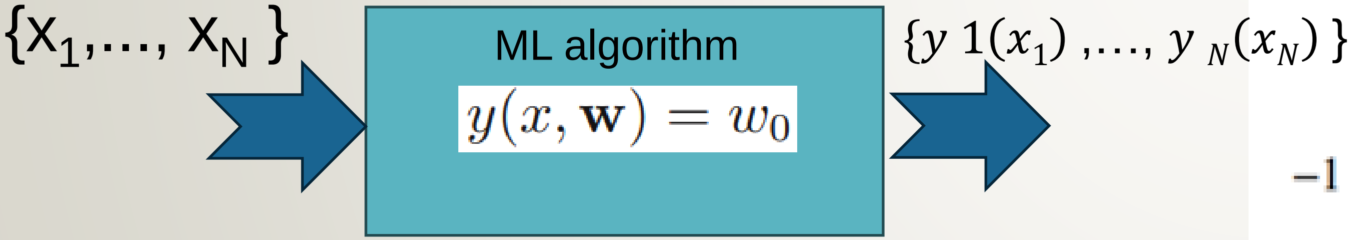
Regression example: Training Phase

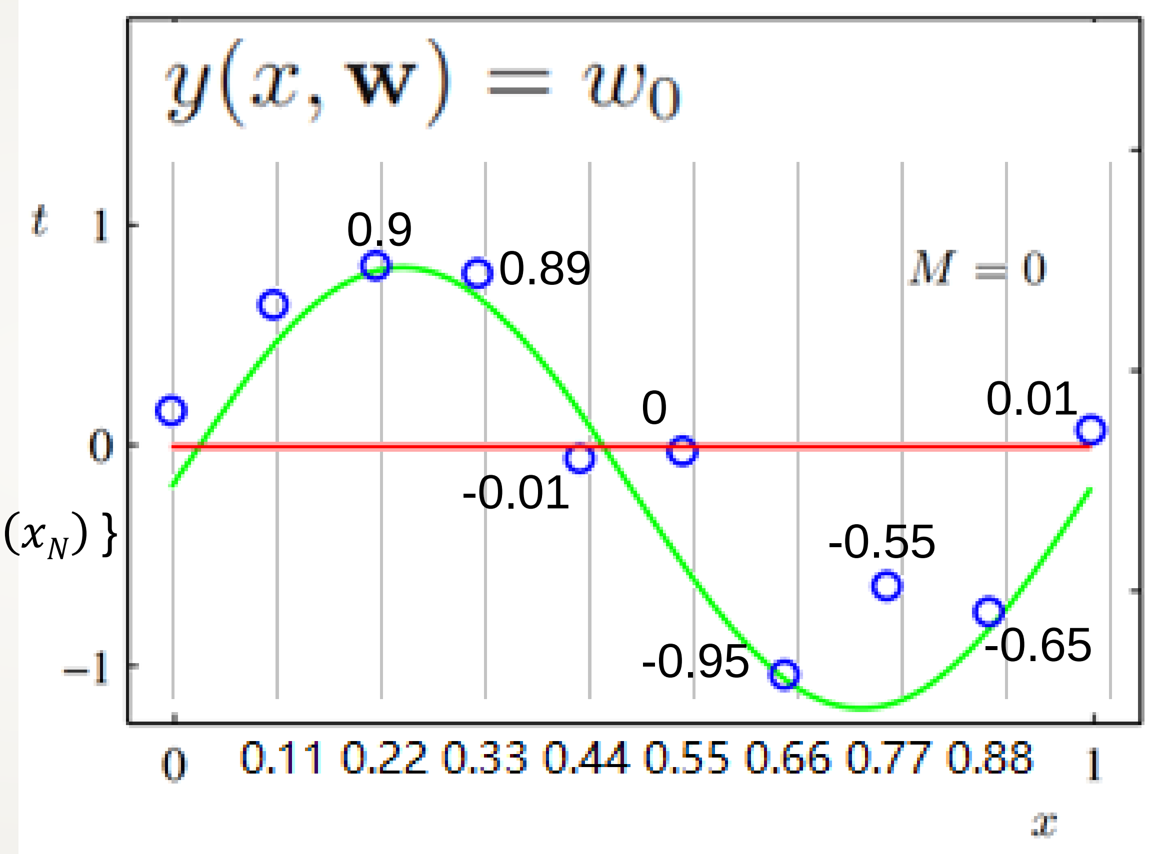
Implicitly trying to discover the underlying function $\sin(2\pi x)$.



Training Phase (Learning Phase):

$y\ (x)\ find\ out\ \ y(training\ set)$

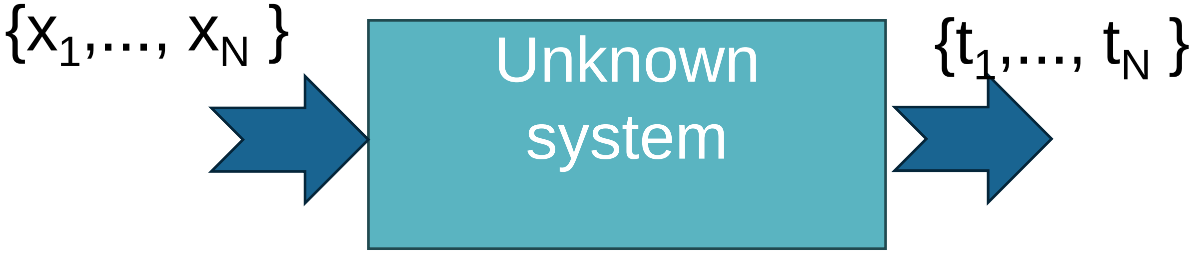




$$y(x, \mathbf{w}) = w_0 + w_1 x + w_2 x^2 + \ldots + w_M x^M = \sum_{j=0}^{M} w_j x^j$$

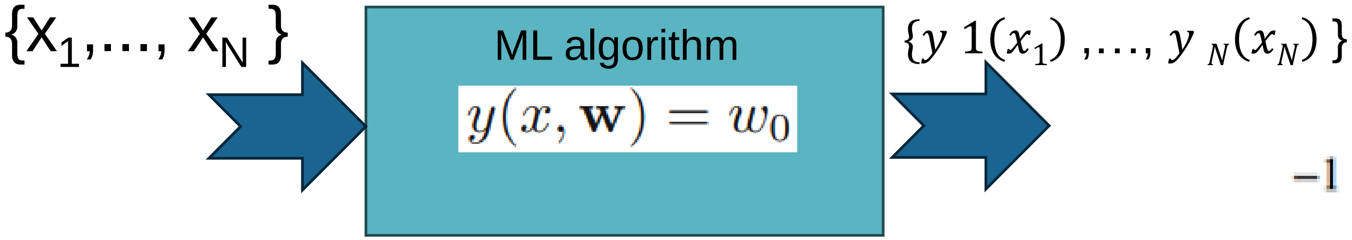
Regression example: Training Phase

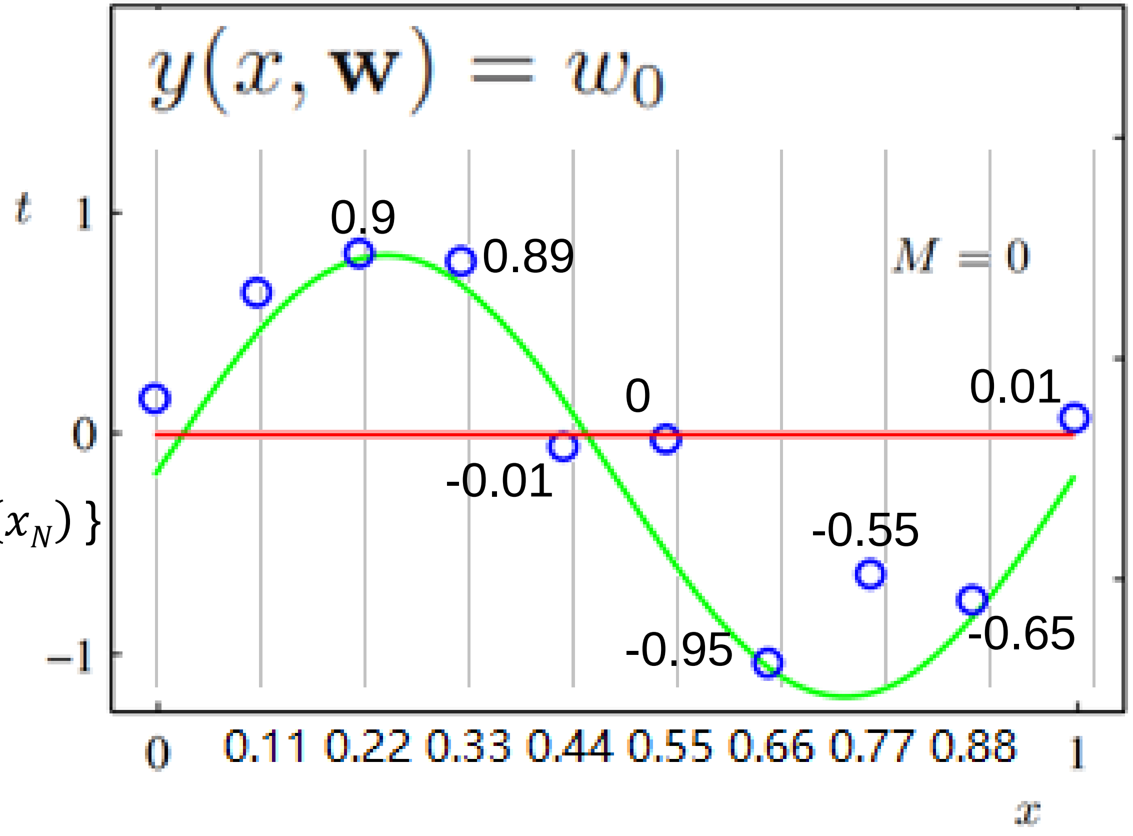
Implicitly trying to discover the underlying function $\sin(2\pi x)$.



Training Phase (Learning Phase):

$y\ (x)\ find\ out\ \ y(training\ set)$





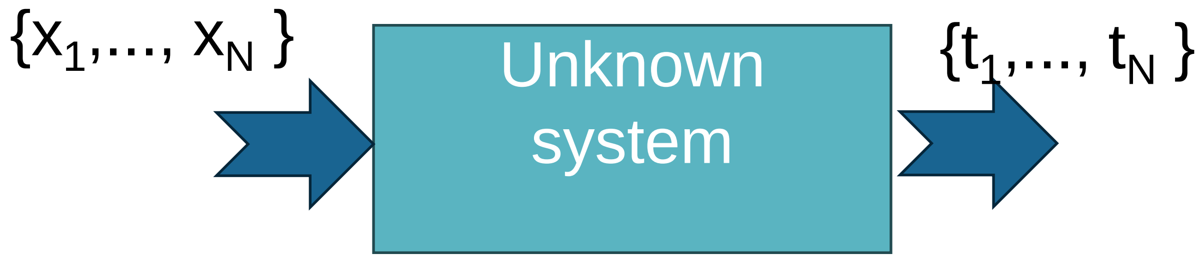
Error function :

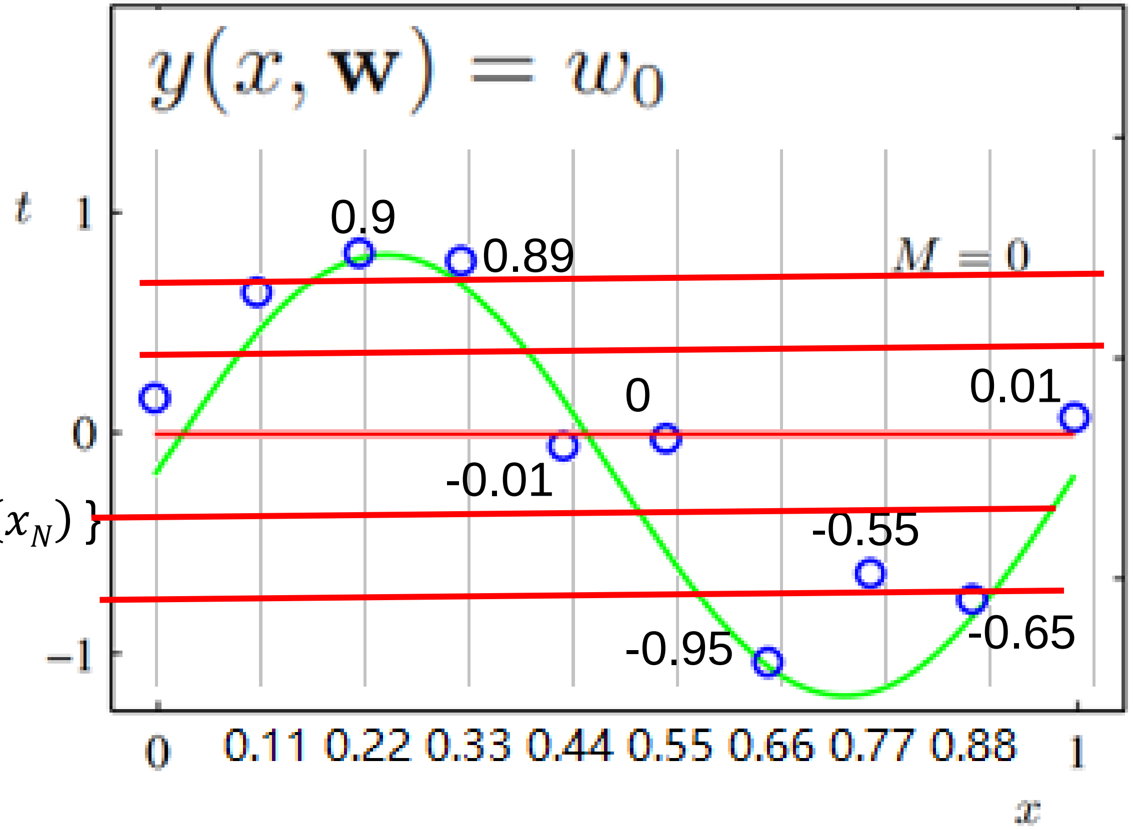
$$E(\mathbf{w}) = \frac{1}{2}\sum_{n=1}^{N}\{y(x_n, \mathbf{w}) - t_n\}^2$$

that measures the misfit between the function y(x, w), for any given value of w, and the training set data points.

Regression example: Training Phase (exploit vs explore)

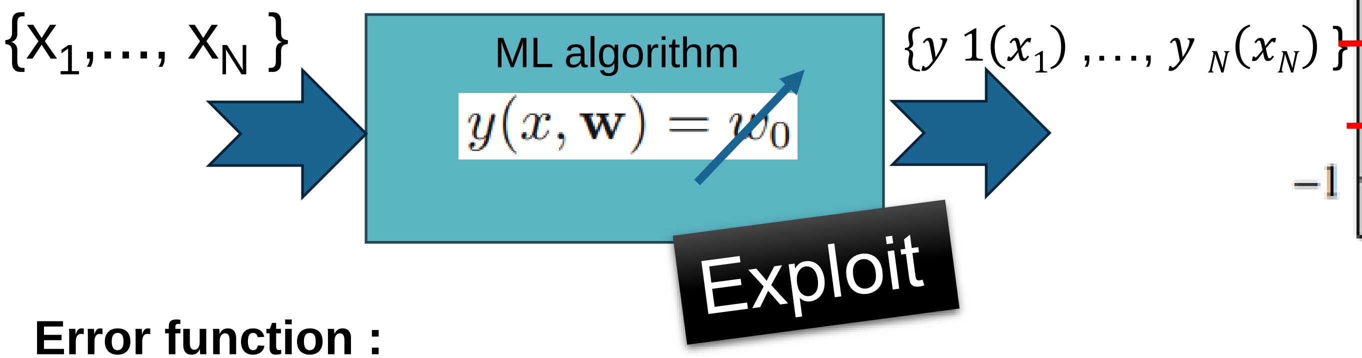
Implicitly trying to discover the underlying function $\sin(2\pi x)$.





Training Phase (Learning Phase):

$y\ (x)\ find\ out\ \ y(training\ set)$



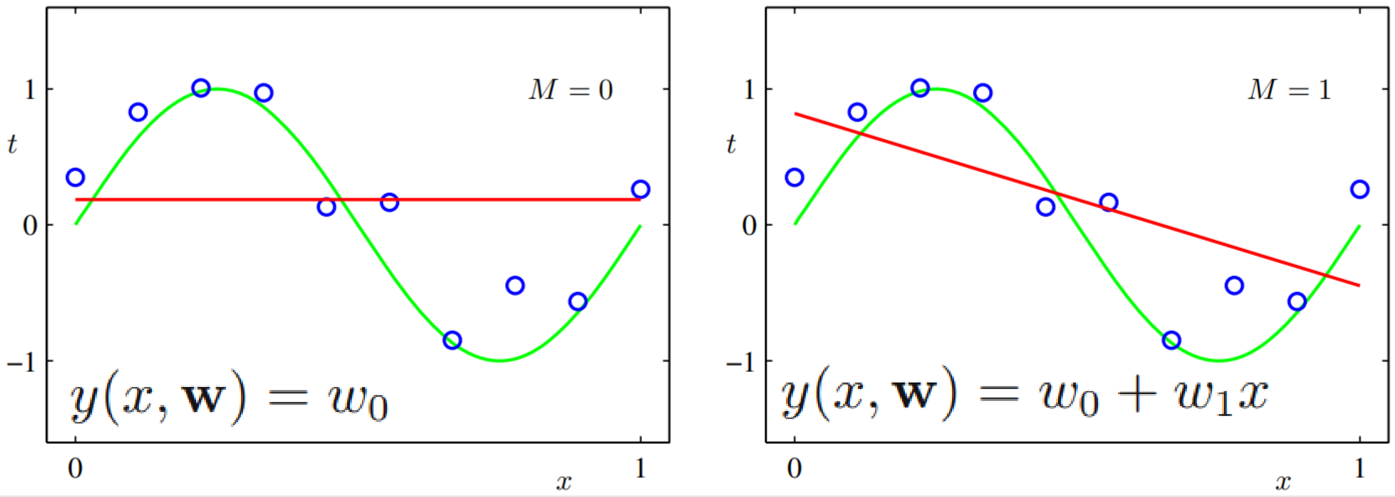
Error function :

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

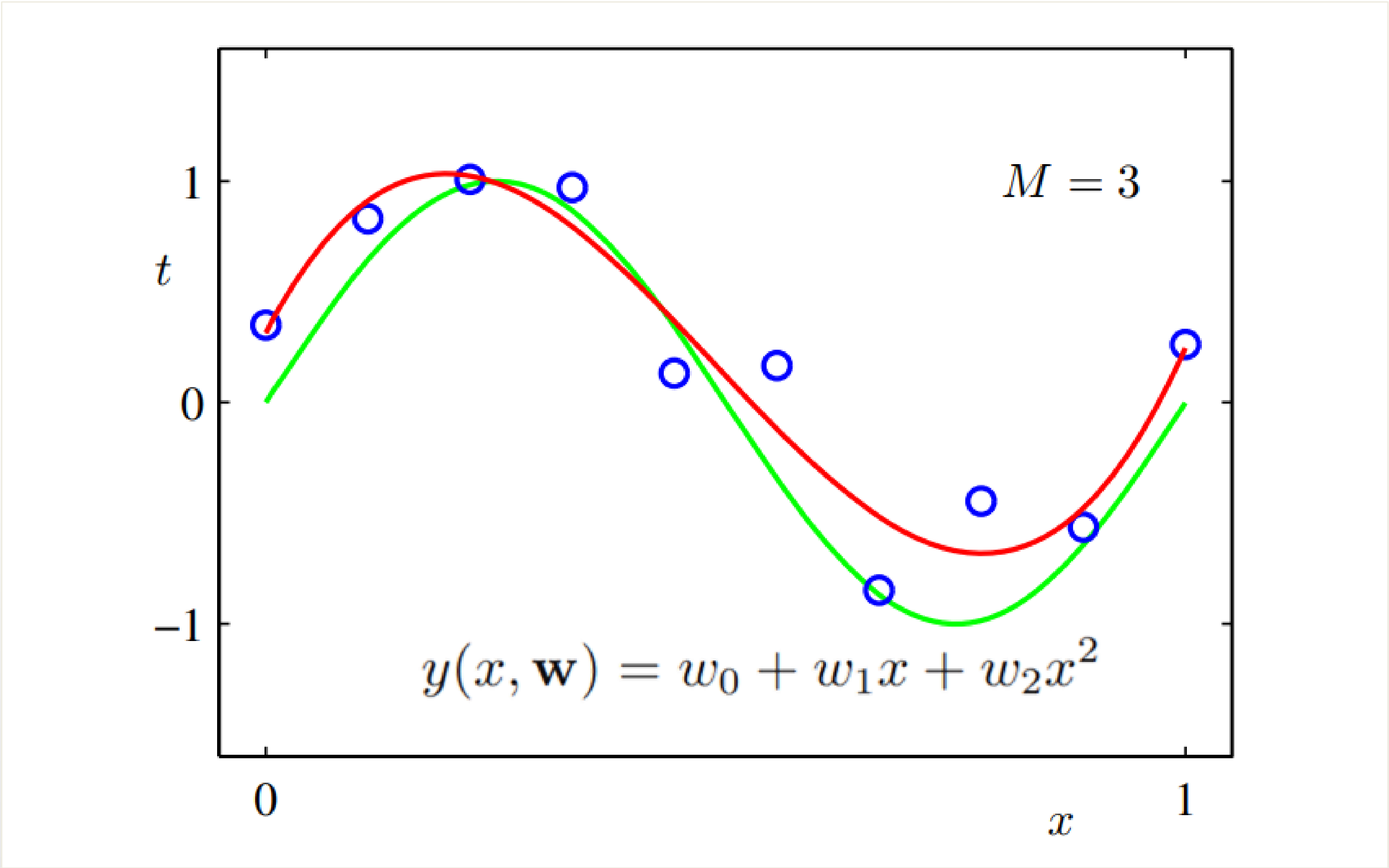
$$y(x, \mathbf{w}) = w_0 + w_1 x + w_2 x^2 + \ldots + w_M x^M = \sum_{j=0}^{M} w_j x^j$$

REGRESSION EXAMPLE: EXPLORING

$$y(x, \mathbf{w}) = w_0 + w_1 x$$



REGRESSION EXAMPLE: CURVE FITTING



REGRESSION EXAMPLE: GENERALIZATION

$y\ (x)\ find\ out\ \ y(training\ set)$

$\{x_1,\ldots, x_N\}$

ML algorithm

$y(x, \mathbf{w}) = w_0$

$\{y\ 1(x_1)\ ,\ldots,\ y\ _N(x_N)\ \}$

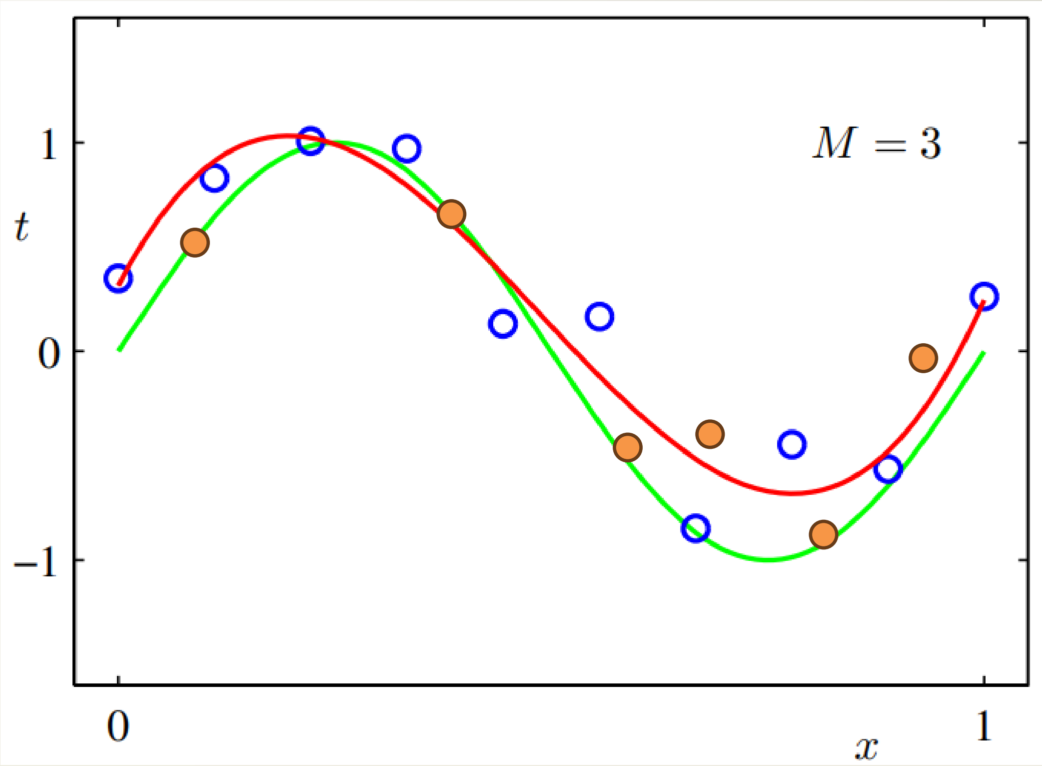
$\{\hat{x}_1,\ldots,\ \hat{x}_N\ \}$

test set

Test Phase

(Use Data set(test set) $\notin$ *training set*):

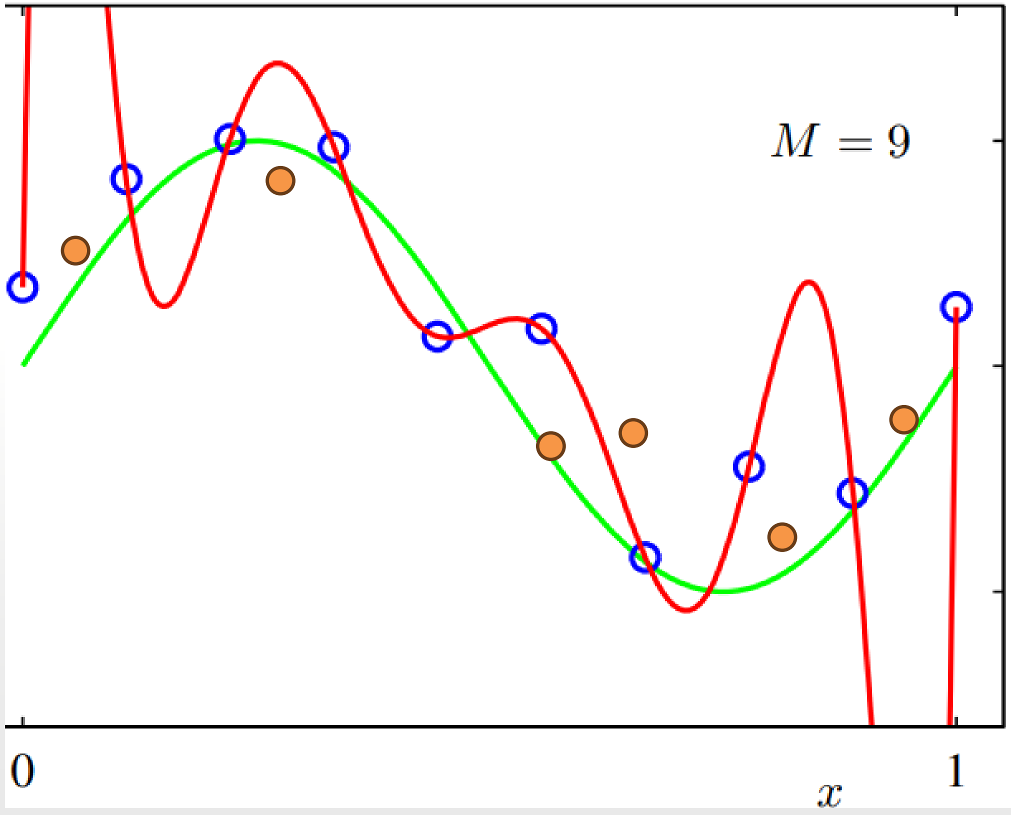
The ability to categorize correctly new examples that differ from those used for training is known as generalization.

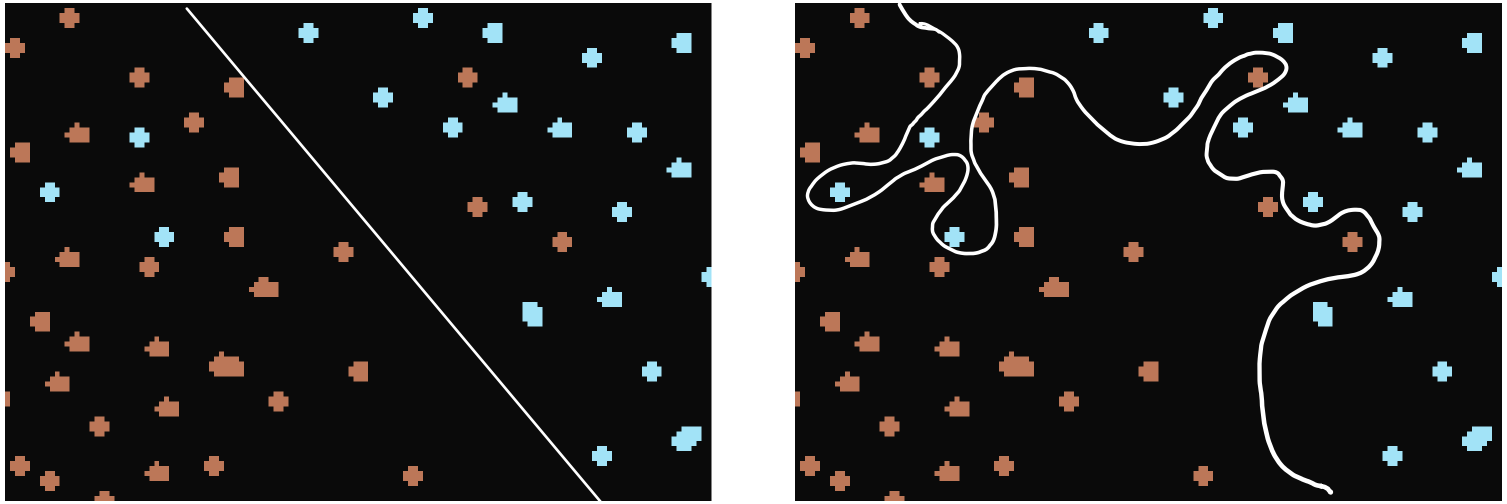


REGRESSION EXAMPLE: OVERFITTING

- Regression is:
  - Predict a number
  - infinitely many possible outputs

Will give good results through training phase and very poor results in testing phase (poor **generalization**.)

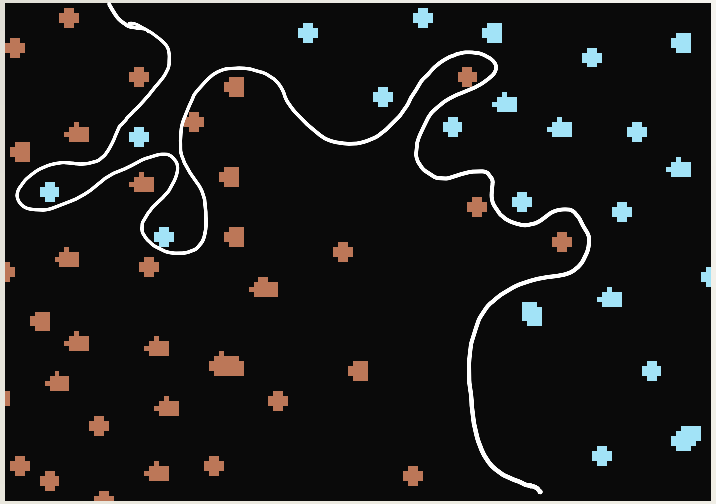


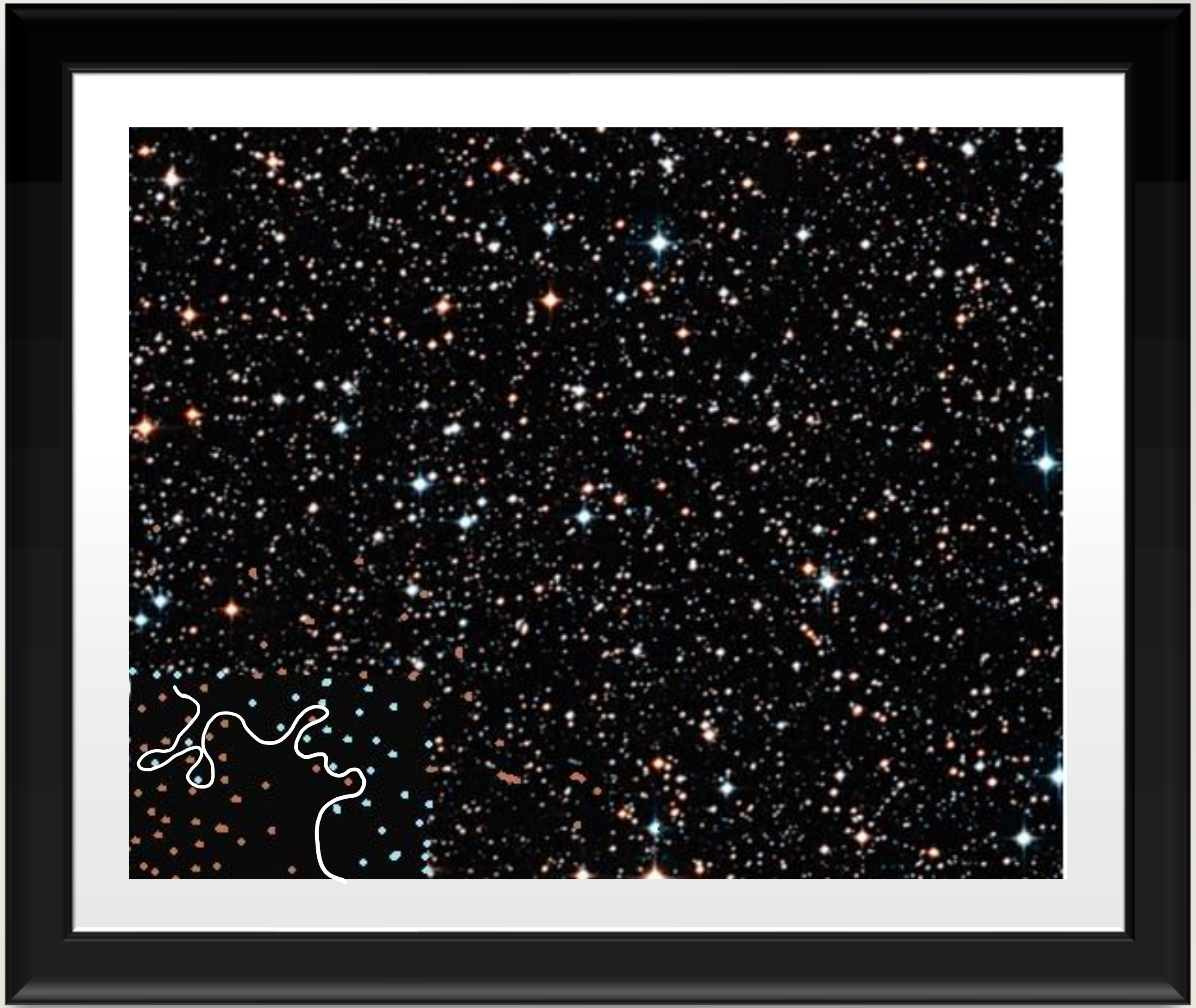


OVERFITTING CONCEPT (GIVEN DATASET)

22

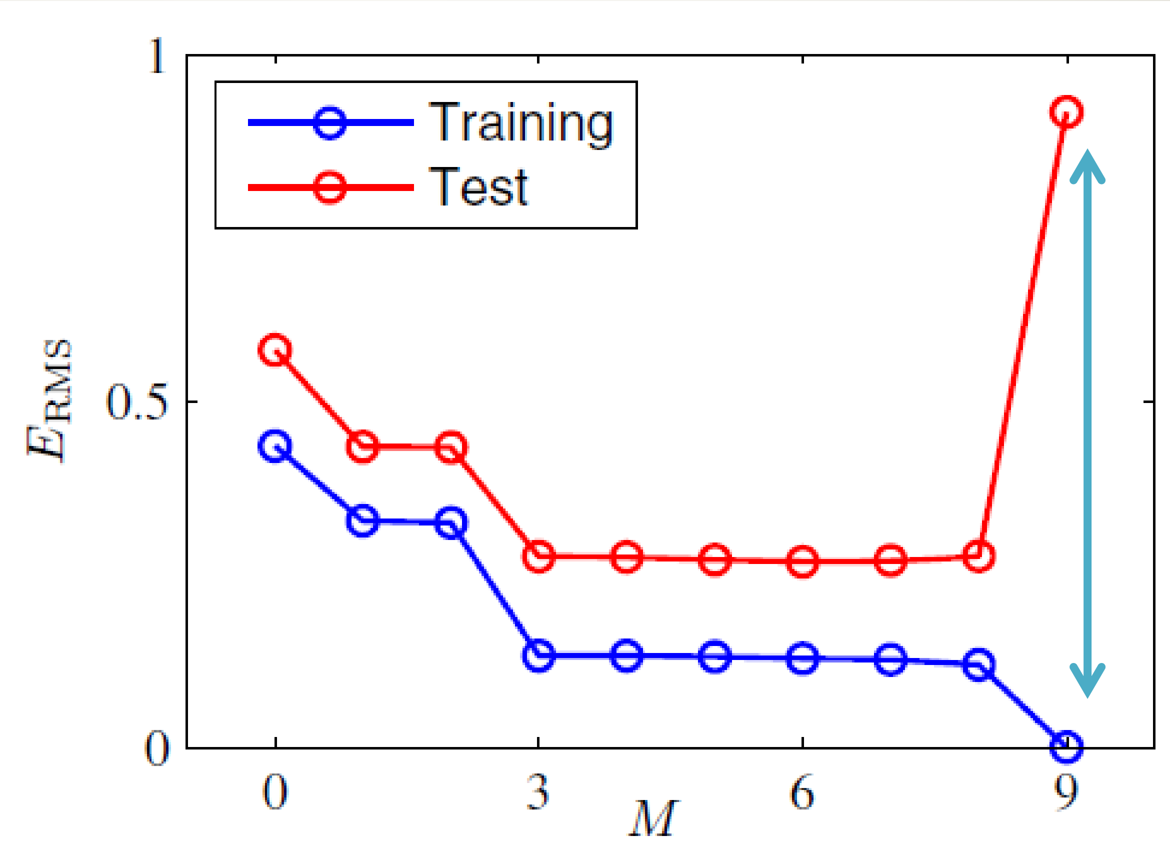
OVERFITTING (WHOLE IMAGE)





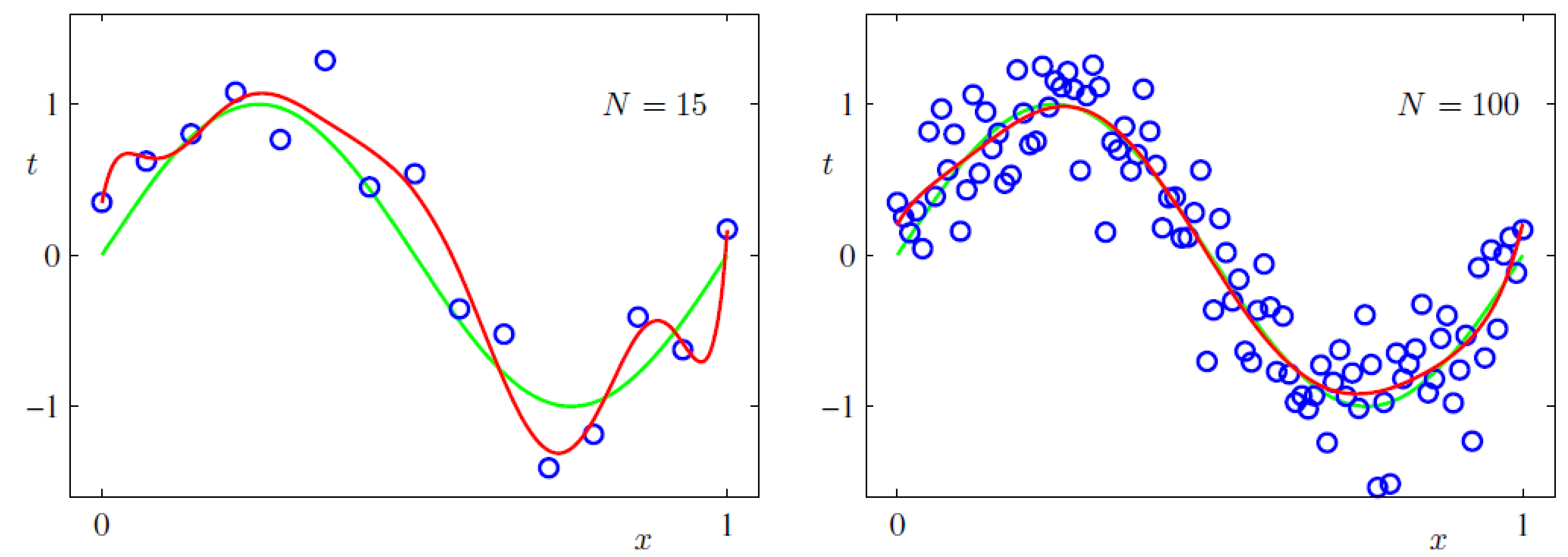
23 OVERFITTING

For M = 9, the training set error goes to zero, However, the test set error has become very large



Over fitting

GENERALIZATION (INCREASING DATASET)



References

Christopher M. Bishop, “Pattern Recognition and Machine Learning “, 2006