

# Introducción al Aprendizaje de Máquina

Adaptado por: Jonnatan Arias Garcia

[jonnatan.arias@utp.edu.co](mailto:jonnatan.arias@utp.edu.co) [jariasg@uniquindio.edu.co](mailto:jariasg@uniquindio.edu.co)

David Cardenas peña - [dcardenasp@utp.edu.co](mailto:dcardenasp@utp.edu.co)

Hernán Felipe Garcia - [hernanf.garcia@udea.edu.co](mailto:hernanf.garcia@udea.edu.co)

# Contenido



# I. Motivación

- Introducción
- Ejemplos de aplicación
- Big data
- DataScience

# Inicio

- Inteligencia artificial
  - Máquinas (programas) que son capaces de imitar (incluso mejorar) comportamientos humanos que etiquetaríamos como “inteligentes”
- Machine Learning (aprendizaje de maquina)
  - Una de las técnicas de la inteligencia artificial
  - La máquina “aprende” a conseguir sus resultados
    - En base a los datos
    - Sin que se expliciten las reglas que debe seguir

# ARTIFICIAL INTELLIGENCE

Early artificial intelligence stirs excitement.



# MACHINE LEARNING

Machine learning begins to flourish.



# DEEP LEARNING

Deep learning breakthroughs drive AI boom.



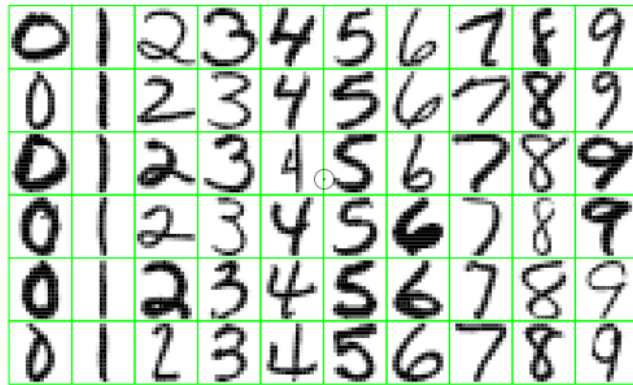
Since an early flush of optimism in the 1950s, smaller subsets of artificial intelligence – first machine learning, then deep learning, a subset of machine learning – have created ever larger disruptions.

# Introducción

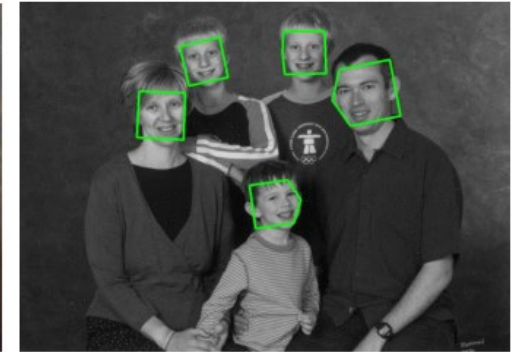
- El curso es sobre aprendizaje a partir de datos
- Se estudian diferentes algoritmos que permiten resolver diferentes problemas de predicción
- El algoritmo se diseña con base a un modelo matemático o función y a la base de datos.
- Se desea extraer conocimiento a partir de datos

# Ejemplo predicción

Reconocimiento de dígitos manuscritos



Detección y reconocimiento de rostros.



Mercado de acciones

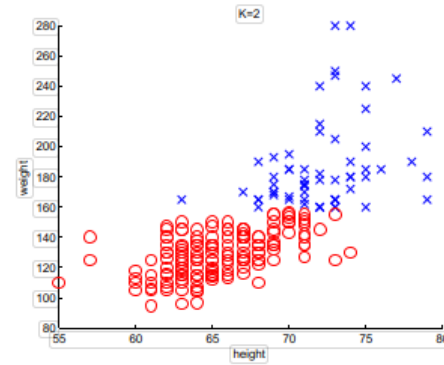
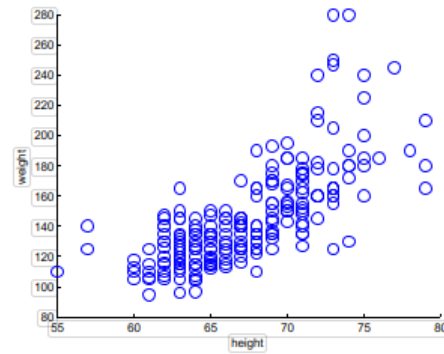


Predecir la edad de una persona que mira un video particular en YouTube.



# Ejemplo predicción

## Agrupamiento



## Sistemas de recomendación

Customers Who Bought This Item Also Bought

Machine Learning: A Probabilistic...  
by Kevin P. Murphy  
★★★★☆ 35  
Hardcover  
\$81.71 ✓Prime

The Elements of...  
by Trevor Hastie  
★★★★☆ 40  
#1 Best Seller  
Bioinformatics  
Hardcover  
\$84.04 ✓Prime

Probabilistic Graphical Models: Principles and...  
by Daphne Koller  
★★★★☆ 26  
Hardcover  
\$99.75 ✓Prime

Machine Learning with R  
by Brett Lantz  
★★★★☆ 26  
Paperback  
\$49.49 ✓Prime

An Introduction to...  
by Gareth James  
★★★★☆ 37  
#1 Best Seller  
Mathematical & Statistical...  
Hardcover  
\$75.99 ✓Prime

Reinforcement Learning: An Introduction...  
by Richard S. Sutton  
★★★★☆ 17  
Hardcover  
\$64.60 ✓Prime





# Big data

## 1 minuto en internet

- Instagram: se comparten aproximadamente 65.972 fotos y videos, y casi 140.000 usuarios visitan una página de negocio
- TikTok, se ven 625 millones de videos.
- Google procesa más de 6 millones de búsquedas
- Netflix y Hulu, se ven 452.000 horas de contenido cada minuto.
- más de 16 millones de textos enviados cada minuto y 6.000 millones de correos electrónicos enviados y recibidos diariamente.
- Amazon, se gastan 4.722 euros cada segundo



“From the dawn of Civilization until 2003, human kind generated five exabytes (10<sup>18</sup>) bytes of data. Now we produce five exabytes every two days...and the pace is accelerating.” —Eric Schmidt, Executive Chairman, Google.

# Data Science

“Information is the oil of the 21 st century, and analytics is the combustión engine.” — Peter Sondergaard, Gartner Research.

## II. Inicio

- Definiciones
- Tipos de aprendizaje
- Aplicaciones para cada tipo de aprendizaje

# Definiciones básicas

- ❑ **Conjunto de entrenamiento** ( $\{\mathbf{X}, \mathbf{t}\}$ ): conjunto de  $N$  dígitos  $\{\mathbf{x}_1, \dots, \mathbf{x}_N\}$  con sus correspondientes etiquetas  $\{t_1, \dots, t_N\}$ , empleados para sintonizar los parámetros de un modelo de predicción.
- ❑ **Entrenamiento**: el resultado de ejecutar el algoritmo de aprendizaje de máquina puede ser expresado como una función  $y(\mathbf{x}, \theta)$  que toma un dígito  $\mathbf{x}$  y entrega una salida codificada de la misma manera que  $\mathbf{t}$ .
- ❑ **Validación**: probar la función obtenida con un conjunto diferente de dígitos a los utilizados en el entrenamiento (conjunto de validación).
- ❑ **Generalización**: habilidad de clasificar correctamente dígitos de validación.

# Aprendizaje supervisado y no supervisado

- **Aprendizaje supervisado:** se conocen los vectores de características ( $\mathbf{X}$ ), y sus correspondientes etiquetas de salida ( $\mathbf{t}$ ).
  - La variable de salida es discreta: *clasificación*.
  - La variable de salida es continua: *regresión*.
  
- **Aprendizaje no supervisado:** sólo se conocen los vectores de características ( $\mathbf{X}$ )
  - Descubrir grupos de datos similares: *agrupamiento*.
  - Determinar la distribución de los datos: *estimación de densidad*.
  - Proyectar los datos a un espacio de menor dimensión: *reducción de dimensionalidad*.
  
- **Otro tipo de aprendizaje:** semi-supervisado, aprendizaje activo, aprendizaje de múltiples etiquetas, aprendizaje por refuerzo.

# Aprendizaje Supervisado

- Reconocimiento de dígitos (Clasificación)
- Detección y reconocimiento de rostros (Clasificación)
- Predecir edad en YouTube (Regresión)
- Acciones en bolsa de valores (Regresión)

# Aprendizaje No Supervisado

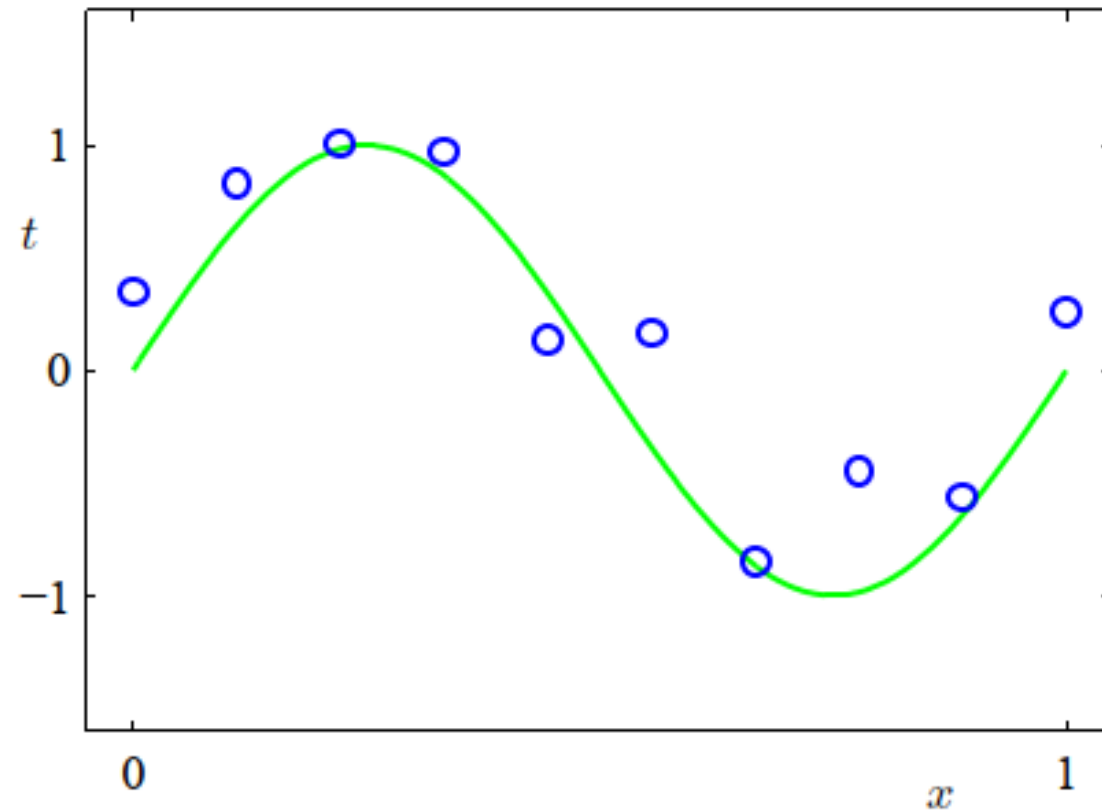
- Llenado de matrices y datos faltantes (Clúster o agrupamiento)
- Clasificar usuarios de e-commerce (Clúster o agrupamiento)
- Descubrir patrones de agrupamiento y similares
- Predecir películas según otras personas que vieron lo mismo

# III. Idea de la regresión

- Idea Base
- Función polinomial (modelo)
- Selección de modelo
- Validación
- Regularización



# Regresión



- ❑ **Regresión:** supongamos una función conocida  $\sin(2\pi x)$  con ruido aleatorio incluido en la variable objetivo  $\mathbf{t}$ .
- ❑ Conjunto de entrenamiento:  $\mathbf{x} \equiv \{x_1, \dots, x_N\}^\top$ ,  $\mathbf{t} \equiv \{t_1, \dots, t_N\}^\top$ .

# Regresión

- ❑ Objetivo: usar el conjunto de entrenamiento para hacer predicciones  $\hat{t}$  para algún valor nuevo de  $\hat{x}$ .
- ❑ **Dificultad:** generalizar  $\sin(2\pi x)$  a partir de un conjunto finito de datos.
- ❑ Debido al ruido hay incertidumbre acerca del verdadero valor de  $\hat{t}$ .

# Función Polinomial

- Usando como modelo de predicción una función polinomial,

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

donde  $\mathbf{w} \equiv \{w_0, w_1, \dots, w_M\}$ .

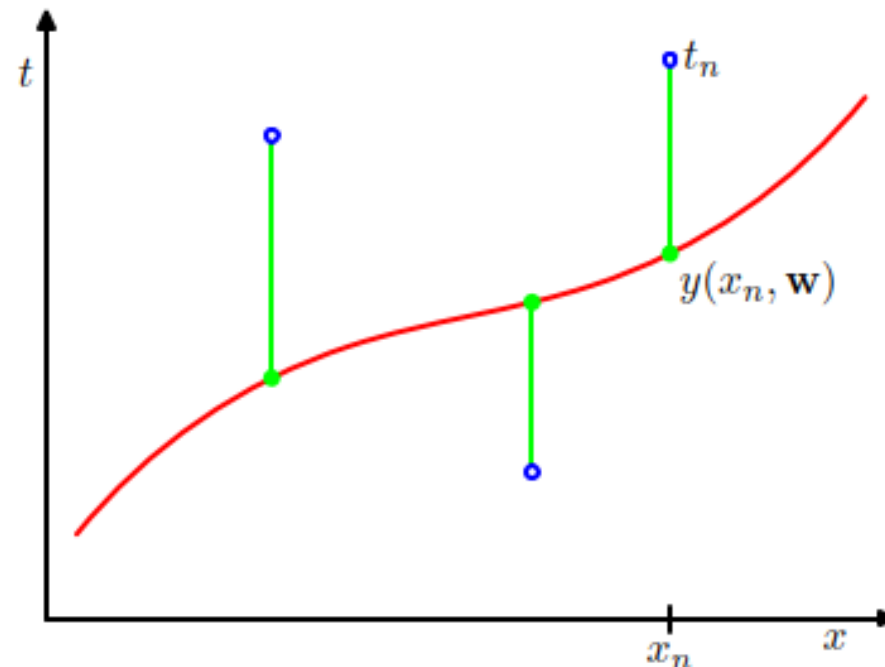
- Nótese que la función es lineal respecto a  $\mathbf{w}$ .
- El proceso de entrenamiento consiste en encontrar los coeficientes  $\mathbf{w}$ , que mejor ajusten el polinomio al conjunto de entrenamiento.

# Función Polinomial

- ❑ Esto se realiza minimizando una función de error

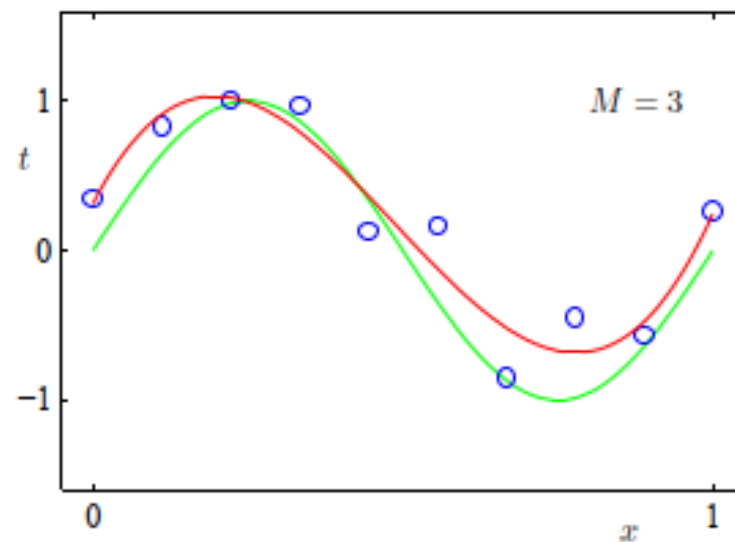
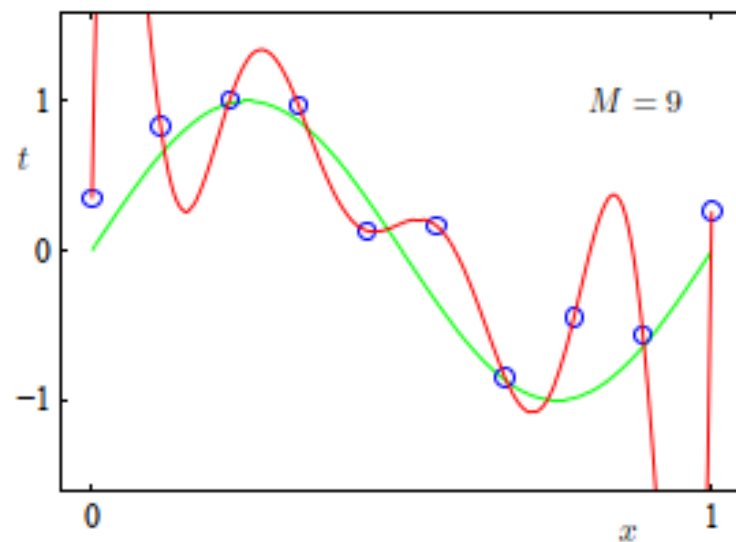
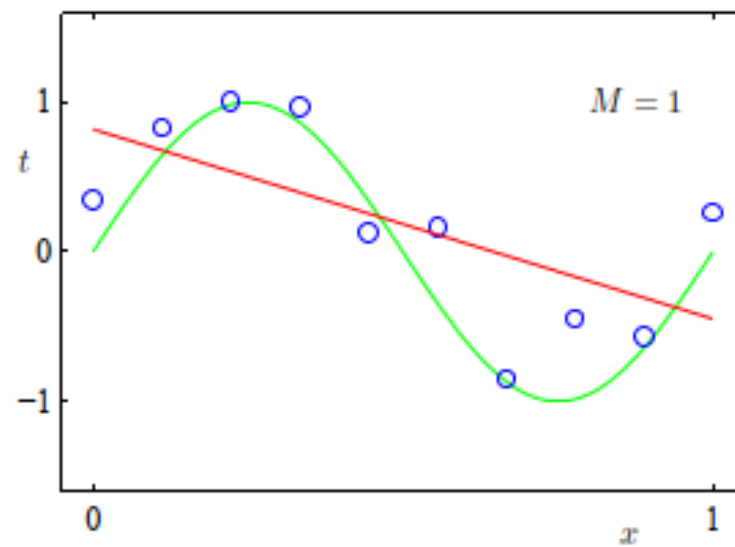
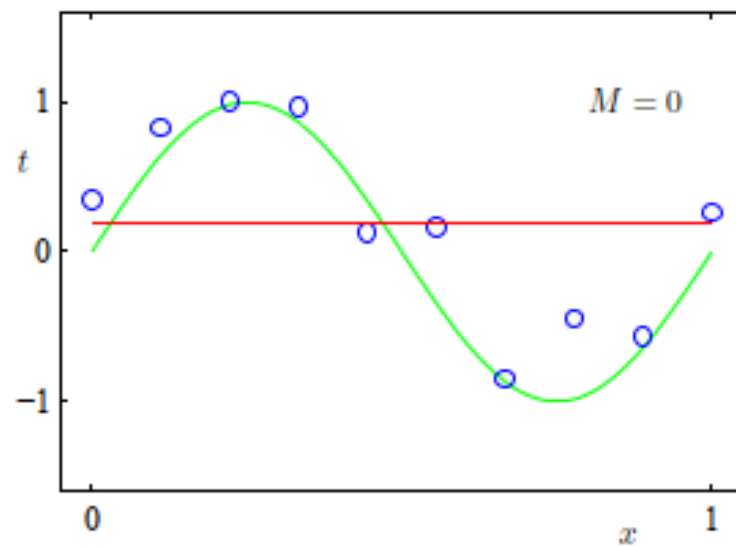
$$E(\mathbf{w}) = \frac{1}{2} \sum_{n=1}^N \{y(x_n, \mathbf{w}) - t_n\}^2.$$

- ❑ La minimización del error tiene solución única  $\mathbf{w}^*$ . El polinomio resultante estará dado por  $y(x, \mathbf{w}^*)$ .



# Selección del modelo

Una pregunta natural: cómo escoger el orden del polinomio,  $M$ ?

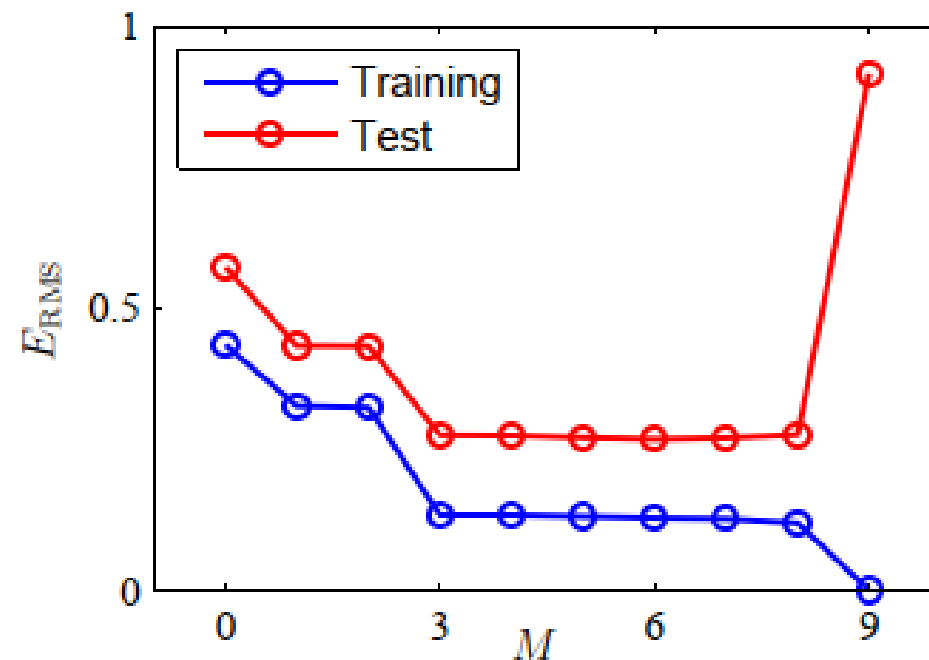


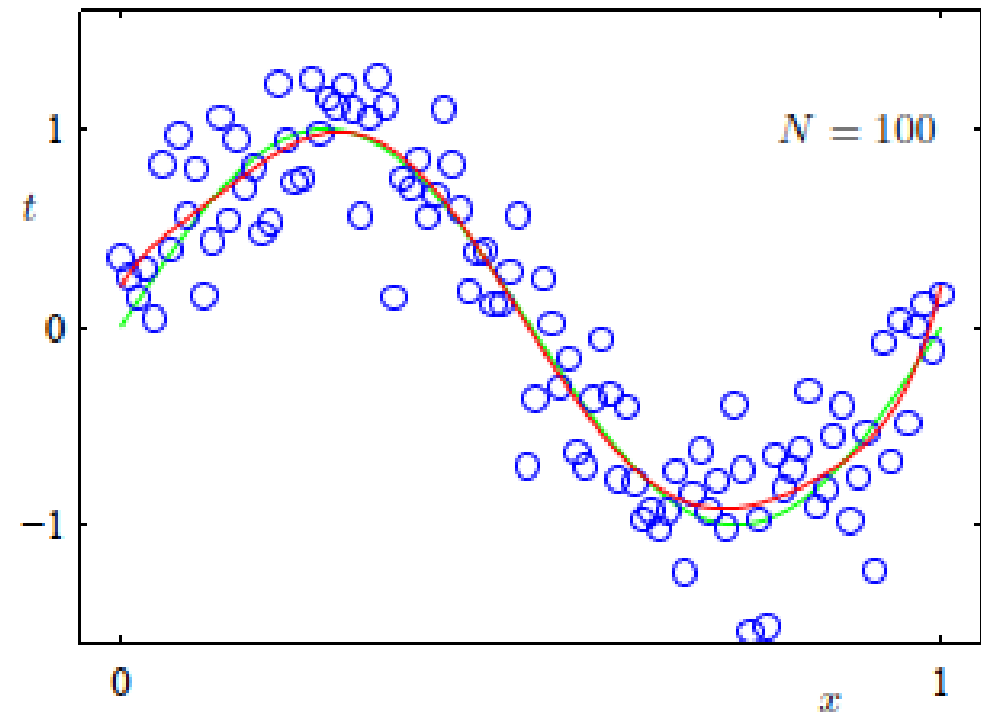
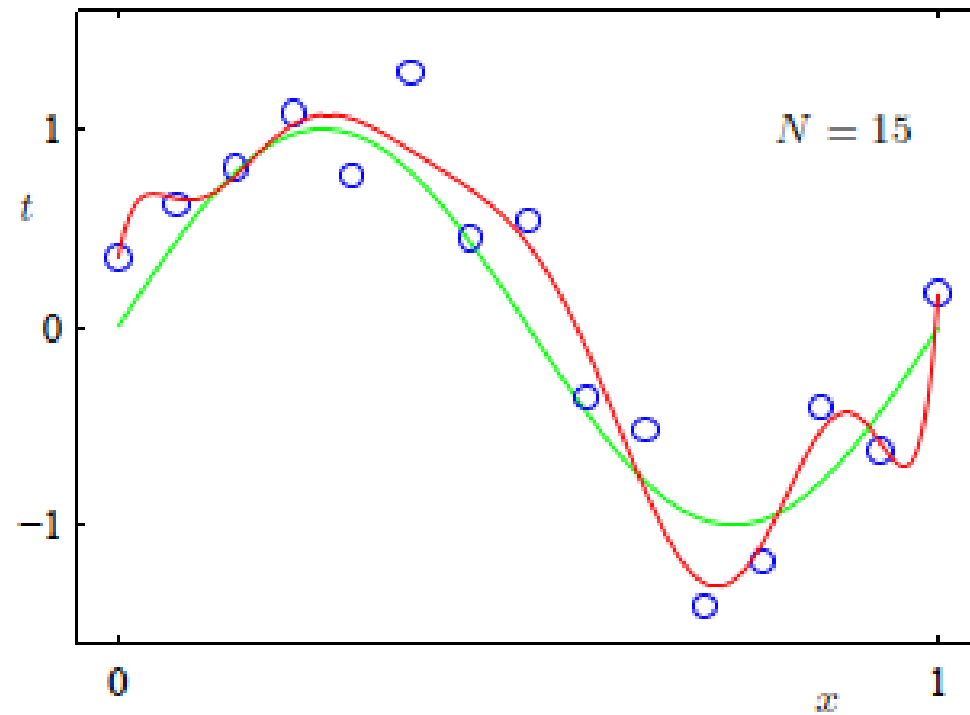
# Validación

- Para lograr una buena generalización, verificar  $E(\mathbf{w}^*)$  sobre un conjunto de validación usando

$$E_{\text{RMS}} = \sqrt{2E(\mathbf{w}^*)/N}.$$

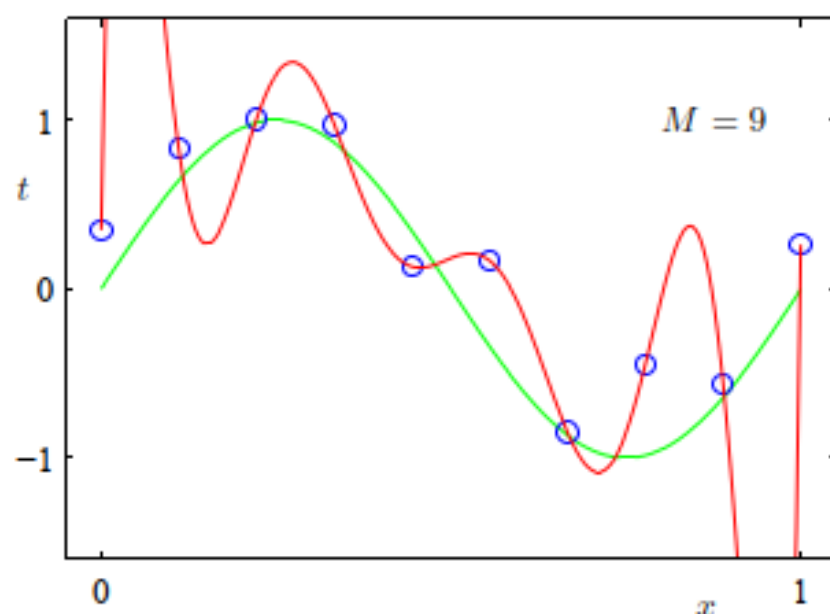
- El error RMS permite comparar errores para conjuntos de diferentes tamaños.





Entre más grande el conjunto de entrenamiento, más complejo el modelo que se puede usar.

## Regularización (I).



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

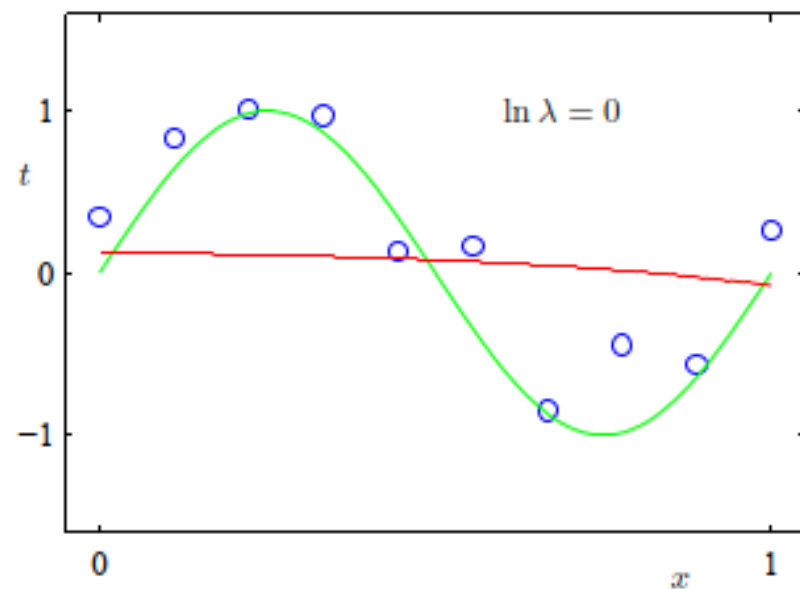
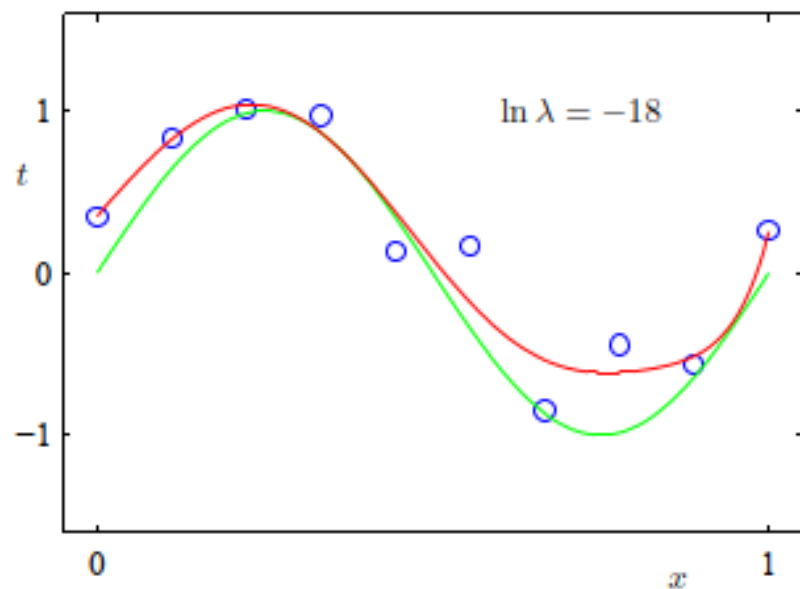
Se puede regularizar la función de error para prevenir que  $\mathbf{w}$  tome valores grandes

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

donde  $\lambda$  es el término de regularización.



## Regularización (II).



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