Regresión lineal con una variable

Representación de modelo

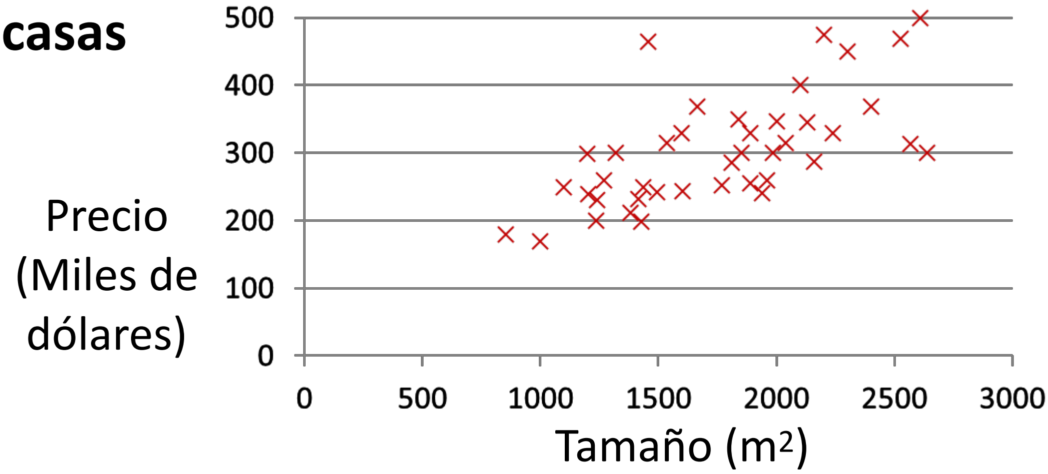
Regresión lineal

- Una pregunta importante en relación a un conjunto de datos es saber ***como las variables se relacionan***. Por ejemplo, como se relaciona la estatura y la edad, o las notas y el tiempo de estudio.

- En este case nos centraremos en la mas simple relación es decir la lineal, lo cual se representar mediante la regresión lineal.

- Esta herramienta se puede representar a nivel computacional. A pesar de su simplicidad es sorprendentemente poderosa.

Precio de casas



<u>Aprendizaje supervisado</u>

Utiliza la respuesta “correcta” para cada ejemplo en data.

<u>Problema de regresión</u>

Predice salida real.

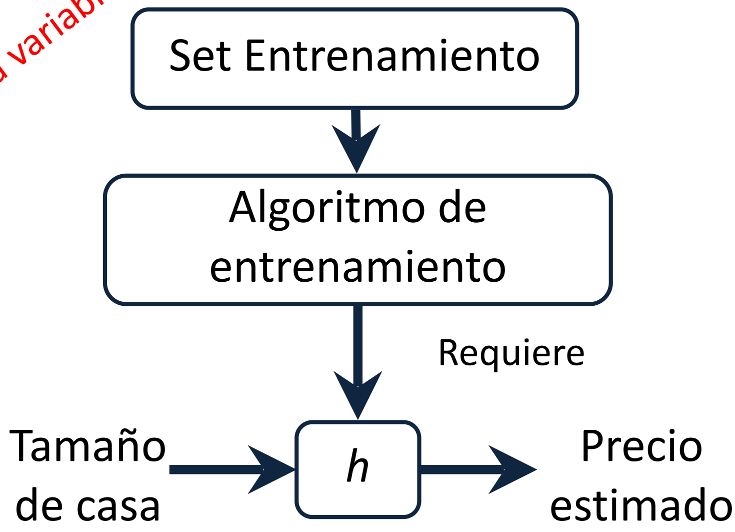
Precio de casas

Tamaño en m^2 (x)	Precio ($) en 1000's (y)
2104	460
1416	232
1534	315
852	178
…	…

Notación:

- **n** = Número de ejemplos de entrenamiento
- **x** = variable “input” / features
- **y** = variable “output” / variable “target”

Regresión lineal con una variable



¿Como representar *h* ?

Caso de Test 1

-Dada una empresa de turismo interesada en usar Machine Learning. ¿Cuáles son ejemplos de problemas de regresión *plausibles*?. Asuma que un cliente típicamente deja sus datos personales y una foto. Además que un cliente en general participa en tours de playa o montaña.

- ◯ Predecir el número de clientes que tendrá la empresa.
- ◯ Predecir si un cliente participará en un tour playero, dado su registro.
- ◯ Predecir la edad de un cliente a partir de fotos del mismo.
- ◯ Predecir cuanto gastará cliente en el siguiente mes dado su historial de consume y datos.

Regresión lineal con una variable

Función de costo

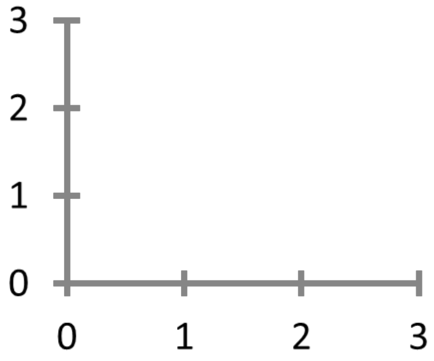
	Tamaño en metros² (x)	Precio ($) en 1000's (y)
Set de entrenamiento	2104	460
	1416	232
	1534	315
	852	178
	…	…

Hipótesis: $h_\theta(x) = \theta_0 + \theta_1 x$

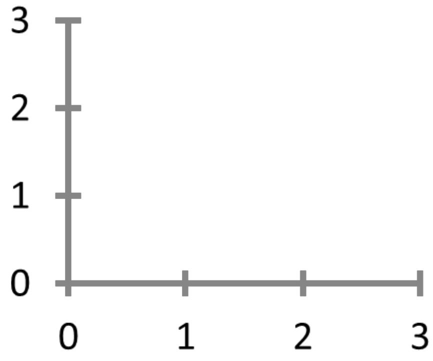
θ_i's: Parámetros

¿Como elegir θ_i's ?

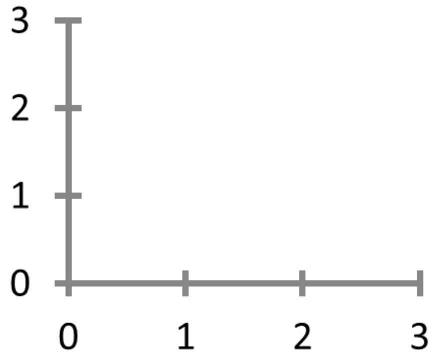
$h_\theta(x) = \theta_0 + \theta_1 x$



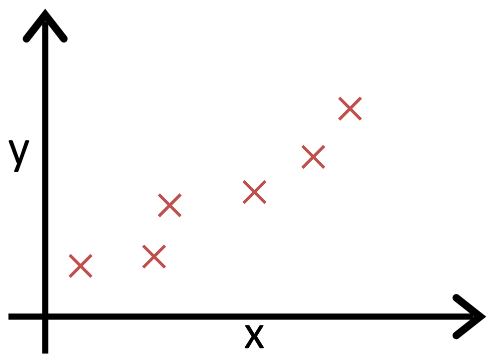
$\theta_0 = 1.5$
$\theta_1 = 0$



$\theta_0 = 0$
$\theta_1 = 0.5$



$\theta_0 = 1$
$\theta_1 = 0.5$



Idea: Escoger θ_0, θ_1 tal que $h_\theta(x)$ sea cercana a y para datos de entrenamiento (x, y)

$$\min_{\Theta_0,\Theta_1} \frac{1}{2m}\sum_{i=1}^{m}\left(h_\Theta(x^{(i)}) - y^{(i)}\right)^2$$

$$h_\Theta(x^{(i)}) = \Theta_0 + \Theta_1 x^{(i)}$$

$$J(\Theta_0,\Theta_1) = \frac{1}{2m}\sum_{i=1}^{m}\left(h_\Theta(x^{(i)}) - y^{(i)}\right)^2$$

$$\min_{\Theta_0,\Theta_1} J(\Theta_0,\Theta_1)$$

Función de costo
(error cuadrático)

Regresión lineal con una variable

Intuición de función de costo

.:.

Simplificado

Hipotesis:

$$h_\theta(x) = \theta_0 + \theta_1 x$$

$$h_\theta(x) = \theta_1 x$$

Parametros:

$$\theta_0, \theta_1$$

$$\theta_1$$

Función de costo:

$$J(\theta_0, \theta_1) = \frac{1}{2m} \sum_{i=1}^{m} \left(h_\theta(x^{(i)}) - y^{(i)}\right)^2$$

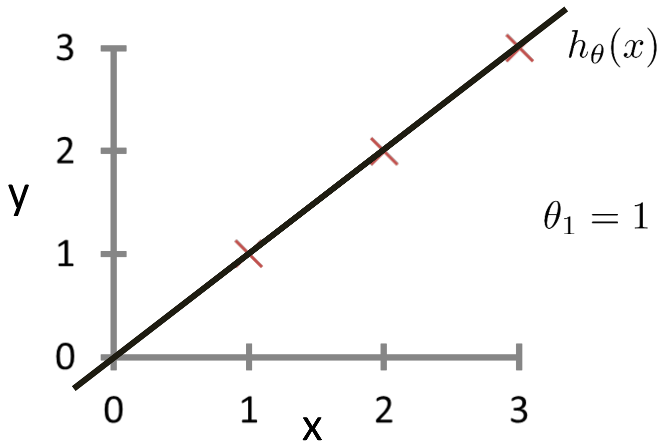
$$J(\theta_1) = \frac{1}{2m} \sum_{i=1}^{m} \left(h_\theta(x^{(i)}) - y^{(i)}\right)^2$$

Goal: $\underset{\theta_0, \theta_1}{\text{minimize}}\ J(\theta_0, \theta_1)$

$$\underset{\theta_1}{\text{minimize}}\ J(\theta_1)$$

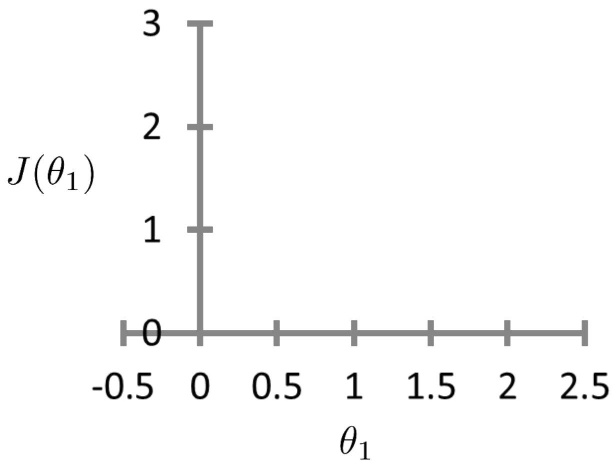
$h_\theta(x)$

(for un fijo θ_1, esta es una función de x)



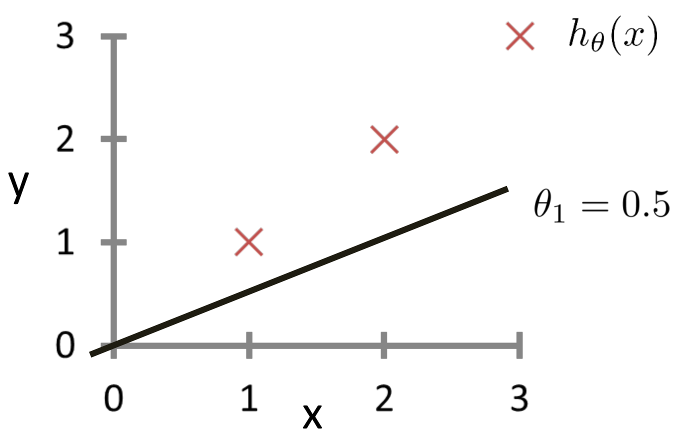
$J(\theta_1)$

(función de parametro θ_1)



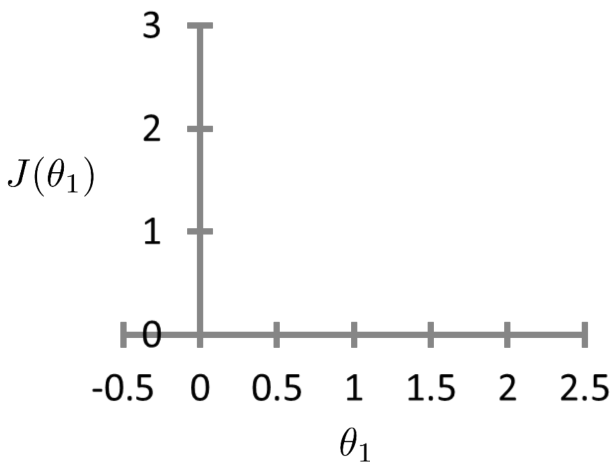
$h_\theta(x)$

(siendo fijo θ_1, esta es una función de x)



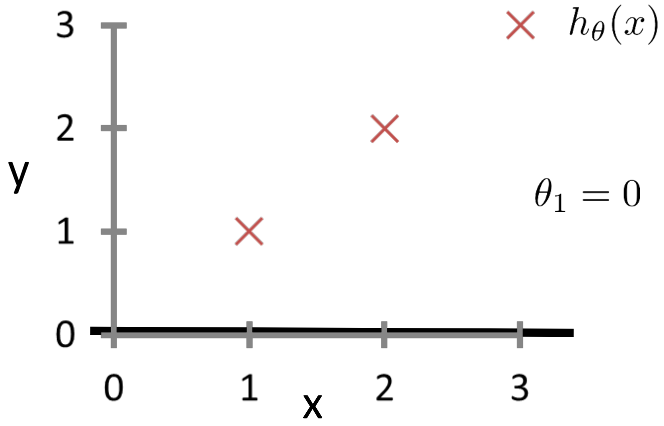
$J(\theta_1)$

(función de parametro θ_1)



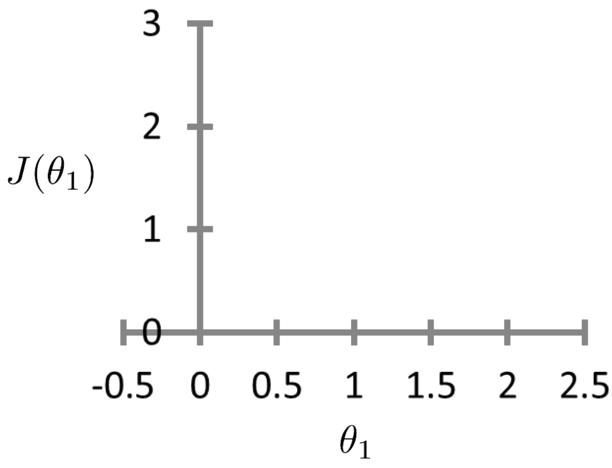
$h_\theta(x)$

(para un fijo θ_1, esta es una función de x)



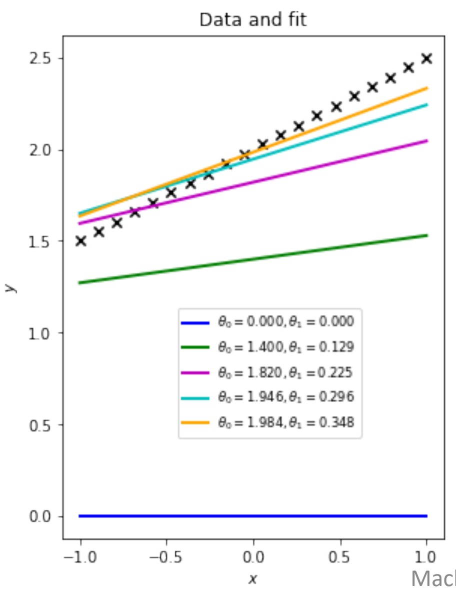
$J(\theta_1)$

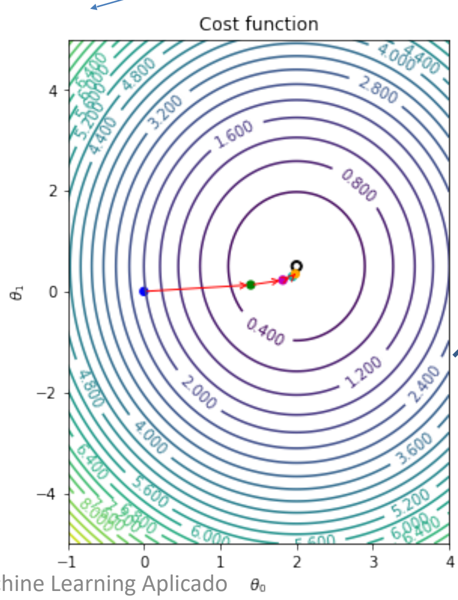
(función de parametro θ_1)

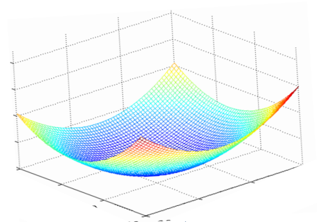


¿y en el caso original?

$$h_\theta(x) = \theta_0 + \theta_1 x$$

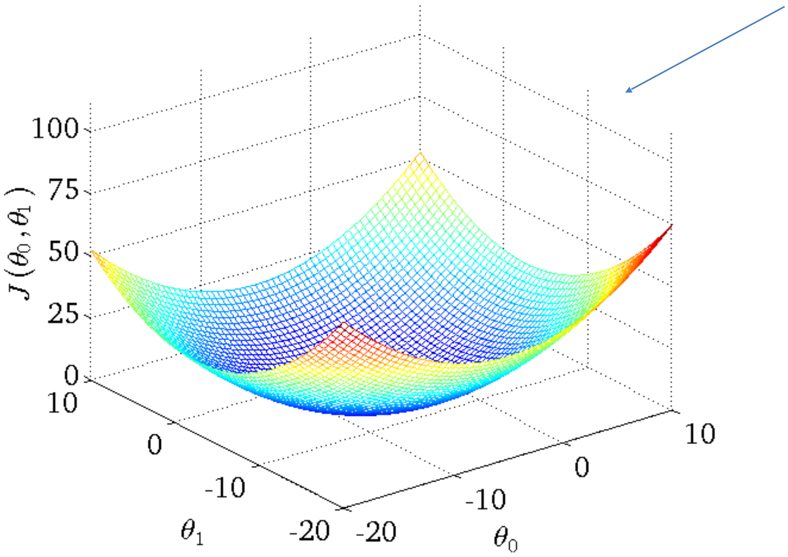






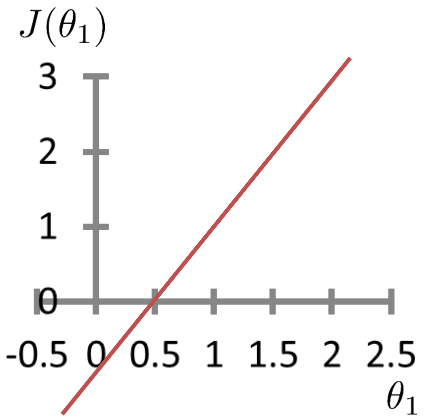
¿y en el caso original?

$$h_\theta(x) = \theta_0 + \theta_1 x$$



Caso de Test 2

-Imagine que un amigo suyo implementa la función de costo cuadrática ; es decir *J.* Pero al plotear obtiene la siguiente Figura. ¿Qué sospecha que pasó? . $J(\theta_1) = \frac{1}{2m} \sum_{i=1}^{m} \left(h_\theta(x^{(i)}) - y^{(i)}\right)^2$



- ◯ Implementó el cubo de error en vez de cuadrado.
- ◯ Implementó el valor absoluto de error.
- ◯ Implementó el error sin elevar al cuadrado.
- ◯ Ninguna de las anteriores

RegresiónLineal con una variable

Descenso de gradiente

Algoritmo de descenso de gradiente

Repetir hasta convergencia

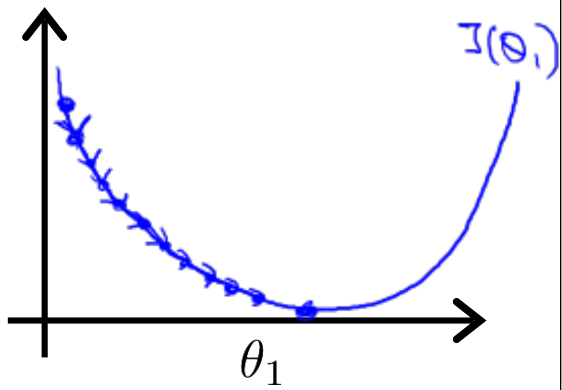
$$\theta_j := \theta_j - \alpha \frac{\partial}{\partial \theta_j} J(\theta_0, \theta_1)$$

De forma simultánea en j=0 y j=1.

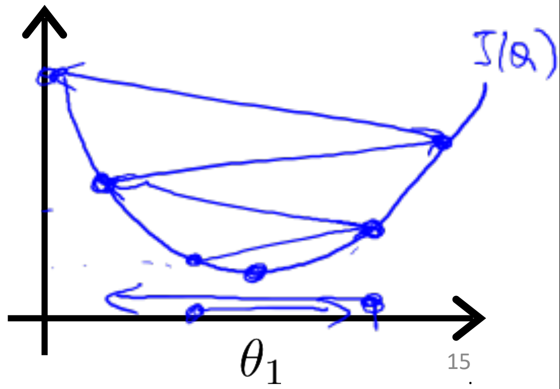
Resulta que llega a mínimo de función de costo. Ver sketch de prueba del descenso de gradiente en: http://www.stat.cmu.edu/~ryantibs/convexopt-F13/scribes/lec6.pdf

$$\theta_1 := \theta_1 - \alpha \frac{\partial}{\partial \theta_1} J(\theta_1)$$

Si α es muy pequeña, descenso de gradiente puede ser lento.



Si α es muy grande, descenso de gradiente puede traspasar el minimo. Puede fallas para converger, incluso diverger.



Caso de Test 3

-Usted al implementar el descenso de gradiente, usted se equivoca e implementa el ascenso de gradiente, es decir cambio signo de resta a suma. ¿Que cree que probablemente sucedería?.

$$\theta_1 := \theta_1 - \alpha \frac{d}{d\theta_1} J(\theta_1)$$

- ◯ El valor de alpha se haría cada vez mas grande.
- ◯ El algoritmo convergería pero a una respuesta errada.
- ◯ Los cambios de valores de parámetros serán cada vez mas grandes.
- ◯ Ninguna de las anteriores

Regresión lineal con una variable

Descenso de gradiente para Regresión lineal

Algoritmo de descenso de gradiente

Repetir hasta convergencia $\{$

$$\theta_j := \theta_j - \alpha \frac{\partial}{\partial \theta_j} J(\theta_0, \theta_1)$$

$$(\text{for } j = 1 \text{ and } j = 0)$$

$\}$

Modelo de regresión lineal

$$h_\theta(x) = \theta_0 + \theta_1 x$$

$$J(\theta_0, \theta_1) = \frac{1}{2m} \sum_{i=1}^{m} \left(h_\theta(x^{(i)}) - y^{(i)}\right)^2$$

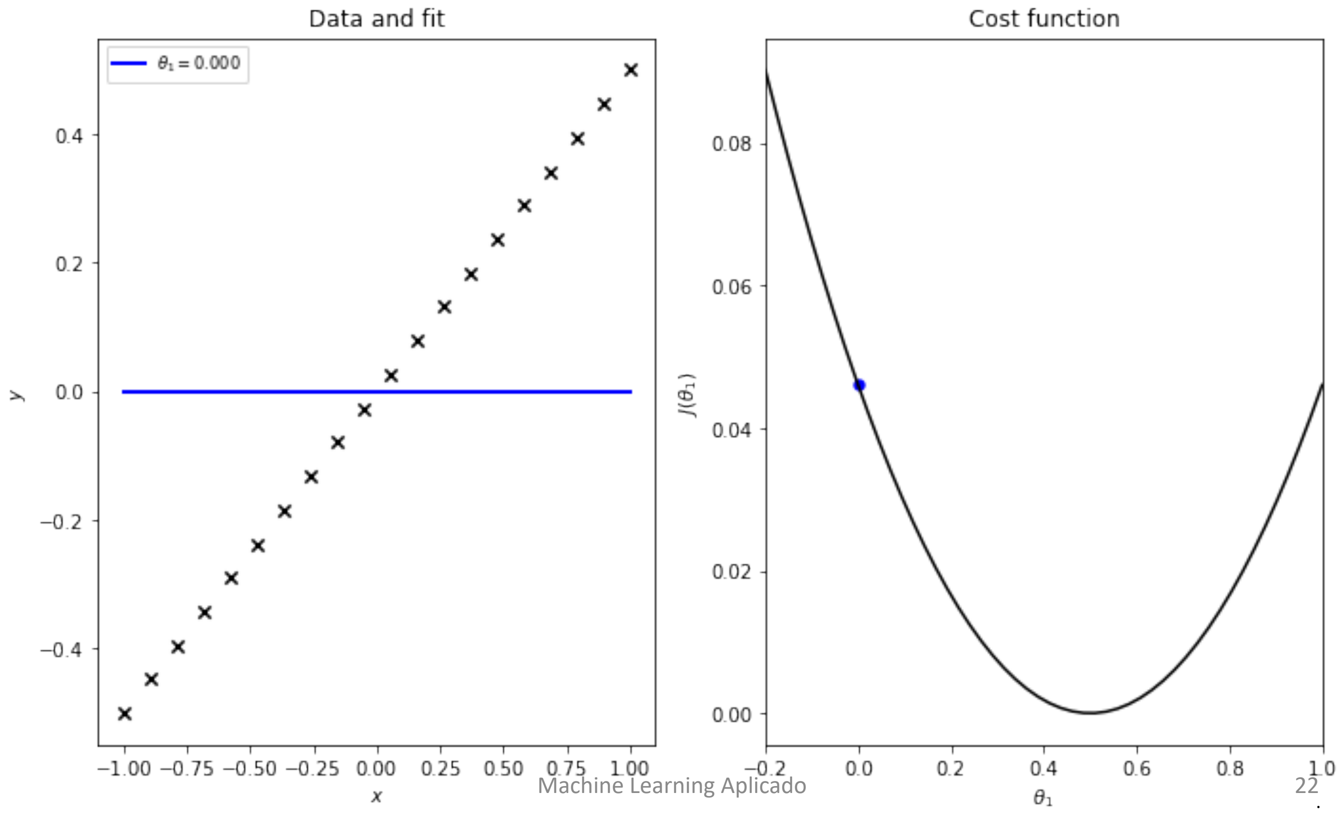
$$\frac{\partial}{\partial\theta_j} J(\theta_0, \theta_1) = \frac{\partial}{\partial\theta_j} \frac{1}{2m} \sum_{i=1}^{m} \left(h_\theta(x^{(i)}) - y^{(i)}\right)^2$$

$$= \frac{\partial}{\partial\theta_j} \frac{1}{2m} \sum_{i=1}^{m} \left(\theta_0 + \theta_1 x^{(i)} - y^{(i)}\right)^2$$

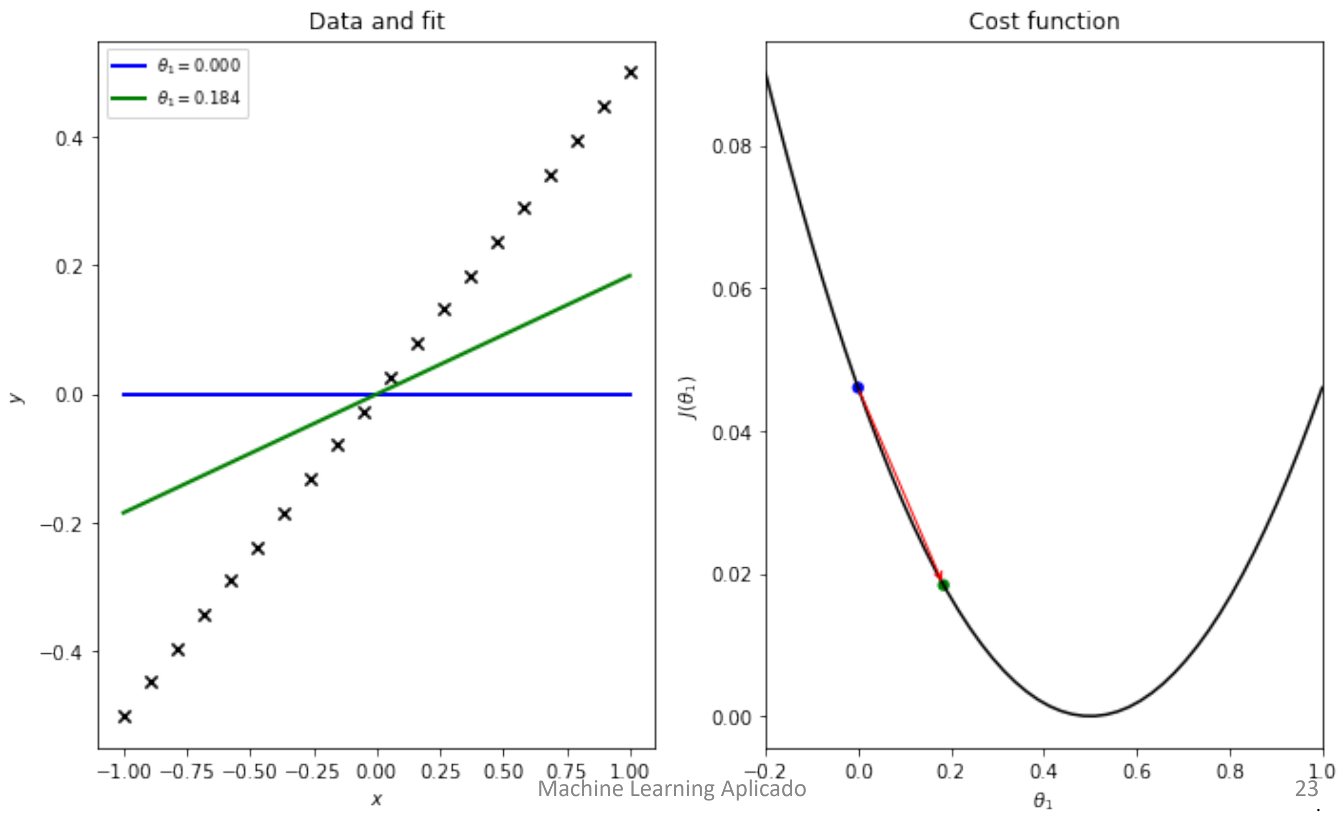
$$j = 0 : \frac{\partial}{\partial\theta_0} J(\theta_0, \theta_1) = \frac{1}{m} \sum_{i=1}^{m} \left(h_\theta(x^{(i)}) - y^{(i)}\right)$$

$$j = 1 : \frac{\partial}{\partial\theta_1} J(\theta_0, \theta_1) = \frac{1}{m} \sum_{i=1}^{m} \left(h_\theta(x^{(i)}) - y^{(i)}\right) \cdot x^{(i)}$$

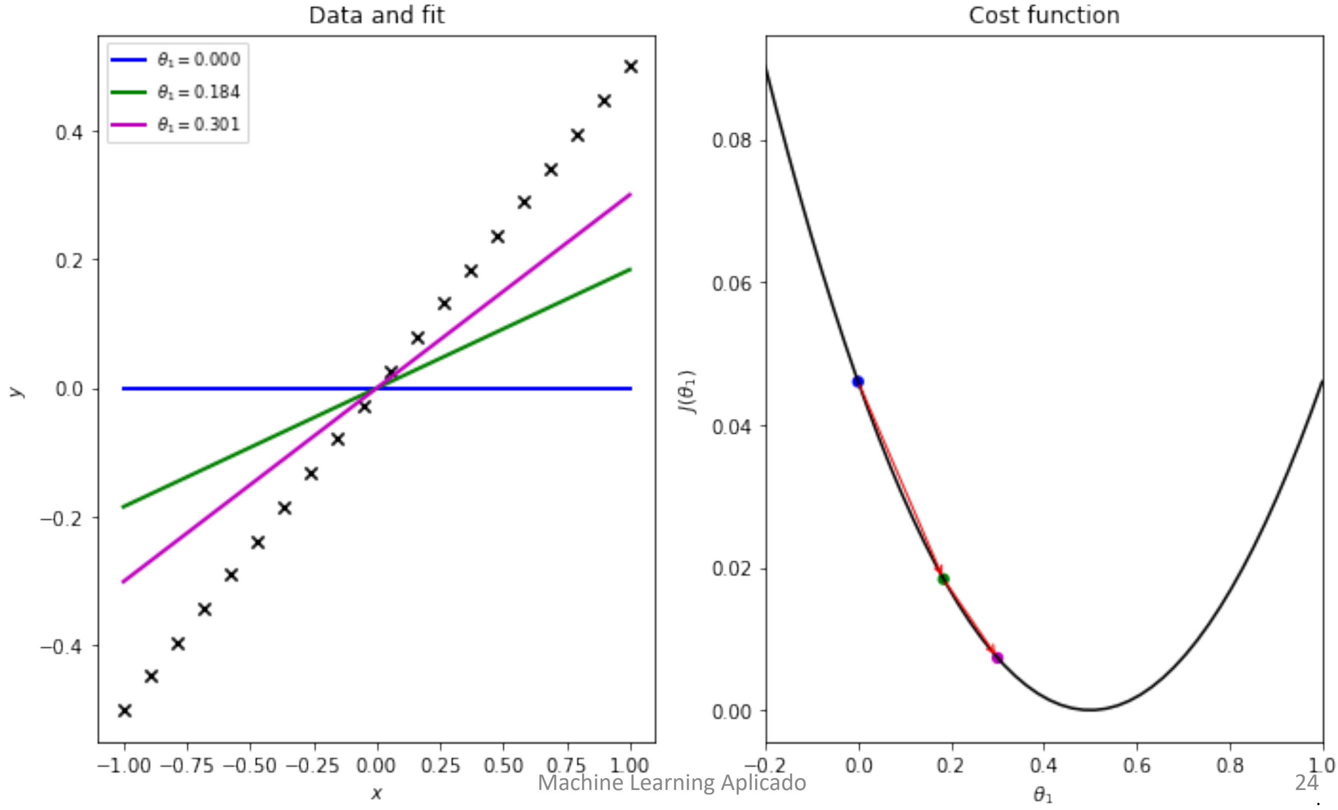
Un ejemplo numérico



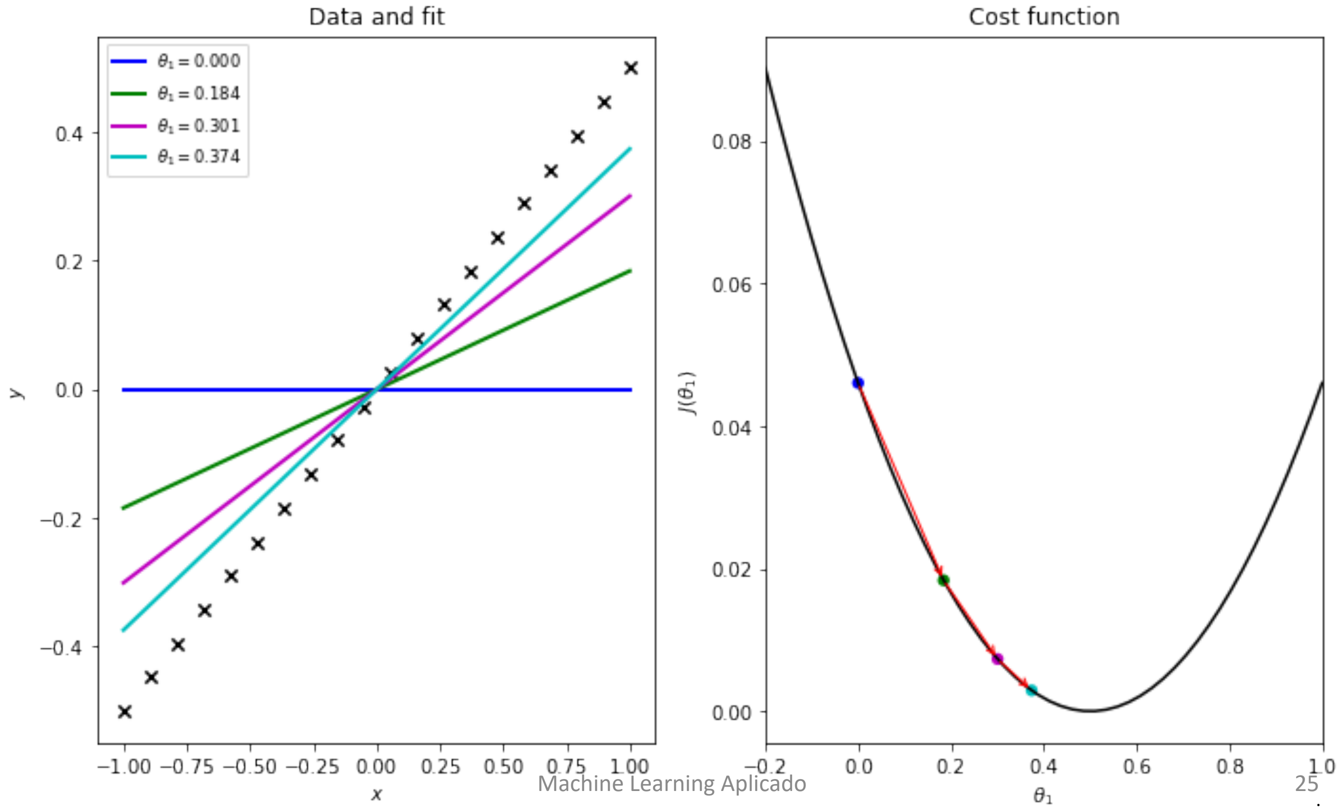
Data and fit
$\theta_1 = 0.000$
y
x
Cost function
$J(\theta_1)$
θ_1



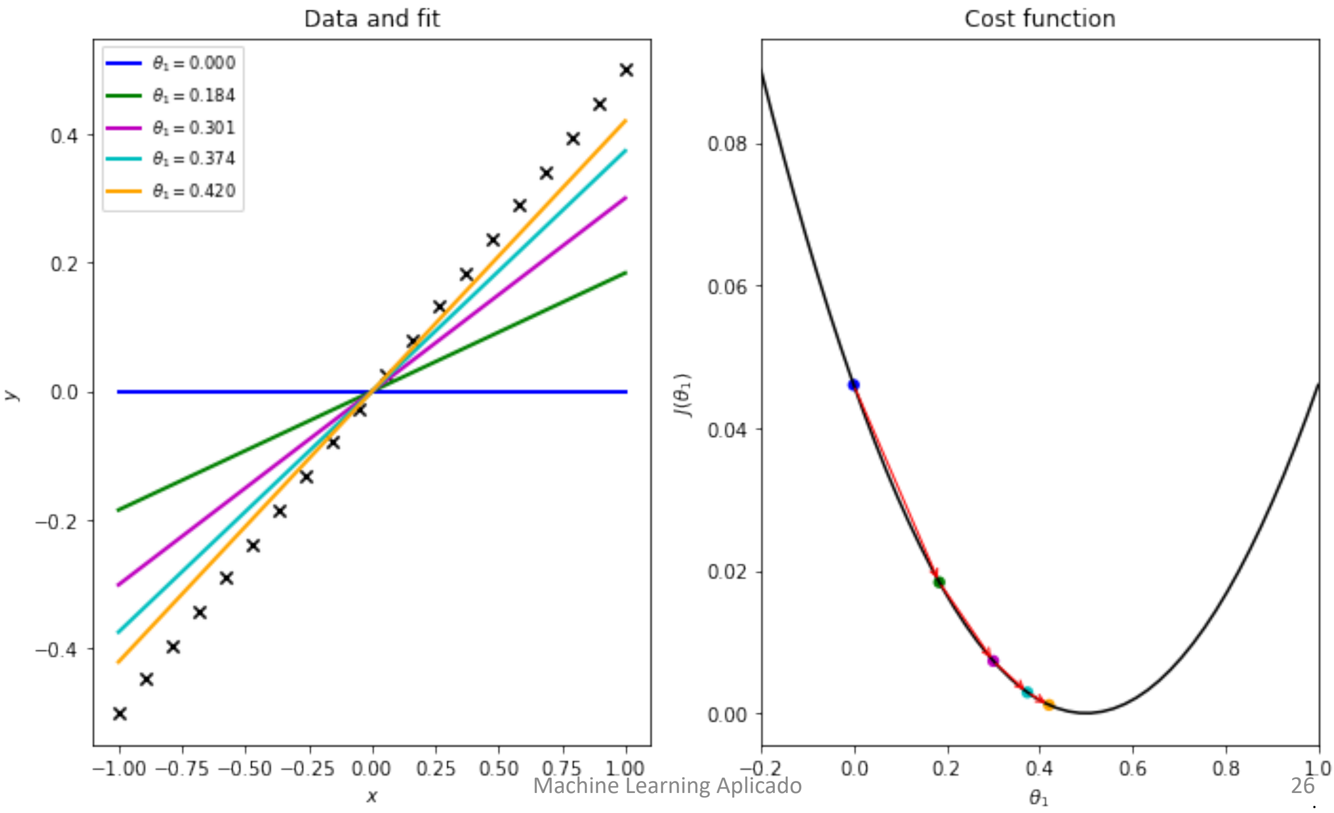
Data and fit
$\theta_1 = 0.000$
$\theta_1 = 0.184$
y
x
Cost function
$J(\theta_1)$
θ_1



Data and fit
$\theta_1 = 0.000$
$\theta_1 = 0.184$
$\theta_1 = 0.301$
y
x
Cost function
$J(\theta_1)$
θ_1



Data and fit
$\theta_1 = 0.000$
$\theta_1 = 0.184$
$\theta_1 = 0.301$
$\theta_1 = 0.374$
y
x
Cost function
$J(\theta_1)$
θ_1



¿y en el caso original?

$$h_\theta(x) = \theta_0 + \theta_1 x$$