Universidad Pública de Navarra
Nafarroako Unibertsitate Publikoa

TEMA 2: REGRESIÓN LINEAL Y LOGÍSTICA

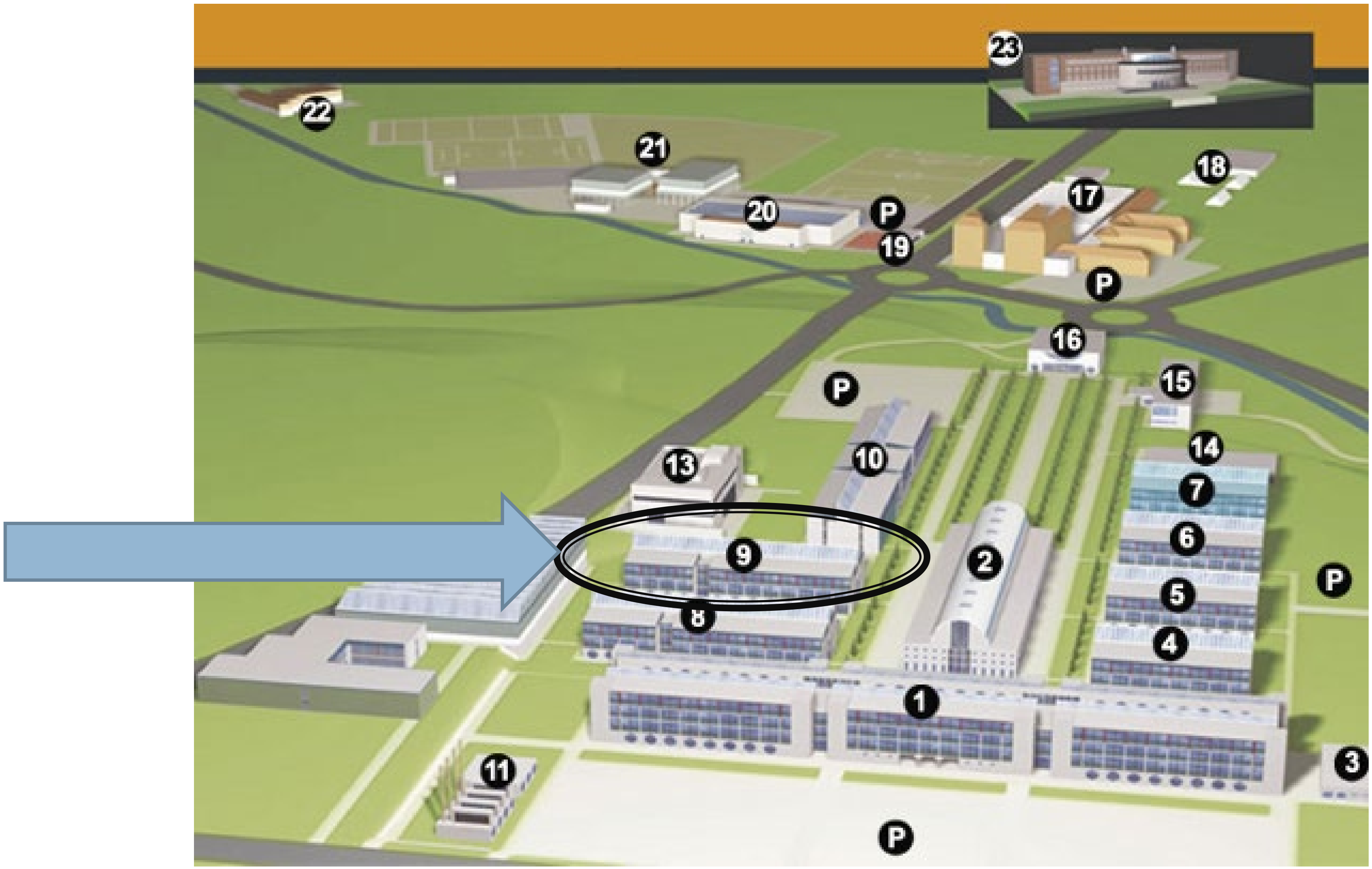
Mikel Galar Idoate
mikel.galar@unavarra.es

Ciencia de datos con técnicas inteligentes
Experto Universitario en Ciencia de Datos y Big Data

Profesor

- Mikel Galar Idoate
  - mikel.galar@unavarra.es
  - Departamento de Estadística, Informática y Matemáticas
  - Edificio de los pinos, 1ª planta, despacho 1009
  - Tutorías
    - https://www.unavarra.es/pdi?uid=8201

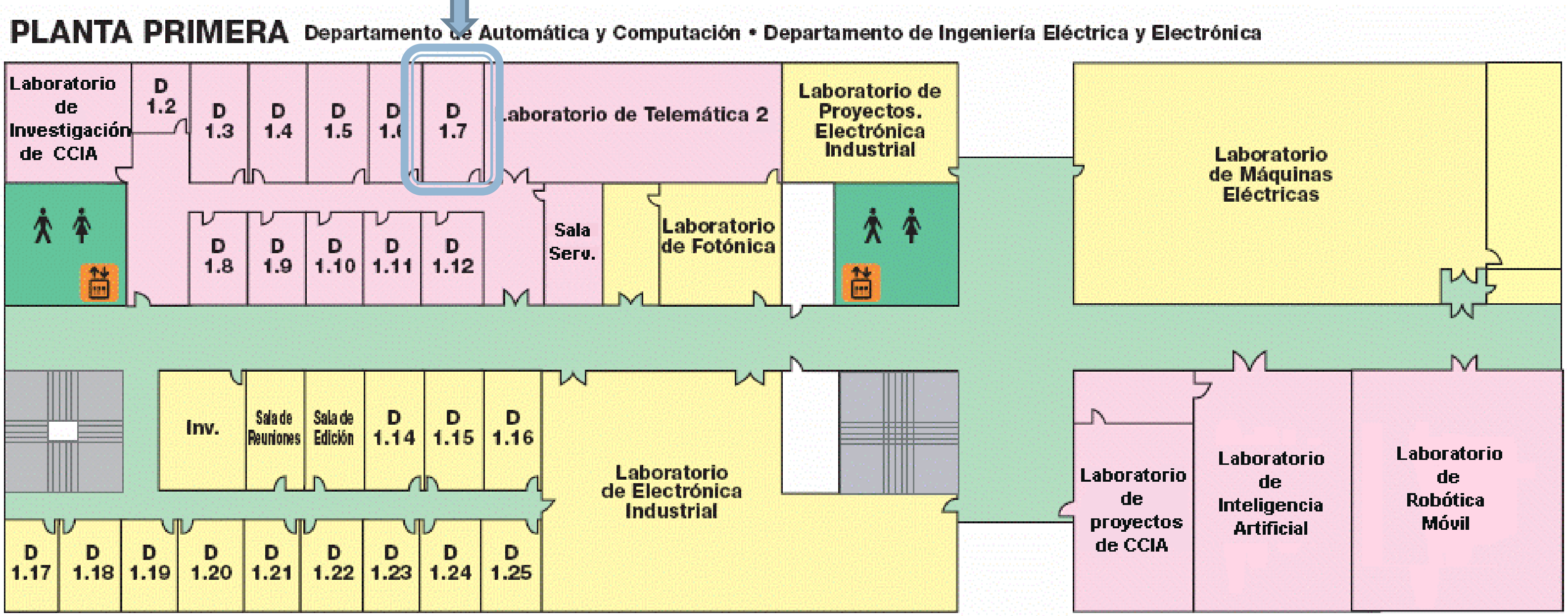
Localización



Localización

Mikel

Entrada



Índice

1. Recordatorio (regresión vs. clasificación)
2. Regresión lineal (regresión)
3. Regresión logística (clasificación)

BREVE RECORDATORIO (REGRESIÓN VS. CLASIFICACIÓN)

Universidad Pública de Navarra
Nafarroako Unibertsitate Publikoa

Mikel Galar Idoate
mikel.galar@unavarra.es

Ciencia de datos con técnicas inteligentes
Experto Universitario en Ciencia de Datos y Big Data para Business Intelligence

¿Qué es machine learning?

- Intuitivamente
  - Máquinas que **imiten a los humanos**
  - Máquinas **inteligentes**
  - Máquinas **capaces de aprender**
    - ¿De dónde? → **¡EJEMPLOS!**

¿Qué es machine learning?

- **No hay una definición establecida**
- Algunas de ellas…
  - Tom Mitchell (1998)
    - *“Problema de aprendizaje: Un programa se dice que aprende de una **experiencia E** respecto a una **tarea T** y alguna **medida de rendimiento P**, si su rendimiento en **T**, medido mediante **P**, mejora con la **experiencia E**”*
    - E = Ejemplos conocidos
    - T = Tarea a realizar
    - P = Forma de evaluar lo bien (o mal) que se ha hecho la tarea

¿Qué es machine learning?

- **No hay una definición establecida**
- Algunas de ellas…
  - Robert E. Schapire (1990)
    - *"Aprender a hacer las **cosas** mejor en el futuro en base a las **experiencias** del pasado"*
    - Cosas = tarea a realizar
    - Experiencias = ejemplos conocidos
- Por eso se suele conocer como…
  - **Aprendizaje basado en ejemplos**

¿Qué es machine learning?

- Diseñar algoritmo que aprendan **patrones** a partir de los **datos**
  - **Ejemplo:** *Dar a la máquina de un conjunto de e-mails que son spam y otros legítimos y dejar que la máquina aprenda a predecir si un **nuevo** e-mail es spam o no*
- Generalmente, aproximaciones basadas en **estadística**
  - **No** sirven **reglas manuales**
  - La máquina debe **aprender las reglas por si misma mirando los datos**
- **Lo que se busca: Generalización**
  - Evitar el **sobre-entrenamiento**
  - La máquina debe generalizar bien a **nuevos datos**
    - E-mails no vistos anteriormente

Tipos de ML

- **Aprendizaje**
  - **Supervisado**
    - Clasificación
    - Regresión
  - **No supervisado**
    - Clustering
  - Semi-supervisado
  - Por refuerzo
  - Activo
  - Transferencia del aprendizaje
  - De similitudes
  - …

En este tema

Tipos de ML

Clasificación

Ejemplos de ML. Clasificación

Filtrado de Spam

Datos

Predicción

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, Michel — 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.
You can contact the instructors by emailing: 10615-instructors@cs.cmu.edu

Natural _LoseWeight SuperFood Endorsed by Oprah Winfrey, Free Trial 1 bottle, pay only $5.95 for shipping mfw rlk Spam | X

Jaquelyn Halley to nherrlein, bcc: thehorney, bcc: ang — show details 9:52 PM (1 hour ago) — Reply

=== Natural WeightL0SS Solution ===

Vital Acai is a natural WeightL0SS 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 WeightL0SS
* Increased metabolism - BurnFat & calories easily!
* Better Mood and Attitude
* More Self Confidence
* Cleanse and Detoxify Your Body
* Much More Energy
* BetterSexLife
* A Natural Colon Cleanse

Spam

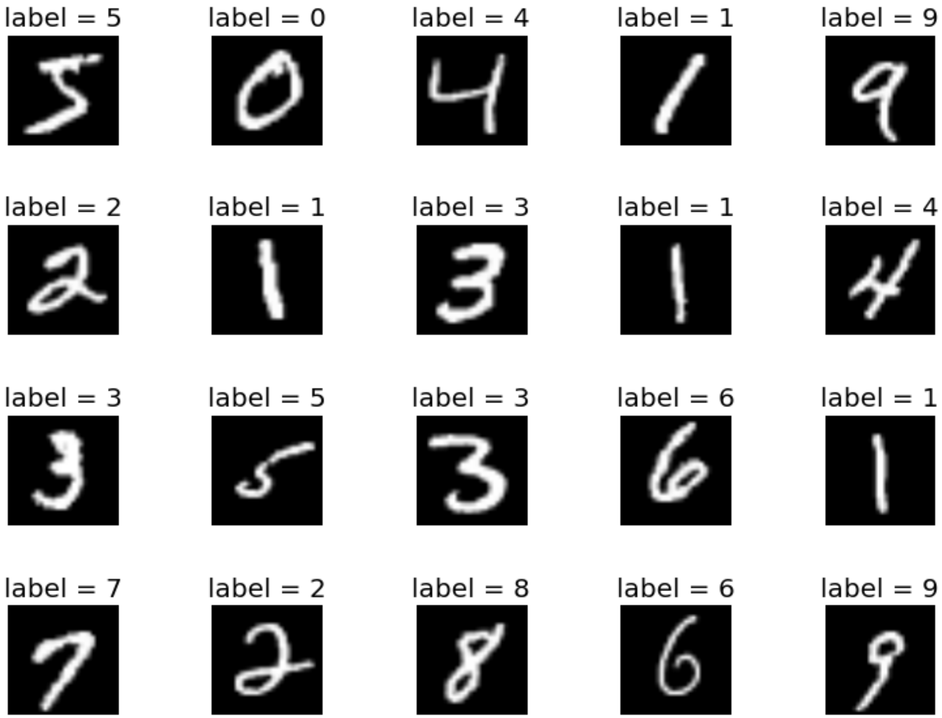
vs.

No spam

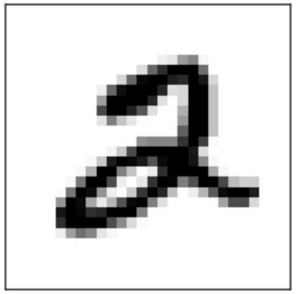
Ejemplos de ML. Clasificación

- **Reconocimiento de dígitos**

Datos de entrenamiento



Predicción (test)
¿A qué clase pertenece?



Ejemplos de ML. Clasificación

- **Detección de fraude en transacciones bancarias**

Tipos de ML

Regresión

Ejemplos de ML

- **Predicción del valor de las acciones**

Ejemplos de ML

- **Predicción de la temperatura**

Temperatura
30°

Ejemplos de ML

- **Valoración de bienes inmuebles**

480.000€

Tipos de ML

Test

¿De qué tipo de tarea hablamos?

EXPERIMENT

[redacted] Twitter sentiment analysis

This experiment demonstrates the use of the Execute R Script, Feature Selection, Feature Hashing modules to train a...

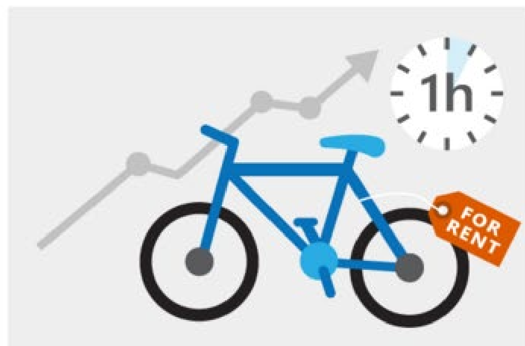
[redacted]

3409 1321 3 months ago

Microsoft

EXPERIMENT

[redacted] Demand estimation



This experiment demonstrates demand estimation using [redacted] with UCI bike rental data.

[redacted]

3808 1622 3 months ago

Microsoft

EXPERIMENT

[redacted] Credit risk prediction

This sample demonstrates how to

[redacted]

[redacted]

6400 1679 3 months ago

Microsoft

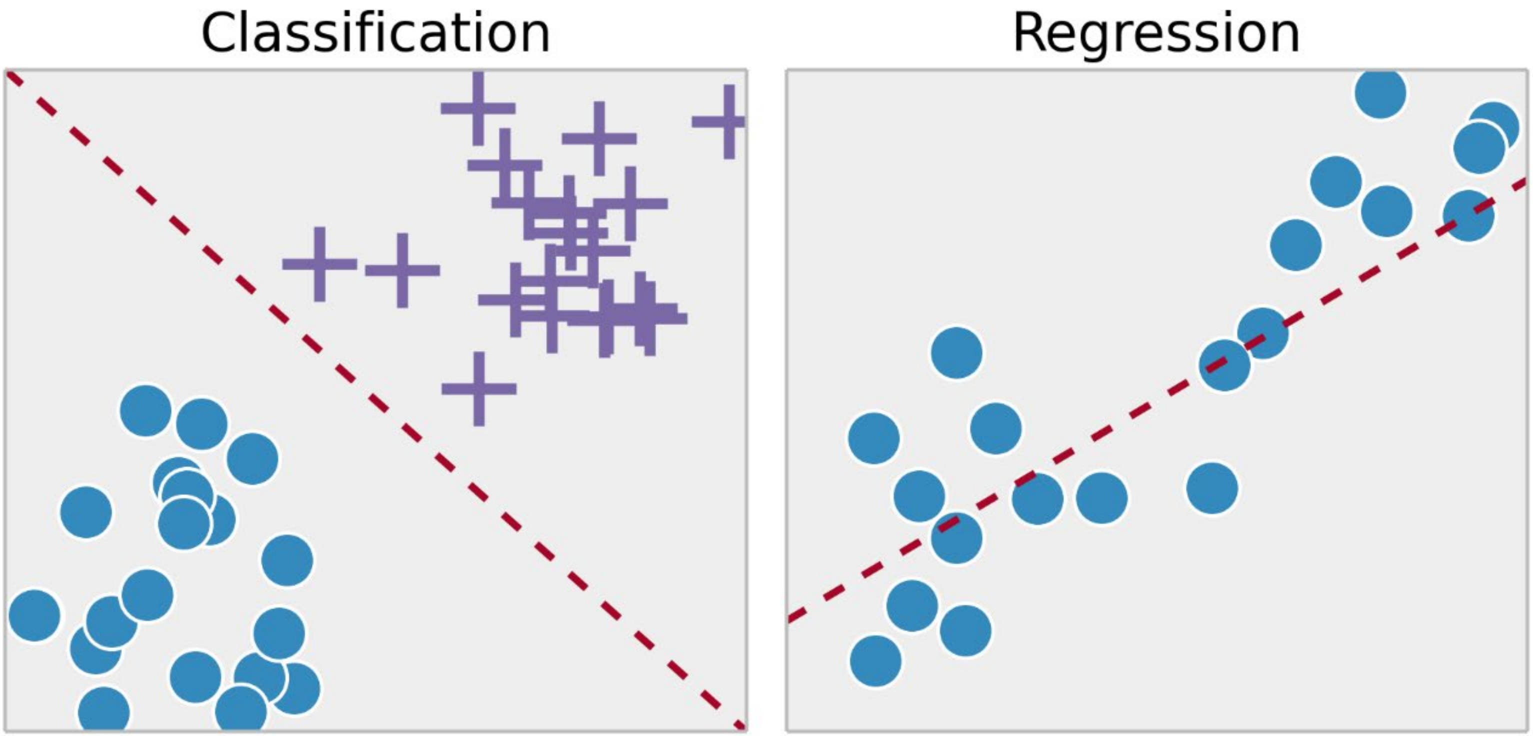
¿De qué tipo de tarea hablamos?

- **Diagnóstico médico**
  - Queremos establecer el tipo de cáncer (maligno/benigno)

¿De qué tipo de tarea hablamos?

- **Predicción de los beneficios**

Regresión vs. Clasificación



REGRESIÓN LINEAL

Mikel Galar Idoate
mikel.galar@unavarra.es

Ciencia de datos con técnicas inteligentes
Experto Universitario en Ciencia de Datos y Big Data para Business Intelligence

Índice

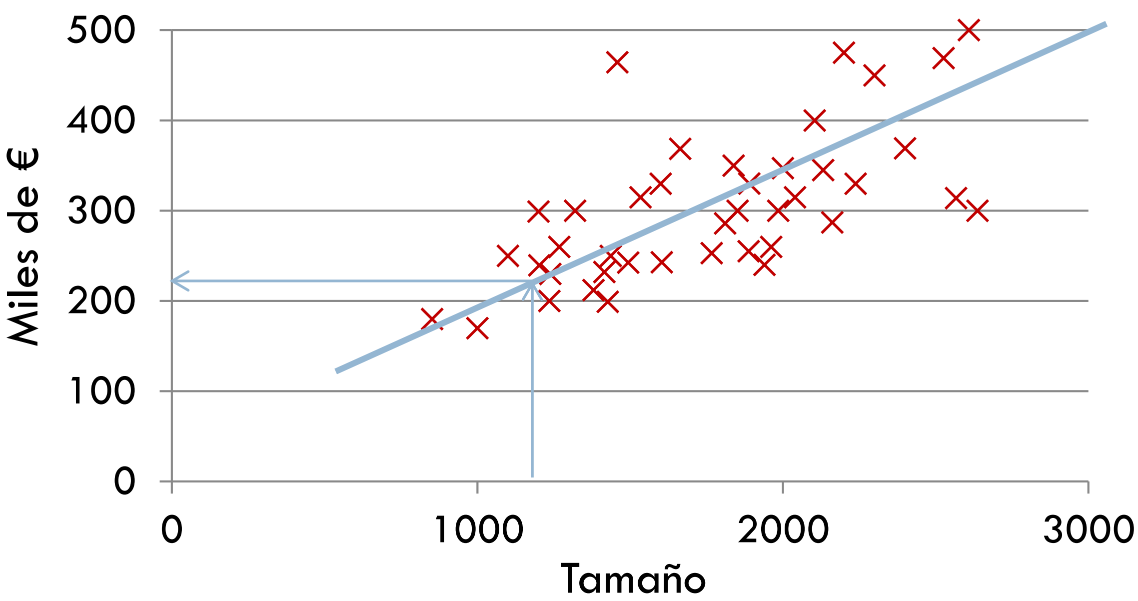
1. Regresión lineal (regresión)
   - Objetivo
   - Modelo
   - Ajuste de parámetros
     - Función de coste
     - Directo
     - Descenso por gradiente
2. Regresión logística (clasificación)

Regresión lineal con una variable

- **Objetivo**
  - Encontrar un **modelo** que represente un conjunto de datos observados
  - El modelo servirá para **predecir valores para nuevos datos**
- En **regresión**…
  - La salida es un valor numérico

Regresión lineal con una variable

- Trabajaremos con un ejemplo
  - Predicción del precio de una casa en base a su tamaño
    - Variable de **entrada:** Tamaño
    - Variable de **salida:** Precio



Regresión lineal con una variable

- **Objetivo** (formal)
  - Estimar un **modelo** en el que la variable de salida (y) se expresa linealmente en términos de las variables del modelo (x) y de un conjunto de parámetros (θ)

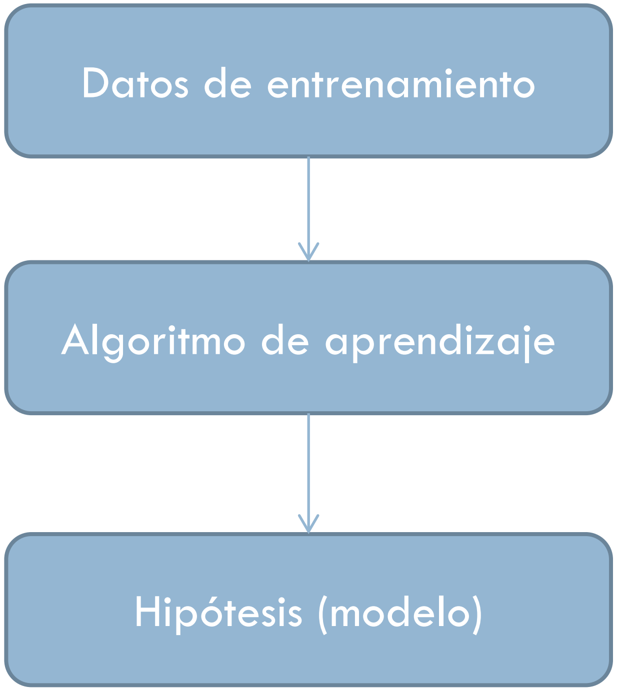
$$y = h_{\theta}(x) = \theta_0 + \theta_1 x$$

  - Es decir… un modelo en el que el precio se obtiene en términos del tamaño de la casa en base a los parámetros θ

$$precio = h_{\theta}(tamaño) = \theta_0 + \theta_1 tamaño$$

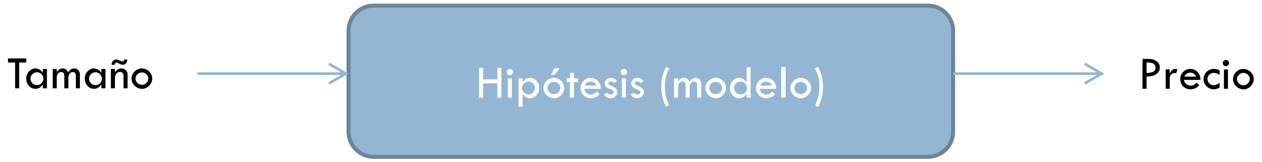
Regresión lineal con una variable

- Proceso



Regresión lineal con una variable

- La **hipótesis** obtenida es una función entre las variables de entrada (x) y la variable de salida (y)



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

Regresión lineal con una variable

- En regresión lineal con una variable, la variable de salida (y) viene dada por una única variable (x)
- La hipótesis que mapea x en y se representa como

$$y = h_{\theta}(x) = \theta_0 + \theta_1 x$$

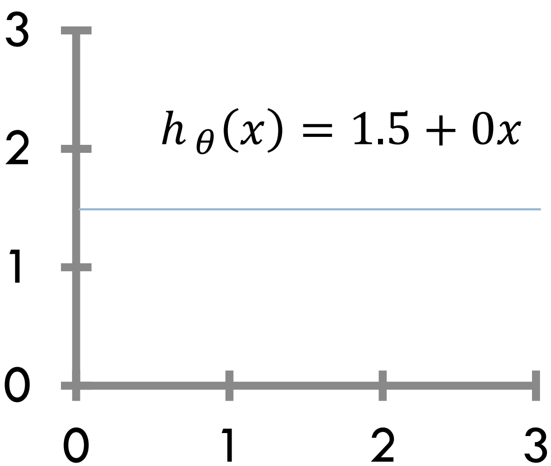
donde $\theta_0, \theta_1 \in \mathbb{R}$

Nota: por comodidad, veremos que podemos expresar la hipótesis como $h_{\theta}(x) = \theta_0 x_0 + \theta_1 x_1$ donde $x_0 = 1$

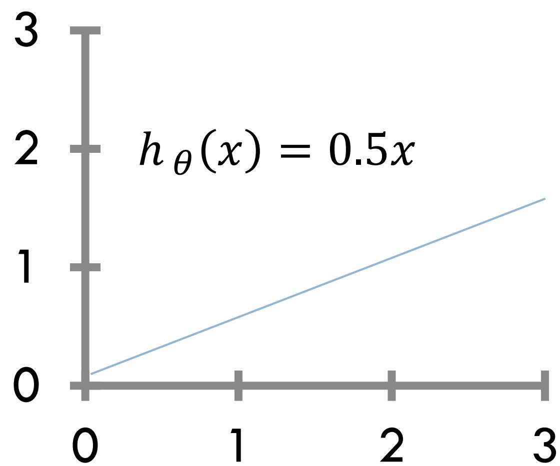
Regresión lineal con una variable

- ¿Cómo afectan los parámetros θ al modelo?

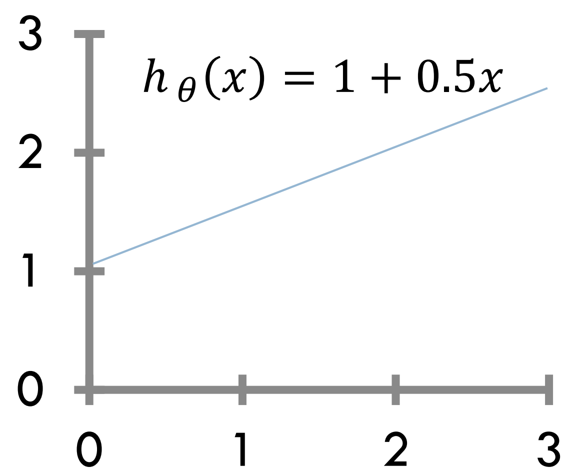
$$h_\theta(x) = \theta_0 + \theta_1 x$$



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



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



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

Regresión lineal con multivariable

- **Objetivo**
  - Encontrar un **modelo** que represente un conjunto de datos observados
  - El modelo servirá para **predecir valores para nuevos datos**
- En **regresión**…
  - La salida es un valor numérico
- En **regresión con múltiples variables**…
  - Cada ejemplo viene representado por más de una variable

$$x = \{x_1, \dots, x_n\}$$

Regresión lineal con multivariable

- **Ejemplos con una sola característica (variable)**

Tamaño (x)	Precio (y)
2104	460
1416	232
1534	315
852	178
…	…

- Hipótesis

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

Regresión lineal con multivariable

- **Ejemplos con múltiples características (variables)**

Tamaño (x_1)	Nº habitaciones (x_2)	Nº pisos (x_3)	Antigüedad (x_4)	Precio (y)
2104	5	1	45	460
1416	3	2	40	232
1534	3	2	30	315
852	2	1	36	178
…	…	…	…	…

- Hipótesis??
  - **Más parámetros!**

$$h_\theta(x) = \theta_0 + \theta_1 x_1 + \theta_2 x_2 + \theta_3 x_3 + \theta_4 x_4$$

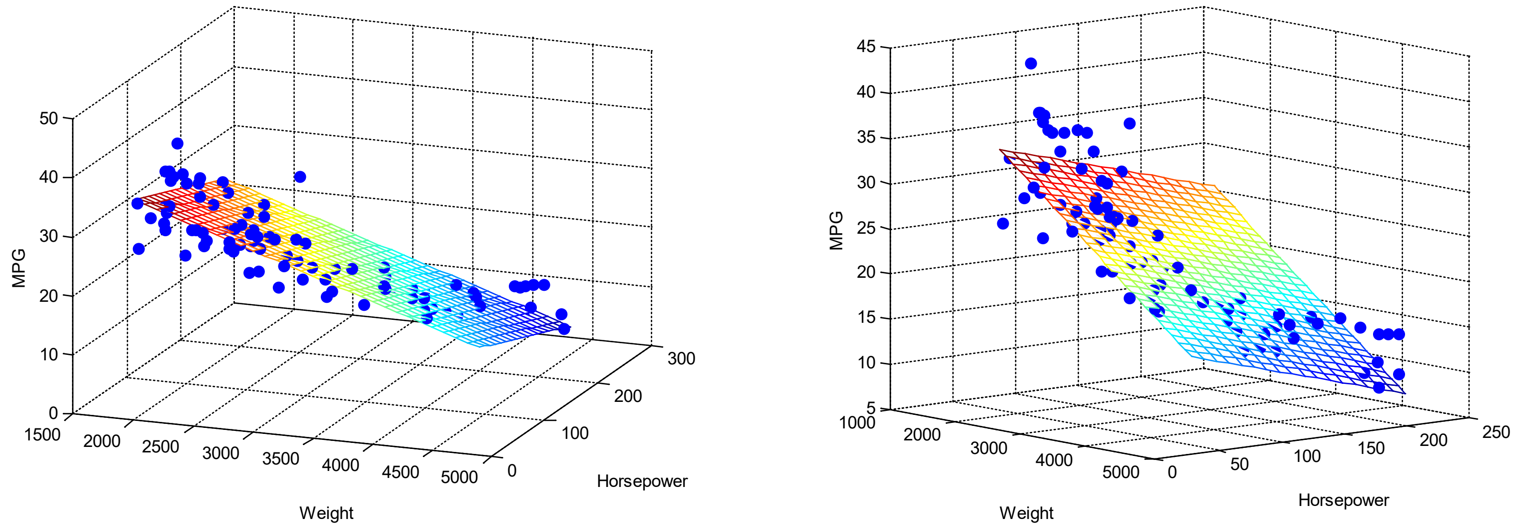
Regresión lineal con multivariable

- **Notación**
  - $h_{\theta}(x) =$ hipótesis
  - $m =$ número de ejemplos
  - $\boldsymbol{n} =$ **número de atributos (características/variables)**
  - $x^{(i)} =$ vector de características del ejemplo i-ésimo
  - $\boldsymbol{x_j^{(i)}} =$ **valor de la característica $\boldsymbol{j}$-ésima del ejemplo $\boldsymbol{i}$-ésimo**

Regresión lineal con multivariable

- Si tenemos más de una dimensión
  - Ya **no es fácil mostrar los ejemplos**
  - **Con 2 características todavía es posible**
    - La regresión representa un **plano**

Ejemplos de planos de regresión sobre 2 variables



Regresión lineal con multivariable

- Cuando tenemos **más de 2 características**
  - Obtenemos un **híper-plano**
- **Objetivo**
  - **Buscar el híper-plano que más se parezca a los datos de entrenamiento**

Regresión lineal con multivariable

- **Objetivo** (formal)
  - Estimar un **modelo** en el que la variable de salida (y) se expresa linealmente en términos de las variables del modelo ($x = \{x_1, \dots, x_n\}$) y de un conjunto de parámetros ($\theta = \{\theta_0, \theta_1, \dots, \theta_n\}$)

$$y = h_\theta(x) = \theta_0 + \theta_1 x_1 + \theta_2 x_2 + \cdots + \theta_n x_n$$

Regresión lineal con multivariable

- **Objetivo** (formal)
  - Es decir… un modelo en el que el precio se obtiene en términos del tamaño de la casa, nº habitaciones, nº pisos y antigüedad en base a los parámetros θ

$$precio = h_{\theta}(tamaño)$$
$$= \theta_0 + \theta_1 tamaño + \theta_2 habitaciones$$
$$+ \theta_3 pisos + \theta_4 antigüedad$$

Regresión lineal con multivariable

- **Podemos representar** $h_{\theta}(x)$ de manera sencilla e independiente de n
  - Definimos $x_0 = 1$
    - Añadimos una variable 0 a todos los ejemplos cuyo valor es siempre 1

$$h_{\theta}(x) = \theta_0 + \theta_1 x_1 + \theta_2 x_2 + \cdots + \theta_n x_n$$

$$\Downarrow$$

$$h_{\theta}(x) = \boxed{\theta_0 x_0} + \theta_1 x_1 + \theta_2 x_2 + \cdots + \theta_n x_n$$

Regresión lineal con multivariable

- **Podemos representar** $h_\theta(x)$ de manera sencilla e independiente de n

$$h_\theta(x) = \theta_0 x_0 + \theta_1 x_1 + \theta_2 x_2 + \cdots + \theta_n x_n$$

- Siendo x el vector de características y θ el vector con los parámetros
  - La regresión con una variable es un caso particular $x = \{x_0, x_1\}$

$$x = \begin{pmatrix} x_0 \\ x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix} \qquad \theta = \begin{pmatrix} \theta_0 \\ \theta_1 \\ \theta_2 \\ \vdots \\ \theta_n \end{pmatrix} \Rightarrow \boxed{h_\theta(x) = \theta^T x}$$

Regresión lineal con multivariable

$$h_{\theta}(x) = (\theta_0\ \theta_1\ \theta_2 \ldots \theta_n)\begin{pmatrix} x_0 \\ x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix} =$$

$$\theta_0 x_0 + \theta_1 x_1 + \theta_2 x_2 + \cdots + \theta_n x_n$$

$$\boxed{h_{\theta}(x) = \theta^T x}$$

Regresión lineal

¿Cómo obtenemos los parámetros θ_i?

Regresión lineal

- **Ejemplos con múltiples características (variables)**

Tamaño (x_1)	**Nº habitaciones (x_2)**	**Nº pisos (x_3)**	**Antigüedad (x_4)**	**Precio (y)**
2104	5	1	45	460
1416	3	2	40	232
1534	3	2	30	315
852	2	1	36	178
…	…	…	…	…

- Hipótesis

$$h_\theta(x) = \theta^T x$$

Regresión lineal

- **Idea**
  - *Buscamos aquellos parámetros* θ *tal que* $h_\theta\,(x)$ *sea lo más parecido posible a los valores observados de* y *en los ejemplos de entrenamiento* (x, y)
    - Intentamos que la **hipótesis se ajuste a los datos observados**
    - Para poder usarla para **predecir nuevos datos**

$$h_\theta(2104, 5, 1, 45) \approx 460$$
$$h_\theta(1416, 3\ , 2, 40) \approx 232$$
$$h_\theta(534, 3, 2, 30) \approx 315$$
$$h_\theta(852, 2, 1, 36) \approx 178$$

Regresión lineal

- Necesitamos una forma de medir el **error**
  - Lo parecidos (o diferentes) que es la salida obtenida $h_\theta(x)$ frente a la deseada y

¡¡Error!!

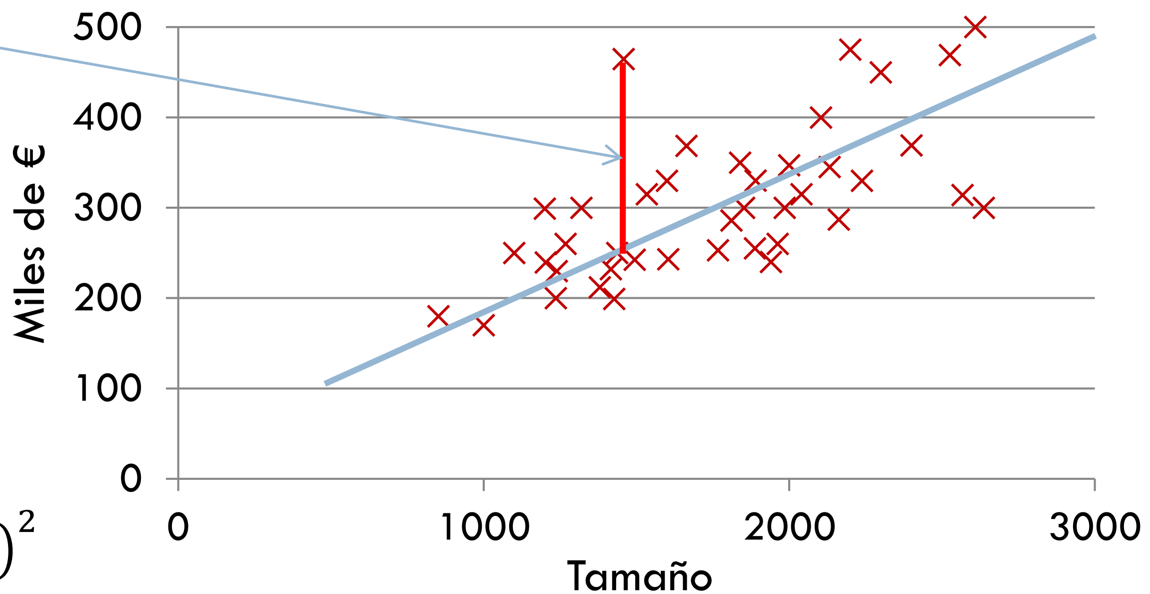
$$\left| h_\theta\left(x^{(i)}\right) - y^{(i)} \right|$$

Utilizamos el error cuadrático

$$\left(h_\theta\left(x^{(i)}\right) - y^{(i)}\right)^2$$

Para todos los ejemplos…

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



Regresión lineal

- Función de **coste** o **error**
  - La suma de los errores que cometemos para todos los ejemplos (tenemos m ejemplos)

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

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

Asumimos $x_0^{(i)} = 1$ para todo $i = 1, \dots, m$

Regresión lineal

- **Objetivo**
  - **Minimizar la función de coste**
    - Encontrar los **parámetros $\boldsymbol{\theta}$** con los que **la función de coste** obtiene el **mínimo valor** para los ejemplos de entrenamiento

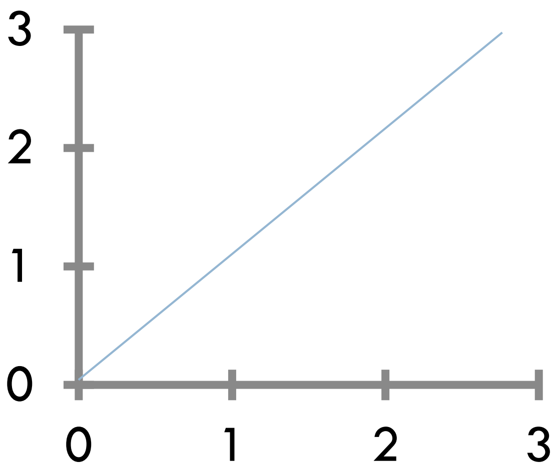
$$\underset{\theta}{\text{Minimizar}} \quad J(\theta) = \frac{1}{2m}\sum_{i=1}^{m}\left(h_\theta\left(x^{(i)}\right) - y^{(i)}\right)^2$$

Regresión lineal

- **Observando el error**
  - Simplificamos la hipótesis (para una variable)
    - Asumimos que tenemos **un único parámetro $\boldsymbol{\theta_1}$**
      - Las rectas siempre pasan por el eje de coordenadas

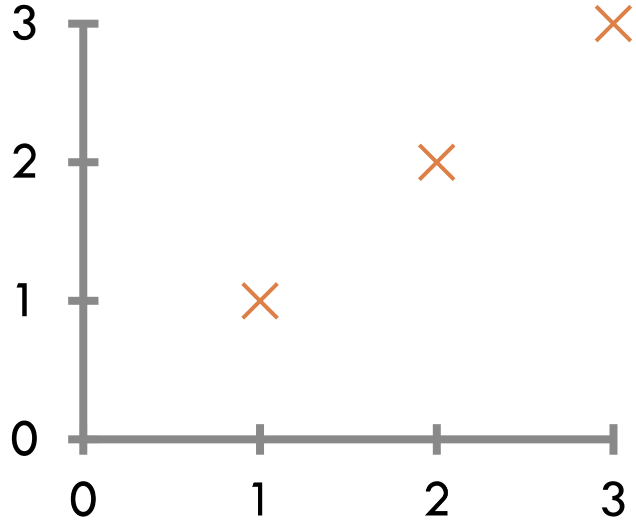
$$h_\theta(x) = \theta_1 x$$

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



Regresión lineal

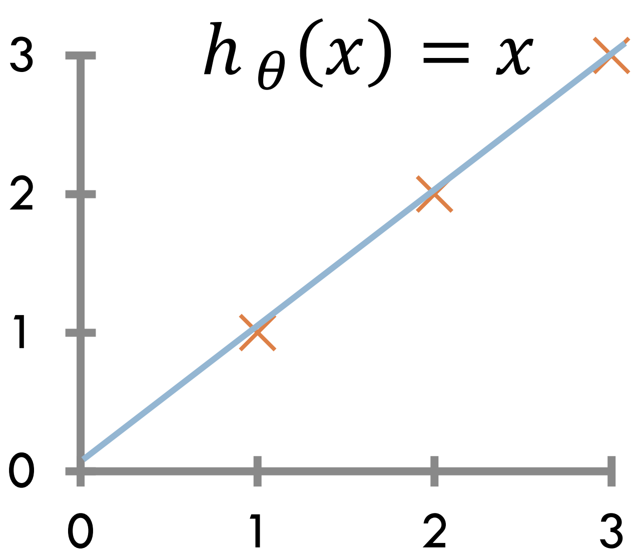
- Supongamos el siguiente conjunto de datos formado por los puntos (1,1) (2,2) y (3,3)



X	Y
1	1
2	2
3	3

Regresión lineal

- Supongamos $\theta_1 = 1$, ¿cuánto es el error?



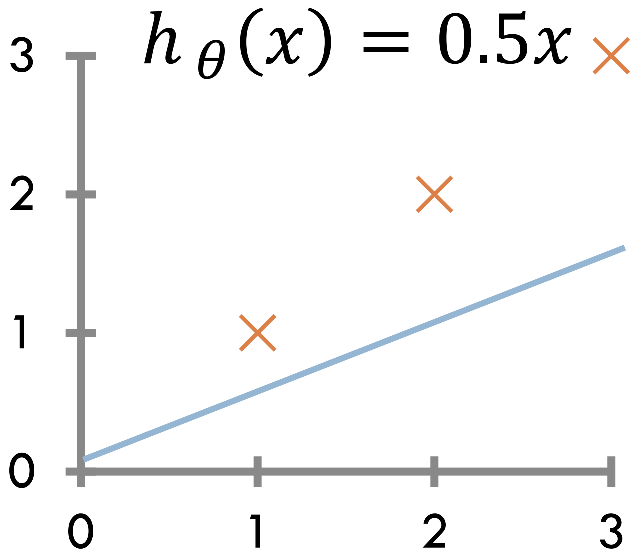
$$J(1) = \frac{1}{2m}\sum_{i=1}^{m}\left(x^{(i)} - y^{(i)}\right)^2$$

$$J(1) = \frac{1}{2\cdot 3}[(1-1)^2 + (2-2)^2 + (3-3)^2]$$

$$J(1) = 0$$

Regresión lineal

- Supongamos $\theta_1 = 0.5$, ¿cuánto es el error?



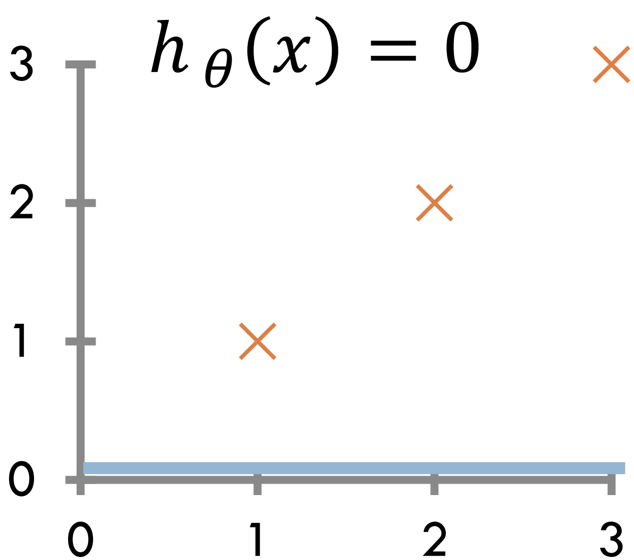
$$J(0.5) = \frac{1}{2m}\sum_{i=1}^{m}\left(0.5x^{(i)} - y^{(i)}\right)^2$$

$$J(0.5) = \frac{1}{2 \cdot 3}[(0.5-1)^2 + (1-2)^2 + (1.5-3)^2]$$

$$J(0.5) = \frac{1}{2 \cdot 3}0.5^2 + 1^2 + 1.5^2 \approx 0.583$$

Regresión lineal

- Supongamos $\theta_1 = 0$, ¿cuánto es el error?



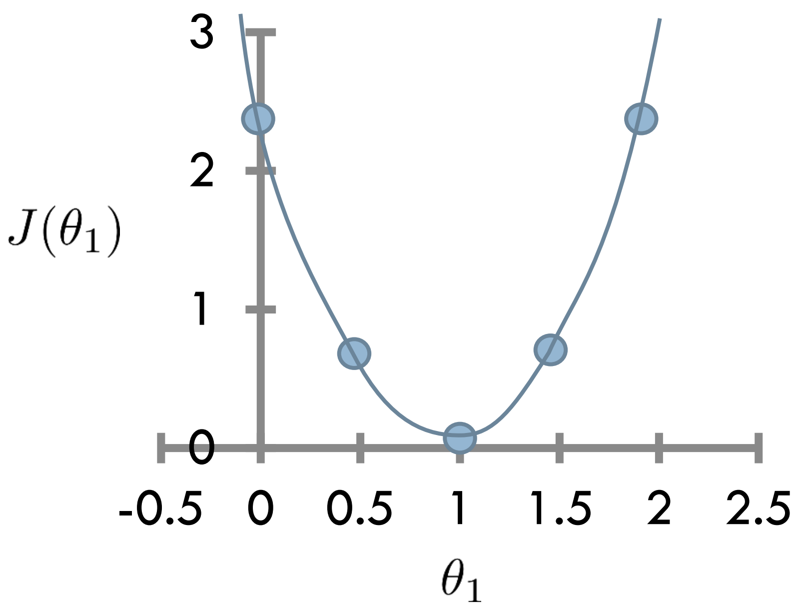
$$J(0) = \frac{1}{2m}\sum_{i=1}^{m}\left(0 - y^{(i)}\right)^2$$

$$J(0) = \frac{1}{2 \cdot 3}[(0-1)^2 + (0-2)^2 + (0-3)^2]$$

$$J(0) = \frac{1}{2 \cdot 3} 1^2 + 2^2 + 3^2 \approx 2.333$$

Regresión lineal

- Si mostramos el error ($J(\theta_1)$) en función del parámetro $\theta_1 \ldots$
  - Podemos observar que obtenemos una función **continua y derivable** que podemos **minimizar**

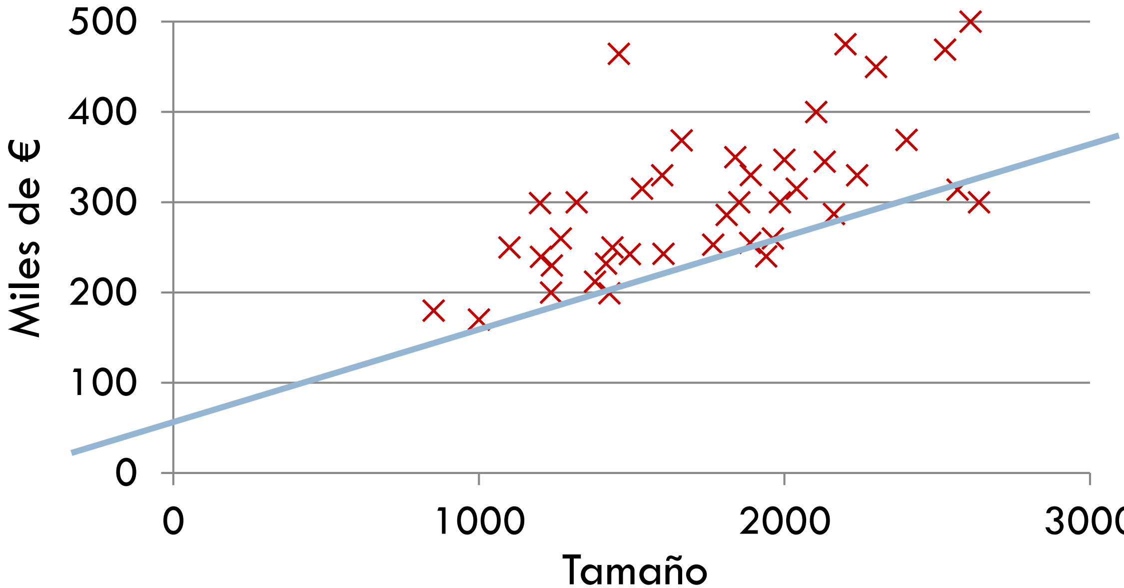


Regresión lineal

- Analicemos el caso con **2 parámetros** (θ_0, θ_1)

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

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

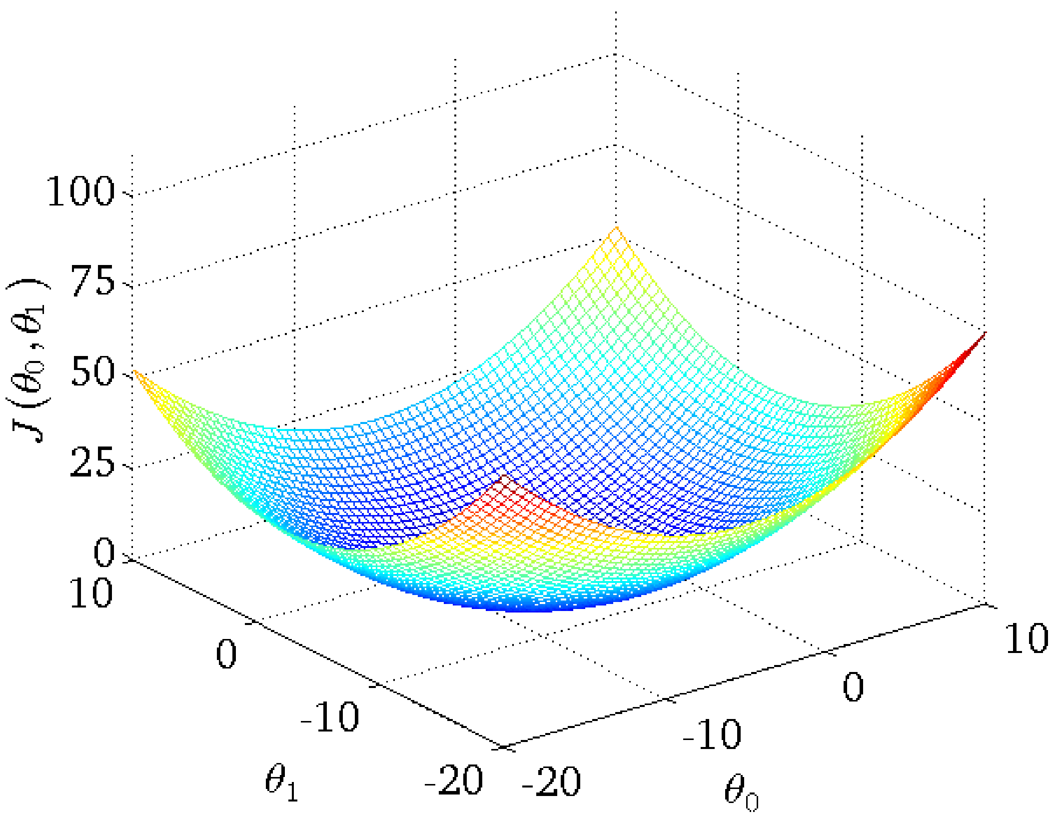


$$h_\theta(x) = 50 + 0.06x$$

$$J(50{,}0.06) \approx 800$$

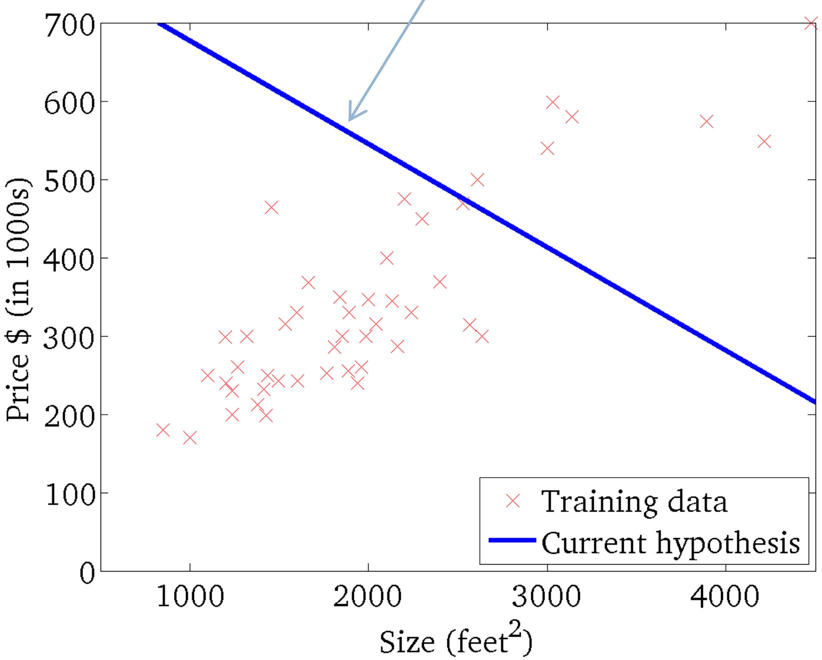
Regresión lineal

- Podemos representar el error en función de (θ_0, θ_1)



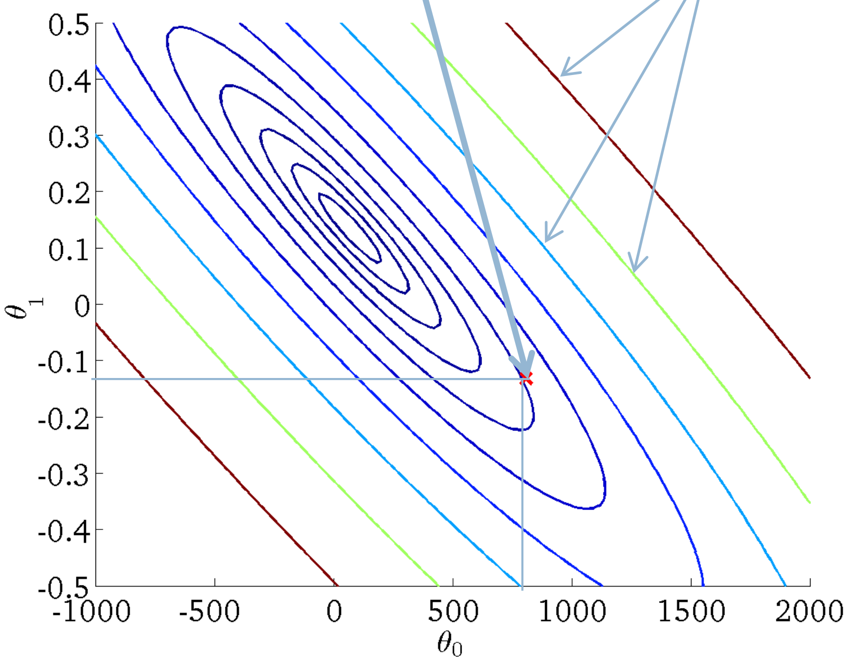
Regresión lineal con una variable

$$h_\theta(x) = 800 - 0.12x$$



$J(800, -0.12)$

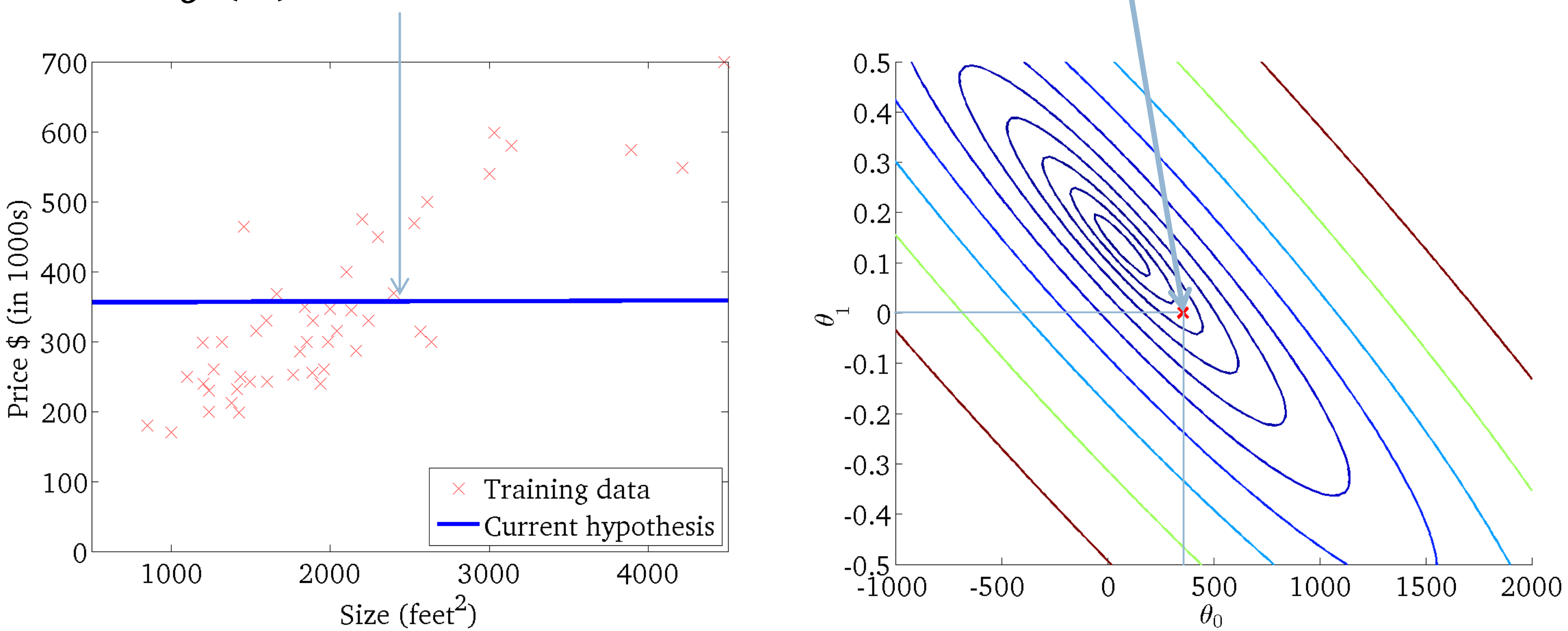
$J(\theta_0, \theta_1)$



Regresión lineal con una variable

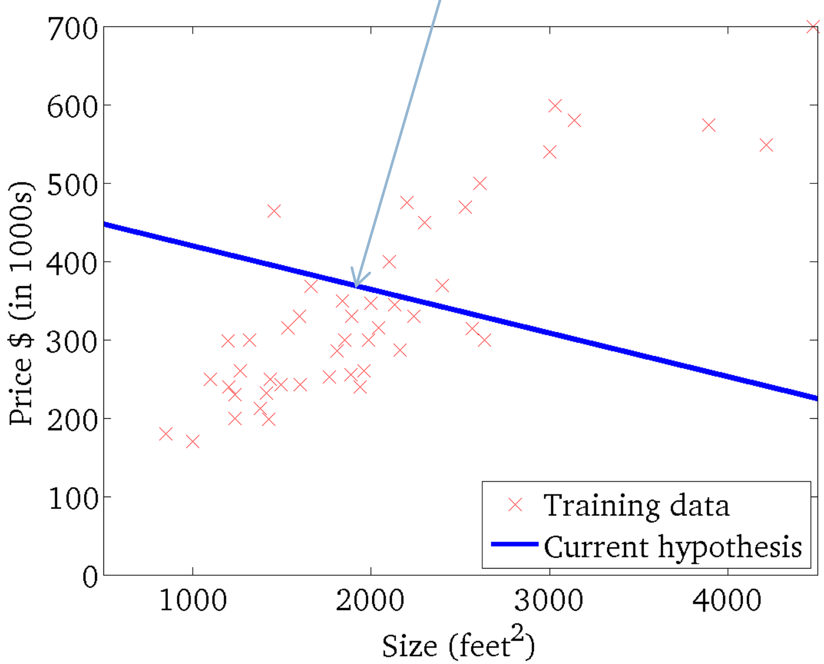
$$h_\theta(x) = 400 + 0x$$

$J(400, 0)$

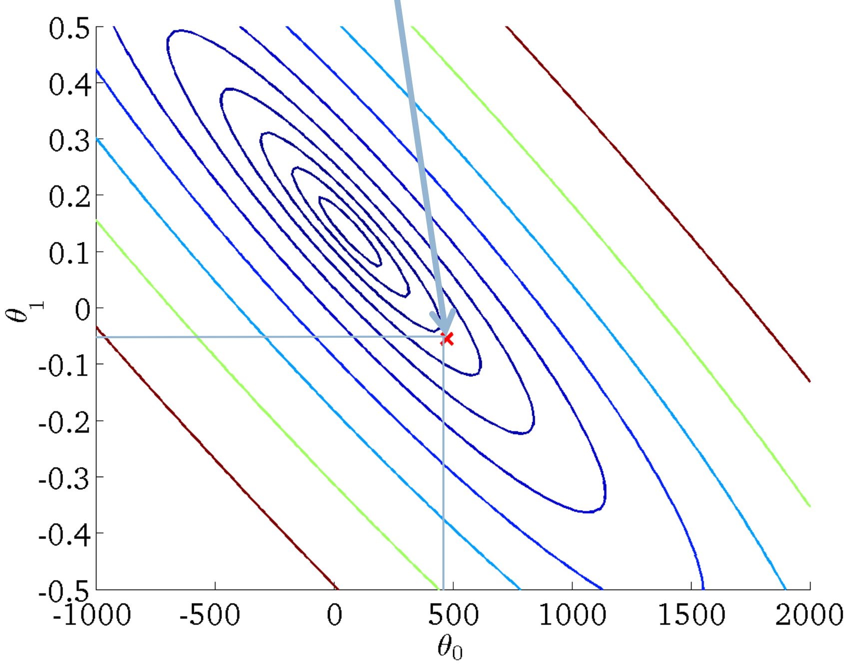


Regresión lineal con una variable

$$h_\theta(x) = 450 + 0.05x$$

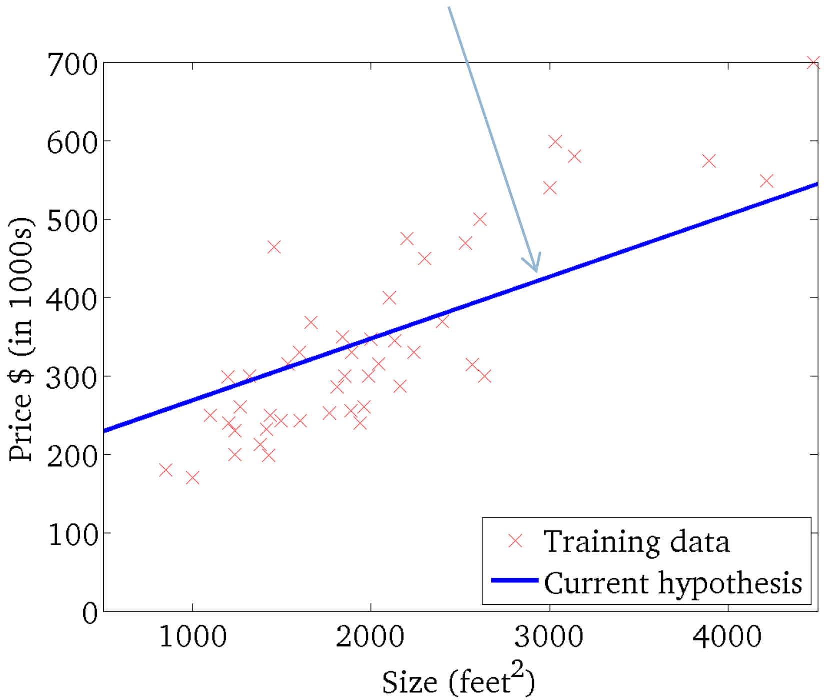


$$J(450, 0.05)$$

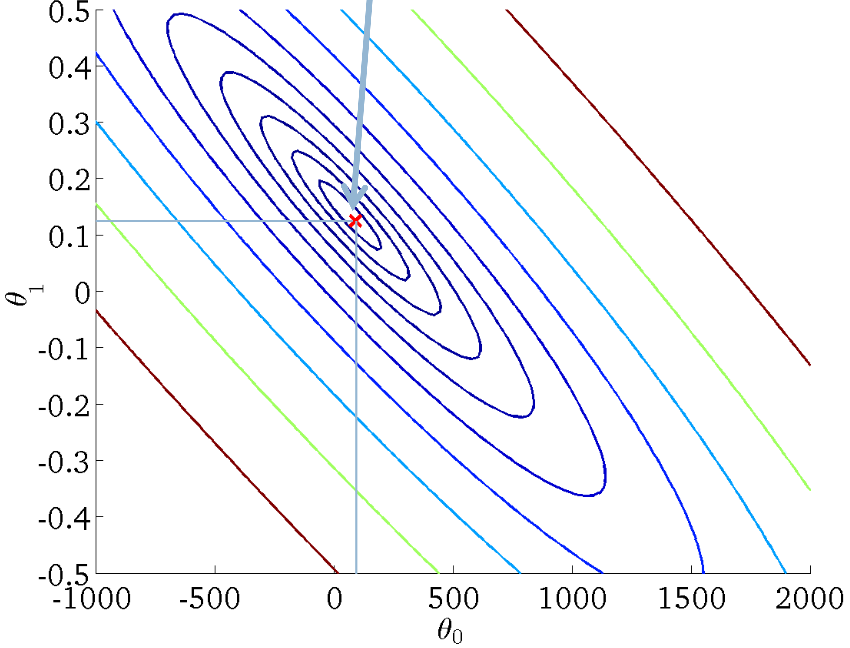


Regresión lineal con una variable

$$h_\theta(x) = 210 + 0.13x$$



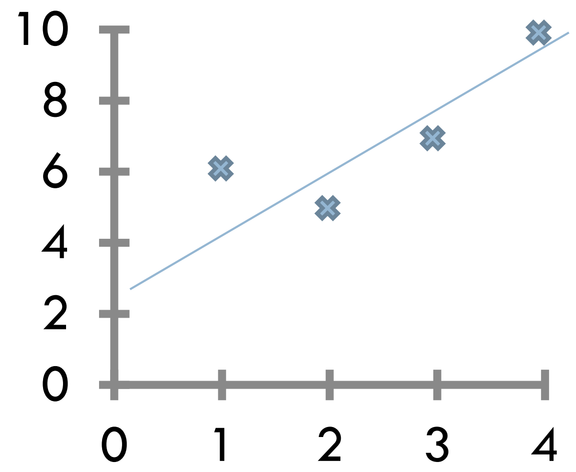
$$J(210, 0.13)$$



Regresión lineal con una variable

- **Cálculo de parámetros**
  - Tenemos un función **continua y derivable** que queremos **minimizar**
    - **Queremos el punto donde la derivada es 0**
  - Un ejemplo
    - Supongamos el conjunto de datos
      - {(1,6), (2,5), (3, 7), (4, 10)}

X	Y
1	6
2	5
3	7
4	10



Regresión lineal con una variable

- Podemos escribir el siguiente sistema de ecuaciones

X	Y
1	6
2	5
3	7
4	10

$$\theta_0 + \theta_1 = 6$$
$$\theta_0 + 2\theta_1 = 5$$
$$\theta_0 + 3\theta_1 = 7$$
$$\theta_0 + 4\theta_1 = 10$$

- Queremos los θ_0, θ_1 que verifiquen todas las ecuaciones
  - **¡¡¡Pero no existen!!!**
  - Debemos aproximar la solución
    - Queremos aquella que obtenga el menor error

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

Regresión lineal

- **Cálculo de parámetros**
  - Tenemos un función **continua y derivable** que queremos **minimizar**
    - **Queremos el punto donde la derivada es 0**
      - Derivadas parciales → sistema de ecuaciones
- **Dos métodos**
  - Solución analítica (directa)
  - Descenso por gradiente

Regresión lineal

- **Solución analítica (directa)**
  - Escribamos los datos en forma **matricial**
  - $X \in \mathbb{R}^{m\times(n+1)}$ que contiene el valor de las variables (x) en los ejemplos del conjunto de entrenamiento
    - m: número de ejemplos
    - n: es el número de variables del modelo
      - En regresión con una variable $n = 1$
  - $\theta \in \mathbb{R}^{n+1}$ que contiene el valor de los parámetros
  - $y \in \mathbb{R}^{m}$ que contiene el valor de la variable de salida (y) en los ejemplos del conjunto de entrenamiento

Regresión lineal

- **Solución analítica (directa)**
  - Escribamos los datos en forma matricial

Tamaño (x_1)	Nº habitaciones (x_2)	Nº pisos (x_3)	Antigüedad (x_4)	Precio (y)
2104	5	1	45	460
1416	3	2	40	232
1534	3	2	30	315
852	2	1	36	178

$$X = \begin{pmatrix} 1 & 2104 & 5 & 1 & 45 \\ 1 & 1416 & 3 & 2 & 40 \\ 1 & 1534 & 3 & 2 & 30 \\ 1 & 852 & 2 & 1 & 36 \end{pmatrix}, \theta = \begin{pmatrix} \theta_0 \\ \theta_1 \\ \theta_2 \\ \theta_3 \\ \theta_4 \end{pmatrix}, y = \begin{pmatrix} 460 \\ 232 \\ 315 \\ 178 \end{pmatrix}$$

Regresión lineal

- Podemos expresar el sistema de ecuaciones en forma matricial de la siguiente forma

$$(X^T X)\theta = X^T y$$

- Y por tanto

$$\boxed{\theta = (X^T X)^{-1} X^T y}$$

- Si $X^T X$ no es invertible
  - Podemos utilizar otras técnicas de minimización
    - **Descenso por gradiente**

Regresión lineal

- ¿Cuándo $X^T X$ no es invertible?
  - **Características redundantes** (linealmente dependientes)
    - Ejemplo
      - x_1 = tamaño en pies^2
      - x_2 = tamaño en metros^2
      - $x_1 = (3.28)^2 x_2$
    - Solución
      - **Eliminar las características redundantes**

1m = 3.28pies

Regresión lineal

- ¿Cuándo $X^T X$ no es invertible?
  - **Demasiadas características** (p.e., $m \leq n$)
    - Ejemplo
      - $m = 10$ (10 ejemplos)
      - $n = 100$ (100 atributos, 101 parámetros θ a estimar)
      - ¡Demasiados parámetros para tan pocos ejemplos!
        - Muchas soluciones…
    - Solución
      - Eliminar algunas características
      - Usar regularización (lo veremos más adelante)

Regresión lineal

- **Descenso por gradiente**
  - Algoritmo de minimización
- **Idea**
  - Empezamos con **valores aleatorios de $\boldsymbol{\theta}$**
  - Cambiamos los valores de θ con el objetivo de **reducir el error $\boldsymbol{J(\theta)}$**
    - Hasta que alcanzamos un mínimo

Regresión lineal

- **Descenso por gradiente**
  - Asignar a $\theta = \{\theta_0, \dots, \theta_n\}$ valores aleatorios
  - Repetir hasta convergencia
    - (n° iteraciones o |error – error_anterior| < umbral)

$$\boxed{\theta_j := \theta_j - \alpha \frac{\partial J(\theta)}{\partial \theta_j} \quad \text{para } j = 0, \dots, n}$$

ACTUALIZAR TODOS LOS θ_j SIMULTÁNEAMENTE

En regresión lineal con múltiples variables…

$$h_\theta(x) = \theta^T x$$

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

Regresión lineal

- **Descenso por gradiente**
  - Repetir hasta convergencia
    - (nº iteraciones o error < umbral)

$$\underline{\theta_j := \theta_j} - \boxed{\alpha} \boxed{\frac{\partial J(\theta)}{\partial \theta_j}}$$ Simultáneamente para $j = 0$ y $j = 1$

Factor de aprendizaje:

Mide cuánto modificamos el valor del parámetro

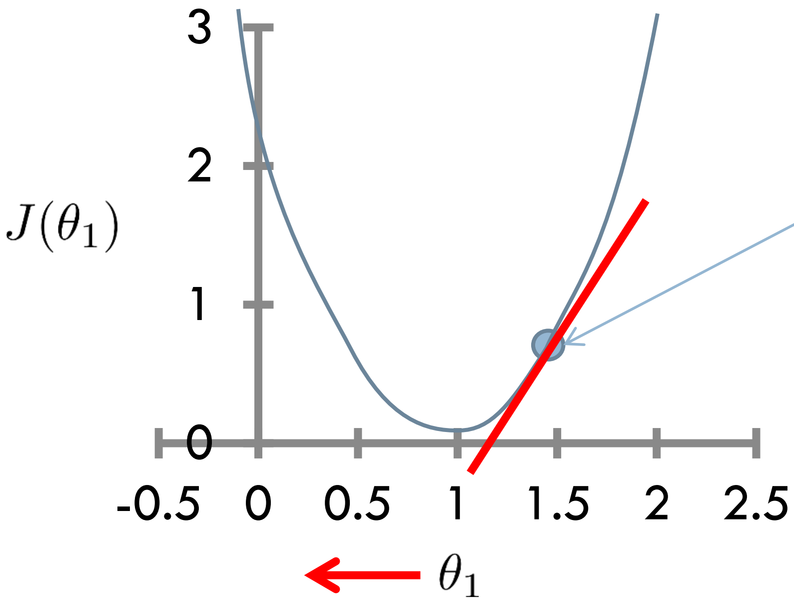
Derivada:

Indica el valor de la pendiente con mayor variación en el punto actual (θ_0, θ_1)

Modificamos el valor de cada parámetro, desplazándolo en la **dirección en la que se reduce el error más rápidamente**

Regresión lineal con una variable

- Comprendiendo el descenso por gradiente



El **gradiente** en este punto indica la pendiente de la recta tangente a la función de coste respecto al parámetro θ_1

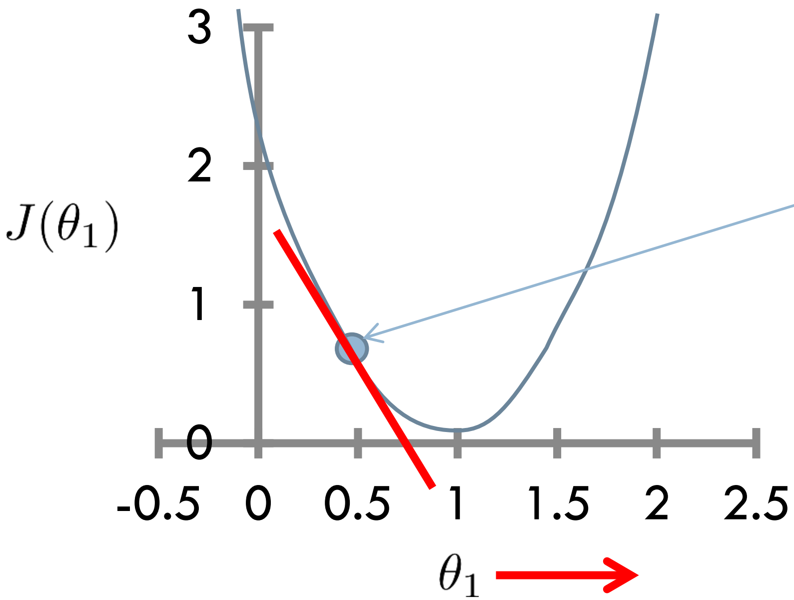
$$\frac{\partial J(\theta)}{\partial \theta_1} \geq 0 \rightarrow \downarrow \theta_1$$

Por eso, lógicamente, si queremos reducir el error, debemos ir en la dirección contraria (negativa) a la de la pendiente (gradiente)

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

Regresión lineal con una variable

- Comprendiendo el descenso por gradiente



El **gradiente** en este punto indica la pendiente de la recta tangente a la función de coste respecto al parámetro θ_1

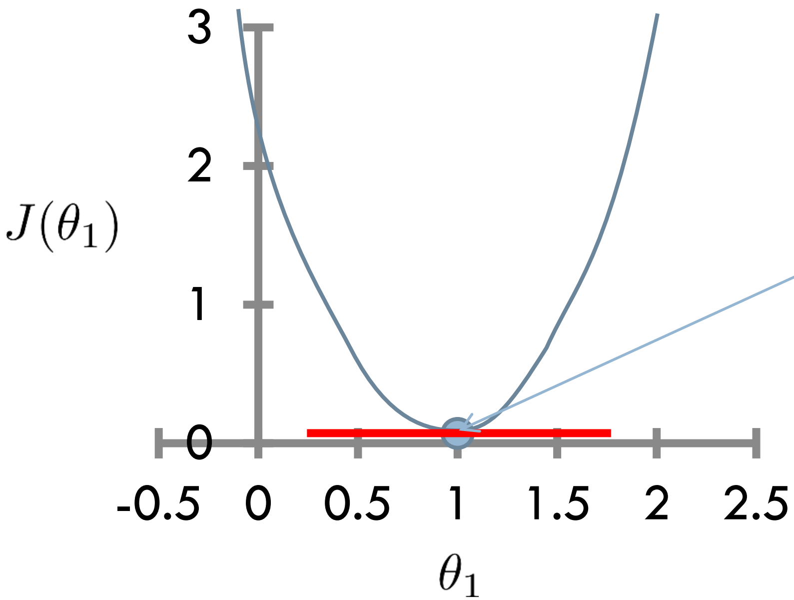
$$\frac{\partial J(\theta)}{\partial \theta_1} \leq 0 \rightarrow \ \uparrow \theta_1$$

Por eso, lógicamente, si queremos reducir el error, debemos ir en la dirección contraria (negativa) a la de la pendiente (gradiente)

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

Regresión lineal con una variable

- Comprendiendo el descenso por gradiente



El **gradiente** en este punto indica la pendiente de la recta tangente a la función de coste respecto al parámetro θ_1 $\dfrac{\partial J(\theta)}{\partial \theta_1} = 0$

¡Si estamos en el mínimo no hay pendiente!

Por eso, lógicamente, si queremos reducir el error, debemos ir en la dirección contraria (negativa) a la de la pendiente (gradiente)

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

¡¡Si no hay pendiente, no se modifica!!

Regresión lineal

- **Descenso por gradiente**
  - Asignar a $\theta = \{\theta_0, \dots, \theta_n\}$ valores aleatorios
  - Repetir hasta convergencia
    - (nº iteraciones o |error – error_anterior| < umbral)

$$\boxed{\theta_j := \theta_j - \alpha \frac{\partial J(\theta)}{\partial \theta_j} \quad \text{para } j = 0, \dots, n}$$

ACTUALIZAR TODOS LOS θ_j SIMULTÁNEAMENTE

En regresión lineal con múltiples variables…

$$h_\theta(x) = \theta^T x$$

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

Regresión lineal

- **Descenso por gradiente**
  - Asignar a θ valores aleatorios (o a ceros)
  - Repetir hasta convergencia
    - (n° iteraciones o |error – error_anterior| < umbral)

$$\boxed{\theta_j := \theta_j - \alpha \frac{1}{m} \sum_{i=1}^{m} \left(h_\theta\left(x^{(i)}\right) - y^{(i)}\right) \cdot x_j^{(i)} \quad \textbf{para todo } \boldsymbol{j = 0, \dots, n}}$$

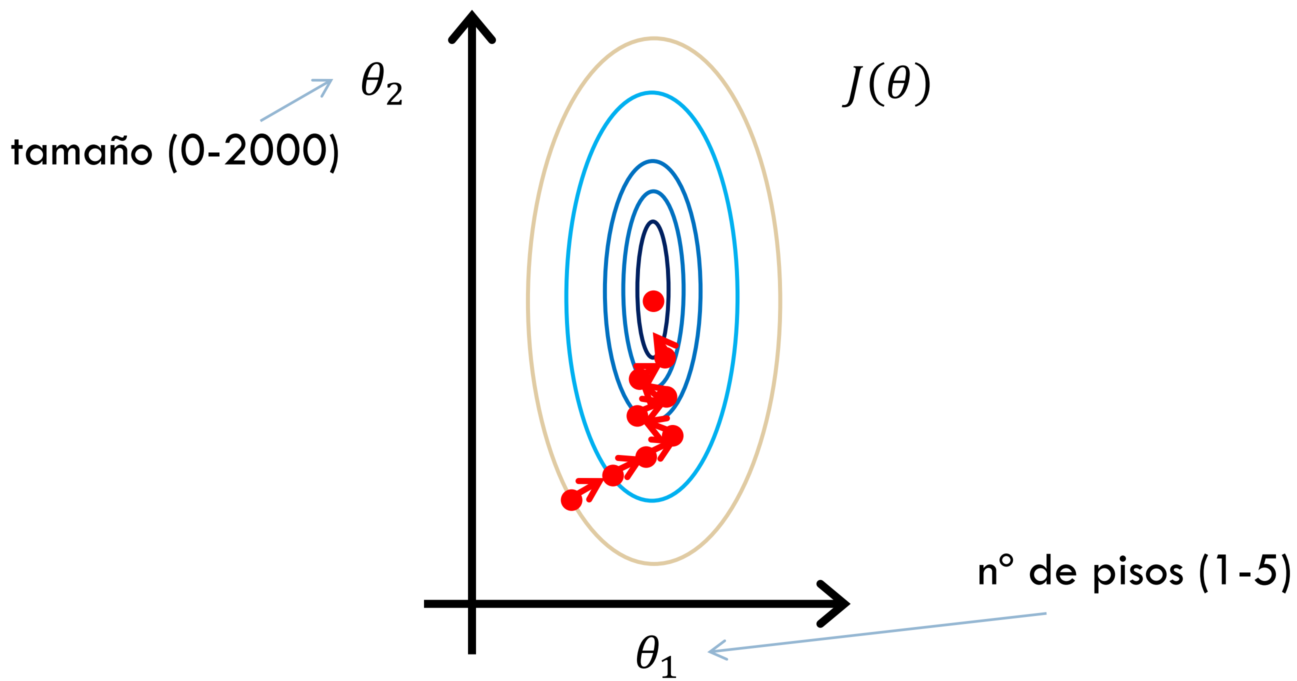
ACTUALIZAR TODOS LOS θ_j SIMULTÁNEAMENTE

Regresión lineal

- Descenso por gradiente por "**lotes**"
  - Usamos todos los ejemplos en cada iteración
  - Puede ser costoso si tenemos muchos ejemplos
- Alternativa
  - **Descenso por gradiente incremental**
  - Actualizamos los parámetros para cada ejemplo
    - Suele ser más rápido (mejor convergencia)
    - Pero puede no converger nunca al mínimo global
      - Puede corregirse haciendo decrecer α según avanza el método

Regresión lineal

- **Problemas con el descenso por gradiente**
  - Al tener diferentes características
    - Suelen estar en escalas (rangos) diferentes
    - Ralentiza la convergencia del gradiente



Regresión lineal

- **Solución**
  - **Escalado de los rangos**
    - Normalizamos todos los atributos
      - Sus valores sigan una distribución normal
        - Media 0 $(\mu = 0)$
        - Desviación estándar 1 $(\sigma = 0)$
        - Todos los valores estarán **aproximadamente** entre [-1, 1]
      - Otras posibles normalizaciones
        - Hacer que todos los valores estén en [0, 1]

Regresión lineal

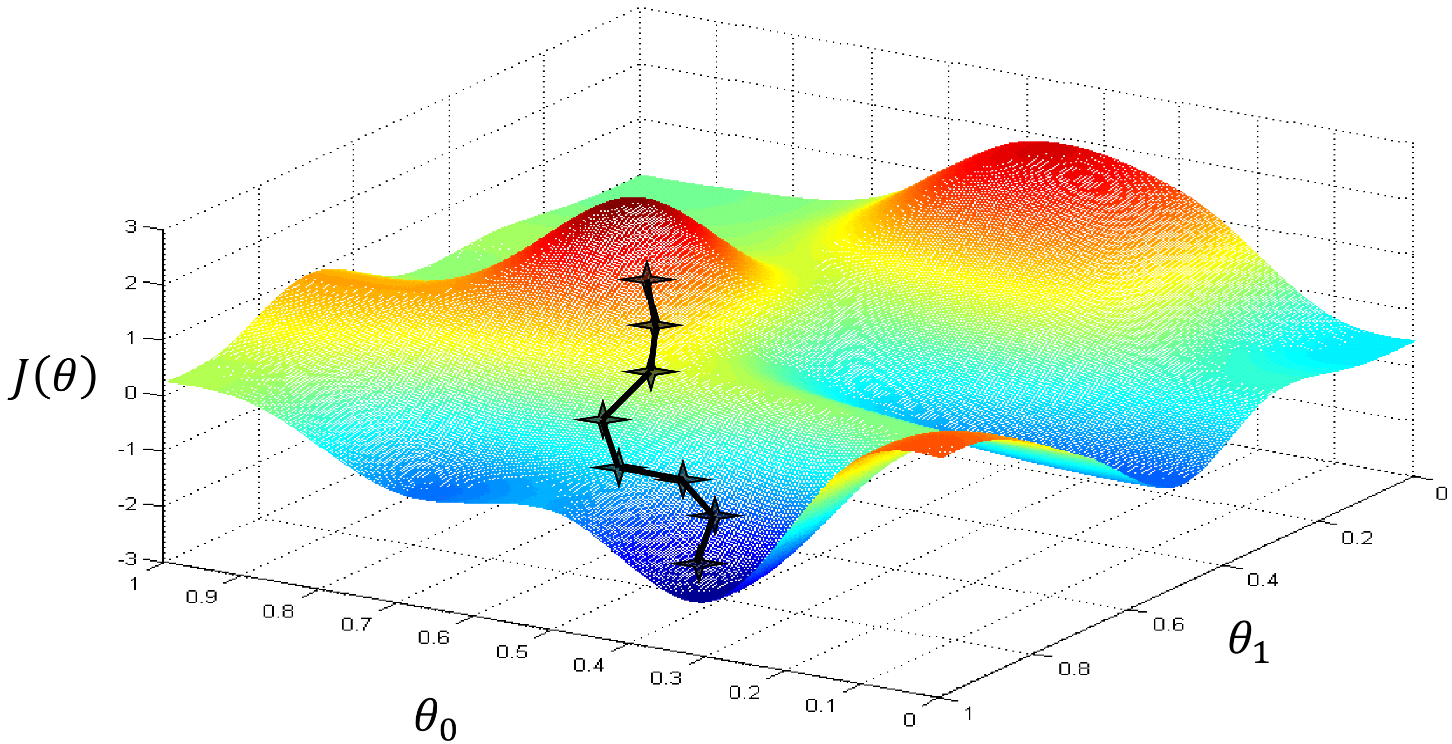
- **Normalización de la media**
  - Para cada atributo j reemplazamos el valor $x_j^{(i)}$ en cada ejemplo i de la siguiente forma

$$x_j^{(i)} = \frac{x_j^{(i)} - \mu_j}{\sigma_j} \qquad \mu_j = \frac{1}{m}\sum_{i=1}^{m} x_j^{(i)} \qquad \sigma_j = \sqrt{\frac{1}{m}\sum_{i=1}^{m}\left(x_j^{(i)} - \mu_j\right)^2}$$

  - **Nota: La normalización no es necesaria para obtener la solución por el método directo**

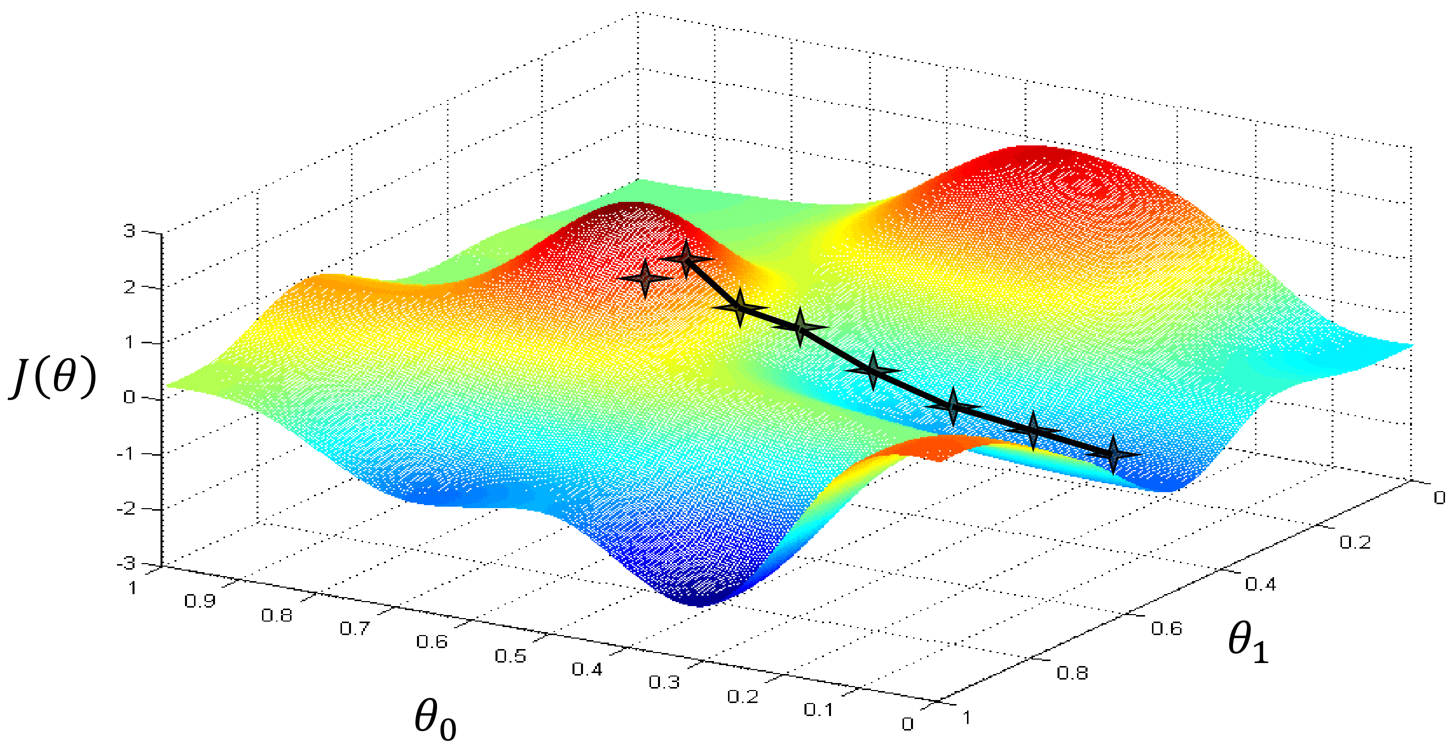
Regresión lineal

- Gráfica del descenso por gradiente



Regresión lineal

- El resultado puede variar según la inicialización de los parámetros

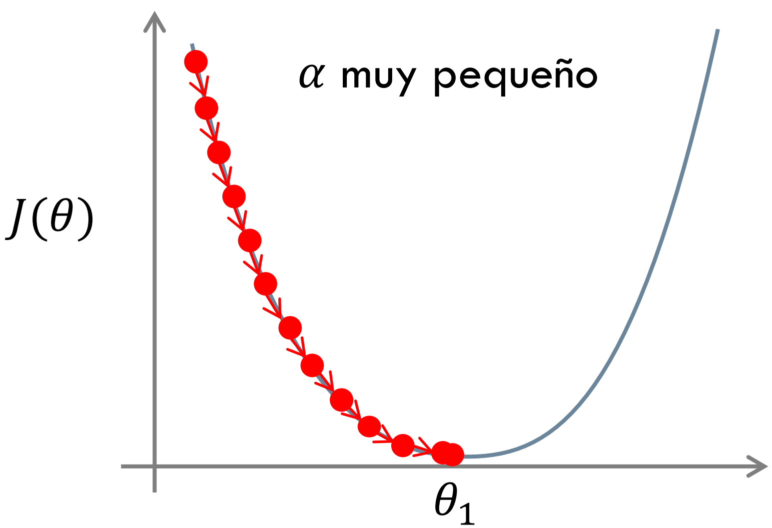


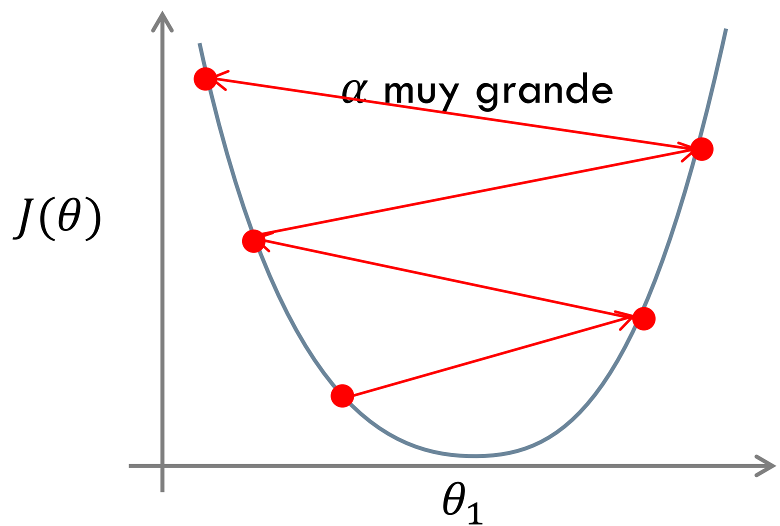
Nota: En la regresión lineal la función a minimizar es siempre convexa, por lo que la inicialización no tiene una influencia importante en el resultado

Regresión lineal

- Efecto del **factor de aprendizaje**
  - Si α es muy pequeño
    - El descenso puede ser **muy lento**
  - Si α es muy grande
    - Podemos saltarnos el mínimo
    - Puede **no converger** (incluso divergir)

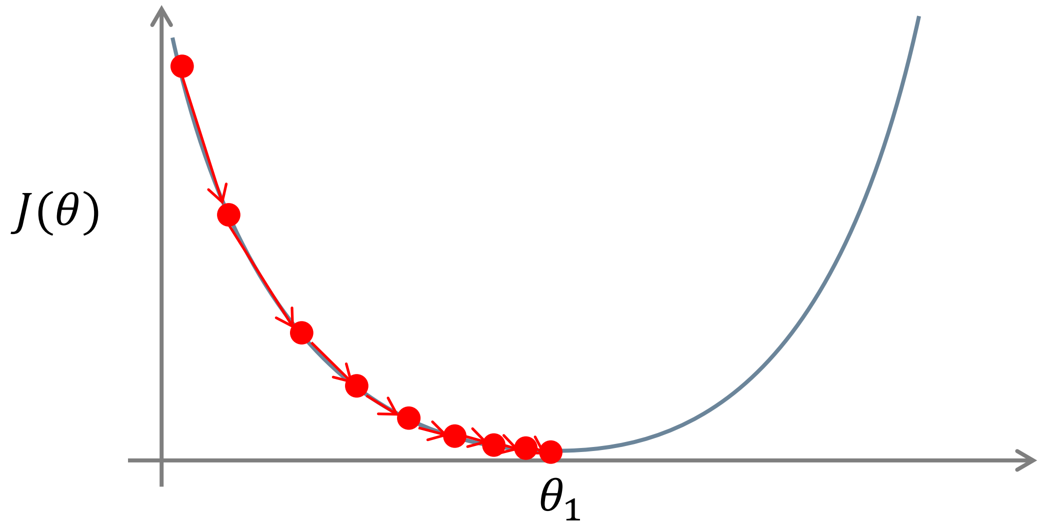
$$\theta_1 := \theta_1 - \alpha \frac{\partial J(\theta)}{\partial \theta_1}$$





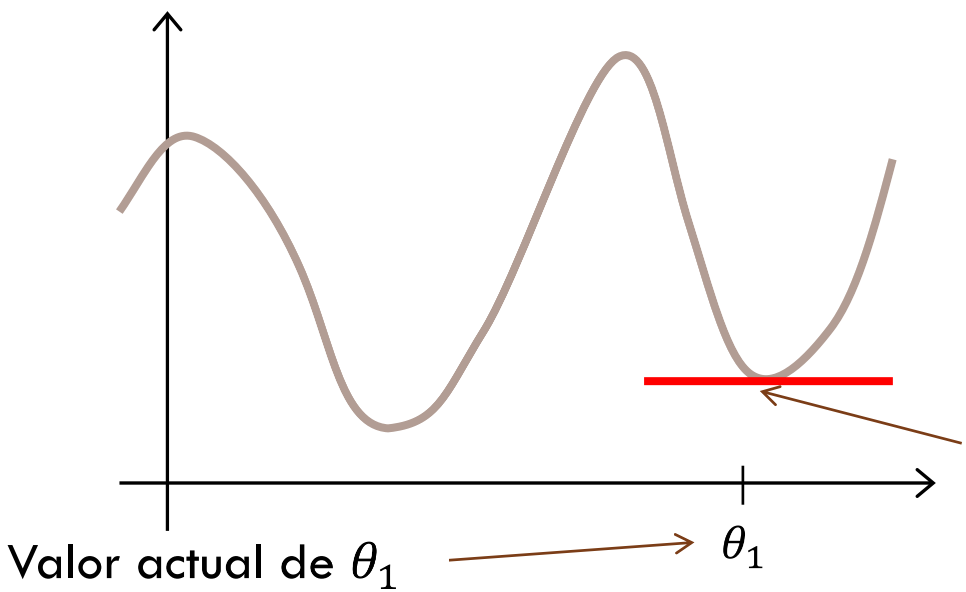
Regresión lineal

- Efecto del **factor de aprendizaje**
  - **Nota**: El método del gradiente puede converger a un mínimo, incluso fijando α
    - El paso es proporcional al gradiente
      - Por tanto, tiende a reducirse según nos acercamos al mínimo



Regresión lineal

- **Mínimos locales**
  - El gradiente es 0
  - El descenso queda atrapado
    - ¡Pero en regresión lineal esto no ocurre!
      - Tenemos una función objetivo con un único mínimo global



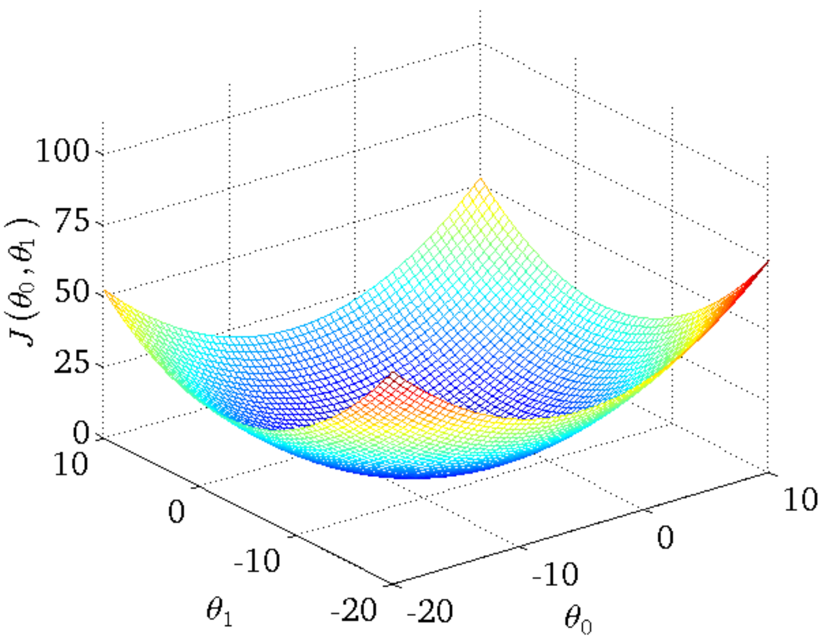
$$\theta_1 := \theta_1 - \alpha \frac{\partial J(\theta)}{\partial \theta_1}$$

$$\Downarrow$$

$$\theta_1 := \theta_1 - 0$$

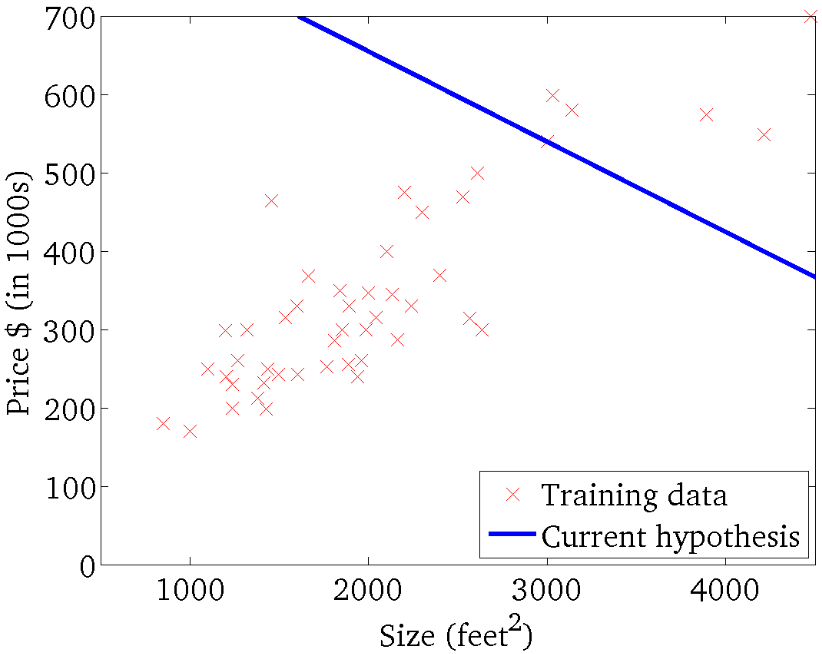
Regresión lineal

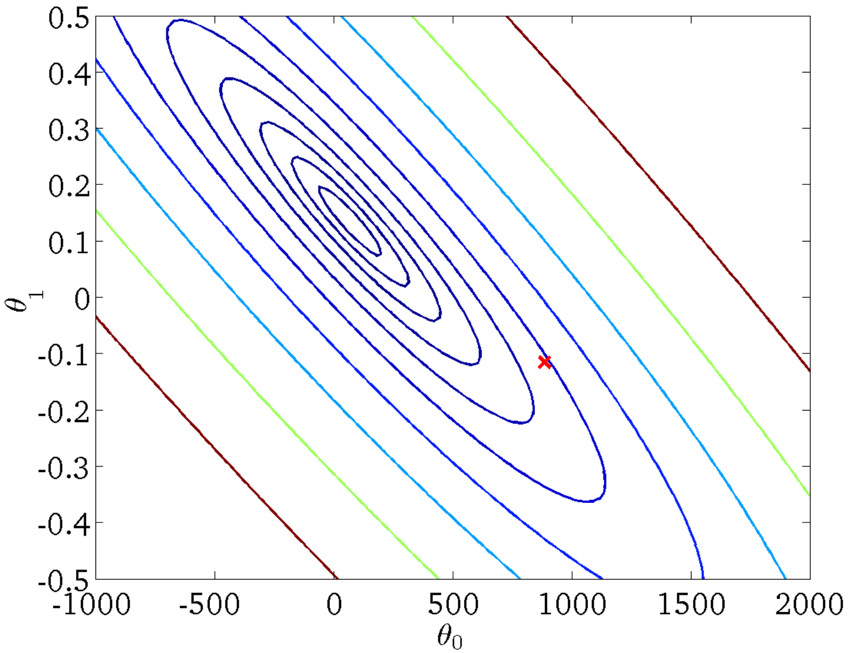
- Pero en regresión lineal…
  - La función de coste siempre es **convexa**
    - Tiene un único mínimo local que es el global
  - Siempre encontramos la solución (con α no muy grande)



Regresión lineal con una variable

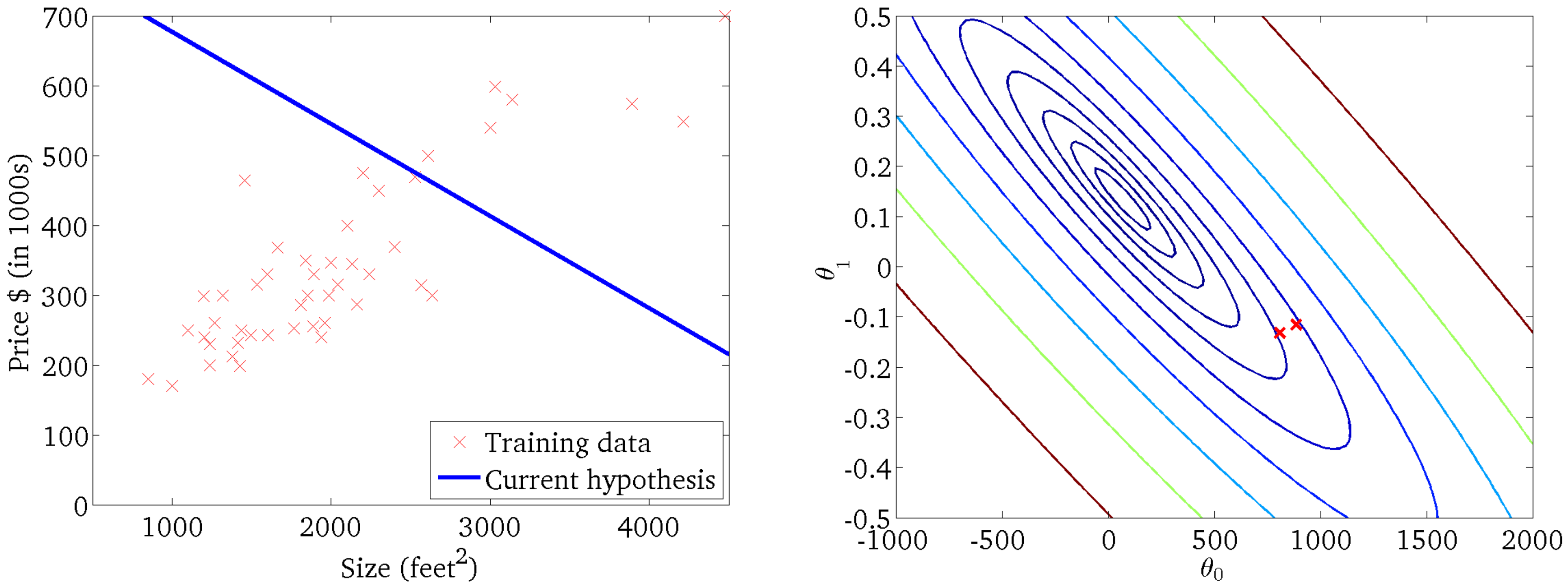
- Gradiente
  - Iteración 1, θ_0, θ_1 aleatorios





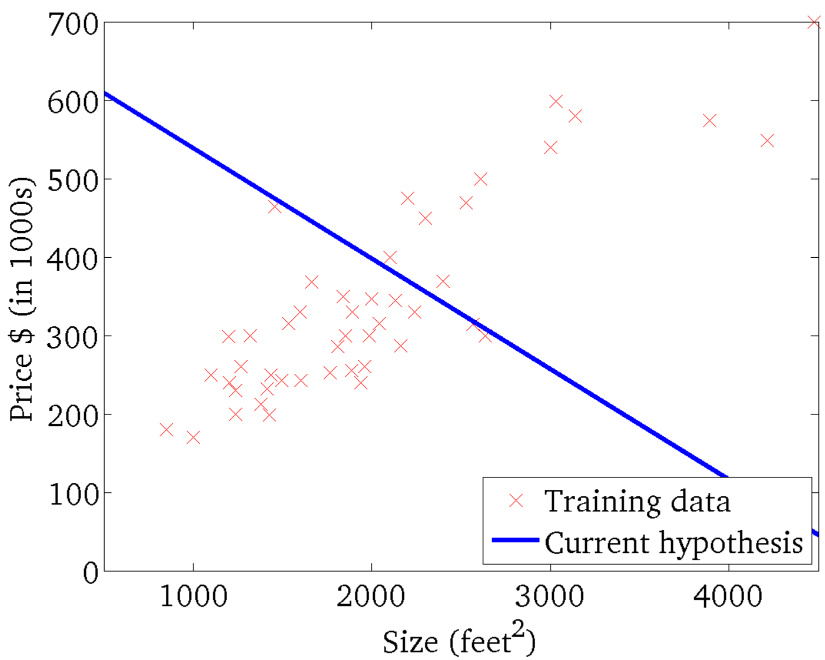
Regresión lineal con una variable

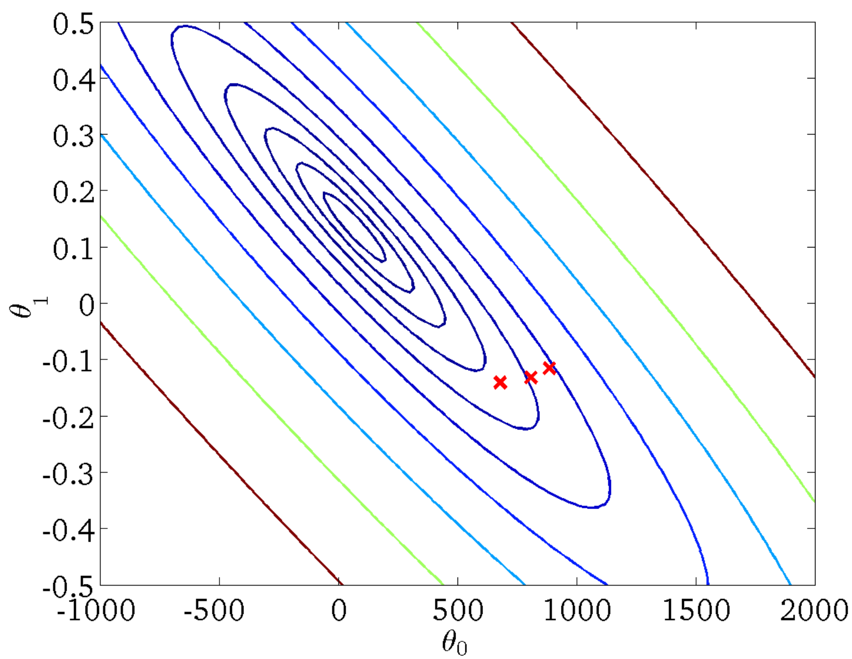
- Gradiente
  - Iteración 30



Regresión lineal con una variable

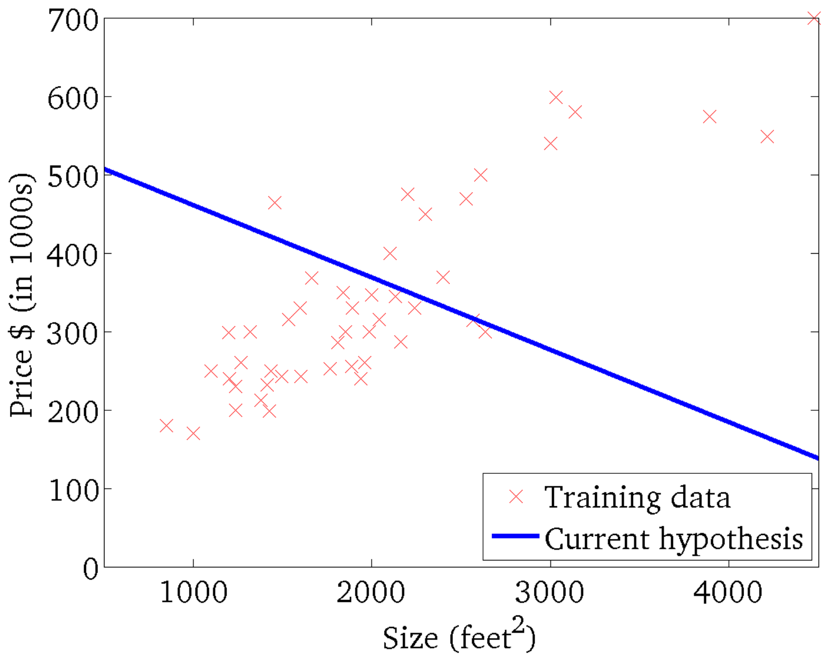
- Gradiente
  - Iteración 60

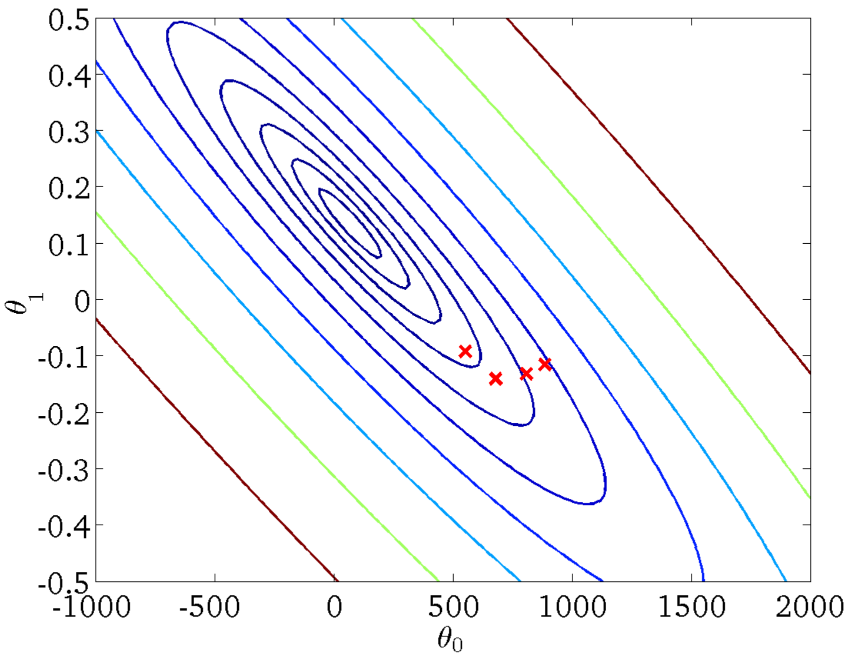




Regresión lineal con una variable

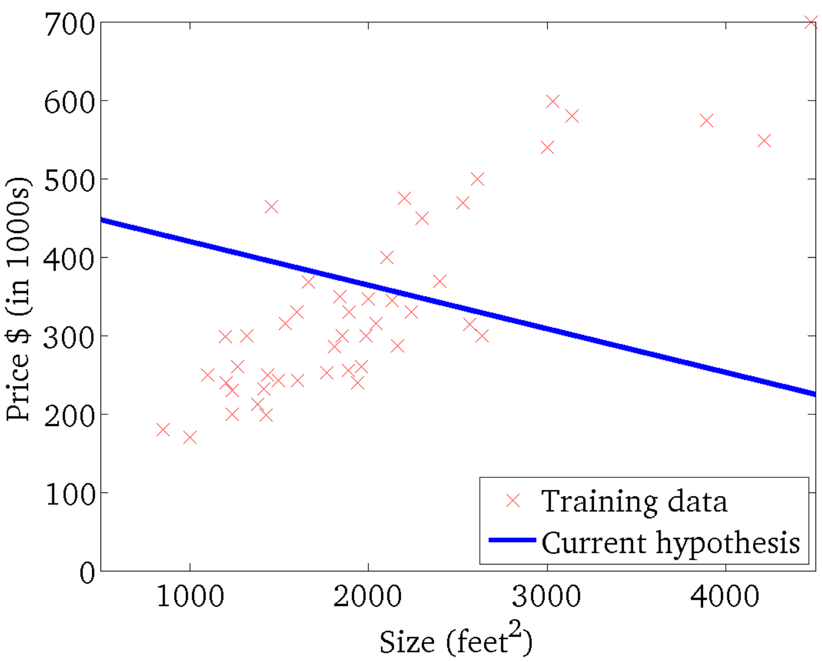
- Gradiente
  - Iteración 90

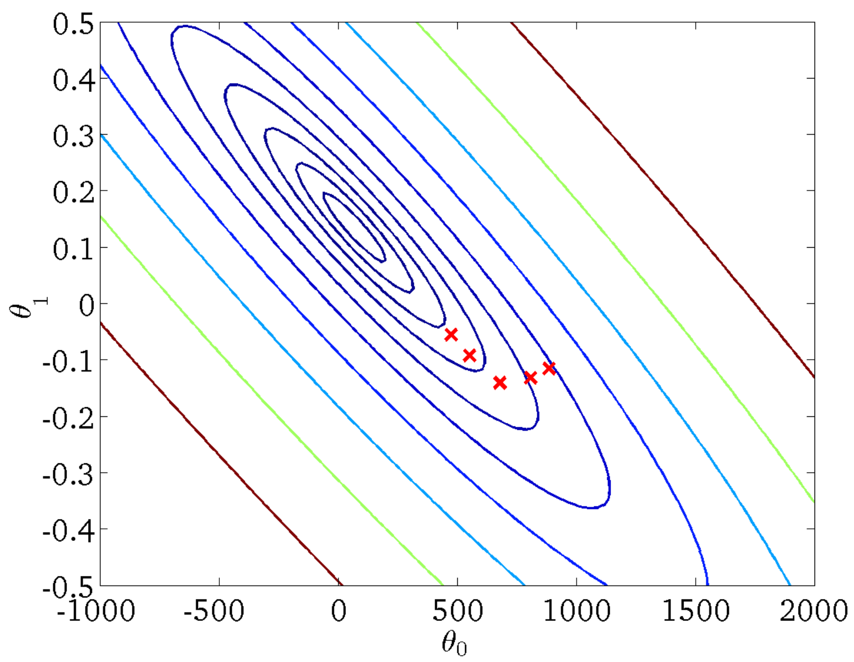




Regresión lineal con una variable

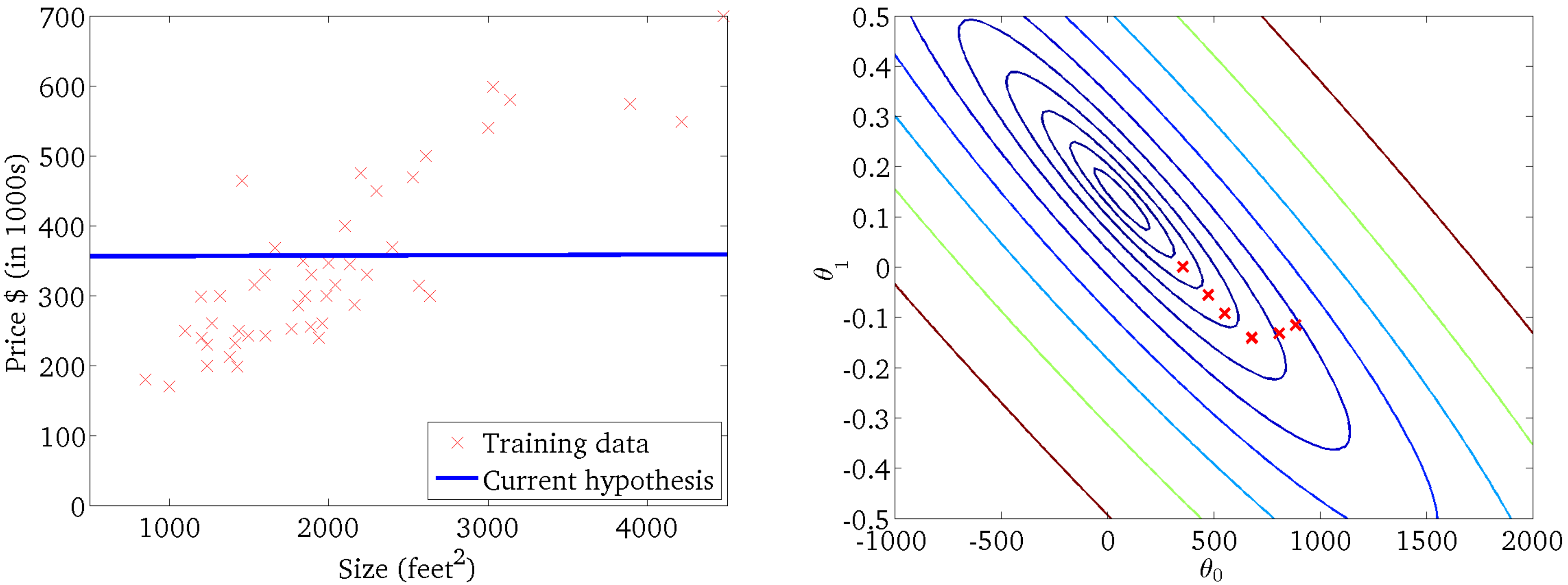
- Gradiente
  - Iteración 120





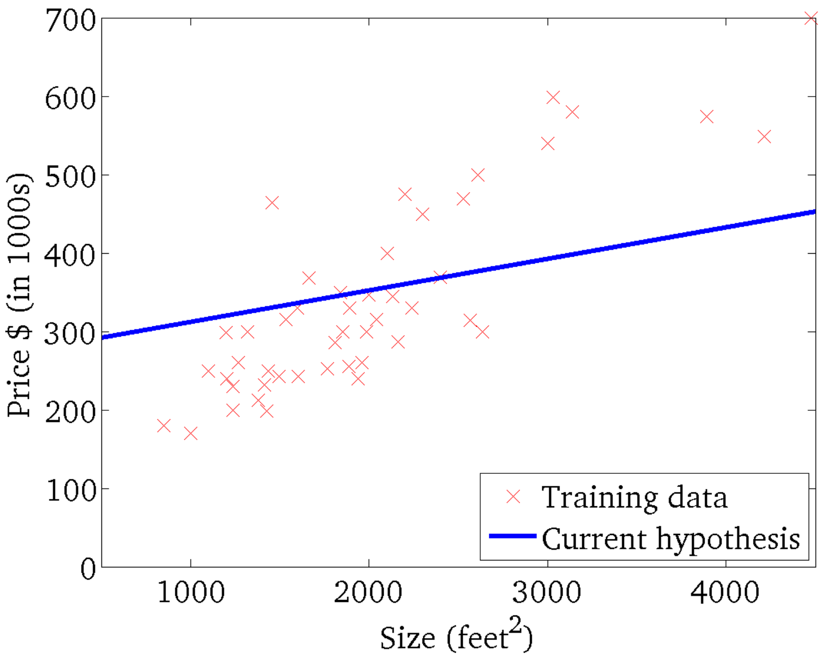
Regresión lineal con una variable

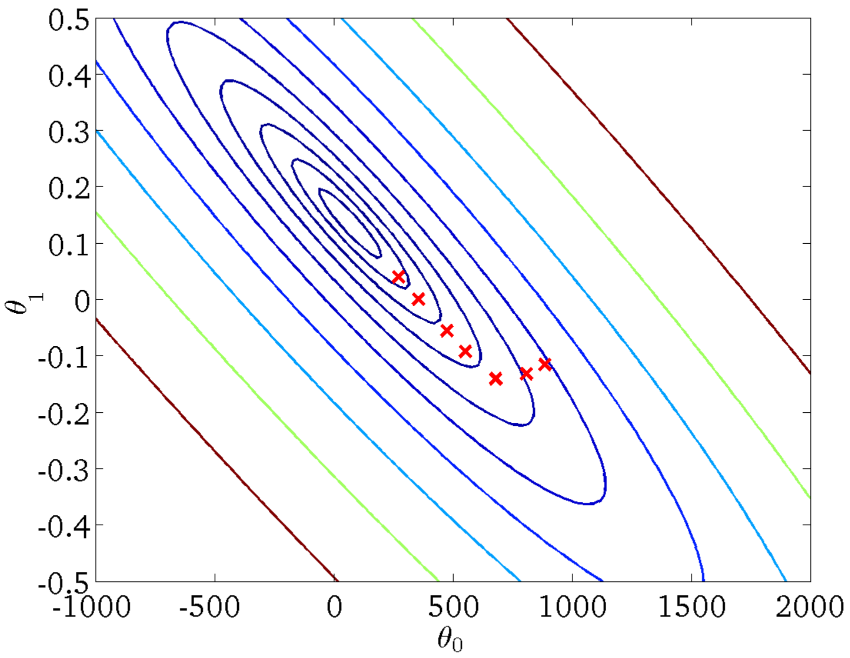
- Gradiente
  - Iteración 150



Regresión lineal con una variable

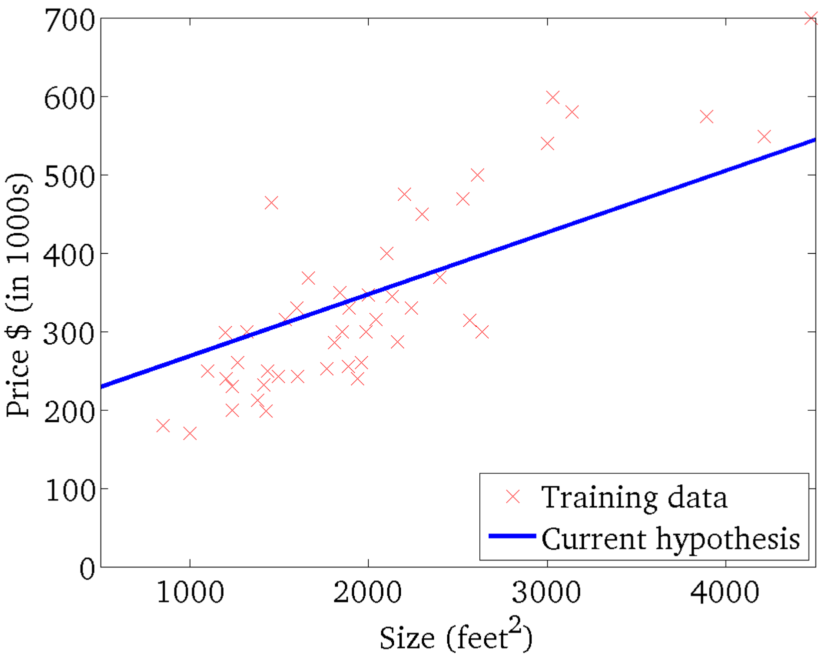
- Gradiente
  - Iteración 180

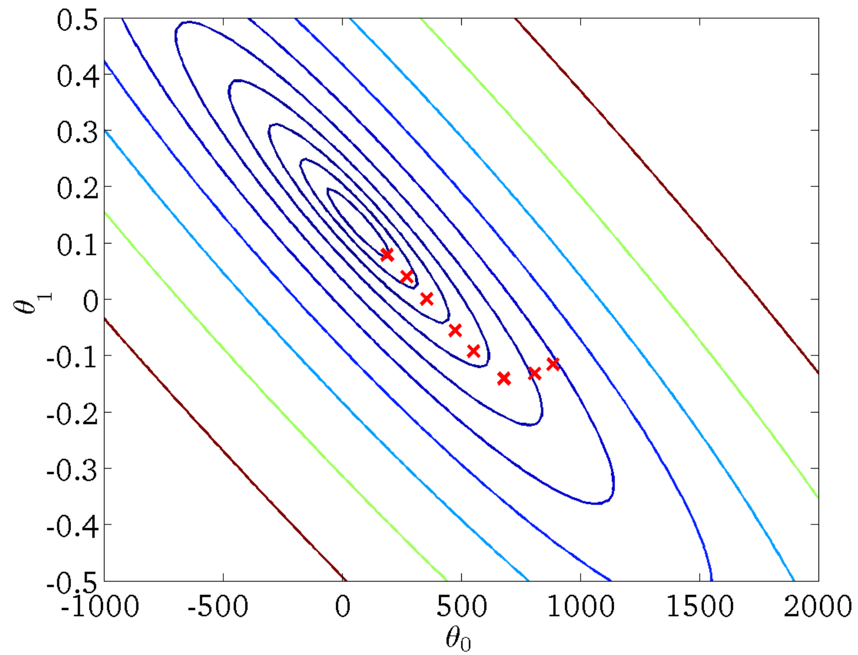




Regresión lineal con una variable

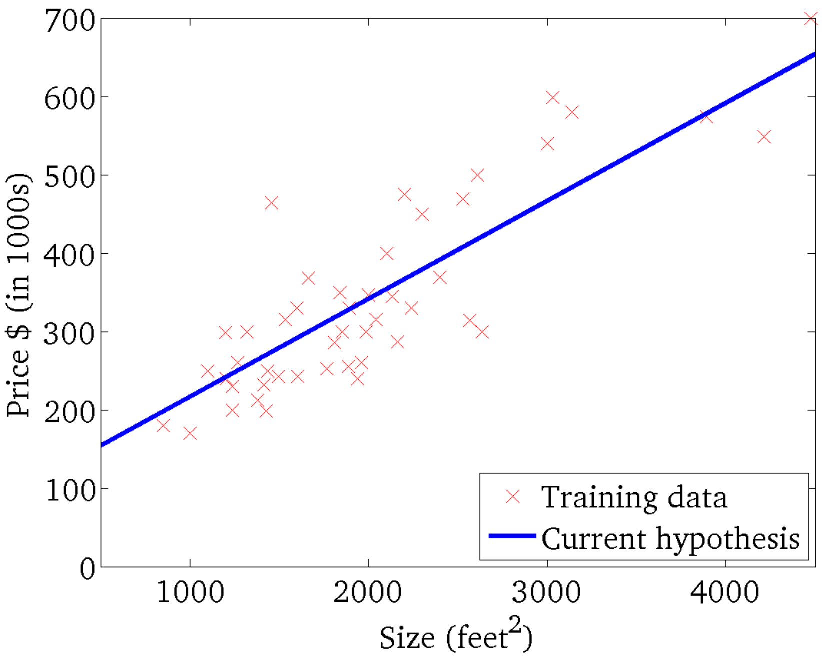
- Gradiente
  - Iteración 210

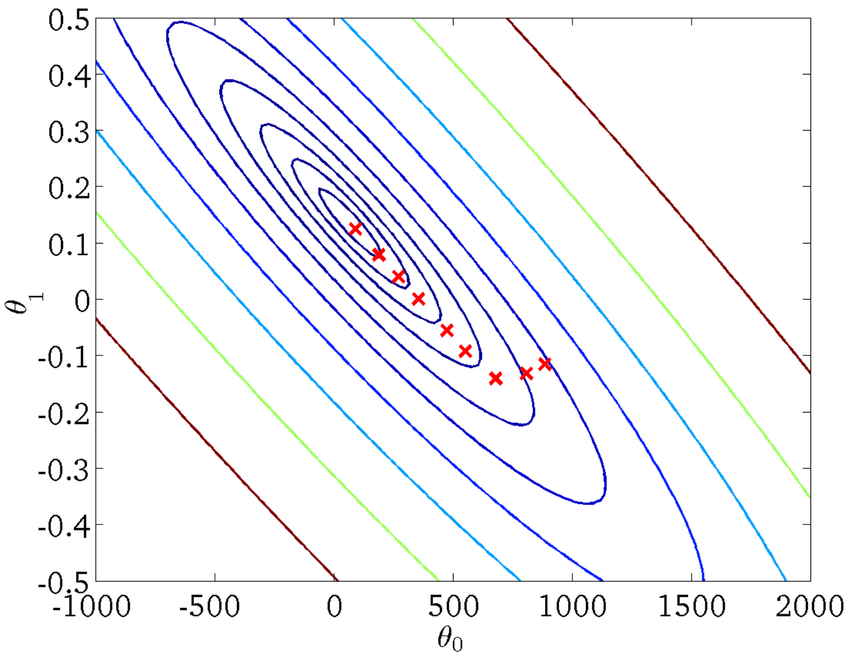




Regresión lineal con una variable

- Gradiente
  - Iteración 240



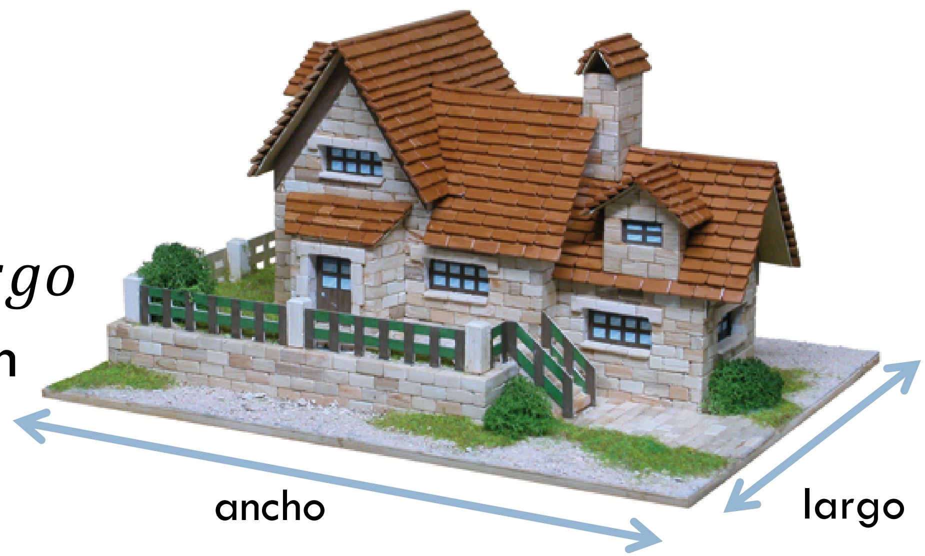


Regresión lineal

- Importancia de las características
  - **Elección de las características importante**
    - Puede ser la diferencia entre **un buen y un mal modelo**
  - Ejemplo
    - Tasación de inmueble

$$h_{\theta}(x) = \theta_0 + \theta_1 ancho + \theta_2 largo$$

Quizás tenga más sentido tasarlos en base al área…

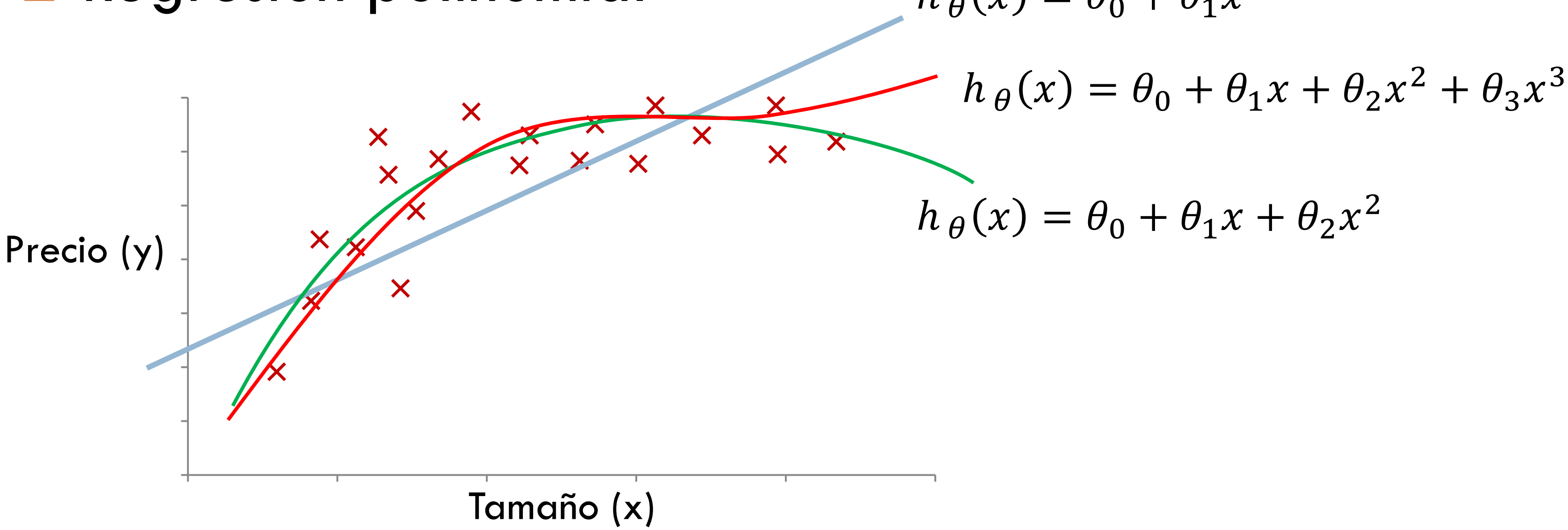


$$h_{\theta}(x) = \theta_0 + \theta_1 area$$

donde $area = ancho \cdot largo$

Regresión lineal

- Regresión polinomial



$$h_\theta(x) = \theta_0 + \theta_1 x_1 + \theta_2 x_2 + \theta_3 x_3$$
$$= \theta_0 + \theta_1(tamaño) + \theta_2(tamaño)^2 + \theta_3(tamaño)^3$$

$x_1 = (tamaño)$

$x_2 = (tamaño)^2$

$x_2 = (tamaño)^3$

Obtenemos características polinomiales a partir de la original

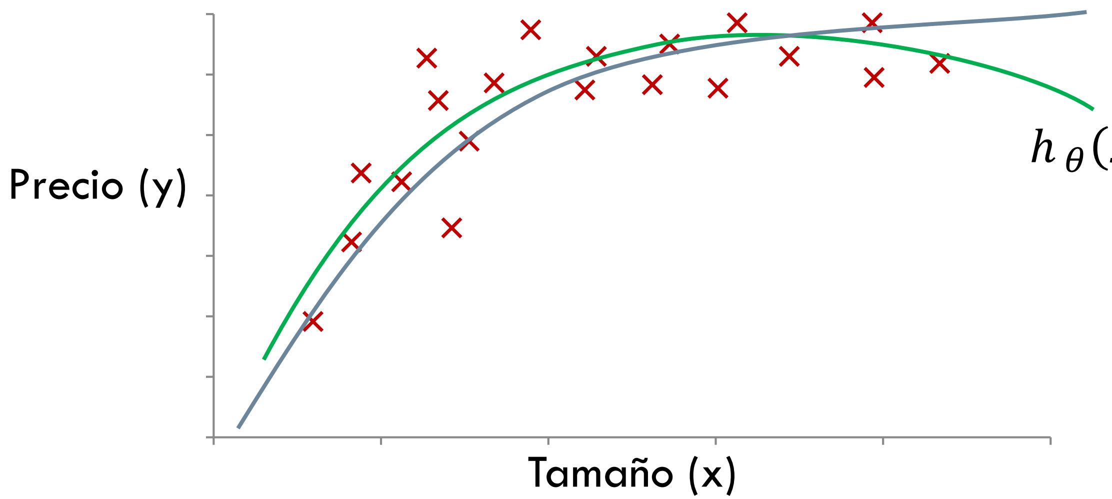
Regresión lineal

- Regresión polinomial
  - **Cuidado!!**
    - **La normalización se vuelve necesaria**
      - Tamaño [0 1000]
      - Tamaño^2 [0 1000 000]
      - Tamaño^3 [0 10^9]

Regresión lineal

- Regresión polinomial
  - Elegir las características adecuadas no es sencillo
  - Depende del problema

$$h_\theta(x) = \theta_0 + \theta_1 x + \theta_2 \sqrt{x}$$



$$h_\theta(x) = \theta_0 + \theta_1 x + \theta_2 x^2$$

$$h_\theta(x) = \theta_0 + \theta_1 (tamaño) + \theta_2 (tamaño)^2$$

$$h_\theta(x) = \theta_0 + \theta_1 (tamaño) + \theta_2 \sqrt{(tamaño)}$$

Los precios no decrecen según aumenta el tamaño

Regresión lineal

- **Descenso por gradiente vs. Solución directa**

- Descenso por gradiente
  - Seleccionar α
  - Algoritmo iterativo
  - Funciona bien incluso con n grande
  - Usar con $n > 10\,000$

- Solución directa
  - No requiere α
  - No iterativo
  - Necesita calcular $(X^T X)^{-1}$
    - Coste $O(n^3)$
  - Lento si n es muy grande
  - Usar con $n < 10\,000$

CLASIFICACIÓN MEDIANTE REGRESIÓN LOGÍSTICA

Mikel Galar Idoate
mikel.galar@unavarra.es

Ciencia de datos con técnicas inteligentes
Experto Universitario en Ciencia de Datos y Big Data para Business Intelligence

Índice

1. Regresión lineal (regresión)
2. Regresión logística (clasificación)
   - Objetivo
   - Modelo
   - Ajuste de parámetros
     - Función de coste
     - Descenso por gradiente

Clasificación

- **Objetivo**
  - Encontrar un **modelo** que represente un conjunto de datos observados
  - El modelo servirá para **predecir valores para nuevos datos**
- En **clasificación**…
  - **La salida es un valor discreto (clase/etiqueta)**
  - Ejemplo
    - Datos de entrada: tamaño del tumor, tonalidad del tumor
    - **Salida: Benigno o Maligno**

Clasificación

- **Dos tipos de clasificación**
  - Binaria
    - Consideran 2 clases
      - Benigno o Maligno
      - Llueve o no llueve
      - Spam o no spam
  - Multi-clase
    - Consideran más de 2 clases
      - Soleado, nuboso, lluvia, nieve
      - Spam, Trabajo, Personal

Matemáticamente

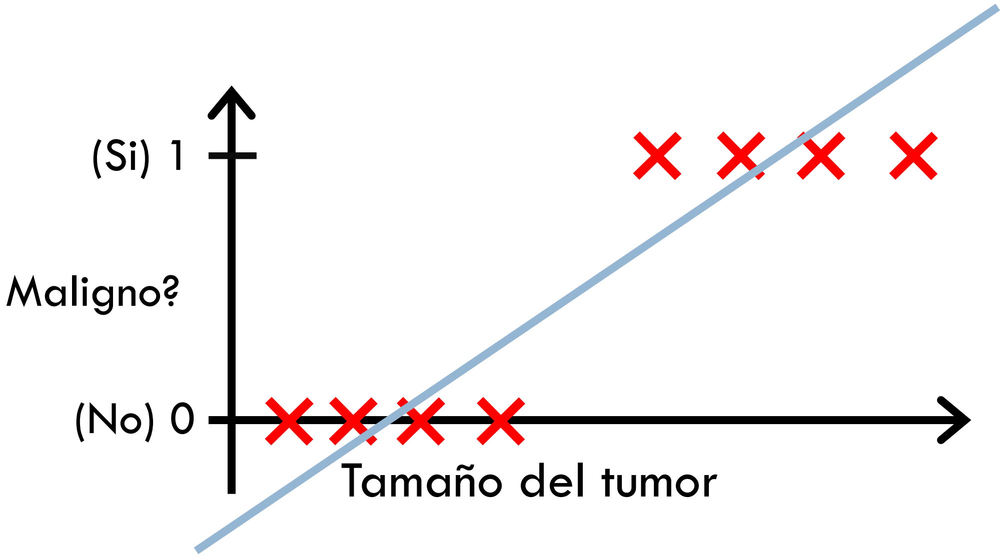
$$y \in \{0,1\}$$

$$y \in \{0,1,2,\dots,N\}$$

La regresión logística funciona para problemas binarios, para los multi-clase se debe utilizar junto con estrategias de descomposición (OVO-OVA)

Clasificación

- **¿Podemos clasificar mediante regresión lineal?**

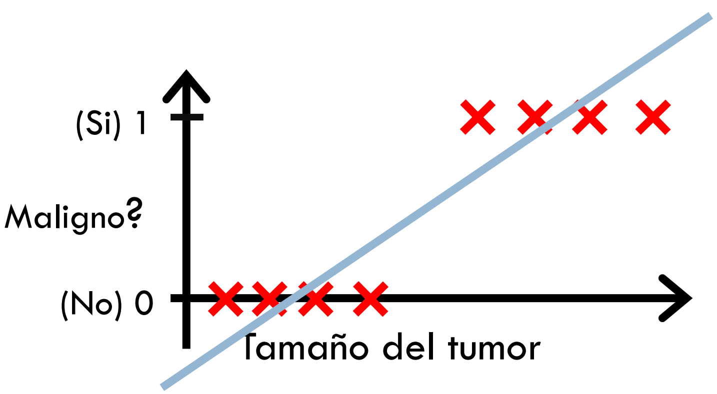


Tamaño (x)	Maligno? (y)
1	0
2	0
3	0
4	0
6	1
7	1
8	1
9	1

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

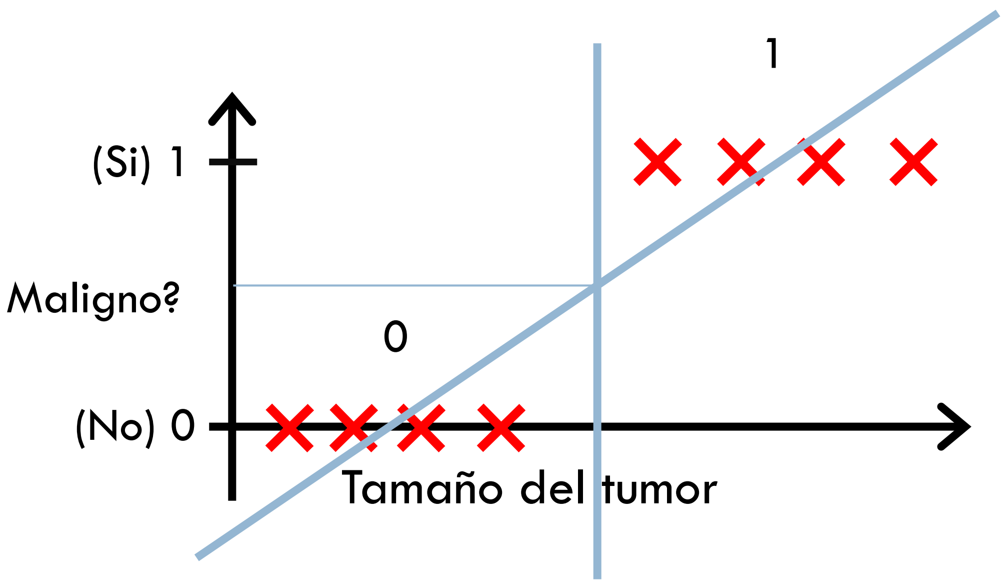
Clasificación

- **¿Podemos clasificar mediante regresión lineal?**
  - El resultado de la hipótesis $h_\theta(x)$ es un número real
    - No necesariamente 0 o 1
  - Tenemos que transformar la hipótesis tal que
    - $h_\theta(x) \in \{0, 1\}$
    - Predecir 0 si $h_\theta(x) < 0.5$
    - Predecir 1 si $h_\theta(x) \geq 0.5$



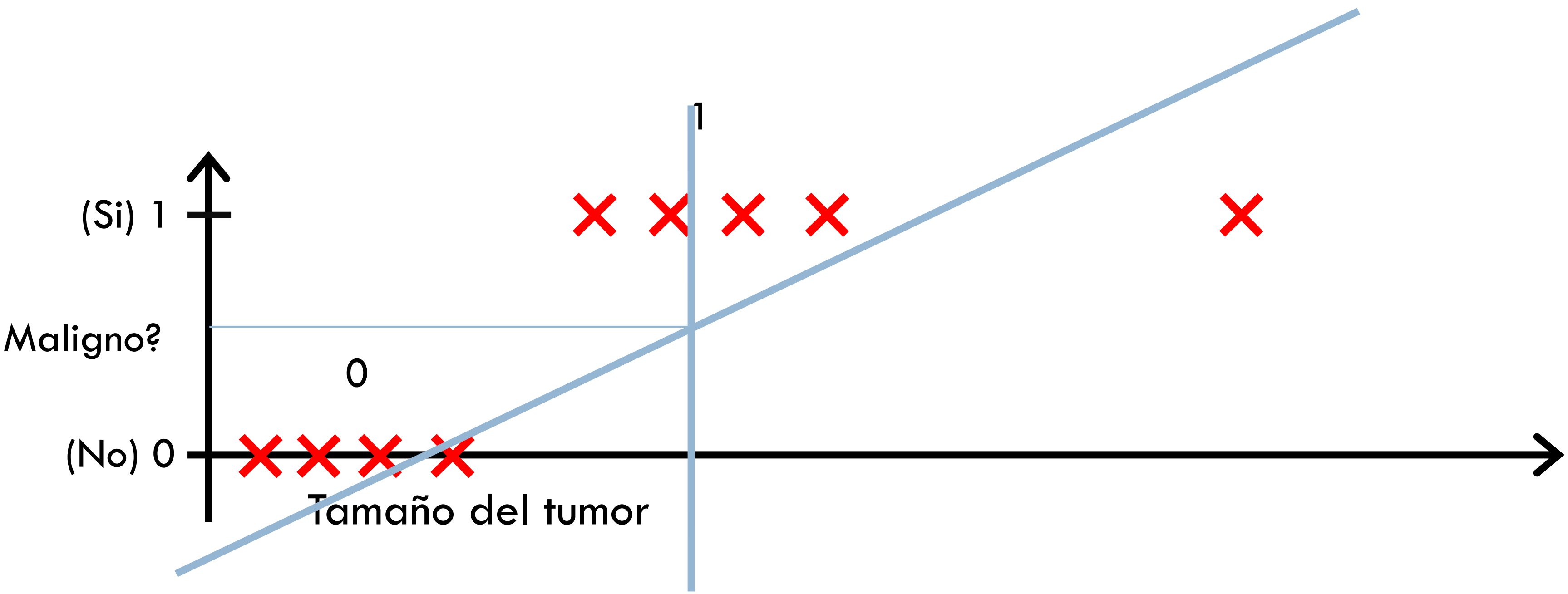
Clasificación

- **¿Podemos clasificar mediante regresión lineal?**
  - Todo lo que queda por debajo de 0.5
    - Lo clasificamos como 0
  - Todo lo que queda por encima de 0.5
    - Lo clasificamos como 1



Clasificación

- **¿Podemos clasificar mediante regresión lineal?**
  - ¿Y si añadimos un nuevo dato? $(12, 1)$
    - ¡Problema!
      - No aporta información realmente relevante
      - $h_\theta(x) < 0$ y $h_\theta(x) > 1$ →No tiene mucho sentido



Clasificación

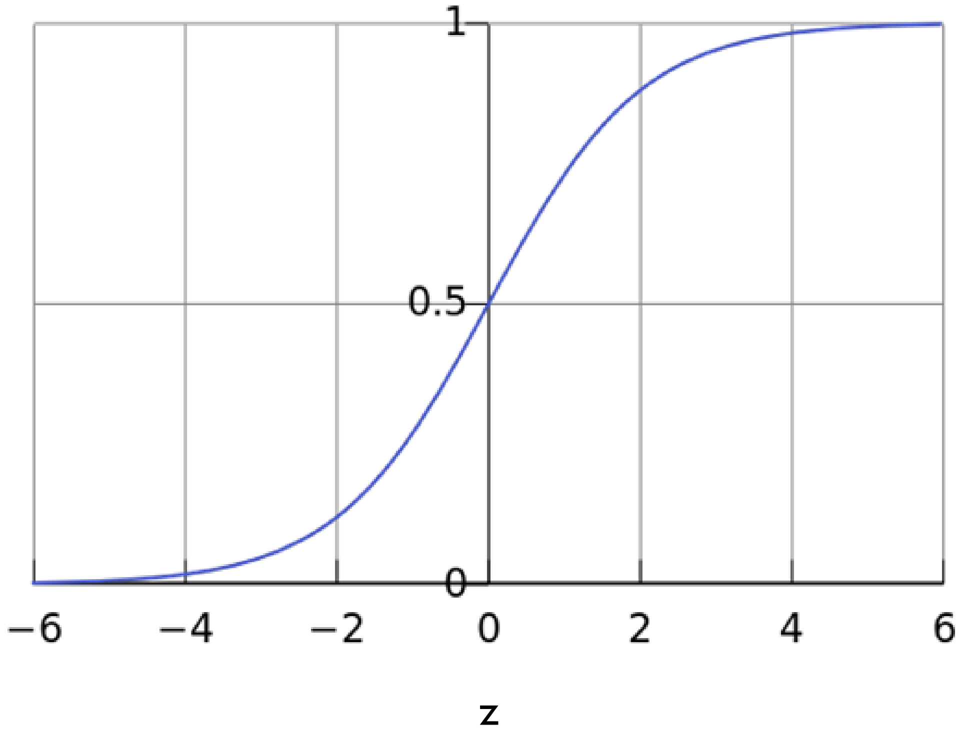
- **¿Podemos clasificar mediante regresión lineal?**
  - Necesitamos una hipótesis tal que

$$0 \leq h_\theta(x) \leq 1$$

  - **Regresión logística**

Clasificación: regresión logística

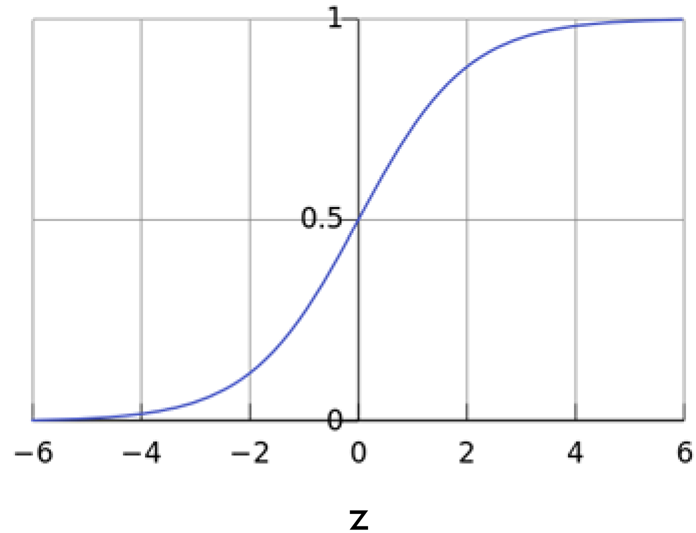
- **Regresión logística**
  - Introduce la función sigmoide (logística)
    - Obtenemos $0 \leq h_{\theta}(x) \leq 1$



Clasificación: regresión logística

- **Regresión logística**
  - Introduce la función sigmoide (logística)
    - Obtenemos $0 \leq h_\theta(x) \leq 1$

$$h_\theta(x) = g(\theta^T x) \qquad g(z) = \frac{1}{1+e^{-z}}$$



$$h_\theta(x) = \frac{1}{1+e^{-\theta^T x}}$$

Clasificación: regresión logística

- **Regresión logística**
  - Ejemplo
    - Modelo con una variable en regresión

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

    - Modelo con una variable en regresión logística

$$h_{\theta}(x) = \frac{1}{1 + e^{-\theta_0 - \theta_1 x}}$$

Clasificación: regresión logística

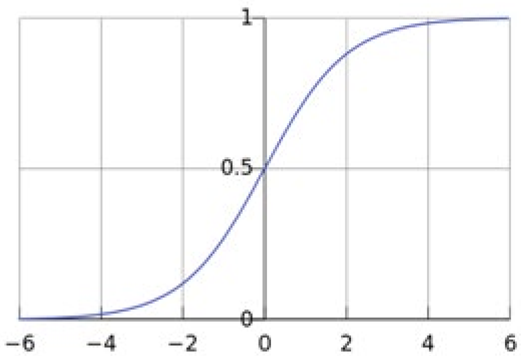
- **Regresión logística**
  - Hipótesis relacionada con la probabilidad
  - Interpretación

$$h_\theta(x) = \frac{1}{1+e^{-\theta^T x}}$$

    - "Dado un valor x, $h_\theta(x)$ es la probabilidad de que $y = 1$, con los parámetros θ"
      - Es decir, $\boldsymbol{h_\theta(x) = P(y = 1|x;\theta)}$
      - Por tanto… $P(y = 1|x;\theta) + P(y = 0|x;\theta) = 1$
        - $\boldsymbol{P(y = 0|x;\theta)} = 1 - P(y = 1|x;\theta) = \boldsymbol{1 - h_\theta(x)}$
    - Ejemplo
      - Si $h_\theta(x) = 0.7$
        - Existe un 70% de probabilidad de que x pertenezca a la clase $y = 1$

Clasificación: regresión logística

- **Frontera de decisión**
  - Dado que $h_\theta(x)$ es una probabilidad
    - Podemos asumir que
      - Predecimos $y = 1$ si $h_\theta(x) \geq 0.5$
      - Predecimos $y = 0$ si $h_\theta(x) < 0.5$
  - ¿Cuándo ocurre esto?
    - $h_\theta(x) \geq 0.5$ si $g(\theta^T x) \geq 0.5$
      - Es decir, si $\theta^T x \geq 0$
    - $h_\theta(x) < 0.5$ si $g(\theta^T x) < 0.5$
      - Es decir, si $\theta^T x < 0$

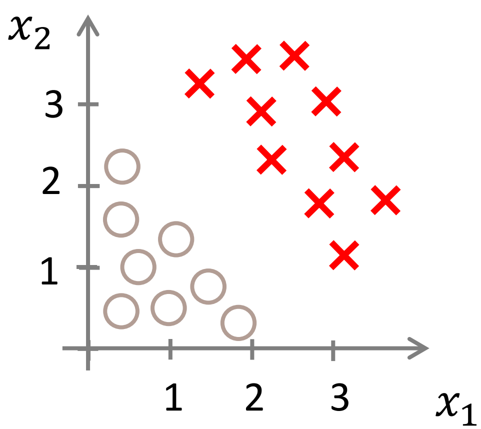


$$h_\theta(x) = g(\theta^T x) = \frac{1}{1 + e^{-\theta^T x}}$$

La frontera de decisión está en los puntos donde $\theta^T x = 0$

Clasificación: regresión logística

- **Frontera de decisión**
  - Un ejemplo
    - Círculos → $y = 0$
    - Aspas→ $y = 1$
    - *Tenemos dos variables x_1, x_2 pero podemos representar los datos en 2 dimensiones con diferentes marcadores para cada clase*
  - ¿Cómo afectan los valores de los parámetros θ al modelo?



Clasificación: regresión logística

- **Frontera de decisión**
  - ¿Cómo afectan los valores de los parámetros θ al modelo?

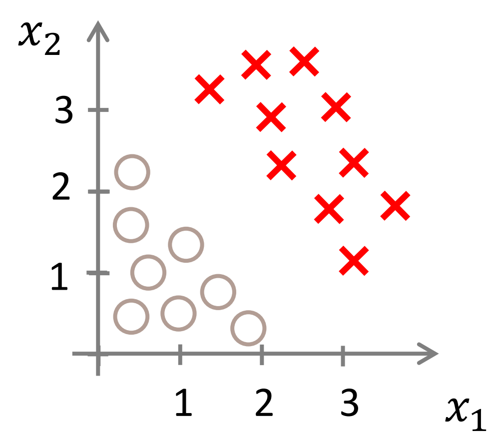
$$h_\theta(x) = g(\theta_0 + \theta_1 x_1 + \theta_2 x_2) = \frac{1}{1 + e^{-\theta_0 - \theta_1 x_1 - \theta_2 x_2}}$$

  - Recordemos que predecimos aspas ($y = 1$) si

$$h_\theta(x) = \frac{1}{1 + e^{-\theta_0 - \theta_1 x_1 - \theta_2 x_2}} > 0.5$$

$$\Downarrow$$

$$\theta_0 + \theta_1 x_1 + \theta_2 x_2 \geq 0$$

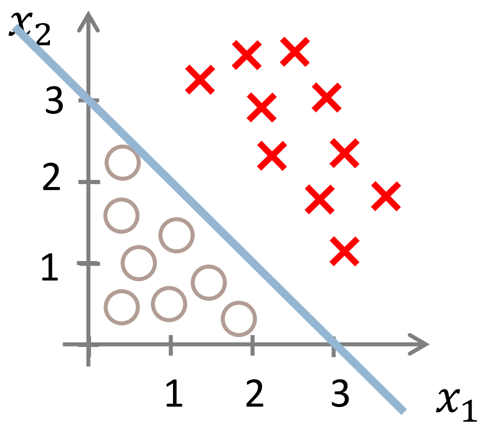


Clasificación: regresión logística

- **Frontera de decisión**
  - ¿Cómo afectan los valores de los parámetros θ al modelo?

$$h_{\theta}(x) = g(\theta_0 + \theta_1 x_1 + \theta_2 x_2) = \frac{1}{1 + e^{-\theta_0 - \theta_1 x_1 - \theta_2 x_2}}$$

  - Tomamos
    - $\theta_0 = -3, \theta_1 = 1, \theta_2 = 1$
    - Predecimos $y = 1$ si $-3 + x_1 + x_2 \geq 0$
      - $x_1 + x_2 \geq 3$



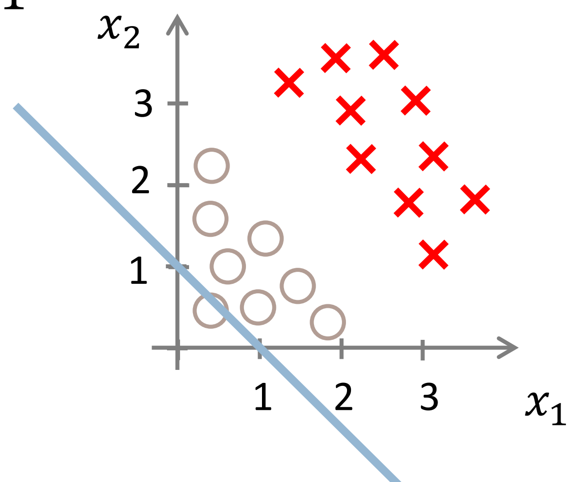
Buena clasificación!

Clasificación: regresión logística

- **Frontera de decisión**
  - ¿Cómo afectan los valores de los parámetros θ al modelo?

$$h_\theta(x) = g(\theta_0 + \theta_1 x_1 + \theta_2 x_2) = \frac{1}{1 + e^{-\theta_0 - \theta_1 x_1 - \theta_2 x_2}}$$

  - Tomamos
    - $\theta_0 = -1, \theta_1 = 1, \theta_2 = 1$
    - Predecimos $y = 1$ si $-1 + x_1 + x_2 \geq 0$
      - $x_1 + x_2 \geq 1$



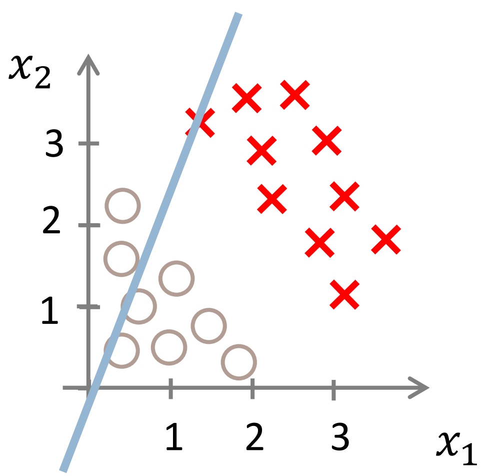
Mala clasificación!

Clasificación: regresión logística

- **Frontera de decisión**
  - ¿Cómo afectan los valores de los parámetros θ al modelo?

$$h_\theta(x) = g(\theta_0 + \theta_1 x_1 + \theta_2 x_2) = \frac{1}{1 + e^{-\theta_0 - \theta_1 x_1 - \theta_2 x_2}}$$

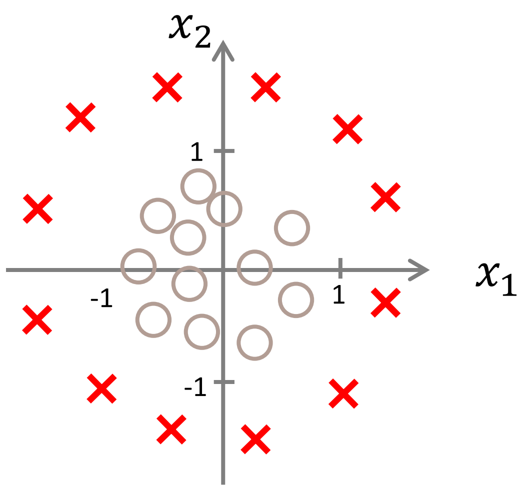
  - Tomamos
    - $\theta_0 = 0, \theta_1 = 2, \theta_2 = -1$
    - Predecimos $y = 1$ si $2x_1 - x_2 \geq 0$
      - $2x_1 - x_2 \geq 0$



Mala clasificación!

Clasificación: regresión logística

- **Frontera de decisión no lineal**



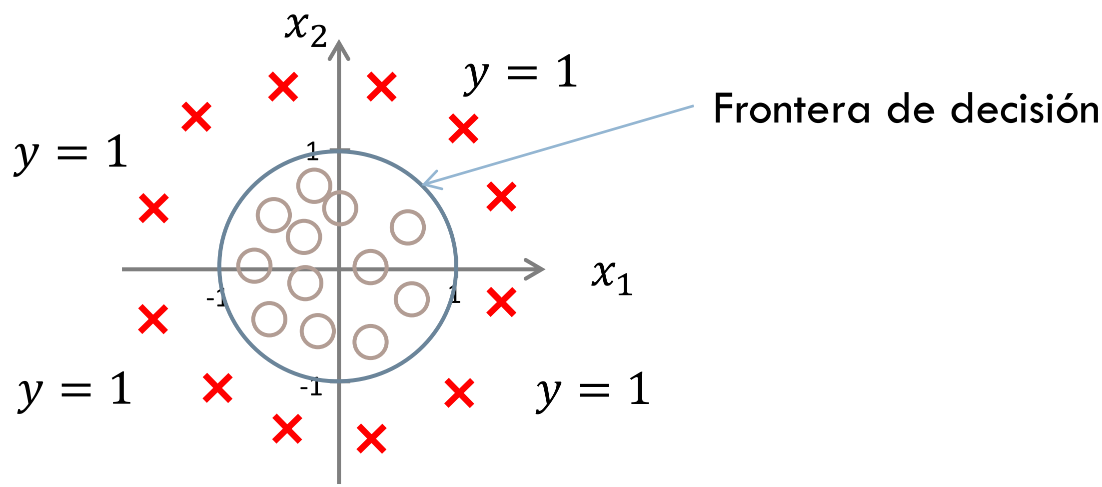
- Necesitamos añadir características polinomiales
  - **Como en regresión lineal**

$$h_\theta(x) = g(\theta_0 + \theta_1 x_1 + \theta_2 x_2 + \theta_3 x_1^2 + \theta_4 x_2^2)$$

Clasificación: regresión logística

- **Frontera de decisión no lineal**

$$h_\theta(x) = g(\theta_0 + \theta_1 x_1 + \theta_2 x_2 + \theta_3 x_1^2 + \theta_4 x_2^2)$$

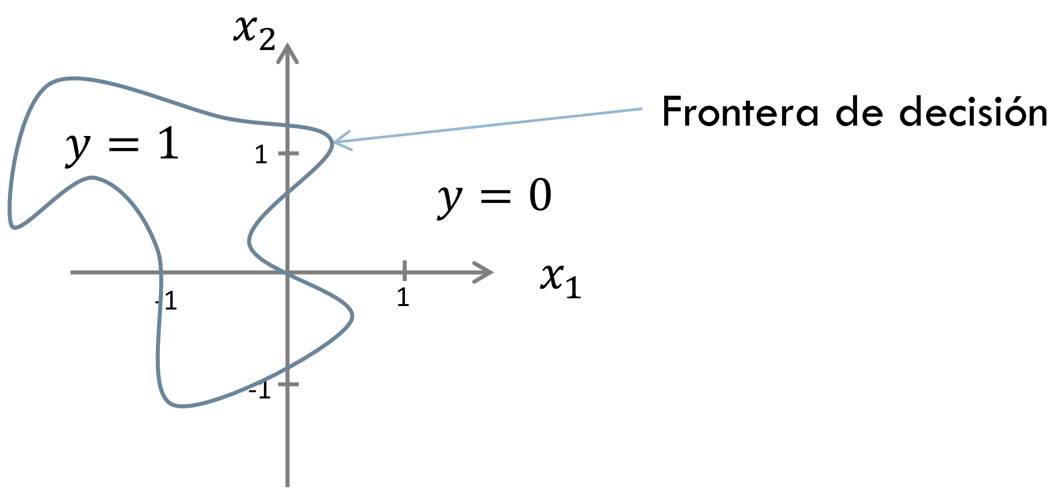


- Tomamos
  - $\theta_0 = -1, \theta_1 = 0, \theta_2 = 0, \theta_3 = 1, \theta_4 = 1$
  - Predecimos $y = 1$ si $-1 + x_1^2 + x_2^2 \geq 0$
    - $x_1^2 + x_2^2 \geq 1$ → Ecuación de la circunferencia! (para radio = 1)

Clasificación: regresión logística

- **Frontera de decisión no lineal**
  - Con polinomios de orden mayor
    - Podemos obtener fronteras más complejas

$$h_\theta(x) = g(\theta_0 + \theta_1 x_1 + \theta_2 x_2 + \theta_3 x_1^2 + \theta_4 x_2^2 + \theta_5 x_1^2 x_x^2 + \theta_6 x_1^3 x_2 + \cdots)$$



Clasificación: regresión logística

¿Cómo obtenemos los parámetros θ?

Clasificación: regresión logística

- **Función de coste**
  - Conjunto de datos

$$\{(x^{(1)}, y^{(1)}), (x^{(2)}, y^{(2)}), \dots , (x^{(m)}, y^{(m)})\}$$

$$x = \begin{bmatrix} x_0 \\ x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} \qquad x_0 = 1, y \in \{0, 1\}$$

$$h_{\theta}(x) = g(\theta^T x) = \frac{1}{1 + e^{-\theta^T x}}$$

Necesitamos una función para medir el error

Clasificación: regresión logística

Función de coste

- ¿Podemos usar la función de coste de la regresión lineal?

$$J(\theta) = \frac{1}{2m}\sum_{i=1}^{m}\left(h_\theta(x^{(i)}) - y^{(i)}\right)^2 \qquad \text{donde} \qquad h_\theta(x) = \frac{1}{1+e^{-\theta^T x}}$$

- Conjunto de datos

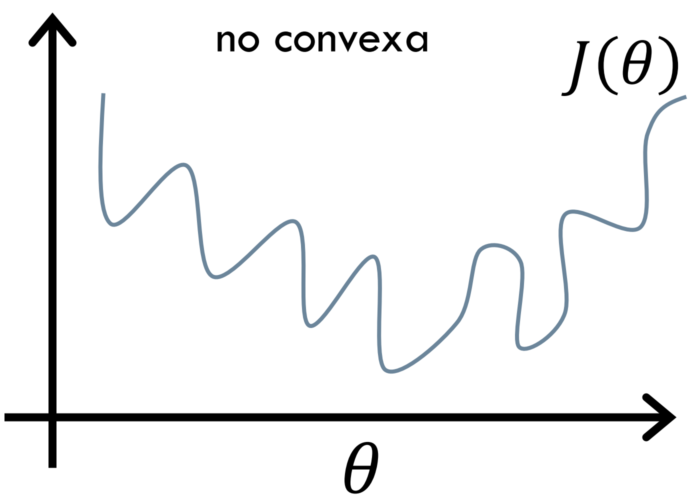
$$\{(1,0), (2,0), (3,1)\}$$

- Salidas de una posible hipótesis (3 fallos)
  - $h_\theta(1) = 0.99, h_\theta(2) = 0.51, h_\theta(3) = 0.38 \quad J(\theta) = \boxed{0.2708}$
- Salidas de una posible hipótesis (2 fallos)
  - $h_\theta(1) = 0.85, h_\theta(2) = 0.05, h_\theta(3) = 0.05 \quad J(\theta) = \boxed{0.2713}$

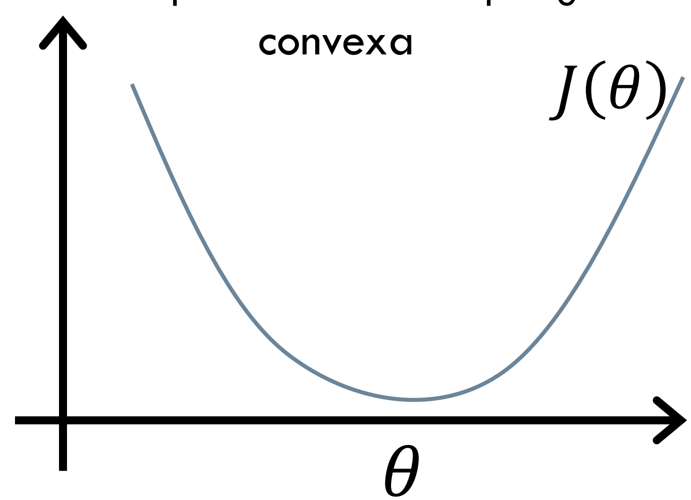
¡¡Más error a pesar de acertar más ejemplos!!

Clasificación: regresión logística

- **Función de coste**
  - El problema surge porque en este caso debido a $h_\theta(x^{(i)}), J(\theta)$ es una función compleja no convexa
    - Con muchos **mínimos locales**
      - ¡No podemos aplicar el descenso por gradiente!



Podemos aplicar el descenso por gradiente



Clasificación: regresión logística

Función de coste

- Por tanto debemos buscar otra función de coste

$$J(\theta) = \frac{1}{2m}\sum_{i=1}^{m}\left(h_\theta(x^{(i)}) - y^{(i)}\right)^2 \quad \text{donde} \quad h_\theta(x) = \frac{1}{1+e^{-\theta^T x}}$$

$$J(\theta) = \frac{1}{2m}\sum_{i=1}^{m} coste\left(h_\theta(x^{(i)}), y^{(i)}\right) \quad \text{donde} \quad h_\theta(x) = \frac{1}{1+e^{-\theta^T x}}$$

  - Vamos a trabajar con un único ejemplo
  - $coste(h_\theta(x), y)$ nos indica cuánto penalizamos el error de la salida de la hipótesis respecto a la salida real (y)

$$coste\left(h_\theta(x^{(i)}), y^{(i)}\right) = \left(h_\theta(x^{(i)}) - y^{(i)}\right)^2$$

En regresión logística debemos utilizar otro coste diferente

Clasificación: regresión logística

- **Función de coste**
  - Queremos una función de coste que cumpla
    - Si $y = 0$
      - Si $h_\theta(x) = 1$ → Aumentar el coste (error!!)
      - Según $h_\theta(x)$ tiende a 0, disminuir el coste (acertamos)
    - Si $y = 1$
      - Si $h_\theta(x) = 0$ → Aumentar el coste (error!!)
      - Según $h_\theta(x)$ tiende a 1, disminuir el coste (acertamos)
  - Coste

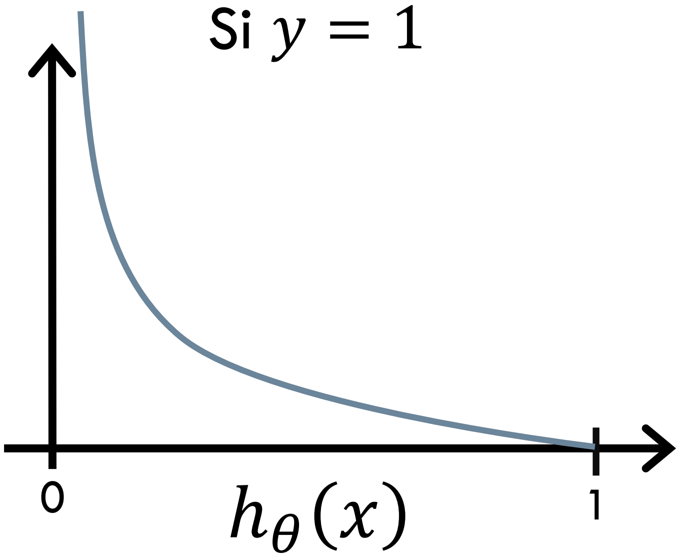
$$coste(h_\theta(x), y) = \begin{cases} -\log(h_\theta(x)) & si\ y = 1 \\ -\log(1 - h_\theta(x)) & si\ y = 0 \end{cases}$$

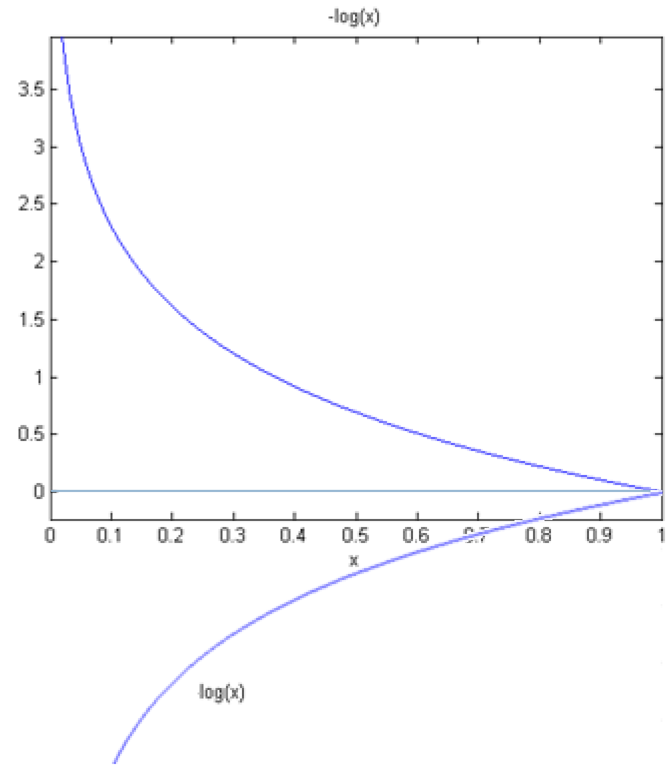
Clasificación: regresión logística

- **Función de coste**

$$coste(h_\theta(x), y) = \begin{cases} -\log(h_\theta(x)) & si\ y = 1 \\ -\log(1 - h_\theta(x)) & si\ y = 0 \end{cases}$$

Si $y = 1$



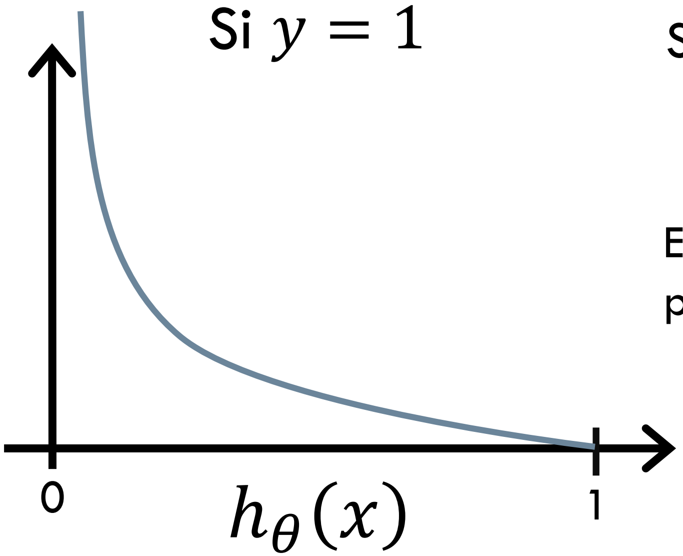


Clasificación: regresión logística

- **Función de coste**

$$coste(h_\theta(x), y) = \begin{cases} -\log(h_\theta(x)) & si\ y = 1 \\ -\log(1 - h_\theta(x)) & si\ y = 0 \end{cases}$$

Si $y = 1$



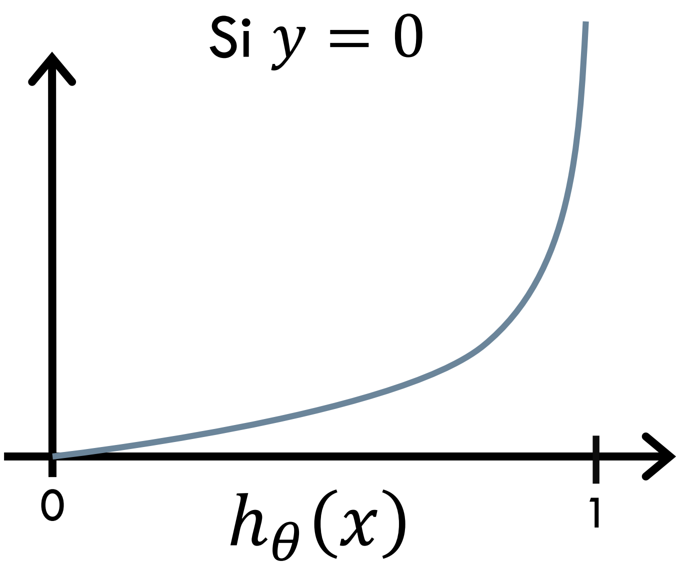
$coste(h_\theta(x), y) = 0$ si $y = 1, h_\theta(x) = 1$
Según $h_\theta(x) \rightarrow 0$
$coste(h_\theta(x), y) \rightarrow \infty$

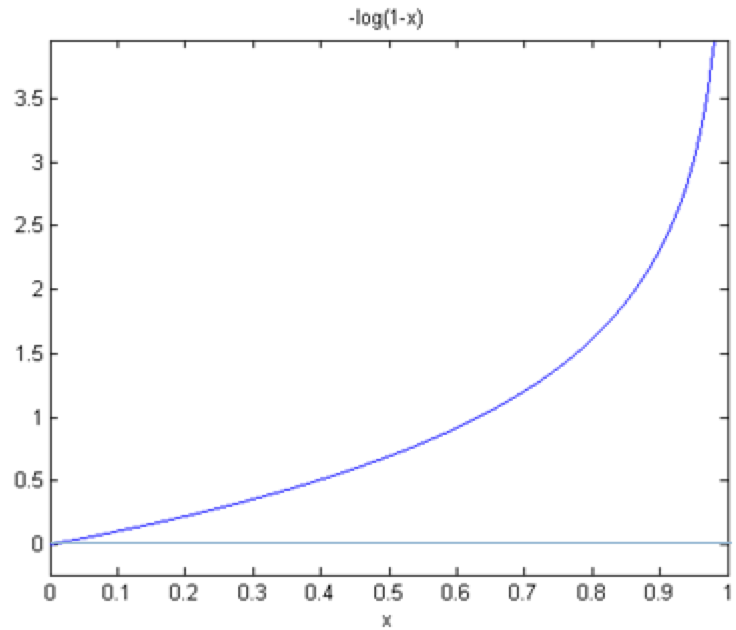
Es decir, si $h_\theta(x) = 0$ (predecimos $P(y = 1|x; \theta) = 0$), pero $y = 1$ lo penalizamos con un alto coste

Clasificación: regresión logística

- **Función de coste**

$$coste(h_\theta(x), y) = \begin{cases} -\log(h_\theta(x)) & si\ y = 1 \\ -\log(1 - h_\theta(x)) & si\ y = 0 \end{cases}$$

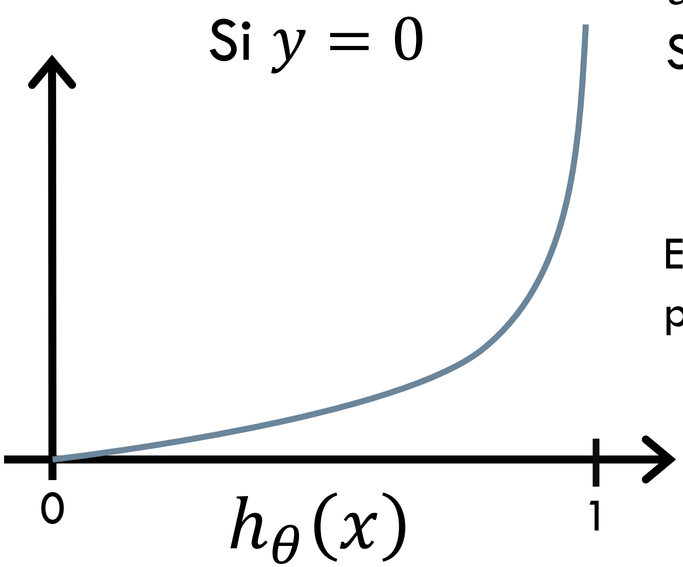




Clasificación: regresión logística

- **Función de coste**

$$coste(h_\theta(x), y) = \begin{cases} -\log(h_\theta(x)) & si\ y = 1 \\ -\log(1 - h_\theta(x)) & si\ y = 0 \end{cases}$$



$coste(h_\theta(x), y) = 0$ si $y = 0, h_\theta(x) = 0$
Según $h_\theta(x) \to 1$
$coste(h_\theta(x), y) \to \infty$

Es decir, si $h_\theta(x) = 1$ (predecimos $P(y = 0|x; \theta) = 0$), pero $y = 0$ lo penalizamos con un alto coste

Clasificación: regresión logística

- **Función de coste**
  - Volvamos al problema anterior
  - Conjunto de datos

$$\{(1,0),\ (2,0),\ (3,1)\}$$

  - Salidas de una posible hipótesis (3 fallos)
    - $h_\theta(1) = 0.99, h_\theta(2) = 0.51, h_\theta(3) = 0.38$
      - $J(\theta) = -\log(0.01) - \log(0.49) - \log(0.38) = 6.2861$
  - Salidas de una posible hipótesis (2 fallos)
    - $h_\theta(1) = 0.85, h_\theta(2) = 0.05, h_\theta(3) = 0.05$
      - $J(\theta) = -\log(0.15) - \log(0.95) - \log(0.05) = 4.9441$

¡¡Ahora sí la función es lógica!!

La función de error es convexa…

Clasificación: regresión logística

- **Función de coste**
  - Podemos escribir la función de una forma más sencilla

$$coste(h_\theta(x), y) = \begin{cases} -\log(h_\theta(x)) & si\ y = 1 \\ -\log(1 - h_\theta(x)) & si\ y = 0 \end{cases}$$

  - Sabemos que $y \in \{0, 1\}$ es decir, $y = 0$ o $y = 1$

$$coste(h_\theta(x), y) = y\left(-\log\left(h_\theta(x)\right)\right) + (1 - y)(-\log(1 - h_\theta(x)))$$

$$coste(h_\theta(x), y) = -y\log\left(h_\theta(x)\right) - (1 - y)\log(1 - h_\theta(x))$$

Clasificación: regresión logística

- **Función de coste**
  - Podemos escribir la función de una forma más sencilla

$$coste(h_\theta(x), y) = \begin{cases} -\log(h_\theta(x)) & si\ y = 1 \\ -\log(1 - h_\theta(x)) & si\ y = 0 \end{cases}$$

  - Sabemos que $y \in \{0, 1\}$ es decir, $y = 0$ o $y = 1$

$$coste(h_\theta(x), y) = -y \log\big(h_\theta(x)\big) - (1 - y)\log(1 - h_\theta(x))$$

Si $y = 0$, $coste(h_\theta(x), y) = -\log\big(1 - h_\theta(x)\big)$
ya que la primera parte se anula

Si $y = 1$, $coste(h_\theta(x), y) = -\log\big(h_\theta(x)\big)$
ya que la segunda parte se anula

Clasificación: regresión logística

- **Función de coste**
  - Ya podemos escribir la función de coste para m ejemplos

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

(el 2 no lo necesitaremos)

$$coste(h_\theta(x), y) = -y\log\left(h_\theta(x)\right) - (1-y)\log\left(1-h_\theta(x)\right)$$

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

Clasificación: regresión logística

- **Función de coste**

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

  - Solo queda ajustar los parámetros θ

$$\min_{\theta} J(\theta)$$

¡Descenso por gradiente!

  - Para realizar nuevas predicciones

$$h_{\theta}(x) = \frac{1}{1 + e^{-\theta^T x}}$$

Clasificación: regresión logística

- **Descenso por gradiente**
  - Asignar a $\theta = \{\theta_0, \dots, \theta_n\}$ valores aleatorios
  - Repetir hasta convergencia
    - (nº iteraciones o |error – error_anterior| < umbral)

$$\boxed{\theta_j := \theta_j - \alpha \frac{\partial J(\theta)}{\partial \theta_j} \quad \text{para } j = 0, \dots, n}$$

ACTUALIZAR TODOS LOS θ_j SIMULTÁNEAMENTE

En regresión logística…

$$h_\theta(x) = \frac{1}{1 + e^{-\theta^T x}}$$

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

Clasificación: regresión logística

- **Descenso por gradiente en regresión logística**
  - Asignar a θ valores aleatorios (o a ceros)
  - Repetir hasta convergencia
    - (nº iteraciones o |error – error_anterior| < umbral)

$$\theta_j := \theta_j - \alpha \frac{1}{m} \sum_{i=1}^{m} \left(h_\theta\left(x^{(i)}\right) - y^{(i)}\right) \cdot x_j^{(i)} \quad \textbf{para todo } \boldsymbol{j} = \mathbf{0}, \dots, \boldsymbol{n}$$

ACTUALIZAR TODOS LOS θ_j SIMULTÁNEAMENTE

Clasificación: regresión logística

- No tenemos ningún problema en calcular

$$\boxed{\frac{\partial J(\theta)}{\partial \theta_j} = \frac{1}{m}\sum_{i=1}^{m}\left(h_\theta\left(x^{(i)}\right) - y^{(i)}\right)\cdot x_j^{(i)} \qquad \textbf{para todo } \boldsymbol{j} = \mathbf{0}, \dots, \boldsymbol{n}}$$

  - Ya que conocemos
    - Los **ejemplos**
    - Los **valores de θ actuales**
  - Calculamos todos los valores de manera simultánea
  - **¡¡Igual que la regresión lineal!!**
  - **Pero ojo…**

$$h_\theta(x) = \frac{1}{1 + e^{-\theta^T x}}$$