Inteligencia Artificial

Dr. Daniel R. Ramírez Rebollo

Introducción

- G.W. Liebnitz (1646-1716) he was first who said – brain is based on mathematics

A. M. Turing - (1912-1954) → *Computers and Intelligence*
“Turing test“

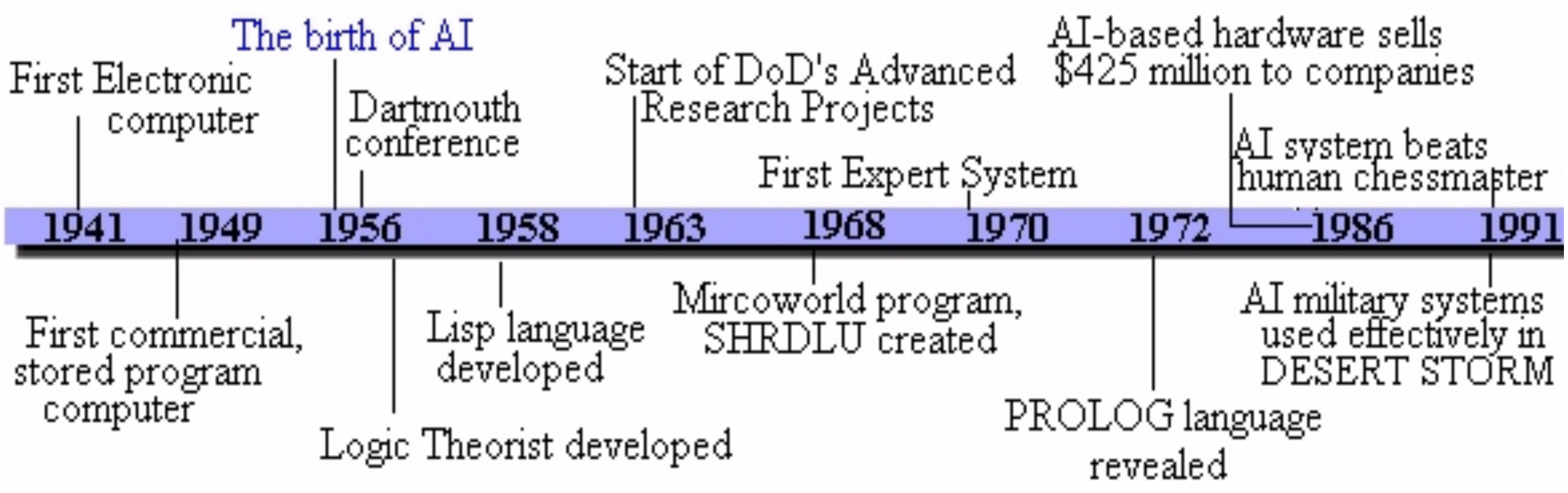
W. McCulloch W. Pitts (´43) → *Artificial neural network*

N. Wiener (1948) → *Book „Cybernetics“ - > AI*

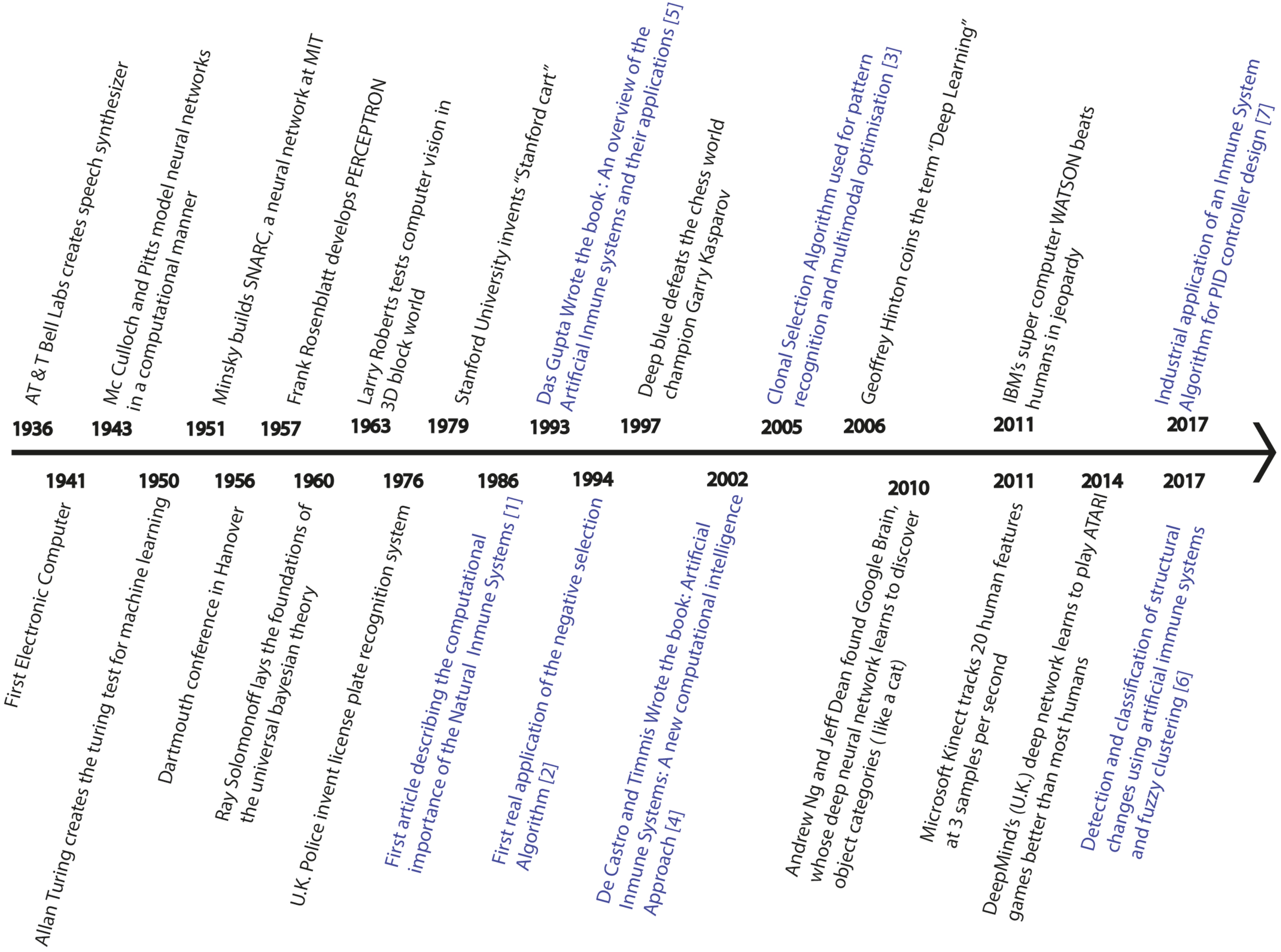
C. Shannon (1953)
Ashby (1952) → *How you program the machine*
To achieve ability to learn ???

Univ. Dartmouthe (´56) → *Establishing AI – as research Field*

Línea del Tiempo



Línea del Tiempo



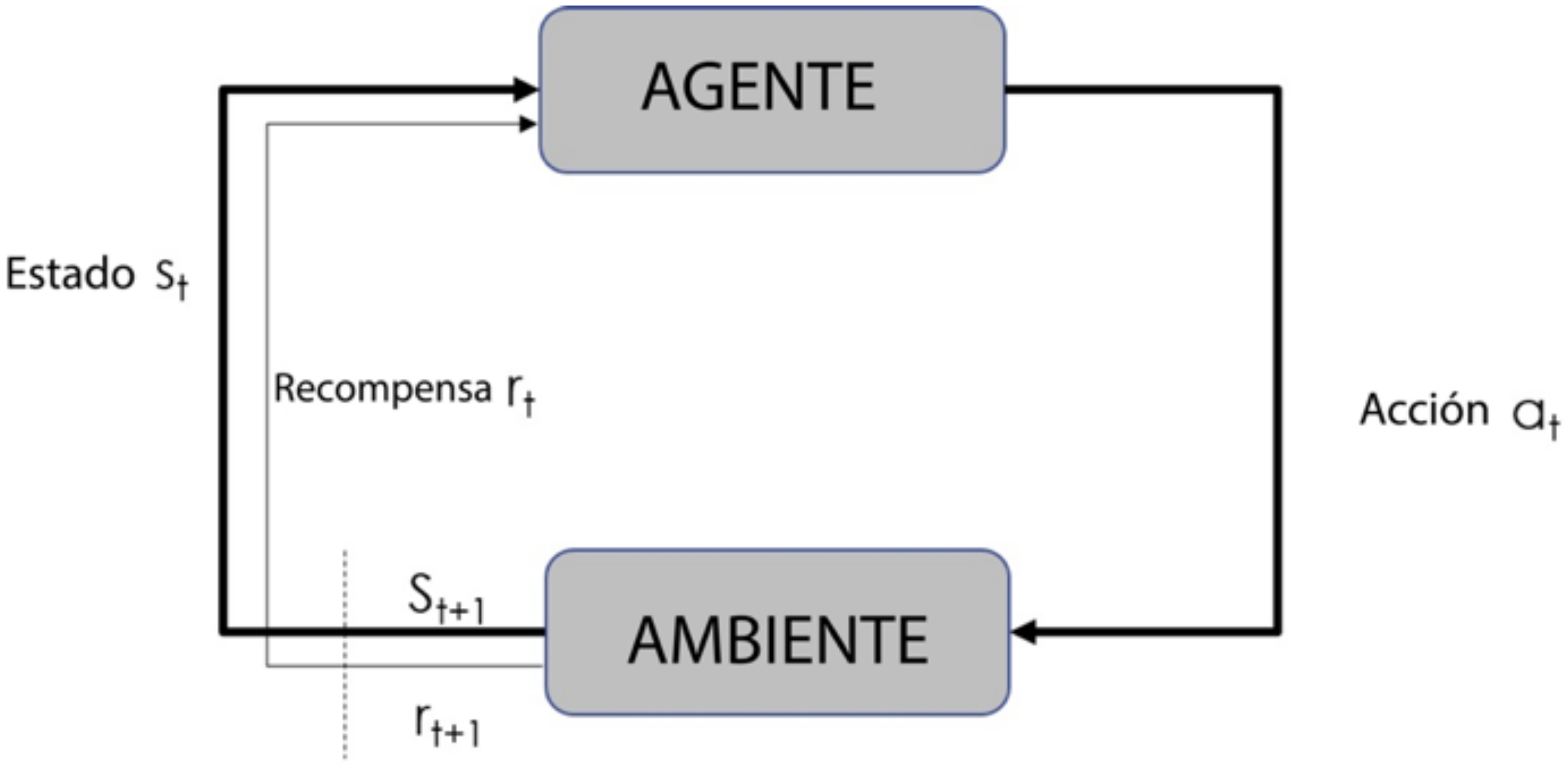
¿Qué es Inteligencia Artificial?

- La **generalización** de una **tarea** a partir de **experiencia**.

- *Arthur Samuel* (1959). Machine Learning: Field of study that gives computers the **ability** to **learn without being explicitly programmed**.

- *Tom Mitchell* (1998) Well-posed Learning Problem: A computer program is said to ***learn*** from **experience E** with respect to some **task** T and some **performance measure P**, if its performance on **T**, as measured by P, **improves** with **experience** E.

Modelo de Aprendizaje



Ejemplo

- Suponga que su software de email observa que correos usted marca o no marca como spam, basandose en esto va aprendiendo a filtrar de major manera aquellos correos que entran en su definición de spam. Basandonos en esta situación hipotética, cual es la tarea T ?

Clasificar los correos; son Spam o No son Spam

Observarlo marcar los correos como Spam o No spam

El número de correos correctamente clasificados como Spam o No spam

Ninguna de las anteriores- este no es un problema de AI

Modelo general

¿Cúando decimos que un sistema es Inteligente?

Inteligente ⟷ **Conocimiento**

- Conocimiento en datos (Redes Neuronales)
- Conocimiento en experiencia (Lógica difusa)
- Conocimiento en espacio de estado, eurística, caos (Computación evolutiva)

Tipos de problemas

- Clasificación de datos y de la experiencia humana
- Modelado a partir de daros o de la experiencia humana
- Predicciones (forecast)
- Optimización (Encontrar los valores óptimos)
- Interfaces Humano-Máquina (diseño basado en humanos)

Aplicaciones

- Credit rating and risk assessment
- Insurance risk evaluation
- Fraud detection
- Insider dealing detection
- Marketing analysis , Mailshot profiling
- Signature verification , Inventory control
- Prediction of prices, electricity load and discharge
- Machinery defect diagnosis
- Signal processing , Character recognition
- Process control & supervision , fault analysis
- Speech , vision and color recognition
- Radar signal classification
- Aircraft control, Car brakes
- Integrated circuit layout
- Image compression
- Prediction of signals and values in engineering

Ejemplos:

?

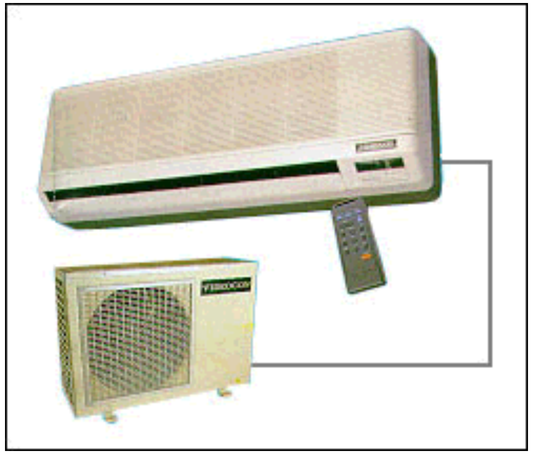
Ejemplos

- Compañia: BPL
- Producto: Lavadora ABS 50F
- Un Sistema difuso decide que tipo de programa y la cantidad de agua así como los ingredients de lavado.

- Compañia : LG
- Producto : Refrigerador
- Un Sistema Neuro difuso controla los procedimientos de refrigeración para optimizer el uso de energía.

- Compañia: Sharp
- Product : Horno de microondas
- A partir de un análisis de aire en el interior la duración de la cocción era controlada.

- Compañia: Videocon
- Product0 : aire condicionado
- Usaba un Sistema Neuro difuso para mantener el control de la temperature en el cuarto.

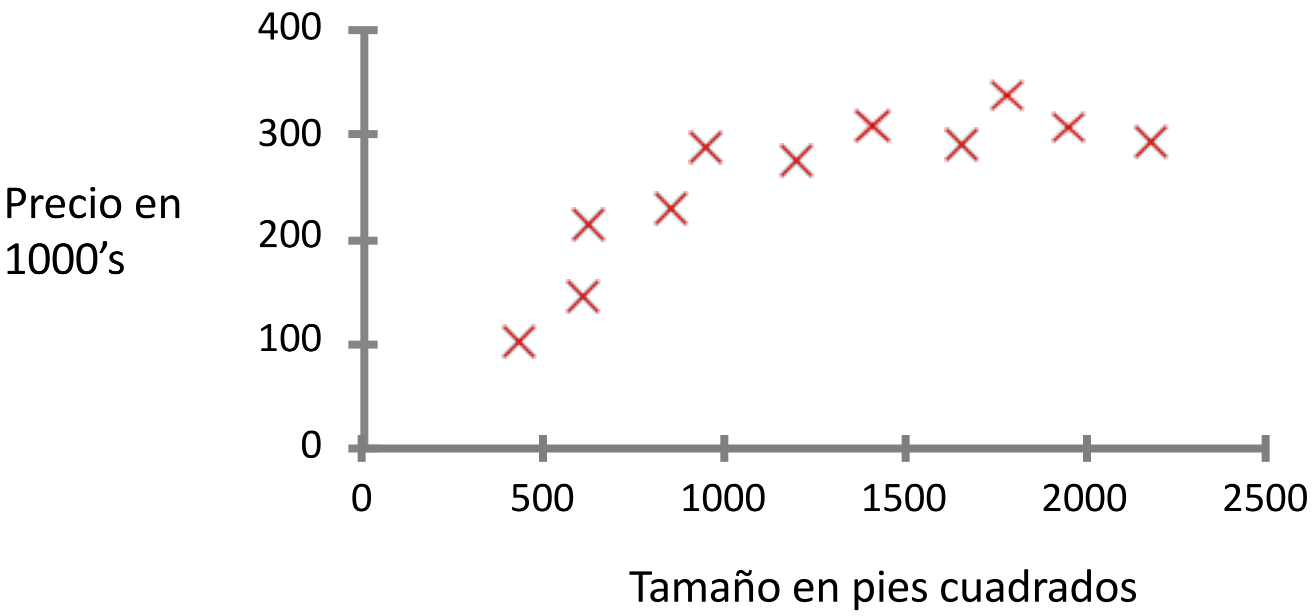


Introducción

Aprendizaje Supervisado

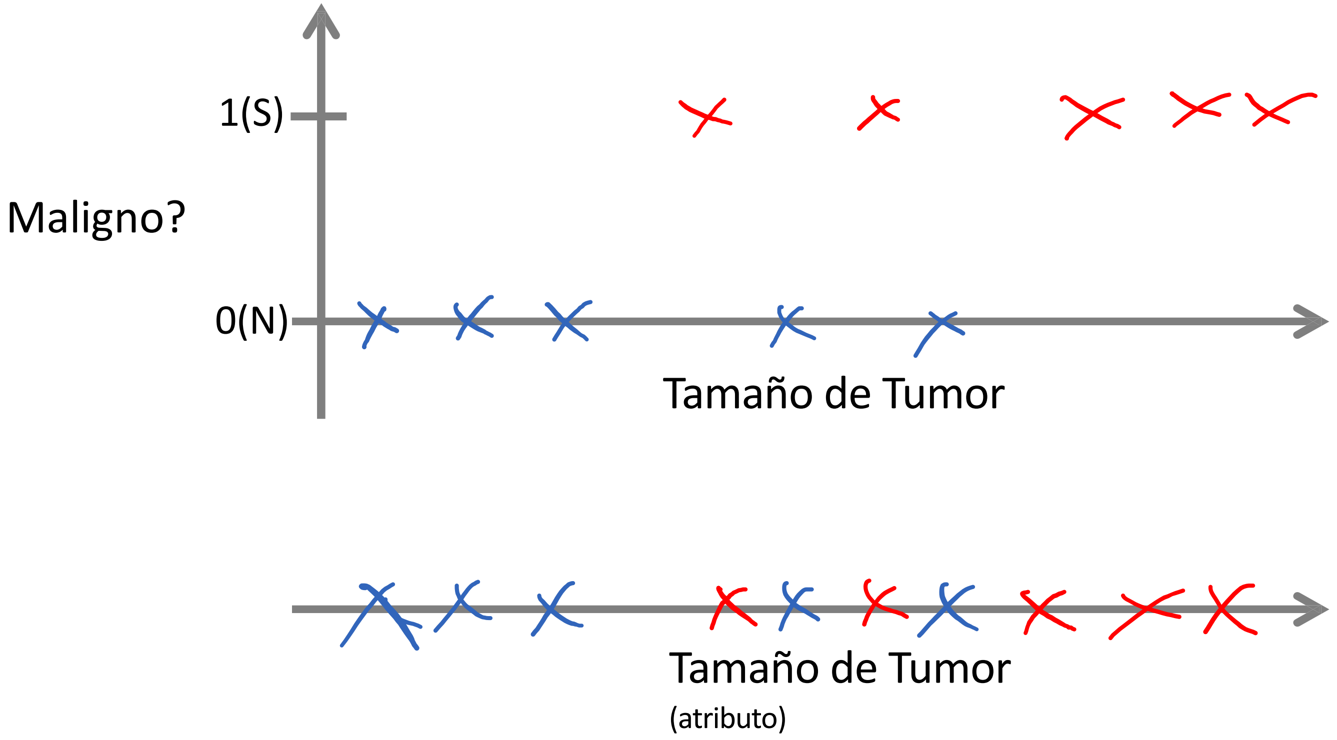
Predicción de precios

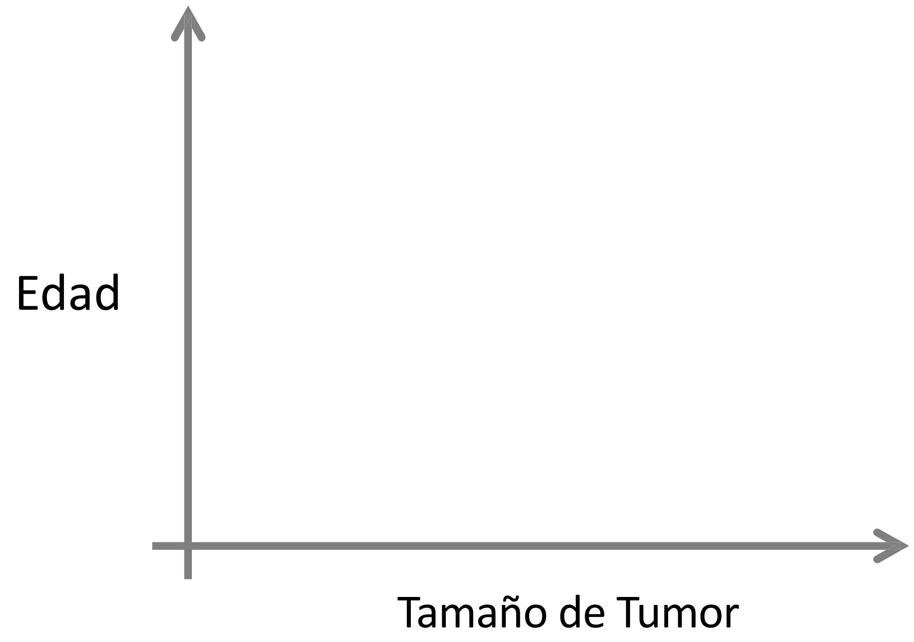
- Aprendizaje supervisado: Le damos las respuestas correctas al algoritmo.
- Regresión: Predecir de manera continua un valor de salida.



Cáncer de Seno (Maligno/Benigno)

- Clasificación: Valor Discreto (0-1)





- Uniformidad del tamaño de la célula
- Uniformidad de forma de la célula
- Espesor del grupo de células
- …

Suponga quye usted tiene una empresa de software.

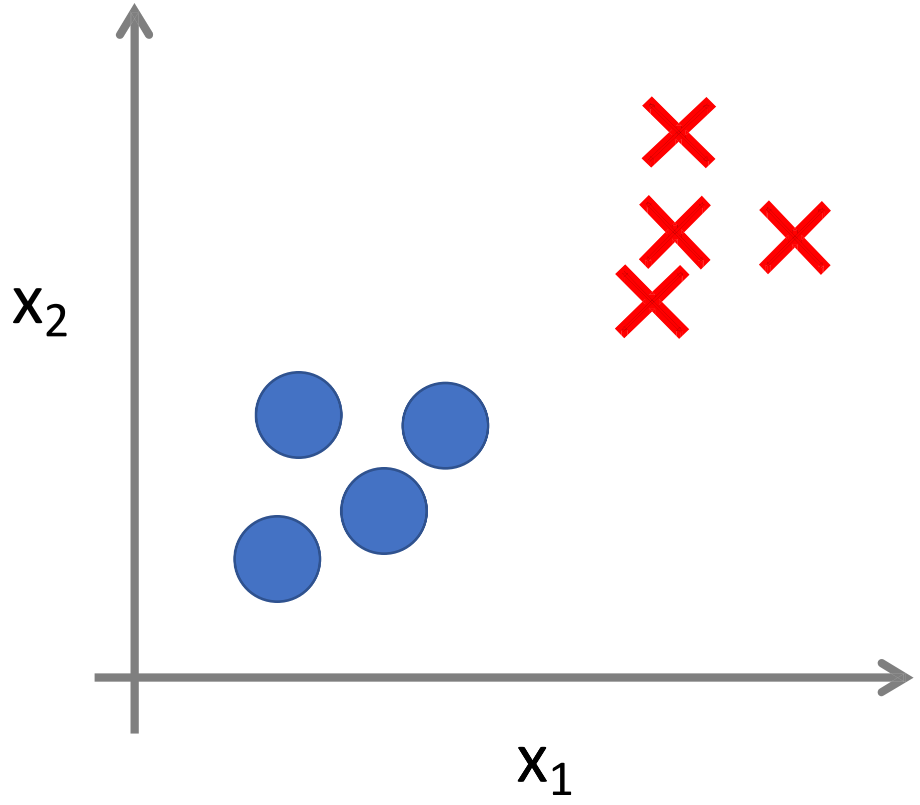
- Problema 1: Usted tiene un gran inventario de objetos similares. Y desea predecir cuantos de estos se venderán en los próximos 5 meses.
- Problema 2: Usted diseña un software para examiner las cuentas de sus clients para saber si esta ha sido hackeada o no.

- ¿Cómo trataría al problema 1 ?
- ¿Cómo trataría al problema 2 ?

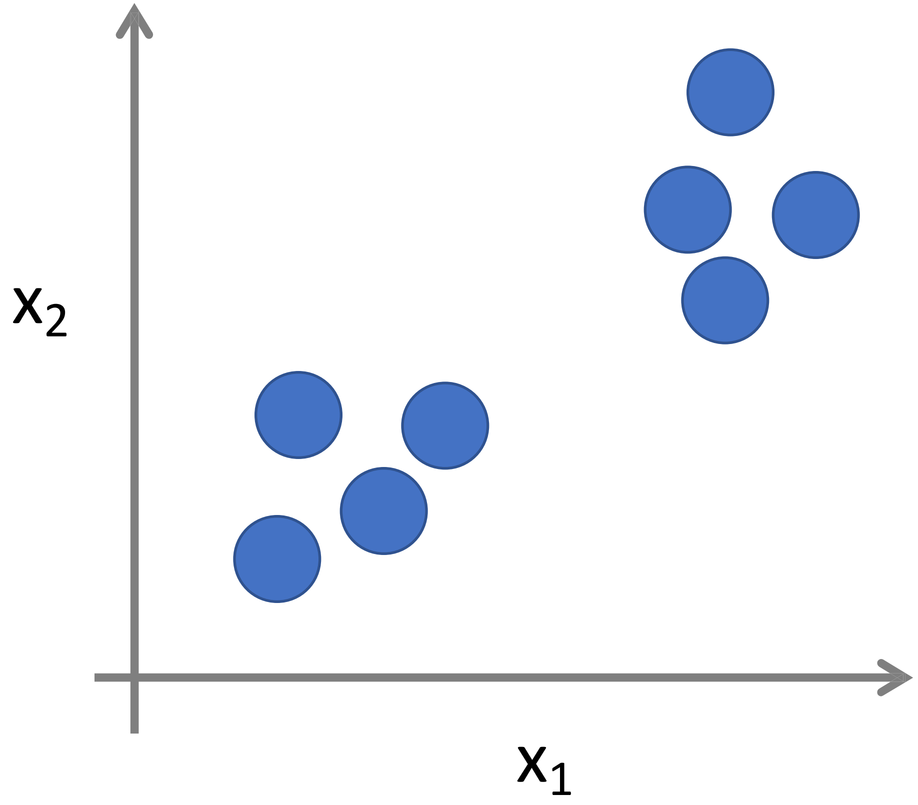
Introducción

Aprendizaje No Supervisado

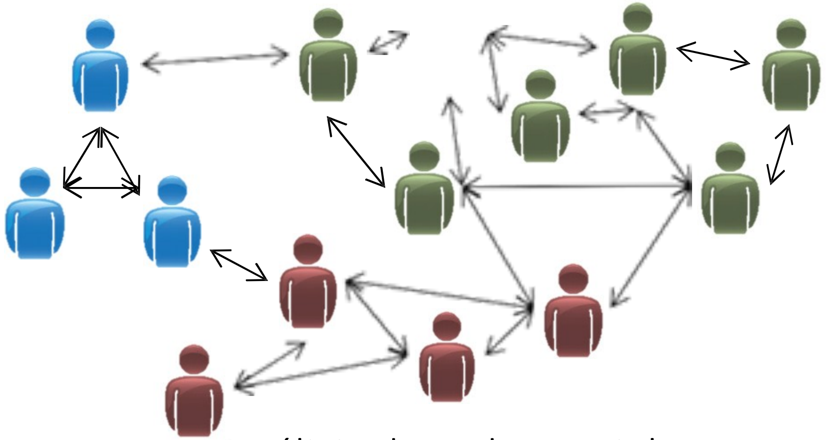
Aprendizaje Supervisado



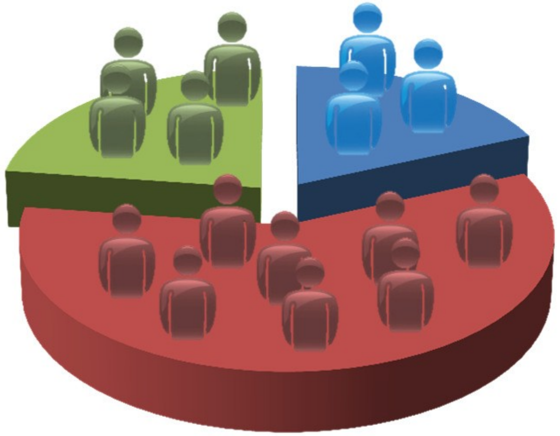
Aprendizaje No Supervisado



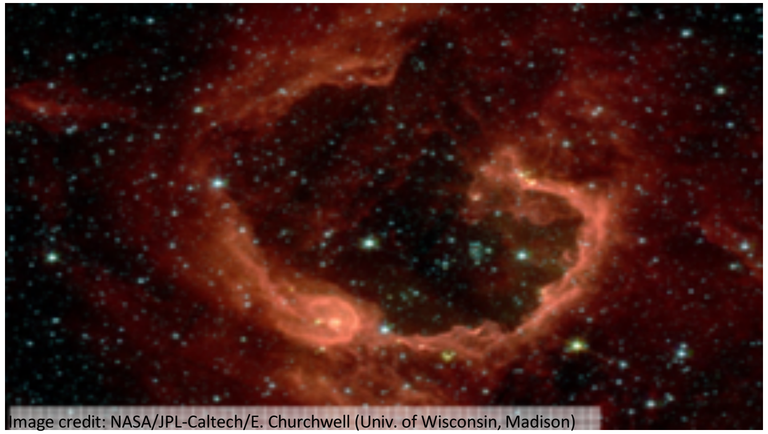
Organización de clusters de Computadoras



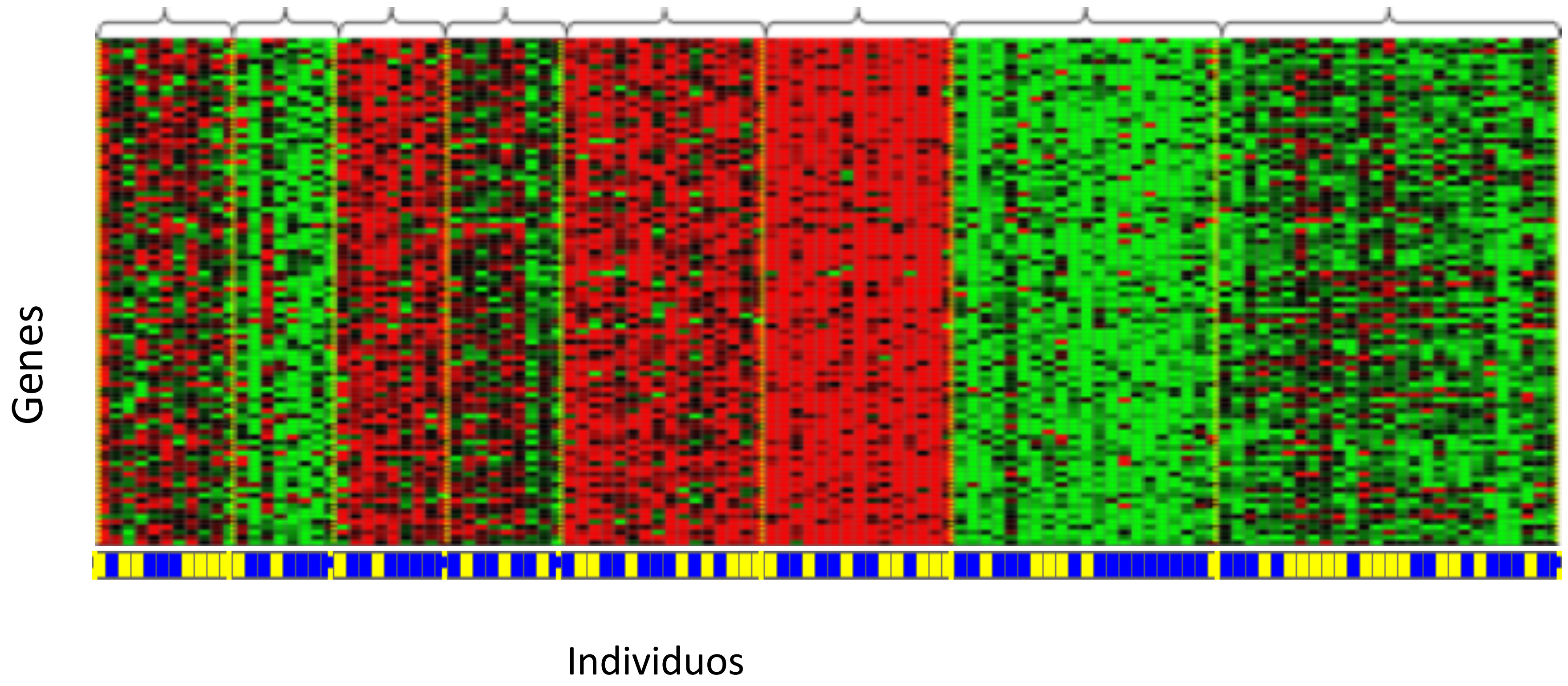
Análisis de redes sociales



Segmentaciones de Mercado



Análisis de datos astronómicos



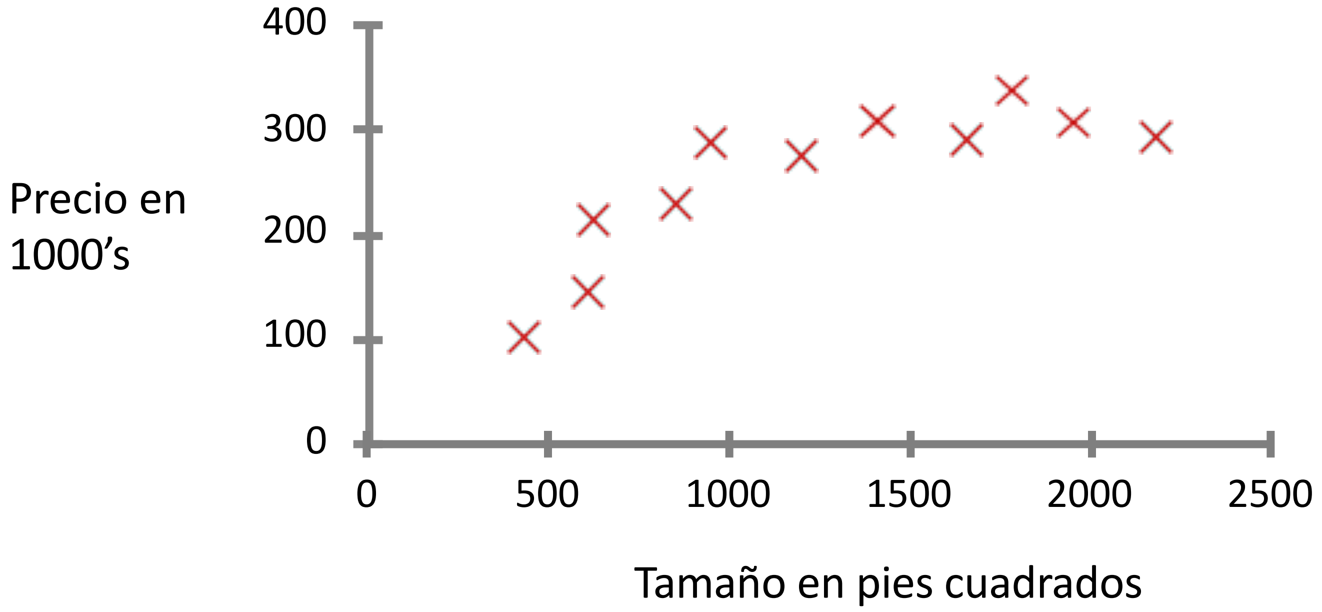
[Fuente: Daphne Koller]

Regresión Lineal Univariable

Representación del modelo

Predicción de precios

- Aprendizaje supervisado: Le damos las respuestas correctas al algoritmo.
- Regresión: Predecir de manera continua un valor de salida.



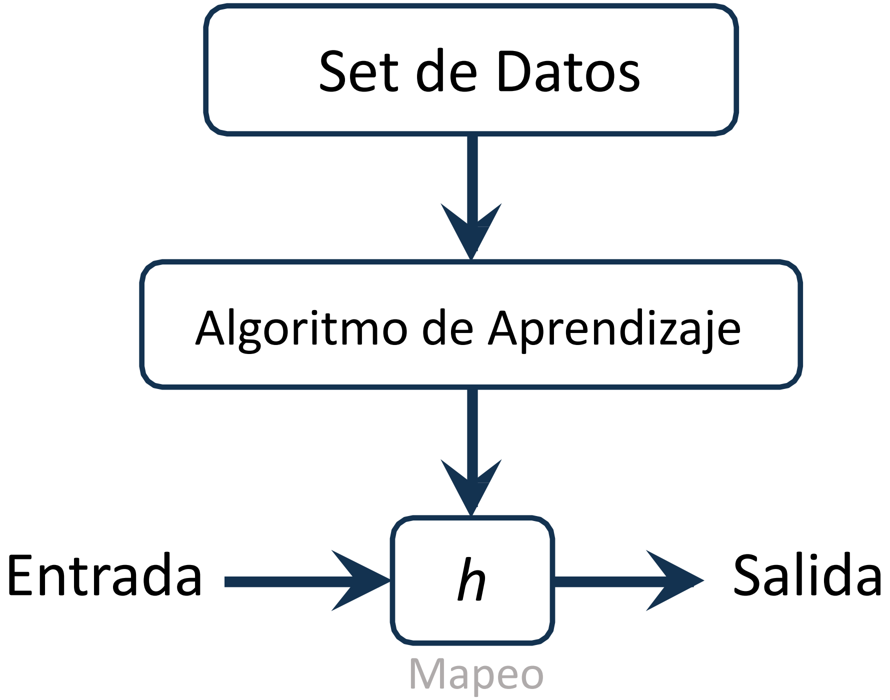
Set de Datos para entrenamiento

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

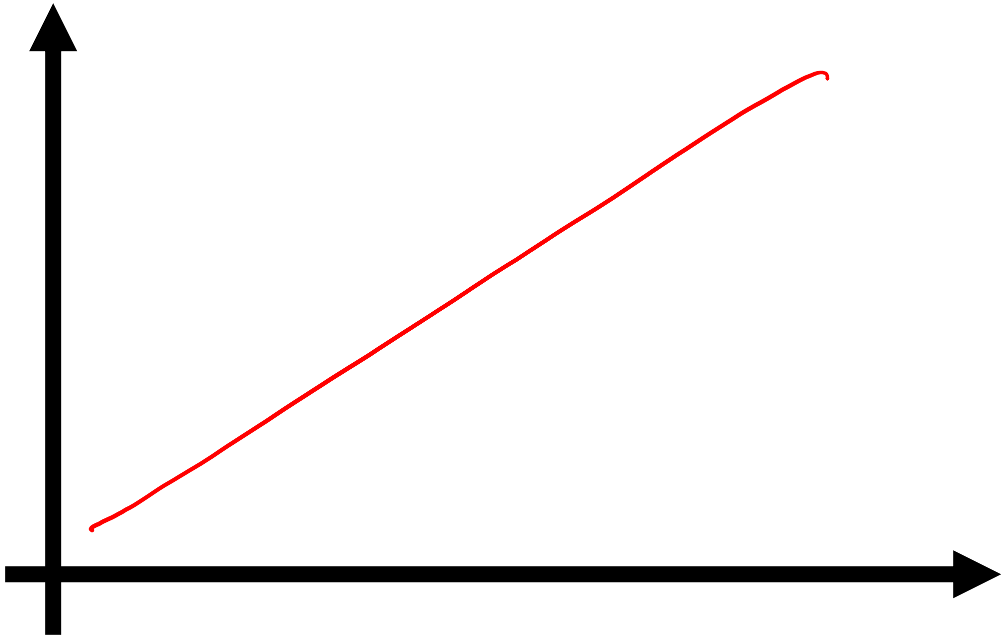
Notación:

m = Número de ejemplos
x’s = “entradas”,“input”, variables, features, atributos
y’s = “salidas”,“output”,variables, objetivo,“target”

Modelo



¿Cómo representariamos a *h*?



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

Set de Entrenamiento

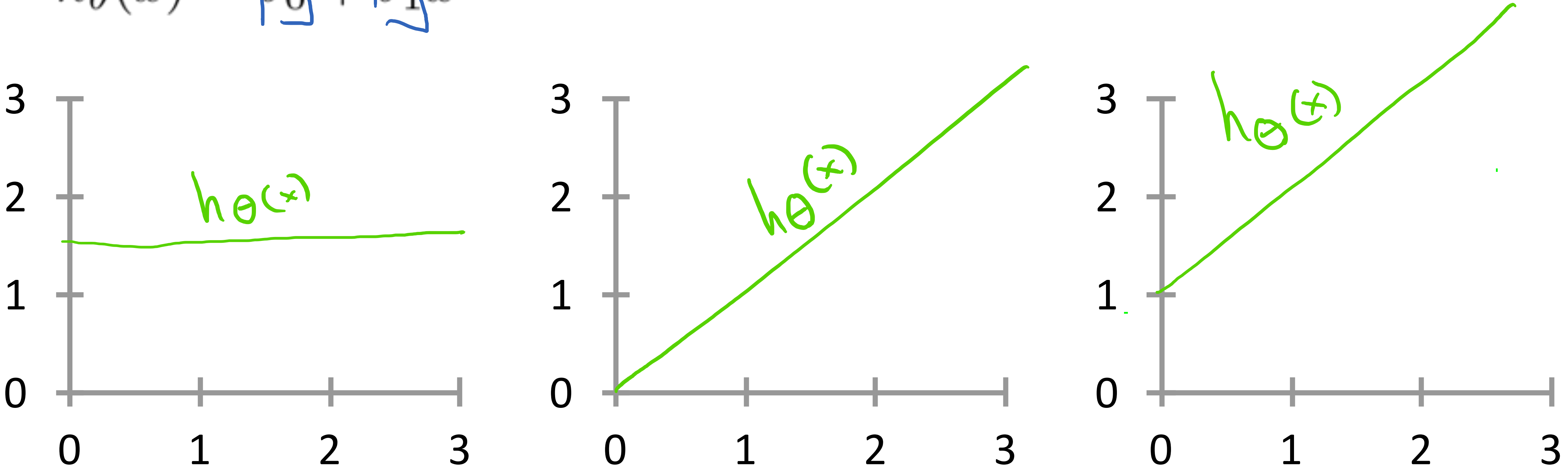
Tamaño en pies^2 (x)	Precio ($) en 1000's (y)
2104	460
1416	232
1534	315
852	178
…	…

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

θ_i's: Parámetros

Como escoger los θ_i's ?

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



$\theta_0 = 1.5$	$\theta_0 = 0$	¿ $\theta_0 = 1$?
$\theta_1 = 0$	$\theta_1 = 0.5$	$\theta_1 = 0.5$

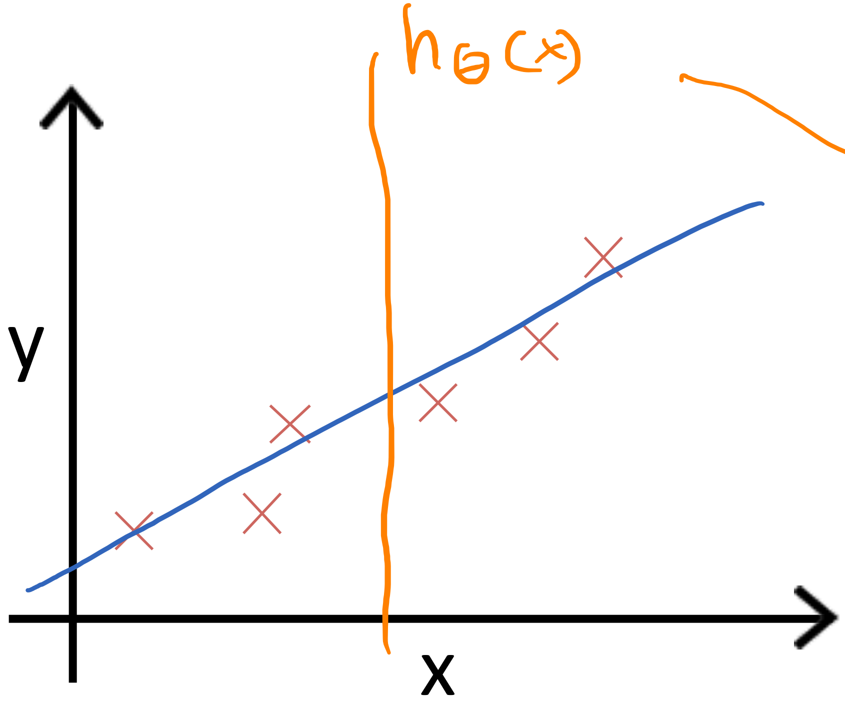
$h_\theta(x) = 1.5 + 0(x)$

$h_\theta(x) = 1.5$ → CTE.

$h_\theta(x) = 0 + .5(x)$

$h_\theta(x) = .5x$

$h_\theta(x) = 1 + .5x$



ERROR CUADRÁTICO

$$\sum_{i=1}^{m} \frac{1}{2m} (\text{RESULTADO} - \text{OBJETIVO})^2$$

e

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

INPUT — OUTPUT

DEBEMOS MINIMIZAR ESTA EXPRESIÓN

MINIMIZAR $J(\theta_0, \theta_1)$

COST FUNCTION

Idea: Escoger θ_0, θ_1 tal que

$h_\theta(x)$ se acerque a y para

nuestros ejemplos (training set) (x, y)

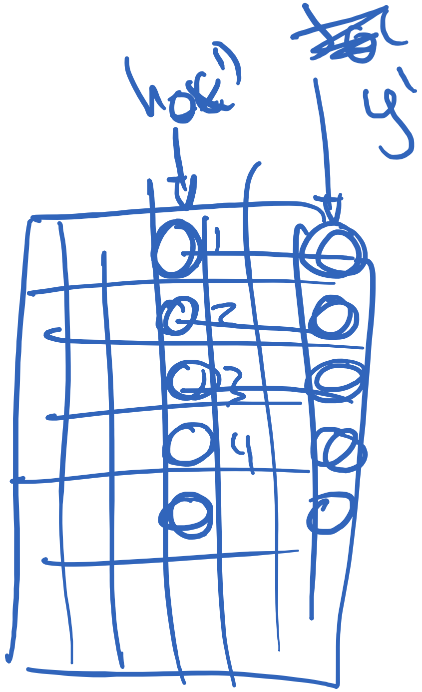
Función de costo

Demostración

Simplificando

Hipótesis:

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



Parámetros:

$$\theta_0, \theta_1$$

Funcion de Costo:

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

Objetivo:

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

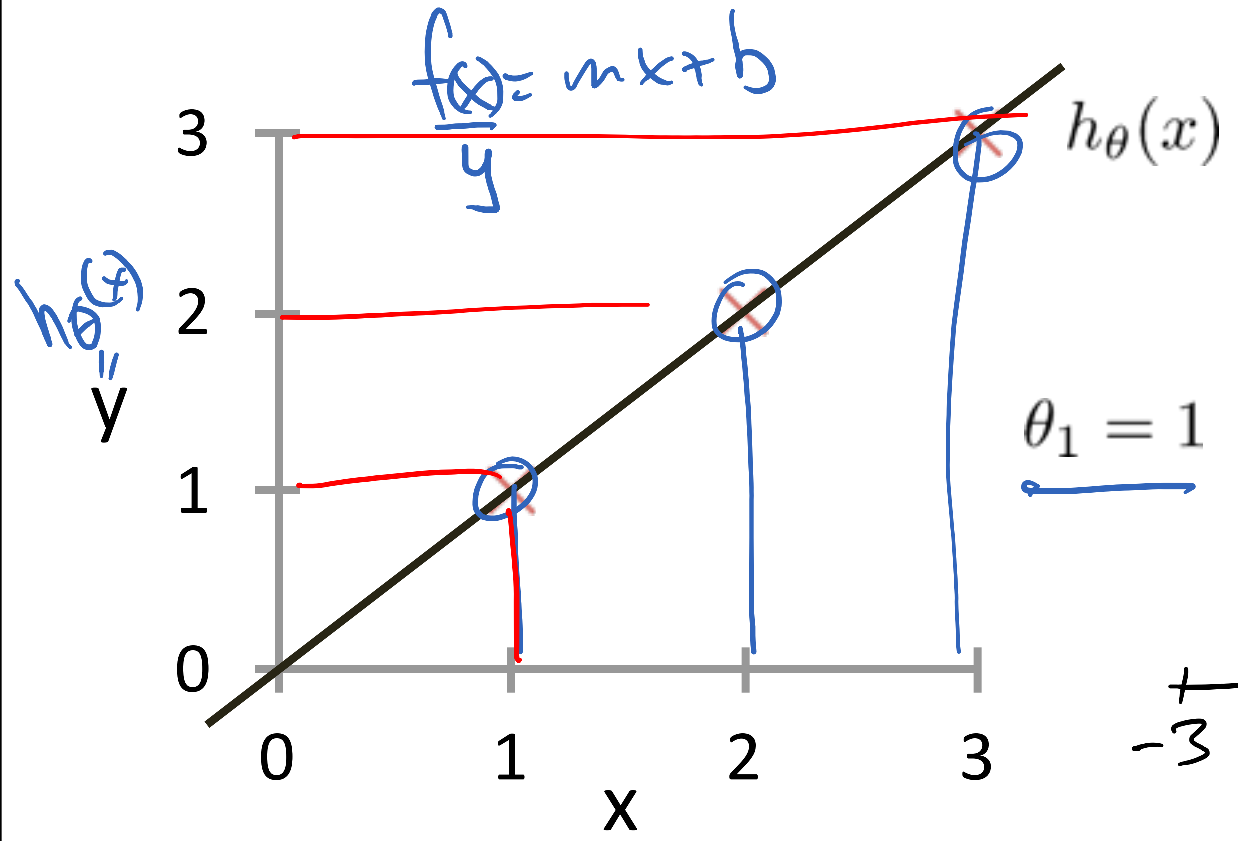
$$h_\theta(x) = \theta_1 x$$

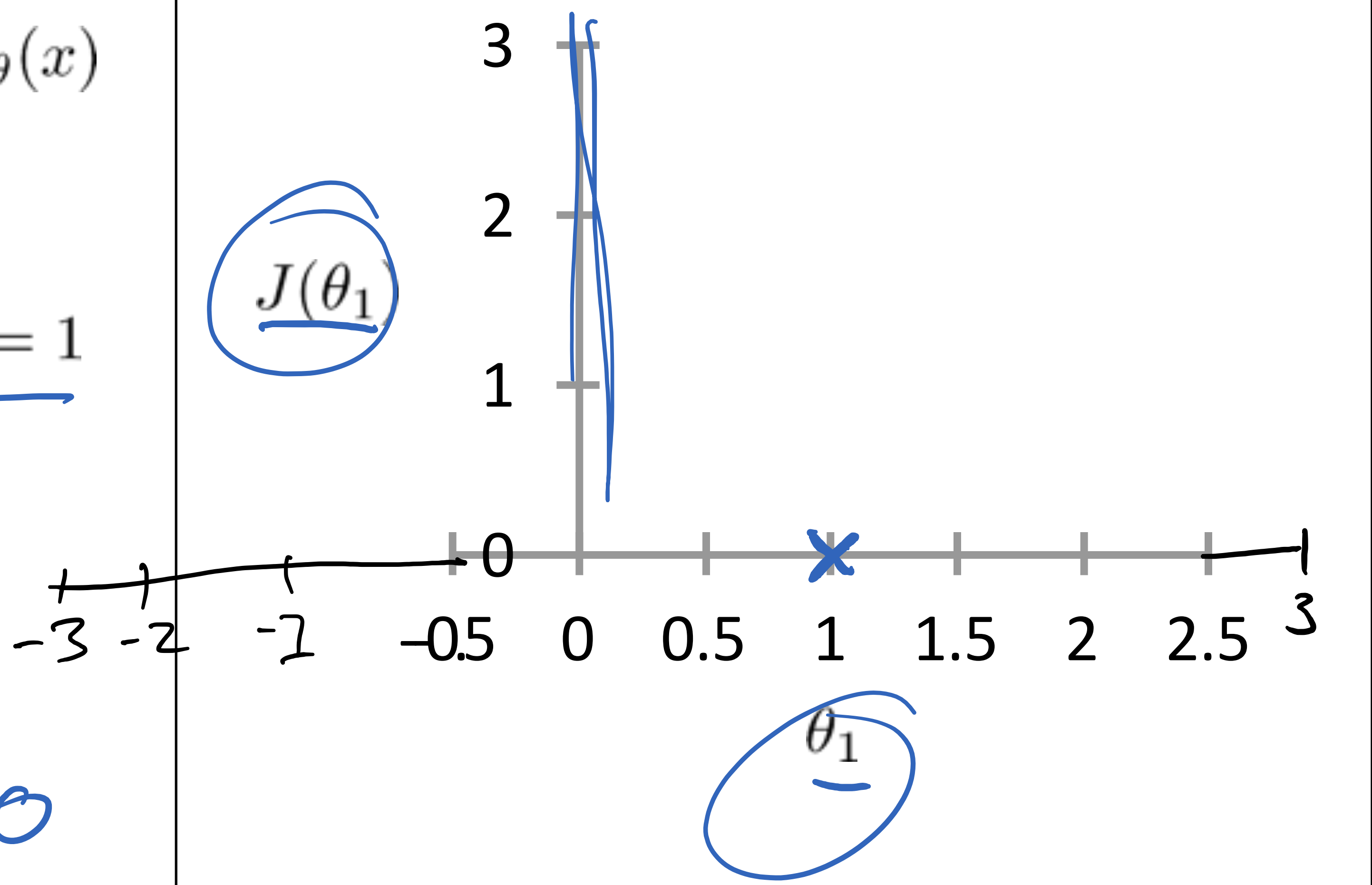
$$\theta_1$$

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

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

$h_\theta(x)$	$J(\theta_1)$
(for fixed θ_1, this is a function of x)	(function of the parameter θ_1)



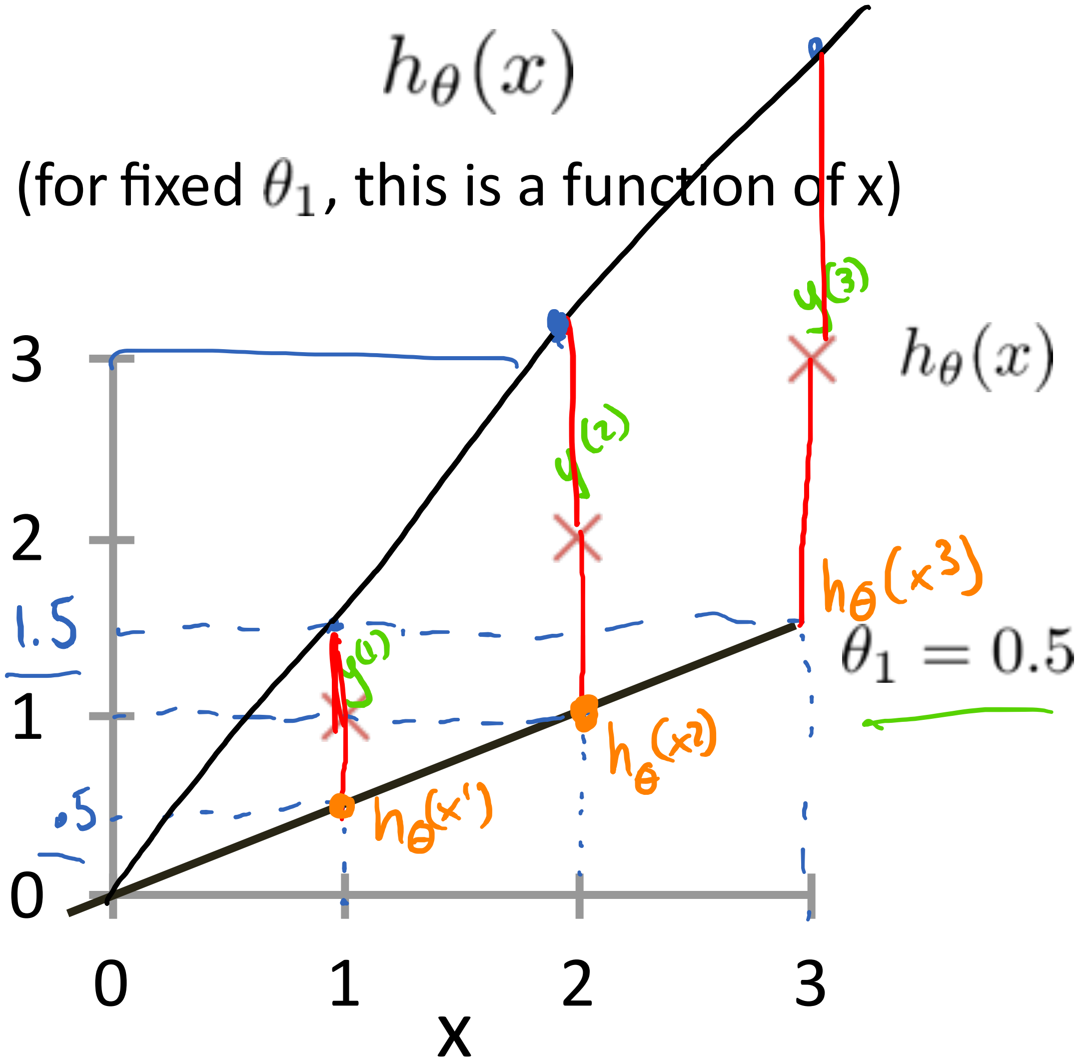


$(1-1)^2 + (2-2)^2 + (3-3)^2 = 0$

$J\theta_1 = 0$

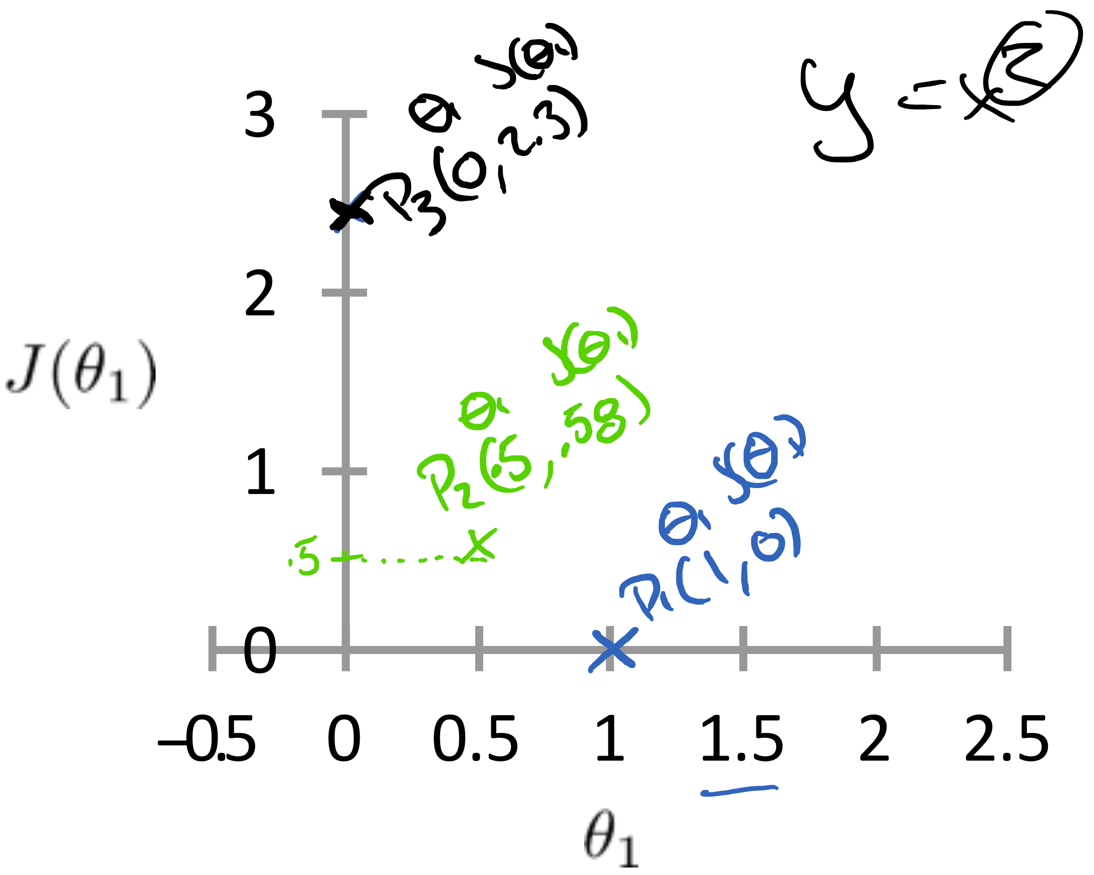
$h_\theta(x)$

(for fixed θ_1, this is a function of x)



$J(\theta_1)$

(function of the parameter θ_1)



$J_\theta = \frac{1}{2m}\left[(.5-1)^2 + (1-2)^2 + (1.5-3)^2\right]$

$\frac{1}{(2)(3)}[3.5] \approx .58$

$y = h_\theta(x) = 1.5x$

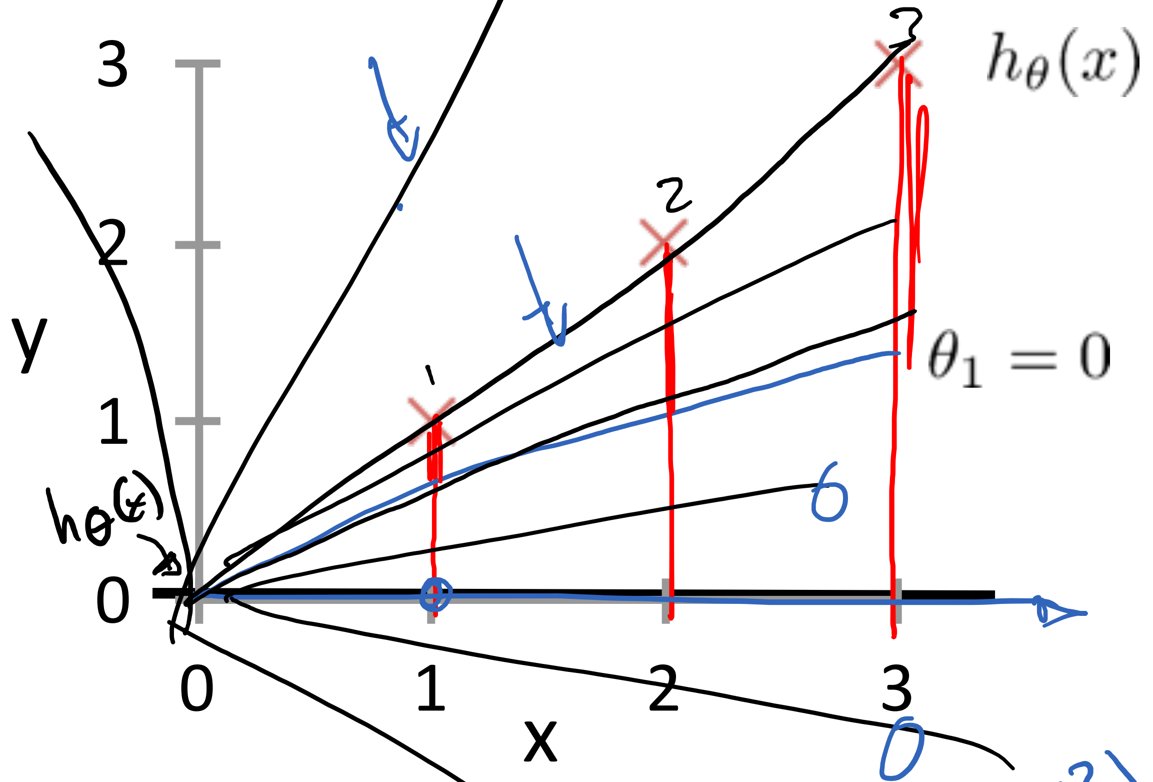
$h_\theta(x) = 2x$

$h_\theta(1) = 1.5(1) = 1.5$

$h_\theta(2) = 3$ $\quad h_\theta(3) = 4.5$

$h_\theta(x)$

(for fixed θ_1, this is a function of x)

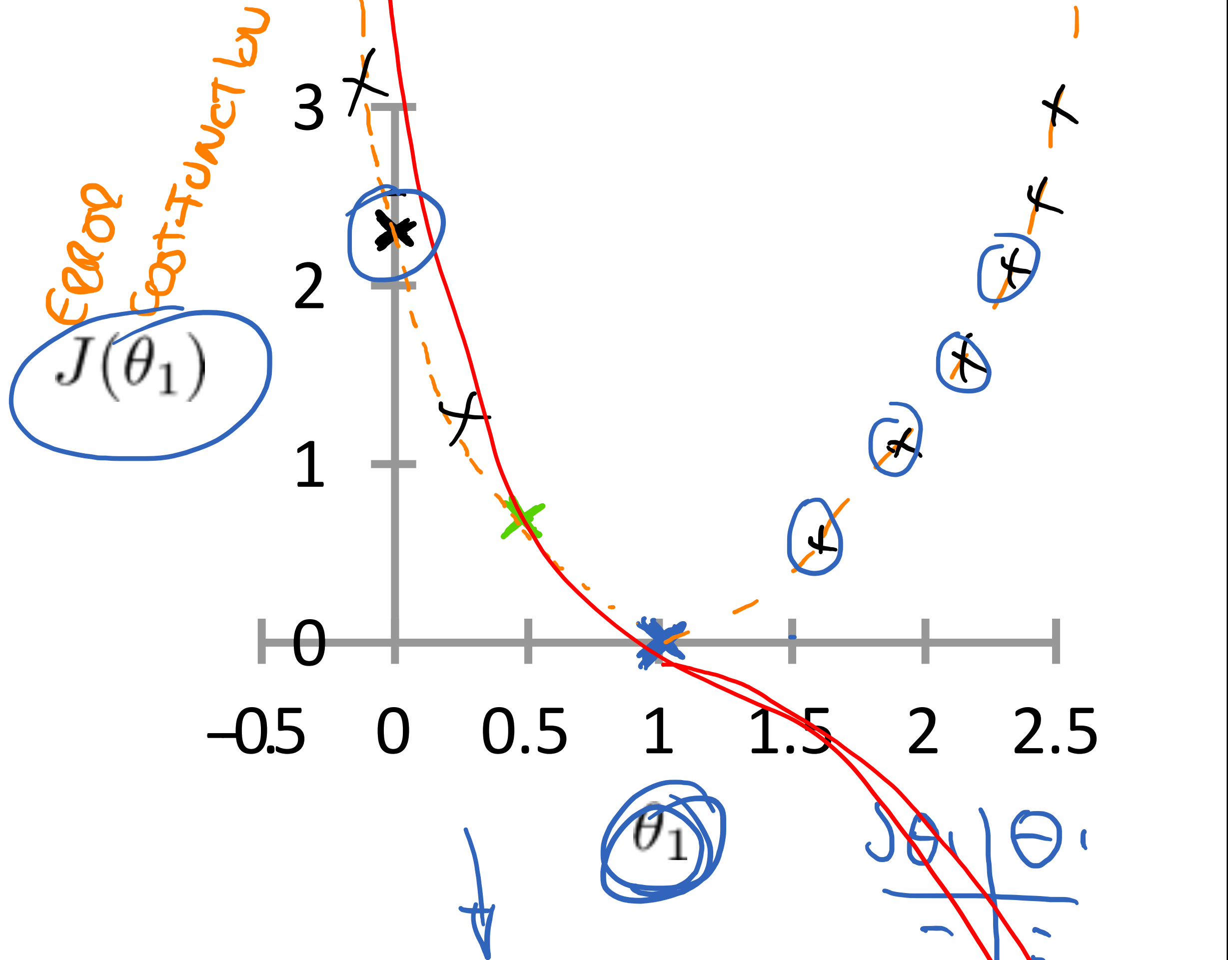


$$J_\theta = \frac{1}{2m}\left((0-1)^2+(0-2)^2+(0-3)^2\right)$$

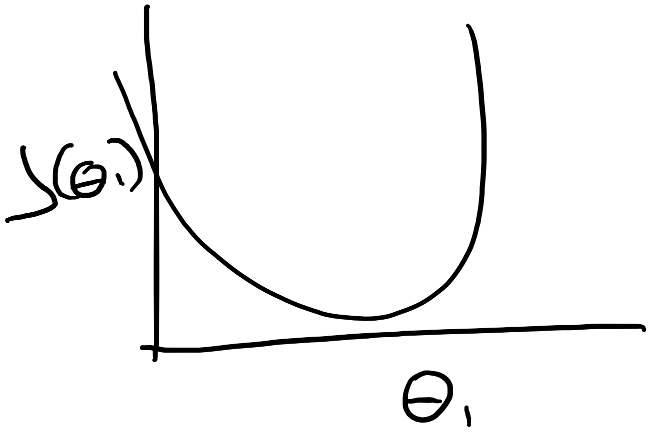
$$= \frac{1}{2(3)}(14) = \frac{14}{6} \approx 2.3$$

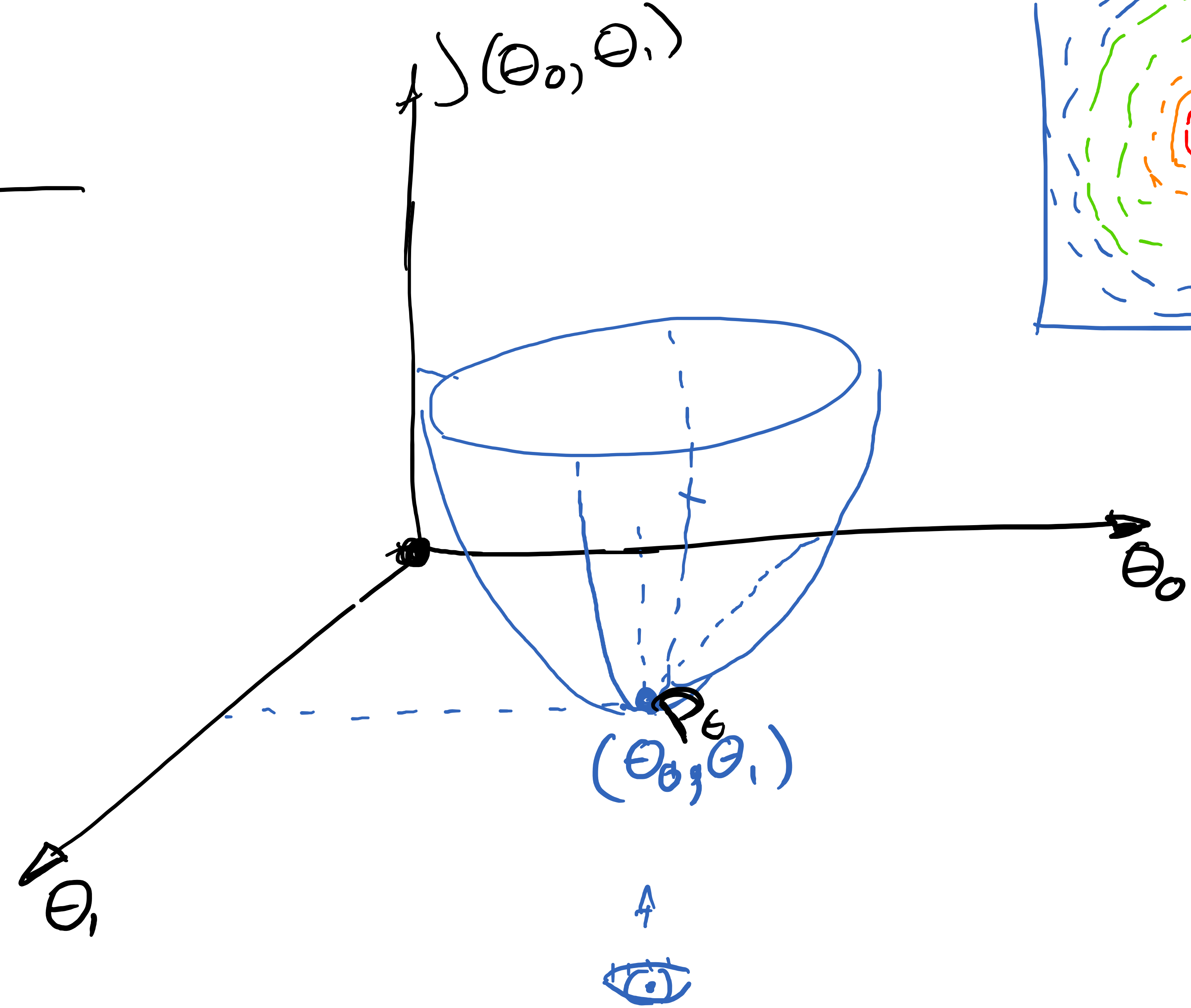
$J(\theta_1)$

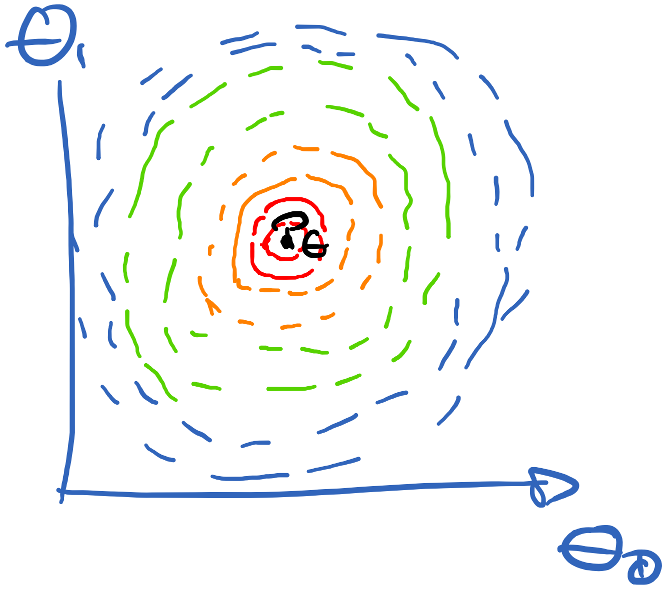
(function of the parameter θ_1)



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


J(θ1)
θ1


J(θ0, θ1)
θ0
θ1
PG
(θ0, θ1)


θ1
PG
θ0

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

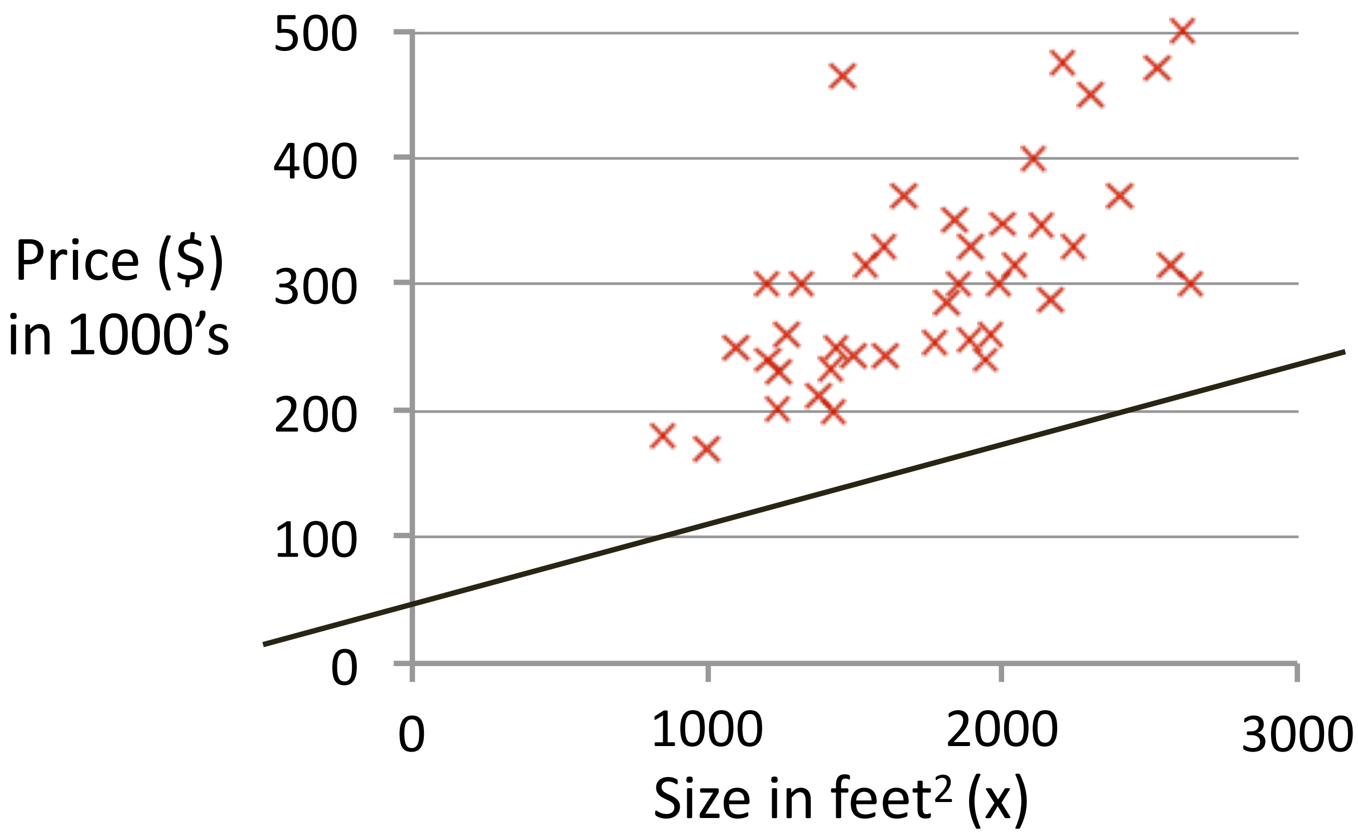
Parameters: θ_0, θ_1

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

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

$h_\theta(x)$

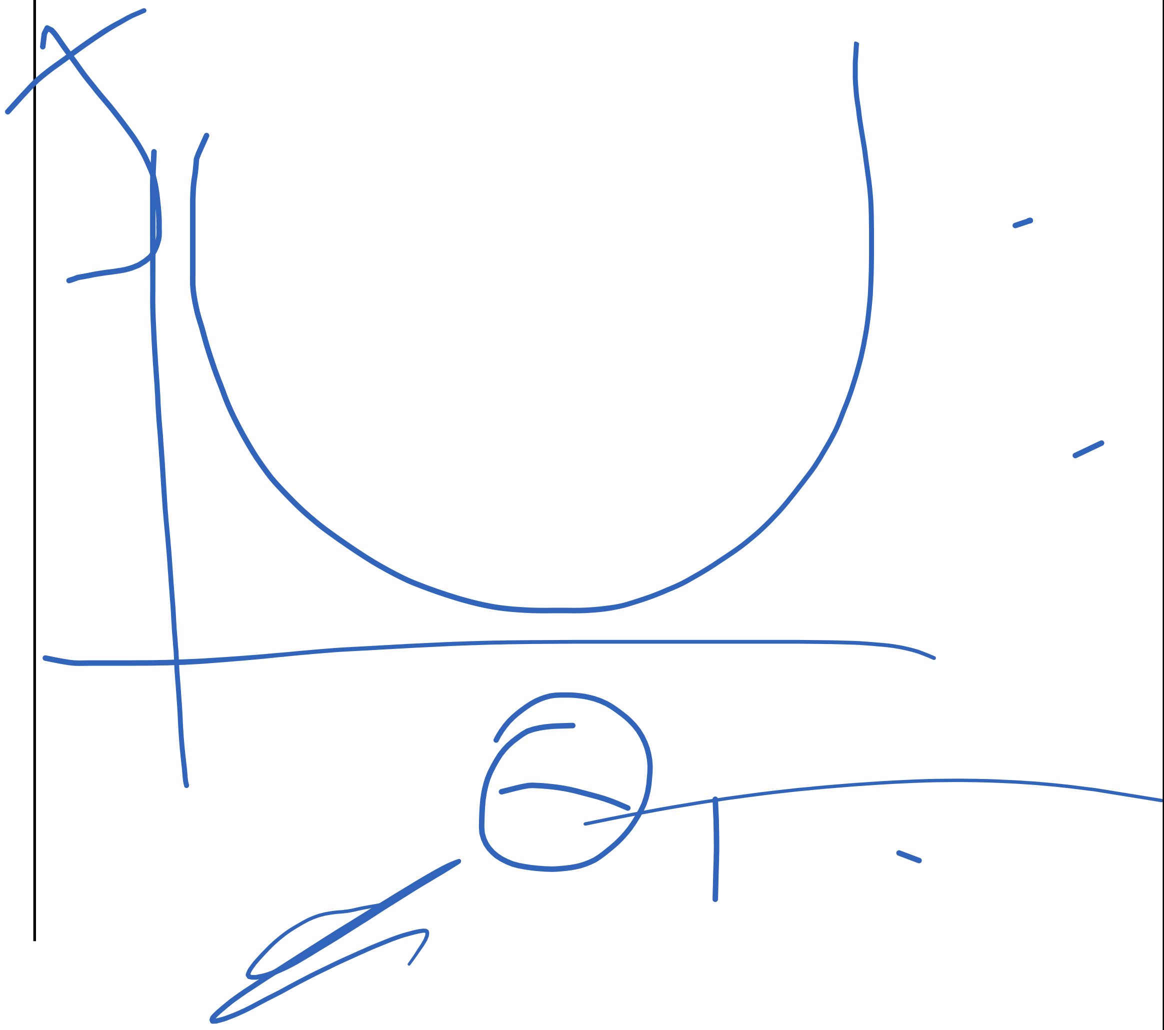
(for fixed θ_0, θ_1, this is a function of x)

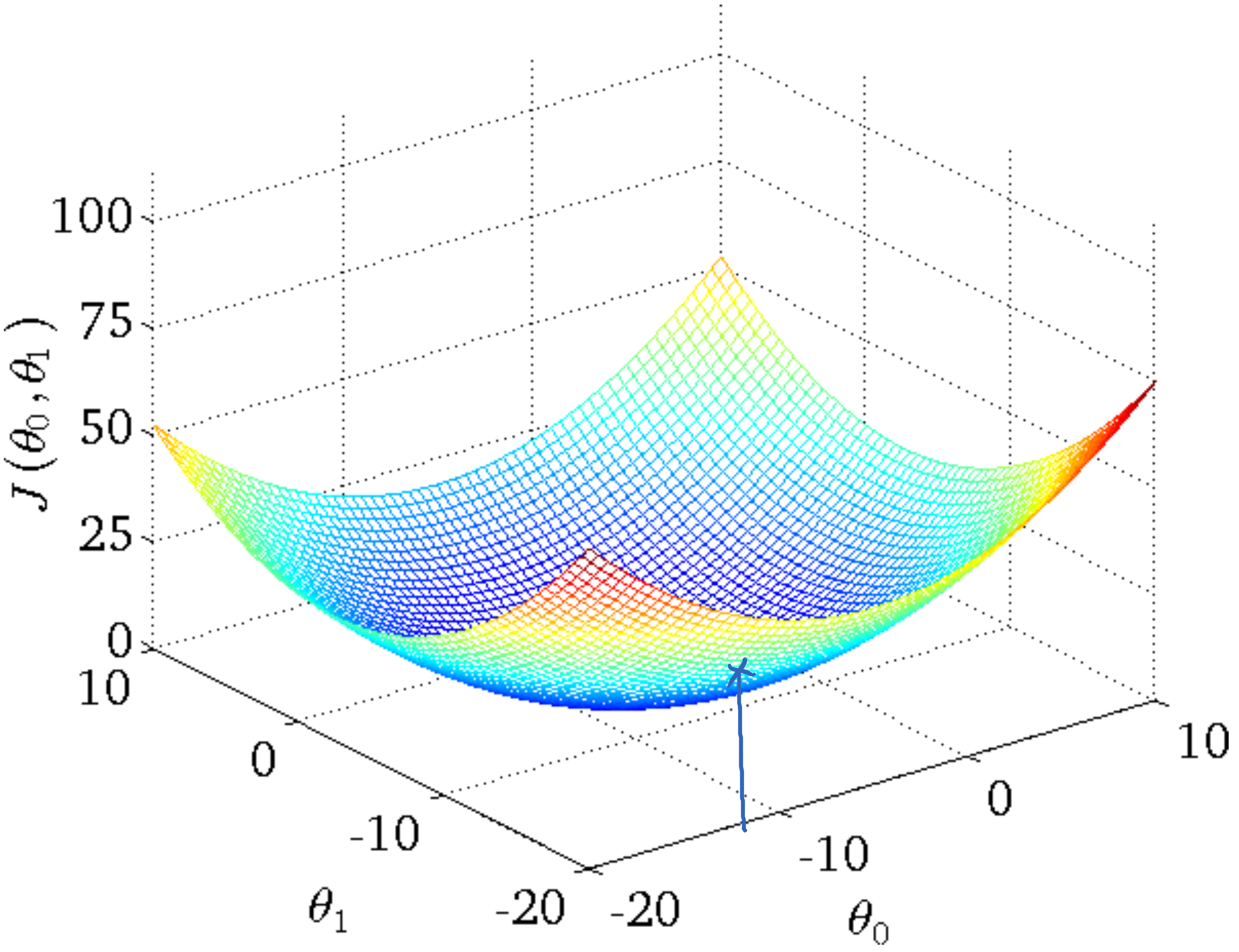


$$h_\theta(x) = 50 + 0.06x$$

$J(\theta_0, \theta_1)$

(function of the parameters θ_0, θ_1)

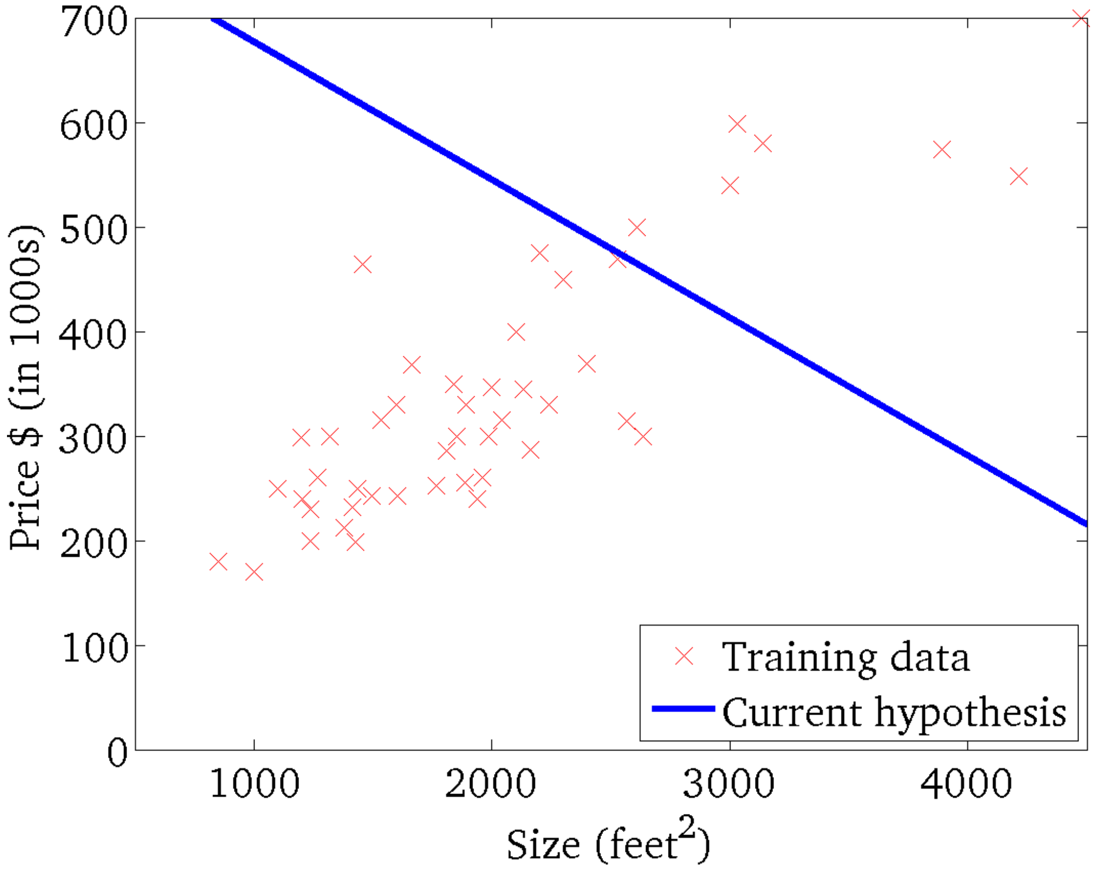




100
75
50
25
0
$J(\theta_0, \theta_1)$
10
0
-10
-20
θ_1
10
0
-10
-20
θ_0

$h_\theta(x)$

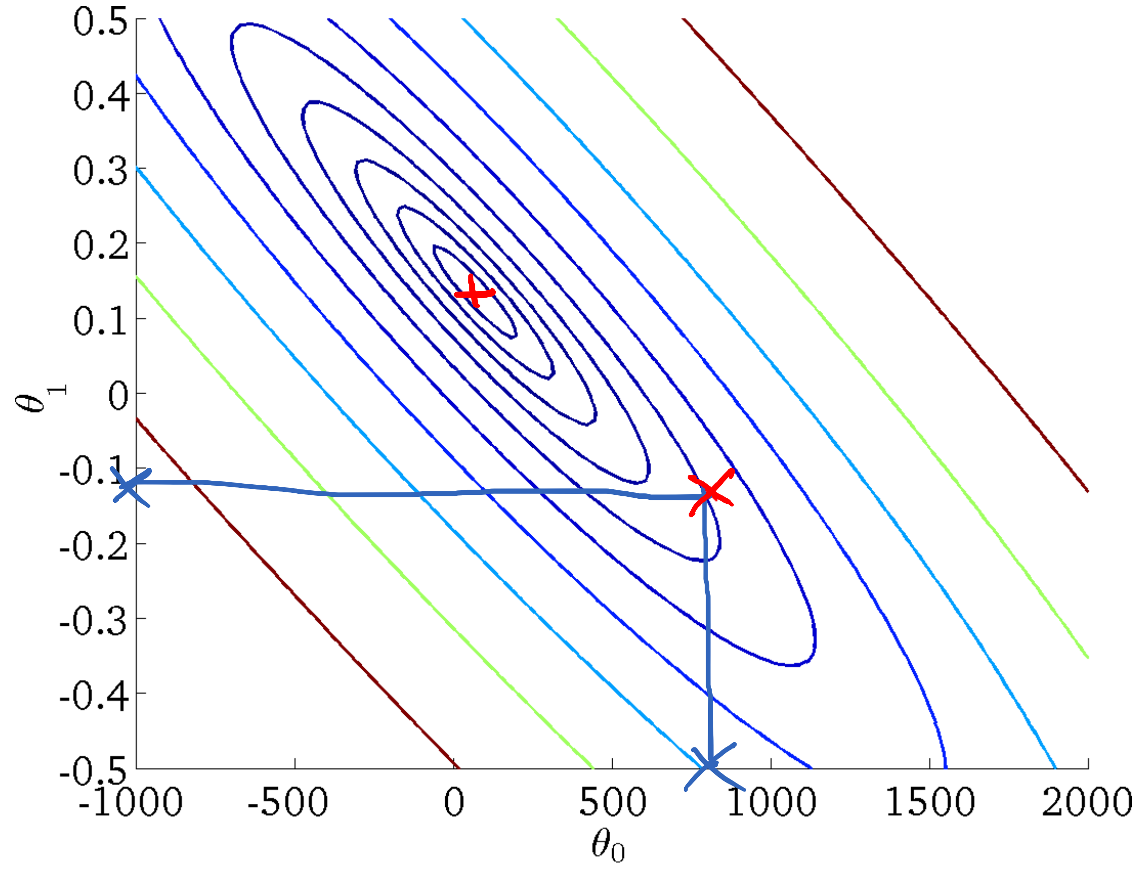
(for fixed θ_0, θ_1, this is a function of x)



$h_\theta(x) = 800 - .13x$

$J(\theta_0, \theta_1)$

(function of the parameters θ_0, θ_1)

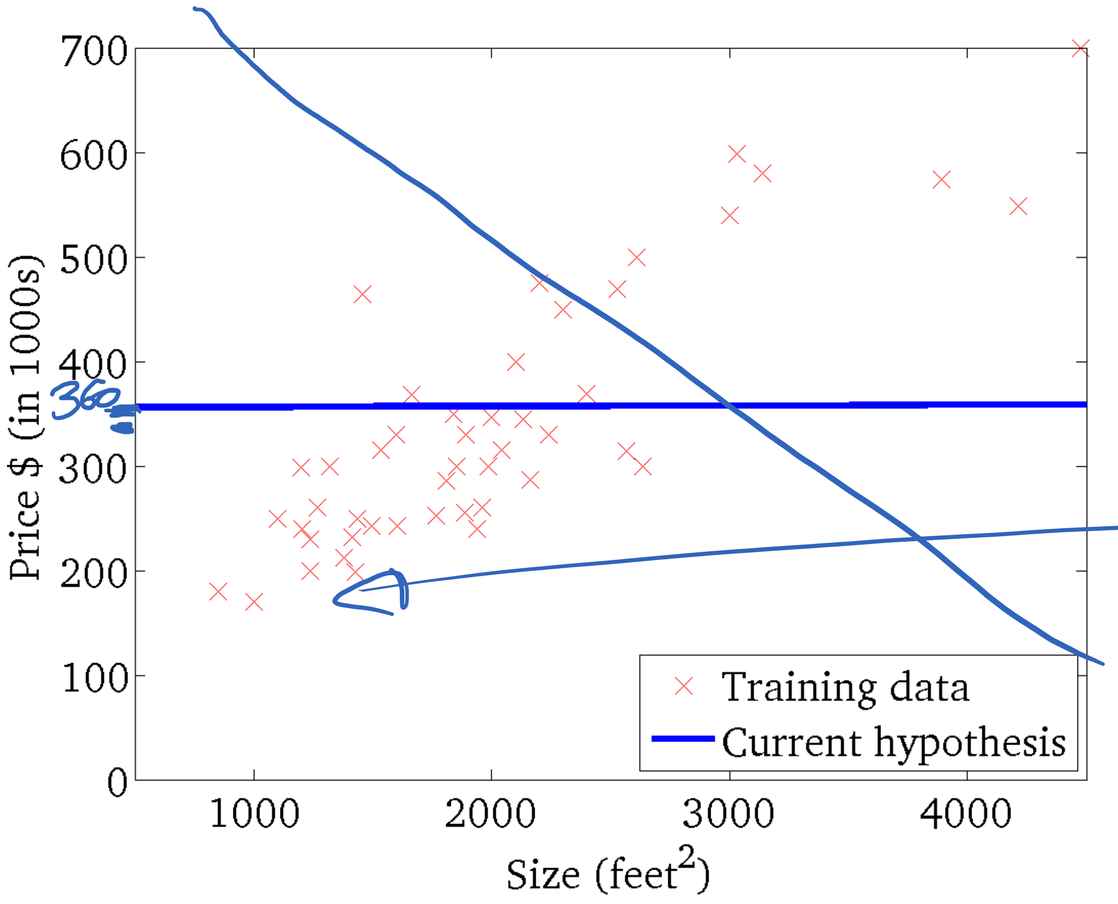


$\theta_0 = 800$

$\theta_1 = -.13$

$$h_\theta(x)$$

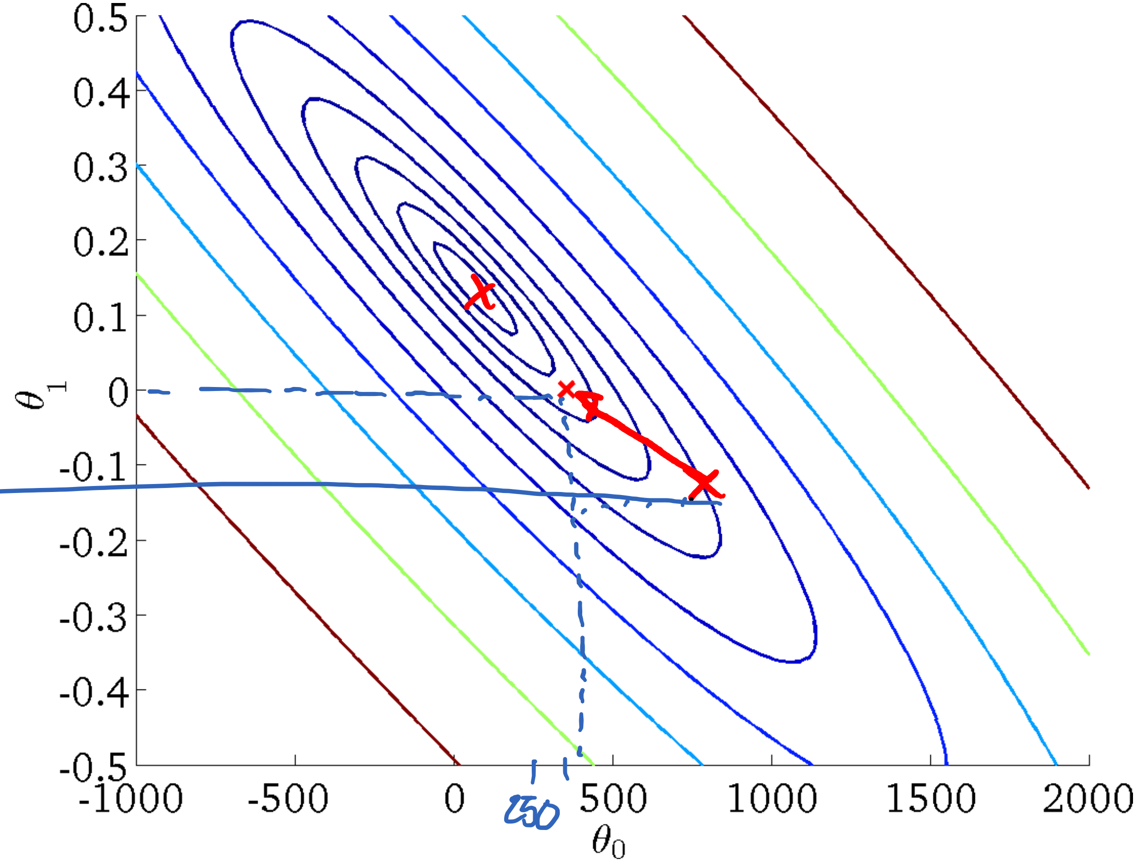
(for fixed θ_0, θ_1, this is a function of x)



$h_\theta(x) = 360$

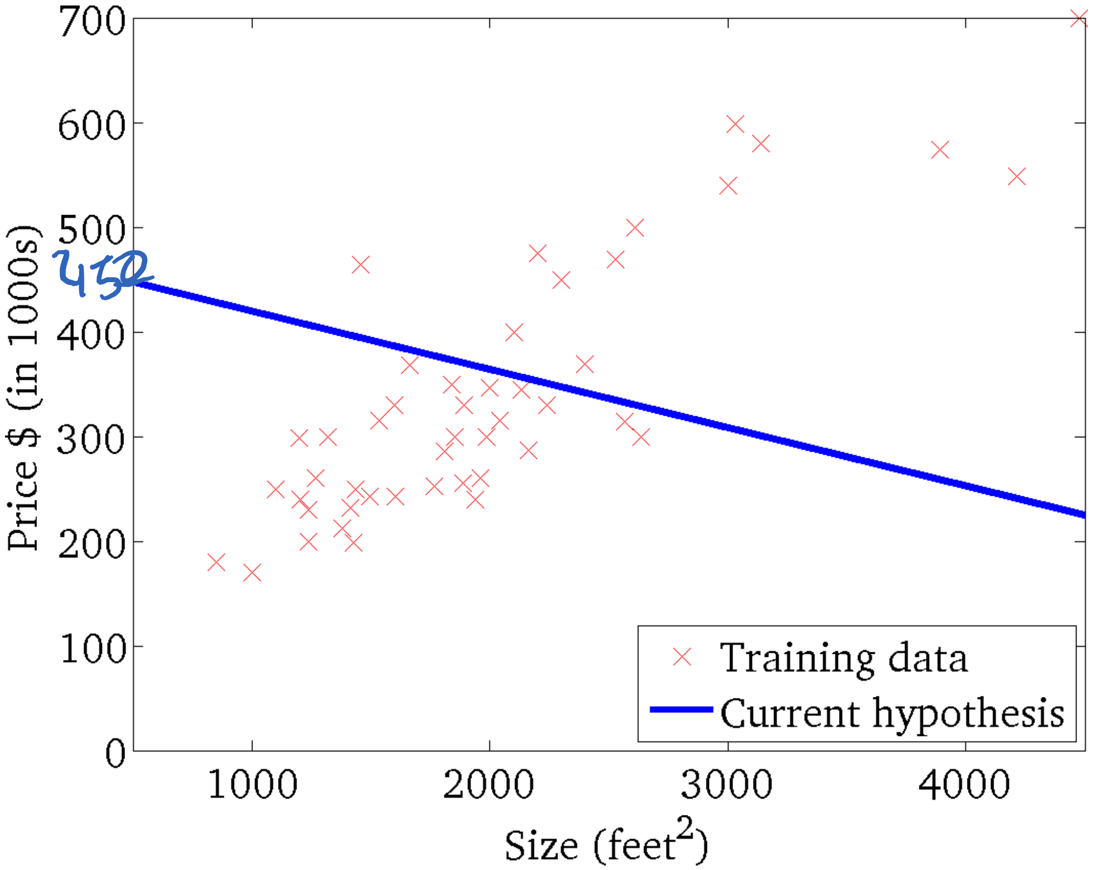
$$J(\theta_0, \theta_1)$$

(function of the parameters θ_0, θ_1)



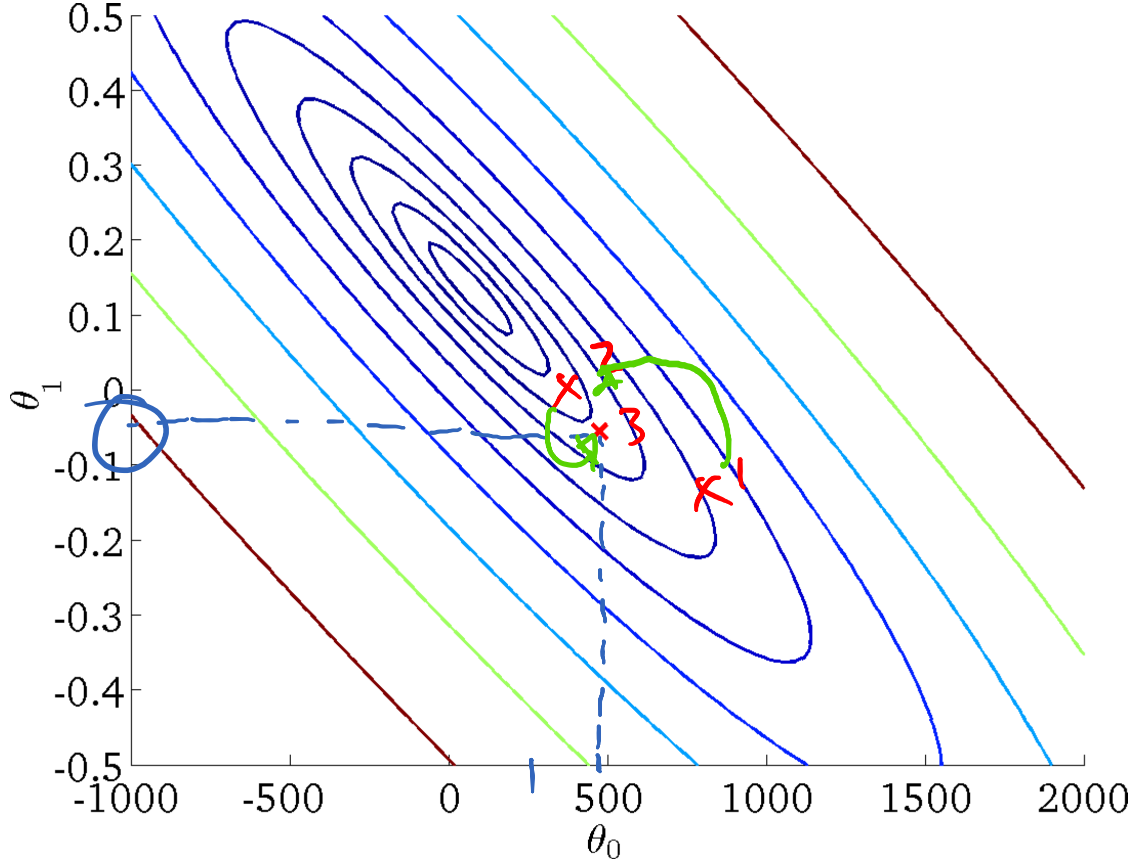
$h_\theta(x)$

(for fixed θ_0, θ_1, this is a function of x)



$J(\theta_0, \theta_1)$

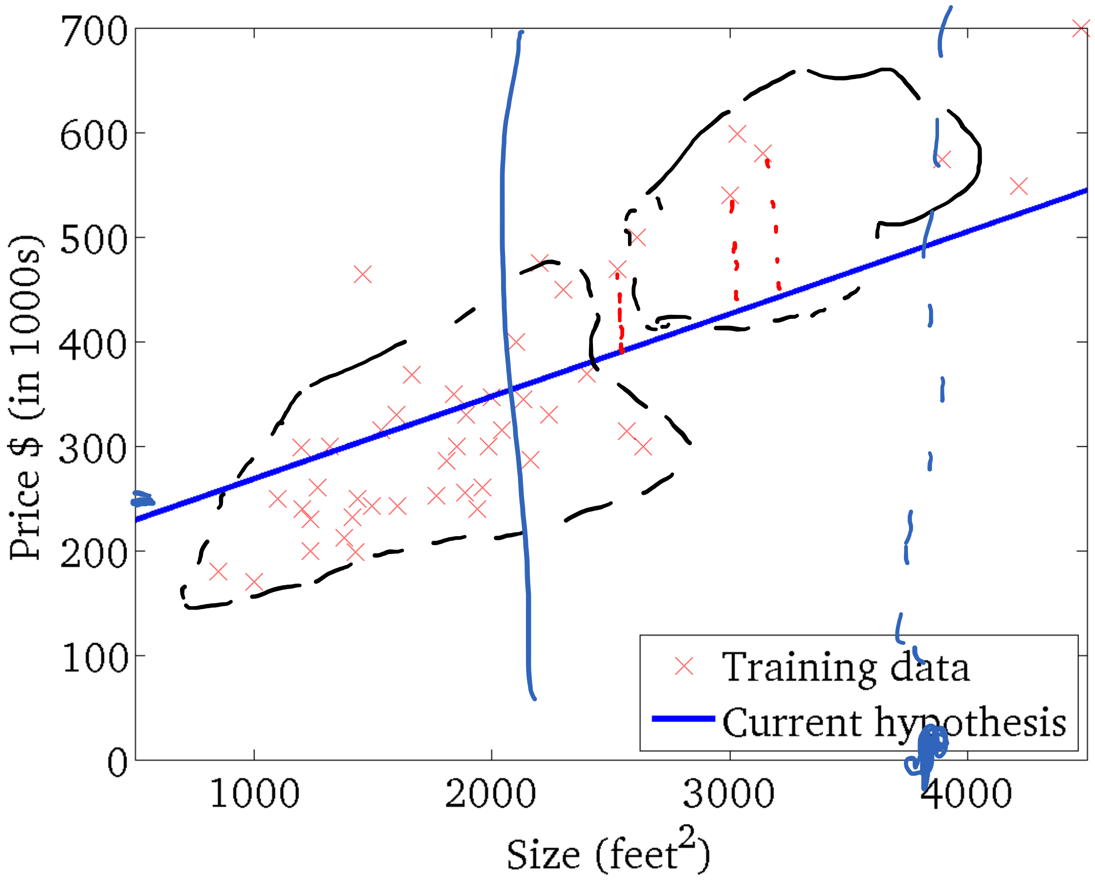
(function of the parameters θ_0, θ_1)



$h_\theta(x) = 450 - .045x$

$h_\theta(x)$

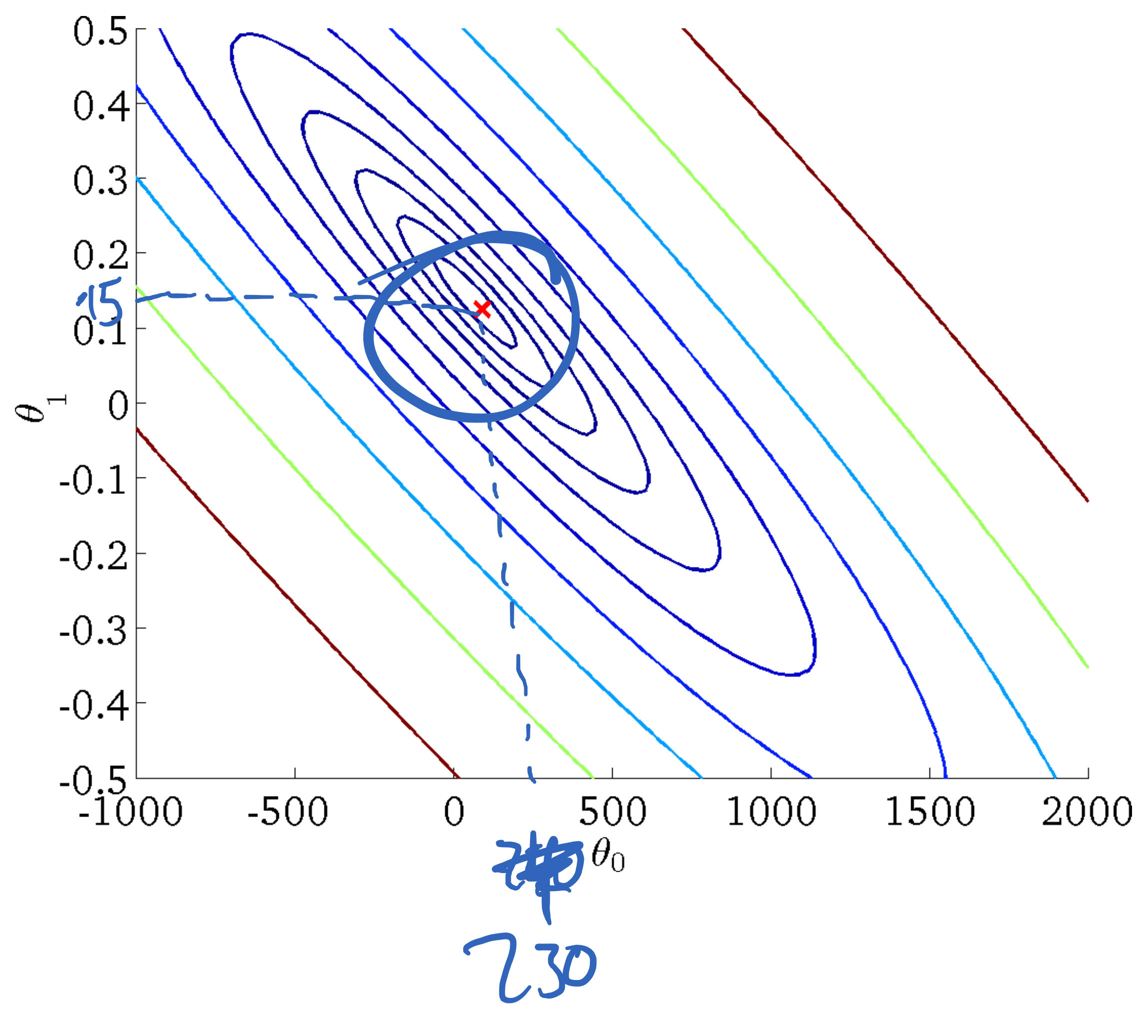
(for fixed θ_0, θ_1, this is a function of x)



$h_\theta(x) = 230 + .15x$

$J(\theta_0, \theta_1)$

(function of the parameters θ_0, θ_1)



Gradiente descendente

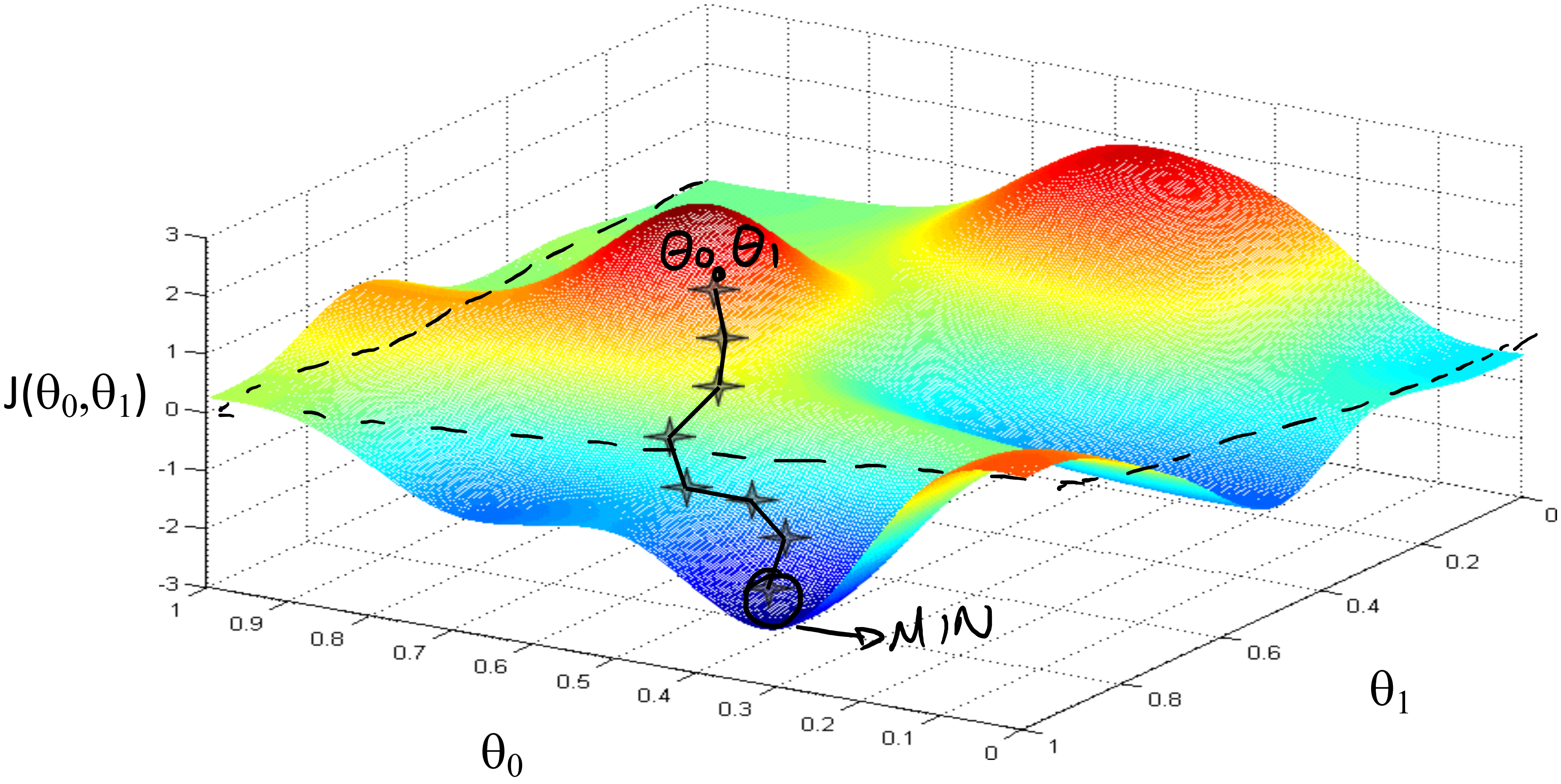
Tenemos la función $J(\theta_0, \theta_1)$

Queremos $\min_{\theta_0, \theta_1} J(\theta_0, \theta_1)$

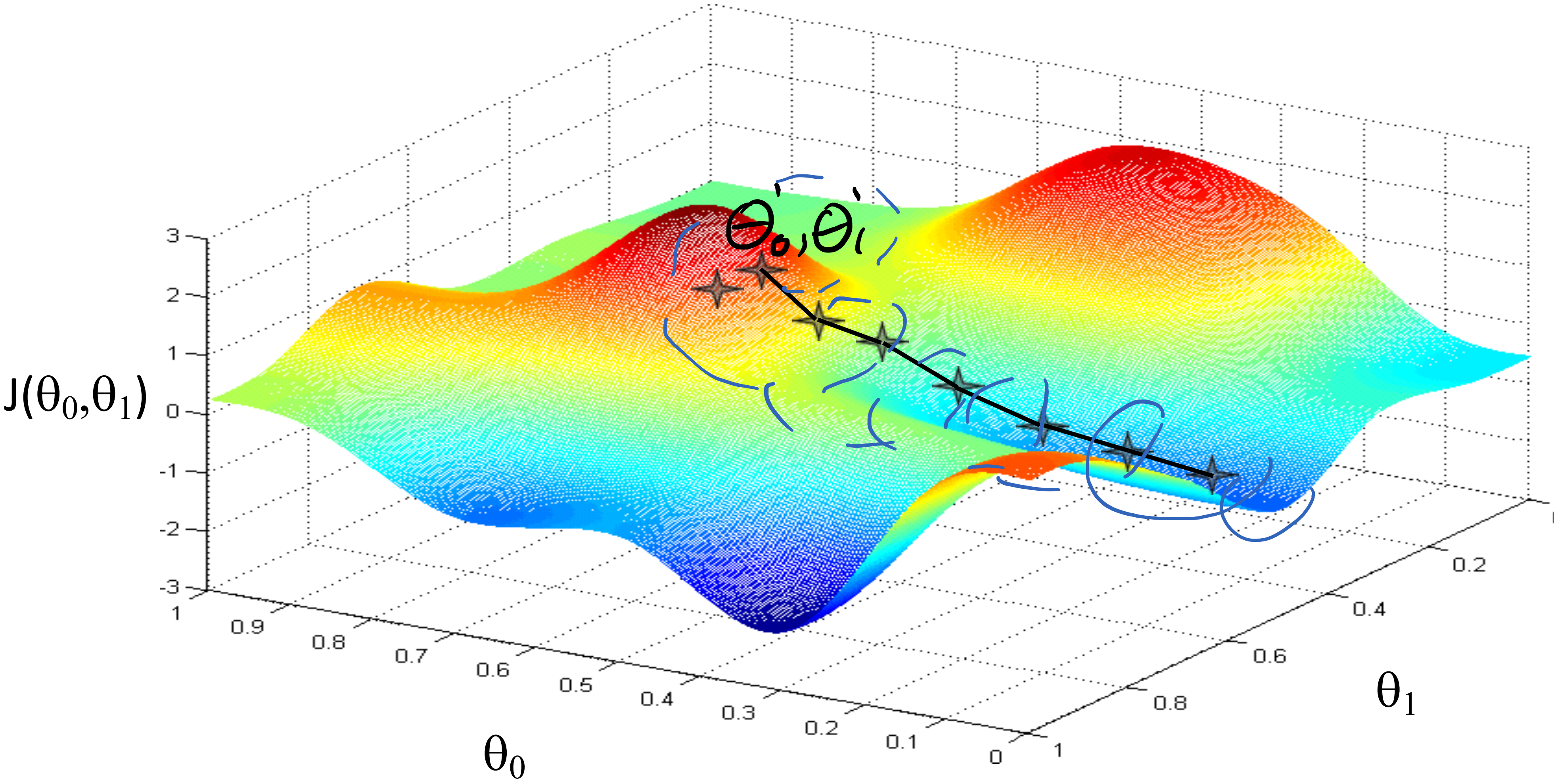
Idea :

- Comenzar con θ_0, θ_1 en algún valor $\theta_0 = 0$, $\theta_1 = 0$
- Cambiar los valores θ_0, θ_1 hasta reducir $J(\theta_0, \theta_1)$

  y se llegue a un valor mínimo (ojalá)



θ_0, θ_1
MIN
$J(\theta_0,\theta_1)$
3
2
1
0
-1
-2
-3
1
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
θ_0
0
0.2
0.4
0.6
0.8
1
θ_1



J(θ0,θ1)
3
2
1
0
-1
-2
-3
1
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
θ0
0
0.2
0.4
0.6
0.8
1
θ1
θ0', θ1'

Gradient descent algorithm

repeat until convergence {

$$\theta_j := \theta_j - \alpha \frac{\partial}{\partial \theta_j} J(\theta_0, \theta_1) \qquad (\text{for } j = 0 \text{ and } j = 1)$$

}

Correct: Simultaneous update

$$\text{temp0} := \theta_0 - \alpha \frac{\partial}{\partial \theta_0} J(\theta_0, \theta_1)$$
$$\text{temp1} := \theta_1 - \alpha \frac{\partial}{\partial \theta_1} J(\theta_0, \theta_1)$$
$$\theta_0 := \text{temp0}$$
$$\theta_1 := \text{temp1}$$

.3

$\theta_0 = .1$ | $\theta_0 = -.2$

$\theta_1 = .3$ | $\theta_1 = .2$

Incorrect:

.1 $\text{temp0} := \theta_0 - \alpha \frac{\partial}{\partial \theta_0} J(\theta_0, \theta_1)$

-.2 $\theta_0 := \text{temp0}$

$\text{temp1} := \theta_1 - \alpha \frac{\partial}{\partial \theta_1} J(\theta_0, \theta_1)$

$\theta_1 := \text{temp1}$

Learning ratio

Algebra lineal

Matrix: Rectangular array of numbers:

Dimension of matrix: number of rows x number of columns

Matrix Elements (entries of matrix)

$$A = \begin{bmatrix} 1402 & 191 \\ 1371 & 821 \\ 949 & 1437 \\ 147 & 1448 \end{bmatrix}$$

$A_{ij} =$ "i, j entry" in the i^{th} row, j^{th} column.

Vector: An n x 1 matrix.

$$y = \begin{bmatrix} 460 \\ 232 \\ 315 \\ 178 \end{bmatrix}$$

$y_i = i^{th}$ element

1--indexed vs 0--indexed:

$$y = \begin{bmatrix} y_1 \\ y_2 \\ y_3 \\ y_4 \end{bmatrix} \qquad y = \begin{bmatrix} y_0 \\ y_1 \\ y_2 \\ y_3 \end{bmatrix}$$

Adición y Multiplicación

Matrix Addition

$$\begin{bmatrix} 1 & 0 \\ 2 & 5 \\ 3 & 1 \end{bmatrix} + \begin{bmatrix} 4 & 0.5 \\ 2 & 5 \\ 0 & 1 \end{bmatrix} =$$

$$\begin{bmatrix} 1 & 0 \\ 2 & 5 \\ 3 & 1 \end{bmatrix} + \begin{bmatrix} 4 & 0.5 \\ 2 & 5 \end{bmatrix} =$$

Scalar Multiplication

$$3 \times \begin{bmatrix} 1 & 0 \\ 2 & 5 \\ 3 & 1 \end{bmatrix} =$$

$$\begin{bmatrix} 4 & 0 \\ 6 & 3 \end{bmatrix} / 4 =$$

Combination of Operands

$$3 \times \begin{bmatrix} 1 \\ 4 \\ 2 \end{bmatrix} + \begin{bmatrix} 0 \\ 0 \\ 5 \end{bmatrix} - \begin{bmatrix} 3 \\ 0 \\ 2 \end{bmatrix} /3$$

Multiplicación de Matrices con Vectores

Example

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

$$\begin{bmatrix} 1 & 3 \\ 4 & 0 \\ 2 & 1 \end{bmatrix} \begin{bmatrix} 1 \\ 5 \end{bmatrix} =$$

Details:

$$A \times x = y$$

$$\begin{bmatrix} \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot \end{bmatrix} \times \begin{bmatrix} \cdot\cdot \\ \cdot\cdot \\ \cdot\cdot \end{bmatrix} = \begin{bmatrix} \cdot\cdot \\ \cdot\cdot \end{bmatrix}$$

m x n matrix (m rows, n columns)

n x 1 matrix (n-dimensional vector)

m-dimensional vector

To get y_i, multiply A's i^{th} row with elements of vector x, and add them up.

Example

$$\mathbb{R}^{3\times 4}$$

$$\begin{bmatrix} 1 & 2 & 1 & 5 \\ 0 & 3 & 0 & 4 \\ -1 & -2 & 0 & 0 \end{bmatrix} \begin{bmatrix} 1 \\ 3 \\ 2 \\ 1 \end{bmatrix} = \begin{bmatrix} 14 \\ 13 \\ -7 \end{bmatrix} \quad \mathbb{R}^3$$

$$(1\times 1)+(2\times 3)+(1\times 2)+(5\times 1)=14$$

$$(0\times 1)+(3\times 3)+(0\times 2)+(4\times 1)=13$$

$$(-1\times 1)+(-2\times 3)+(0\times 2)+(0\times 1)=-7$$

House sizes:

2104
1416
1534
852

$$h_\theta(x) = -40 + 0.25x$$

Multiplicación de matrices

Example

$$\begin{bmatrix} 1 & 3 & 2 \\ 4 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 3 \\ 0 & 1 \\ 5 & 2 \end{bmatrix} =$$

$$\begin{bmatrix} 1 & 3 & 2 \\ 4 & 0 & 1 \end{bmatrix} \times \begin{bmatrix} 1 \\ 0 \\ 5 \end{bmatrix} =$$

$$\begin{bmatrix} 1 & 3 & 2 \\ 4 & 0 & 1 \end{bmatrix} \times \begin{bmatrix} 3 \\ 1 \\ 2 \end{bmatrix} =$$

Details:

$$A \quad \times \quad B \quad = \quad C$$

$$\begin{bmatrix} \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot \end{bmatrix} \times \begin{bmatrix} \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot \end{bmatrix} = \begin{bmatrix} \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot \end{bmatrix}$$

m x n matrix (m rows, n columns)	n x o matrix (n rows, o columns)	m x o matrix

The i^{th} column of the matrix C is obtained by multiplying A with the i^{th} column of B. (for i = 1,2,...,o)

Example

$$\begin{bmatrix} 1 & 3 \\ 2 & 5 \end{bmatrix} \begin{bmatrix} 0 & 1 \\ 3 & 2 \end{bmatrix} =$$

$$\begin{bmatrix} 1 & 3 \\ 2 & 5 \end{bmatrix} \begin{bmatrix} 0 \\ 3 \end{bmatrix} =$$

$$\begin{bmatrix} 1 & 3 \\ 2 & 5 \end{bmatrix} \begin{bmatrix} 1 \\ 2 \end{bmatrix} =$$

House sizes:

2104
1416
1534
852

Have 3 competing hypotheses:

1. $h_\theta(x) = -40 + 0.25x$
2. $h_\theta(x) = 200 + 0.1x$
3. $h_\theta(x) = -150 + 0.4x$

Matrix Matrix

$$\begin{bmatrix} 1 & 2104 \\ 1 & 1416 \\ 1 & 1534 \\ 1 & 852 \end{bmatrix} \times \begin{bmatrix} -40 & 200 & -150 \\ 0.25 & 0.1 & 0.4 \end{bmatrix} = \begin{bmatrix} 486 & 410 & 692 \\ 314 & 342 & 416 \\ 344 & 353 & 464 \\ 173 & 285 & 191 \end{bmatrix}$$

h1 h2 h3

y1 y2 y3

PRED1 PRED2 PRED3

Propiedades de la multiplicación de matrices

$D \times B \times C$ $\quad$ $B \times C \times D$

Let A and B be matrices. Then in general,

$A \times B \neq B \times A$. (not commutative.)

E.g. $$\begin{bmatrix} 1 & 1 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} 0 & 0 \\ 2 & 0 \end{bmatrix} = \begin{bmatrix} 2 & 0 \\ 0 & 0 \end{bmatrix}$$

$$\begin{bmatrix} 0 & 0 \\ 2 & 0 \end{bmatrix} \begin{bmatrix} 1 & 1 \\ 0 & 0 \end{bmatrix} = \begin{bmatrix} 0 & 0 \\ 2 & 2 \end{bmatrix}$$

$A \times B \times C.$

Let $D = B \times C.$ Compute $A \times D.$

Let $E = A \times B.$ Compute $E \times C.$

$4 \times 1 = 4$ $5 \times 1 = 5$ $185{,}000 \times 1 =$

Identity Matrix

Denoted I (or $I_{n \times n}$).

Examples of identity matrices:

$$\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$$
2 x 2

$$\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$
3 x 3

$$\begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$
4 x 4

INFORMAL

For any matrix A,

$$A \cdot I = I \cdot A = A$$

Inversa y Traspuesta

Not all numbers have an inverse.

Matrix inverse:

If A is an m x m matrix, and if it has an inverse,

pinv(

$$AA^{-1} = A^{-1}A = I.$$

$$\begin{bmatrix} 3 & 4 \\ 2 & 16 \end{bmatrix} \begin{bmatrix} .4 & -.1 \\ -.05 & .075 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$$

Matrices that don’t have an inverse are “singular” or “degenerate”

Matrix Transpose

Example:

$$A = \begin{bmatrix} 1 & 2 & 0 \\ 3 & 5 & 9 \end{bmatrix} \qquad A^T = \begin{bmatrix} 1 & 3 \\ 2 & 5 \\ 0 & 9 \end{bmatrix}$$

Let A be an m x n matrix, and let $B = A^T.$

Then B is an n x m matrix, and

$$B_{ij} = A_{ji}.$$