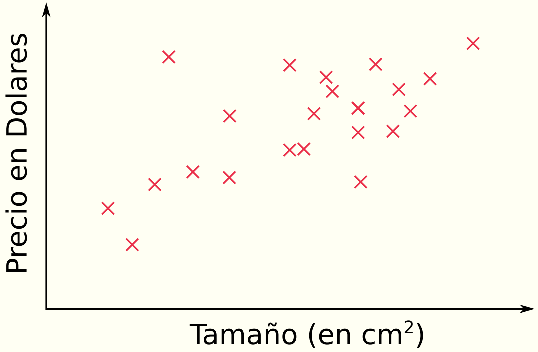
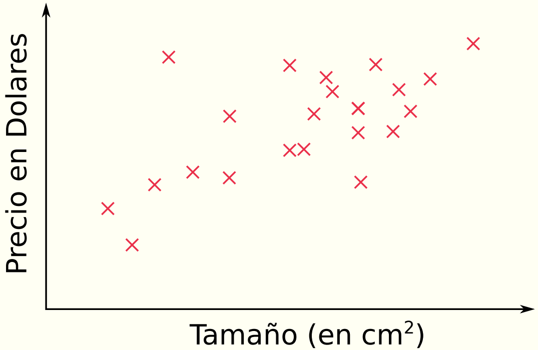
Método del gradiente y regresión lineal

8 de enero de 2017

Set de datos para casas I



Set de datos para casas I



Predecir un valor real utilizando los datos disponibles

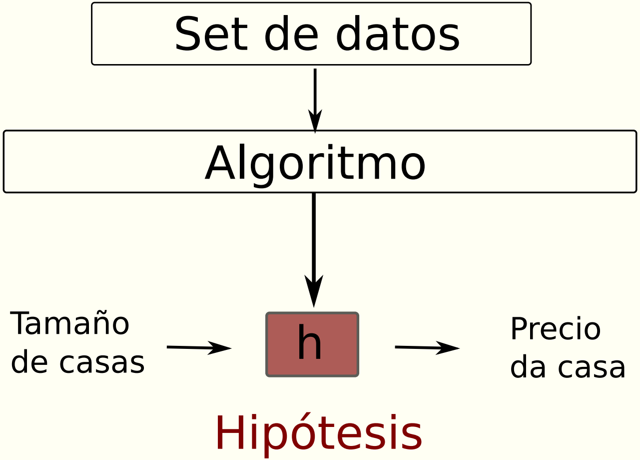
Set de datos para casas II

Tamaño en cm^2	Precio en 1000es de
2014	460
1416	232
1534	178
...	...

Notación:

- $\mathbf{m}$ = Número de puntos de datos
- $\mathbf{x}$'s = variable de entrada
- $\mathbf{y}$'s = variable de salida
- $(\mathbf{x}^{(i)},\mathbf{y}^{(i)})$ = la i-esima dupla de puntos de data

Hipótesis I


Set de datos
Algoritmo
Tamaño
de casas
h
Precio
da casa
Hipótesis

Hipótesis II

Como representamos h?

Hipótesis II

Como representamos h?

$$h_\theta(x) = \theta_0 + \theta_1 x \tag{1}$$

Regresión lineal de una variable

- θ_0 y θ_1 son los parámetros del modelo

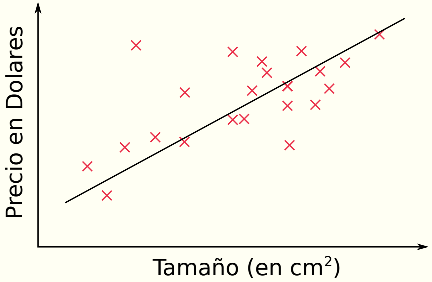
Hipótesis II

Como representamos h?

$$h_\theta(x) = \theta_0 + \theta_1 x \tag{1}$$

Regresión lineal de una variable

- θ_0 y θ_1 son los parámetros del modelo



Parámetros θ

¿Como encuentro los parámetros del modelo?

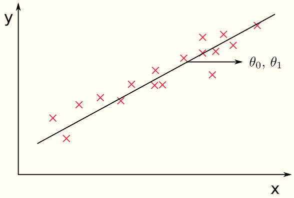
- Idea:
  - Elegir θ_0 y θ_1 tal que $h_\theta(x)$ este cercano a los valores y del set de datos (x, y)

Parámetros θ

¿Como encuentro los parámetros del modelo?

- Idea:
  - Elegir θ_0 y θ_1 tal que $h_\theta(x)$ este cercano a los valores y del set de datos (x, y)
- Se debe minimizar $(h_0 - y)^2$ usando θ_0, θ_1 como parametros

$$\min_{\theta_0,\theta_1} \left[\frac{1}{2m}\sum_{i=1}^{m}((h_0(x^{(i)}) - y^{(i)})^2\right] \tag{2}$$



Función de costo

Es conveniente escribir la función a minimizar como una función de costo

$$J(\theta_0, \theta_1) = \frac{1}{2m} \sum_{i=1}^{m} ((h_0(x^{(i)}) - y^{(i)})^2 \tag{3}$$

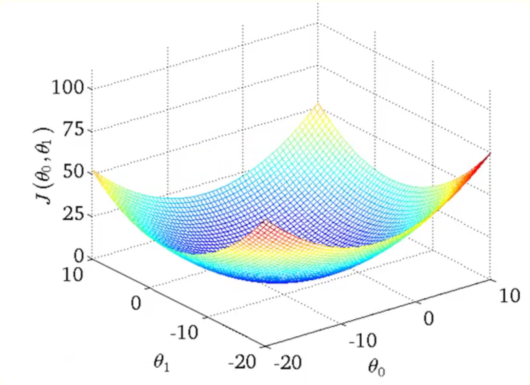
$$\min_{\theta_0, \theta_1} J(\theta_0, \theta_1) \tag{4}$$

- $J(\theta_0, \theta_1) =$ Función de costo, también conocida como función de error cuadrada

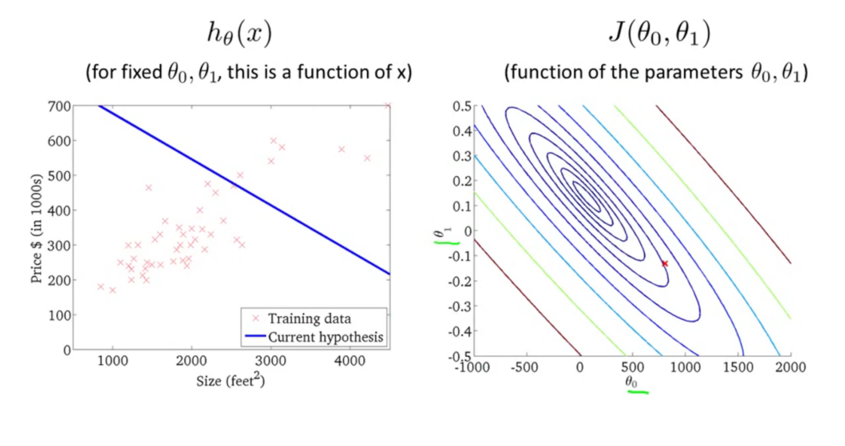
Resumen

- Hipótesis
  - $h_\theta(x) = \theta_0 + \theta_1 x$
- Parametros
  - θ_0, θ_1
- Función de costo:
  - $J(\theta_0, \theta_1) = \frac{1}{2m} \sum_{i=1}^{m}((h_0(x^{(i)}) - y^{(i)})^2$
- Meta:
  - $\min_{\theta_0, \theta_1} J(\theta_0, \theta_1)$

Visualizando la función de costo I



Visualizando la función de costo II



Método del gradiente

Punto de partida

Tenemos una función:

$$J(\theta_0, \theta_1) \tag{5}$$

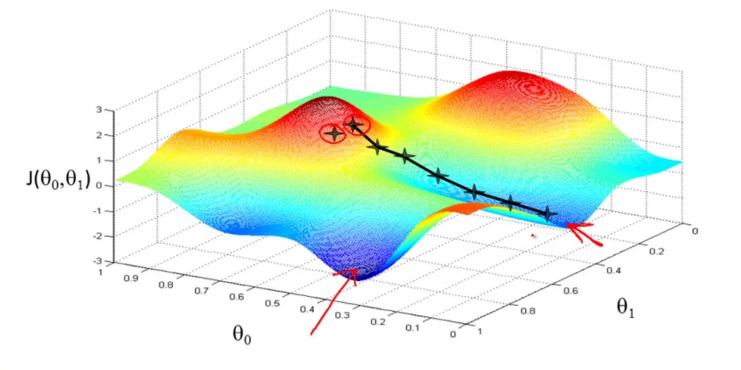
y queremos minimizarla con respecto a los parámetros θ

$$\min_{\theta_0, \theta_1} J(\theta_0, \theta_1) \tag{6}$$

Estrategia:

- Elegir θ_0, θ_1 iniciales
- Cambiar θ_0, θ_1 para reducir $J(\theta_0, \theta_1)$ y llegar a un mínimo.

Descenso a lo largo del gradiente



Algoritmo

$$\text{Repita hasta convergencia} \quad \{\theta_j := \theta_j - \alpha \frac{\partial}{\partial \theta_j} J(\theta_0, \theta_1)\} \quad (\text{Para } j = 0, 1) \qquad (7)$$

El signo $:=$ significa asignación

$\alpha eslarazndeaprendizaje$

Se actualiza θ_0 y θ_1 SIMULTANEAMENTE

El gradiente I

Para poder utilizar el algoritmo que se acaba de mostrar es necesario conocer:

$$\frac{\partial}{\partial \theta_j} J(\theta_0, \theta_1) \tag{8}$$

Si se calculan las derivadas de J con respecto a θ_0 y θ_1 se obtiene que:

$$\frac{\partial}{\partial \theta_0} = \frac{1}{m} \sum_{i=1}^{m} (h_0(x^{(i)}) - y^{(i)}) \tag{9}$$

$$\frac{\partial}{\partial \theta_1} = \frac{1}{m} \sum_{i=1}^{m} (h_0(x^{(i)}) - y^{(i)}) x^{(1)} \tag{10}$$

El gradiente II

Las ecuaciones para poder minimizar la función de costos son entonces:

$$\theta_0 := \theta_0 - \alpha \frac{1}{m} \sum_{i=1}^{m} (h_0(x^{(i)}) - y^{(i)}) \tag{11}$$

$$\theta_1 := \theta_1 - \alpha \frac{1}{m} \sum_{i=1}^{m} (h_0(x^{(i)}) - y^{(i)}) x^{(1)} \tag{12}$$

Estos dos cálculos se repiten hasta que J converge a un mínimo.

Vectorización de la ecuaciones

Muchas veces es conveniente escribir las ecuaciones en forma matricial. Por ejemplo la hipótesis:

$$h_\theta(x) = \theta_0 + \theta_1 x_1 = \theta^T \mathbf{x} \tag{13}$$

$$\mathbf{x} = (x_0, x_1) = (1, x_1) \tag{14}$$