Aprendizaje Automático y Minería de Datos

Regresión lineal

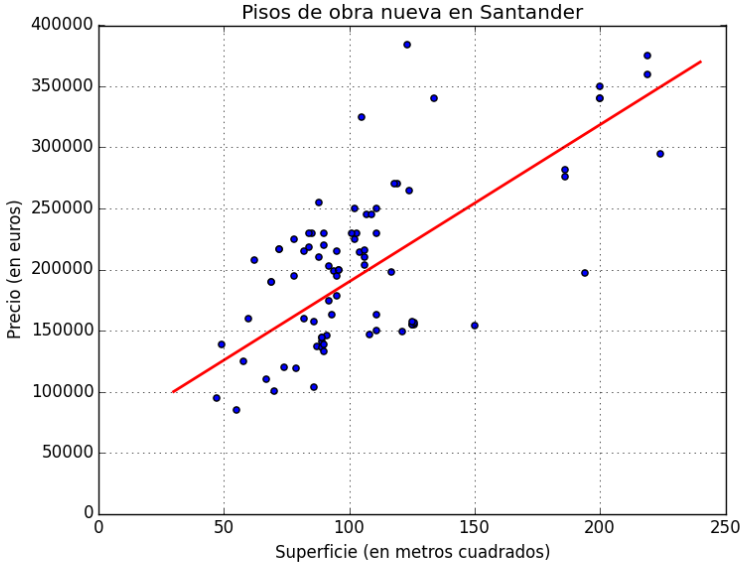
Cristina Tîrnăucă

Dept. Matesco, Universidad de Cantabria

Fac. Ciencias – Grado en Ing. Informática

Regresión lineal univariante

Estimación por mínimos cuadrados ordinarios



Datos de entrada y notación

Datos de entrenamiento:

Superficie en m^2 (x)	Precio en euros (y)
88	210000
90	230000
47	95000
111	230000
...	...

Tabla: Precio (en euros) según superficie (en metros cuadrados)

Notación:
m = número de ejemplos
x = variable de entrada (superficie), $x^{(i)}$
y = variable de salida (precio), $y^{(i)}$

Objetivo: hallar una recta que mejor se aproxime a los datos.

Método iterativo

Algorítmo del gradiente descendente, I

Objetivo: hallar $h(x) = \theta_0 + \theta_1 * x$
Bajo condición: el coste $J(\theta_0, \theta_1) =$ mínimo, donde

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

Idea:

- Empezar con θ_0 y θ_1 al azar
- Modificar θ_0 y θ_1 para reducir $J(\theta_0, \theta_1)$ hasta que (eventualmente) lleguemos a un mínimo.

Método iterativo
Algorítmo del gradiente descendente, II

Idea:

- Empezar con θ_0 y θ_1 al azar
- Modificar θ_0 y θ_1 para reducir $J(\theta_0, \theta_1)$ hasta que (eventualmente) lleguemos a un mínimo.

repetir hasta convergencia $\{$
$\theta_j = \theta_j - \alpha \frac{\partial}{\partial \theta_j} J(\theta_0, \theta_1)$
(para $j = 0$ y $j = 1$)
$\}$
α = ratio de aprendizaje

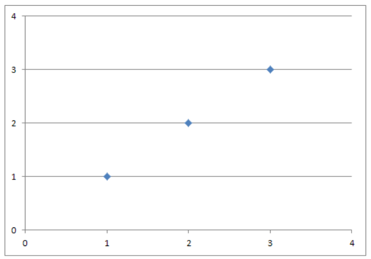
Modelo de regresión lineal:
$h(x) = \theta_0 + \theta_1 * x$
$J(\theta_0, \theta_1) = \frac{1}{2m} \sum_{i=1}^{m} (h(x^{(i)}) - y^{(i)})^2$

Método iterativo
Algorítmo del gradiente descendente, III

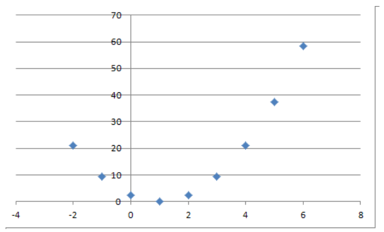
repetir hasta convergencia $\{$
$\theta_0 = \theta_0 -$
$\alpha \frac{\partial}{\partial \theta_0} J(\theta_0, \theta_1) \frac{\partial}{\partial \theta_0} (\frac{1}{2m} \sum_{i=1}^{m} (h(x^{(i)}) - y^{(i)})^2) \frac{\partial}{\partial \theta_0} (\frac{1}{2m} \sum_{i=1}^{m} (\theta_0 + \theta_1 * x^{(i)} - y^{(i)})^2$
$\theta_1 = \theta_1 -$
$\alpha \frac{\partial}{\partial \theta_1} J(\theta_0, \theta_1) \frac{\partial}{\partial \theta_1} (\frac{1}{2m} \sum_{i=1}^{m} (h(x^{(i)}) - y^{(i)})^2) \frac{\partial}{\partial \theta_1} (\frac{1}{2m} \sum_{i=1}^{m} (\theta_0 + \theta_1 * x^{(i)} - y^{(i)})^2$
$\}$
α = ratio de aprendizaje

Función de coste - intuiciones, I
Modelo simplificado: $\theta_0 = 0$

$h(x)$

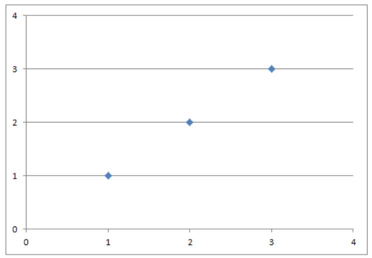


$J(\theta_1)$

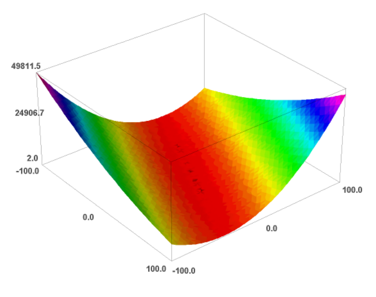


Función de coste - intuiciones, II
Modelo general: θ_0 arbitrario

$h(x)$



$J(\theta_0, \theta_1)$



Función de coste - intuiciones, II

El ratio de aprendizaje α

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

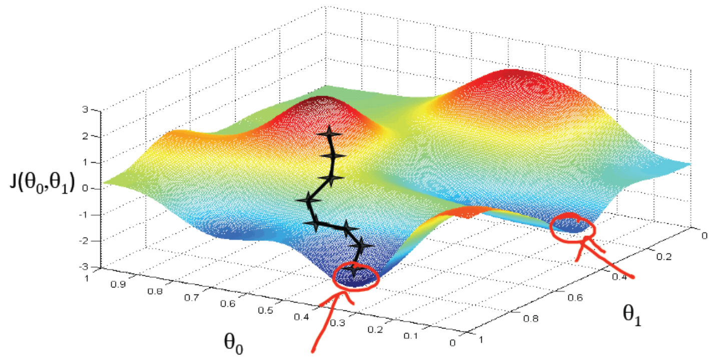
α muy pequeño: el algoritmo tarda mucho en converger

α muy grande: el algoritmo no converge

El algorítmo puede converger a un mínimo local incluso cuando α es fijo.

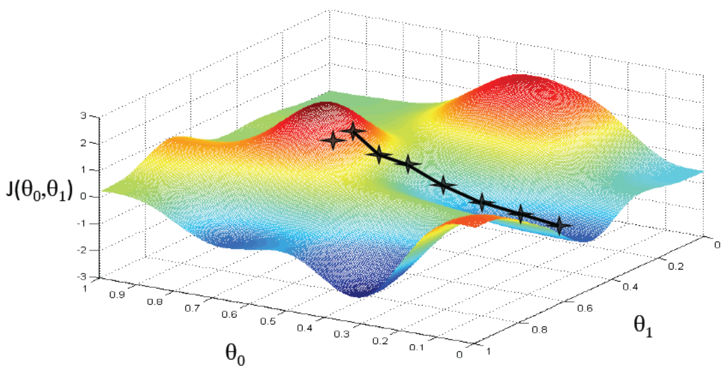
Función de coste - intuiciones, III

Mínimo local / global



Función de coste - intuiciones, IV

Mínimo local / global



Método analítico

Objetivo: hallar $h(x) = \theta_0 + \theta_1 * x$

Bajo condición: el coste $J(\theta_0, \theta_1) =$ mínimo, donde

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

Solución analítica:

$$\theta_0 = \frac{\sum_{i=1}^{m} x^{(i)} * \sum_{i=1}^{m} x^{(i)} y^{(i)} - \sum_{i=1}^{m} x^{(i)} x^{(i)} * \sum_{i=1}^{m} y^{(i)}}{\sum_{i=1}^{m} x^{(i)} * \sum_{i=1}^{m} x^{(i)} - m \sum_{i=1}^{m} x^{(i)} x^{(i)}}$$

$$\theta_1 = \frac{\sum_{i=1}^{m} x^{(i)} * \sum_{i=1}^{m} y^{(i)} - m \sum_{i=1}^{m} x^{(i)} y^{(i)}}{\sum_{i=1}^{m} x^{(i)} * \sum_{i=1}^{m} x^{(i)} - m \sum_{i=1}^{m} x^{(i)} x^{(i)}}$$