Aprendizaje Automático y Minería de Datos

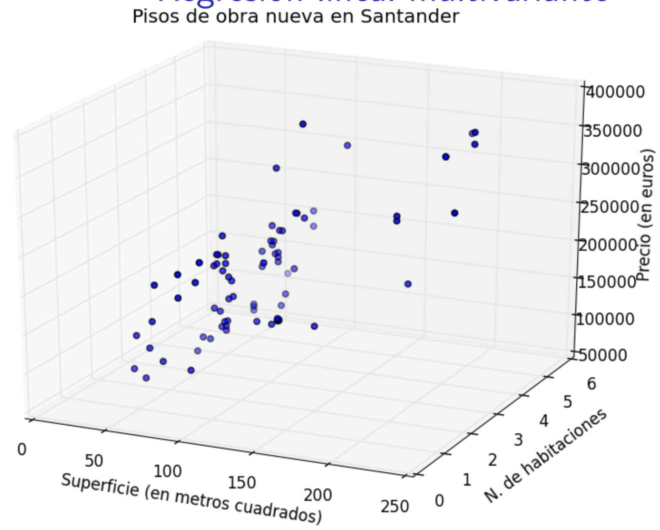
Regresión lineal multivariante

Cristina Tîrnăucă

Dept. Matesco, Universidad de Cantabria

Fac. Ciencias – Grado en Ing. Informática

Regresión lineal multivariante



Datos de entrada y notación

Datos de entrenamiento:

Superficie (x_1)	Número habitaciones (x_2)	Precio (y)
88	2	210000
90	2	230000
47	1	95000
111	4	230000
...	...	...

Notación:
m = número de ejemplos
x_1, x_2 = variables de entrada, $x_1^{(i)}$ y $x_2^{(i)}$
y = variable de salida, $y^{(i)}$

Objetivo: hallar $h_\theta(x_1, x_2) = \theta_0 + \theta_1 * x_1 + \theta_2 * x_2$ tal que el error mínimo cuadrado para los datos de entrenamiento sea mínimo.

Método analítico

Caso particular: 2 variables

Objetivo: hallar $h_\theta(x_1, x_2) = \theta_0 + \theta_1 * x_1 + \theta_2 * x_2$
Bajo condición: el coste $J(\theta)$ = mínimo, donde

$$J(\theta) = \frac{1}{2m} \sum_{i=1}^{m} (h_\theta(x_1^{(i)}, x_2^{(i)}) - y^{(i)})^2$$

Solución analítica:
$\theta = (X^t * X)^{-1} * X^t * y$

$$X = \begin{pmatrix} 1 & x_1^{(1)} & x_2^{(1)} \\ 1 & x_1^{(2)} & x_2^{(2)} \\ \dots & \dots & \dots \\ 1 & x_1^{(m)} & x_2^{(m)} \end{pmatrix} \qquad y = \begin{pmatrix} y^{(1)} \\ y^{(2)} \\ \dots \\ y^{(m)} \end{pmatrix} \quad \theta = \begin{pmatrix} \theta_0 \\ \theta_1 \\ \theta_2 \end{pmatrix}$$

Método analítico

Caso general: p variables

Objetivo: hallar $h_\theta(x_1, \dots, x_p) = \theta_0 + \theta_1 * x_1 + \dots + \theta_p * x_p$
Bajo condición: el coste $J(\theta)$ = mínimo, donde

$$J(\theta) = \frac{1}{2m} \sum_{i=1}^{m} (h_\theta(x_1^{(i)}, \dots, x_p^{(i)}) - y^{(i)})^2$$

Solución analítica:
$\theta = (X^t * X)^{-1} * X^t * y$

$$X = \begin{pmatrix} 1 & x_1^{(1)} & \dots & x_p^{(1)} \\ 1 & x_1^{(2)} & \dots & x_p^{(2)} \\ \dots & \dots & \dots & \dots \\ 1 & x_1^{(m)} & \dots & x_p^{(m)} \end{pmatrix} \quad y = \begin{pmatrix} y^{(1)} \\ y^{(2)} \\ \dots \\ y^{(m)} \end{pmatrix} \quad \theta = \begin{pmatrix} \theta_0 \\ \theta_1 \\ \dots \\ \theta_p \end{pmatrix}$$

Método iterativo

Algorítmo del gradiente descendente, caso general, I

Objetivo: hallar $h_\theta(x) = \theta_0 + \theta_1 * x_1 + \dots + \theta_p * x_p$
Bajo condición: el coste $J(\theta)$ = mínimo, donde

$$J(\theta) = \frac{1}{2m} \sum_{i=1}^{m} (h_\theta(x_1^{(i)}, \dots, x_p^{(i)}) - y^{(i)})^2$$

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

Método iterativo

Algorítmo del gradiente descendente, caso general, II

Objetivo: hallar $h_\theta(x) = \theta_0 + \theta_1 * x_1 + \dots + \theta_p * x_p$
Bajo condición: el coste $J(\theta)$ = mínimo, donde

$$J(\theta) = \frac{1}{2m} \sum_{i=1}^{m} (h_\theta(x_1^{(i)}, \dots, x_p^{(i)}) - y^{(i)})^2$$

repetir hasta convergencia $\{$
  $\theta_0 = \theta_0 - \alpha * \frac{1}{m} \sum_{i=1}^{m} (\theta_0 + \theta_1 * x_1^{(i)} + \dots + \theta_p * x_p^{(i)} - y^{(i)}) * 1$
  $\theta_j = \theta_j - \alpha * \frac{1}{m} \sum_{i=1}^{m} (\theta_0 + \theta_1 * x_1^{(i)} + \dots + \theta_p * x_p^{(i)} - y^{(i)}) * x_j^{(i)}$
  (para $j = 1, \dots, p$)
$\}$
α = ratio de aprendizaje

Método iterativo

Escalación de variables

Re-escalación de variables $\rightarrow$ [0,1]

$$x' = \frac{x - \text{min}}{\text{max} - \text{min}}$$

Estandarización de variables

$$x' = \frac{x - \text{promedio}}{\text{desviación estándar}}$$

Ventajas y desventajas

Iterativo

- Hay que elegir α.
- Hace falta iterar muchas veces.
- Escalación de variables.
- Funciona bien incluso para p muy grande.

Analítico

- No hay que elegir α.
- No hace falta iterar.
- Hay que calcular $(X^t * X)^{-1}$
- Lento, si p es muy grande.

Regresión polinómica

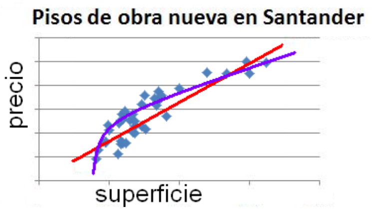


Figure: Precio según superficie

$\theta_0 + \theta_1 * x_1 + \theta_2 * x_2 + \theta_3 * x_3$
$x_1 = superficie$
$x_2 = (superficie)^2$
$x_3 = (superficie)^3$

$\theta_0 + \theta_1 * x + \theta_2 * x^2$

$\theta_0 + \theta_1 * x + \theta_2 * x^2 + \theta_3 * x^3$