



Regresión Lineal Múltiple

Luis Zúñiga

Definición del Modelo

- Modelar el mundo en dos variables es poco útil.
- Necesitamos mejores modelos.
- Nueva idea: un modelo lineal con múltiples variables.

$$y = \theta_0 + \theta_1 x_1 + \theta_2 x_2 + \cdots + \theta_p x_p$$


Diagram illustrating the linear model equation with annotations:

- $\theta_1 x_1$ is labeled "Característica 1" (Characteristic 1).
- $\theta_2 x_2$ is labeled "Característica 2" (Characteristic 2).
- $\theta_p x_p$ is labeled "Característica p " (Characteristic p).

Definición del Modelo

Supongamos que modelamos el precio de una casa de la siguiente manera:

Tamaño	Número de Recámaras	Número de pisos	Antigüedad
--------	---------------------	-----------------	------------


$$\text{Precio} = y = \theta_0 + \theta_1 x_1 + \theta_2 x_2 + \theta_3 x_3 + \theta_4 x_4$$

$p \rightarrow$ # de características

$x_i \rightarrow$ i-ésima característica del modelo

$x^{(j)} \rightarrow$ los valores del j-ésimo ejemplo del conjunto de datos

$x_i^{(j)} \rightarrow$ los valores del j-ésimo ejemplo del conjunto de datos para la i-ésima característica del modelo

Definición del Modelo

Actividad

Supongamos que se tiene el mismo modelo para determinar el precio de una casa. ¿Cómo se interpretaría la hipótesis descrita a continuación?

Tamaño	Número de Recámaras	Número de pisos	Antigüedad
--------	---------------------	-----------------	------------

$$h_{\theta}(x) = 80 + 0.1x_1 + 0.001x_2 + 3x_3 - 2x_4$$

Definición del Modelo

Sea la hipótesis:

$$h_{\theta}(\mathbf{x}) = \theta_0 + \theta_1 x_1 + \theta_2 x_2 + \cdots + \theta_p x_p$$

Entonces, podemos expresarla como

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

donde

$$\mathbf{x} = \begin{bmatrix} x_0 \\ x_1 \\ \vdots \\ x_p \end{bmatrix} = \begin{bmatrix} 1 \\ x_1 \\ \vdots \\ x_p \end{bmatrix}, \quad \Theta = \begin{bmatrix} \theta_0 \\ \theta_1 \\ \vdots \\ \theta_p \end{bmatrix}$$

y $x_0 = 1$, de tal manera que

$$h_{\theta}(\mathbf{x}) = \theta_0 x_0 + \theta_1 x_1 + \theta_2 x_2 + \cdots + \theta_p x_p$$

Definición del Modelo - Resumen

Hipótesis:

$$h_{\boldsymbol{\theta}}(\mathbf{x}) = \theta_0 x_0 + \theta_1 x_1 + \theta_2 x_2 + \cdots + \theta_p x_p$$

$$h_{\boldsymbol{\theta}}(\mathbf{x}) = \boldsymbol{\Theta}^T \mathbf{x}$$

Parámetros:

$$\boldsymbol{\theta} = (\theta_0, \theta_1, \dots, \theta_p)$$

Función de Costo:

$$J(\boldsymbol{\theta}) = \frac{1}{2n} \sum_{i=1}^n (h_{\boldsymbol{\theta}}(\mathbf{x}^{(i)}) - y^{(i)})^2$$

Gradiente Descendiente:

$$\theta_j = \theta_j - \eta \frac{\partial}{\partial \theta_j} J(\boldsymbol{\theta})$$

Gradiente Descendiente

Actividad

Supongan que el modelo tiene la forma

$$h_{\theta}(\mathbf{x}) = \theta_0 x_0 + \theta_1 x_1 + \theta_2 x_2 + \theta_3 x_3.$$

Determinar

$$\frac{\partial}{\partial \theta_j} J(\boldsymbol{\theta})$$

para $j=0,1,2,3$. Recuerden que

$$J(\boldsymbol{\theta}) = \frac{1}{2n} \sum_{i=1}^n (h_{\theta}(\mathbf{x}^{(i)}) - y^{(i)})^2$$

Gradiente Descendiente

Anteriormente:

$$\theta_0 = \theta_0 - \eta \frac{1}{n} \sum (h_{\theta}(x_i) - y_i)$$

$$\theta_1 = \theta_1 - \eta \frac{1}{n} \sum (h_{\theta}(x_i) - y_i) x_i$$

Ahora:

$$\theta_j = \theta_j - \eta \frac{1}{n} \sum_{i=1}^n (h_{\theta}(x^{(i)}) - y^{(i)}) x_j^{(i)}$$

para $j = 1, 2, \dots, p$.

Estimación de Parámetros – Ecuación Normal

Considerando la representación matricial, es posible encontrar de manera analítica los parámetros del modelo. En particular, debemos encontrar la solución para Θ en el sistema de ecuaciones dado por:

$$X^T X \Theta - X^T Y = 0$$

donde

$$X = \begin{bmatrix} 1 & x_{11} & \dots & x_{1p} \\ 1 & x_{21} & \dots & x_{2p} \\ \vdots & \vdots & \vdots & \vdots \\ 1 & x_{n1} & \dots & x_{np} \end{bmatrix}, \quad Y = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}, \quad \Theta = \begin{bmatrix} \theta_0 \\ \theta_1 \\ \vdots \\ \theta_p \end{bmatrix}$$

y cuya solución es:

$$\Theta = (X^T X)^{-1} X^T Y$$

Para finalizar...

Gradiente Descendiente

- Se debe elegir la razón de aprendizaje η .
- Es un proceso iterativo. Puede tardar en converger a un valor.
- Se puede optimizar, por lo que funciona «bien» para n grande.

Ecuación Normal

- No es necesario elegir un parámetro adicional con la razón de aprendizaje η .
- Solución analítica. Nada de iteraciones.
- Se debe calcular $(X^T X)^{-1}$, cuya complejidad es $O(n^3)$.
- Es lento para n grande.

Tarea

- Demostrar todo en la lámina 9.
- ¿Qué pasa cuando, en la ecuación normal, $(X^T X)$ no es invertible?
 - Tip: ¿Qué es la inversa de Moore-Penrose?
- Investigar qué es el *feature scaling* aplicado en gradiente descendiente.
 - Investigar para qué sirve y qué ventajas ofrece normalizar las características de un modelo de aprendizaje.
 - Explicar en qué consiste la *mean normalization*.



Gracias

Luis Zúñiga

Correo: p40887@correo.uia.mx

Web: <https://lzun.github.io>