Tecnológico de Monterrey

Aprendizaje automático

Regresión lineal

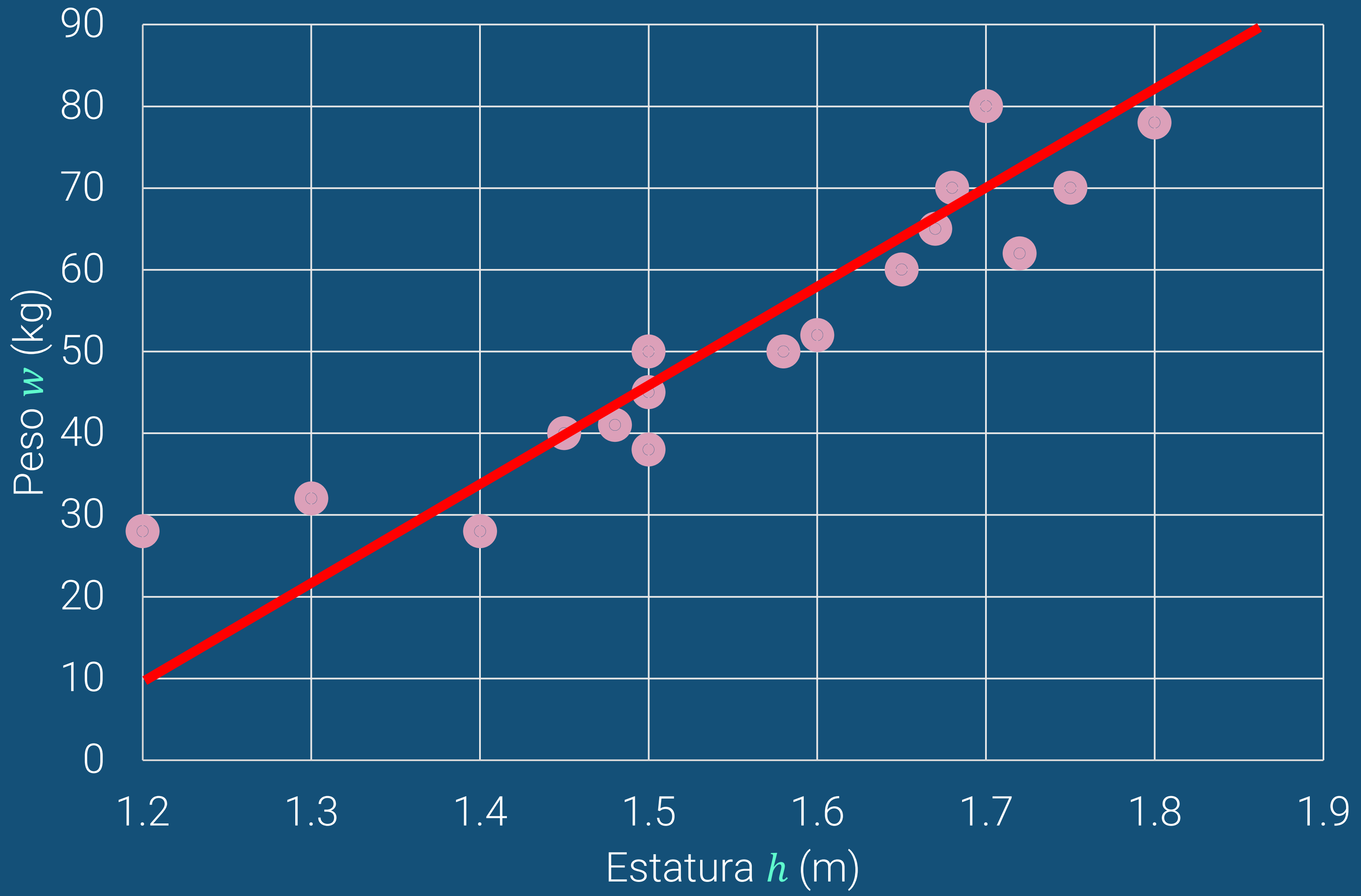
Dr. Omar Mendoza Montoya

14/08/2024

Regresión es el proceso de estimar la relación que hay entre variables de entrada y salida.

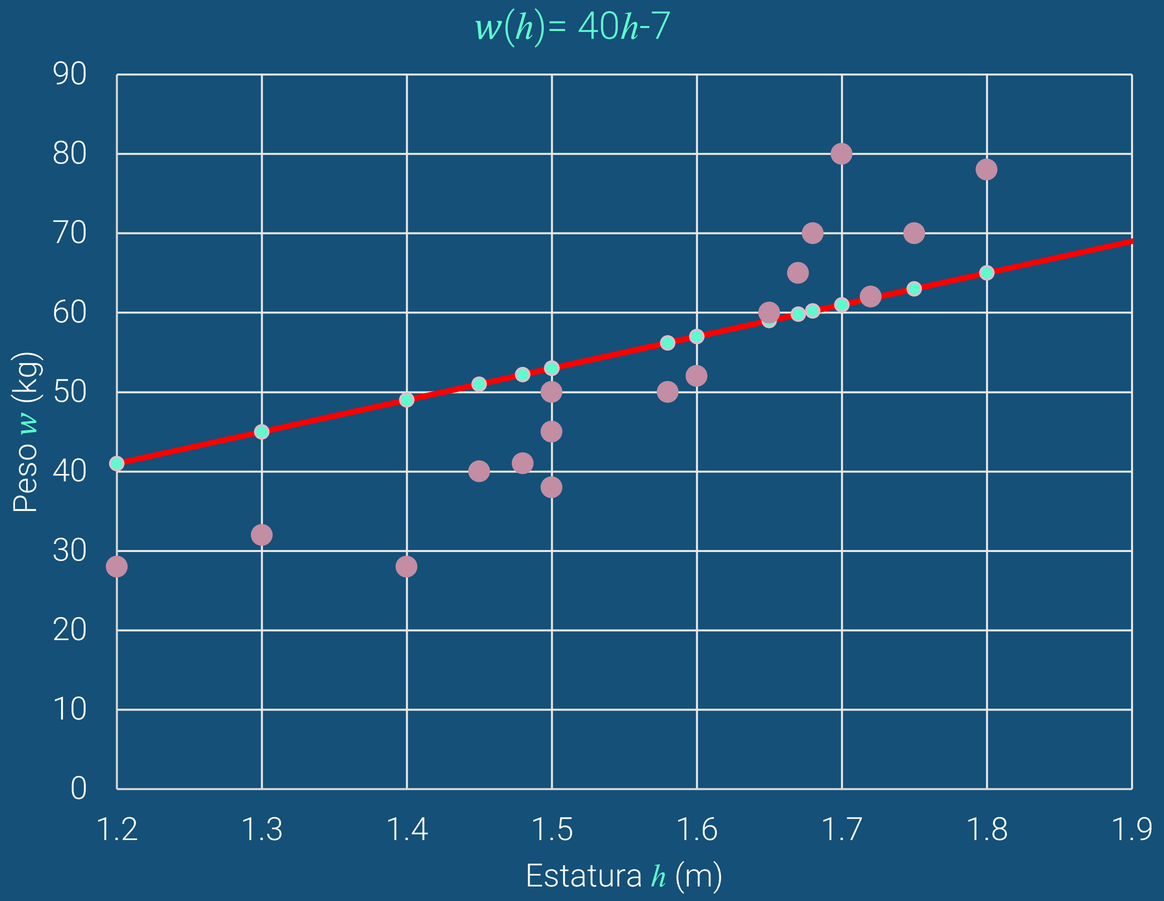
Típicamente, las variables de salida son números reales, mientras que las variables de entrada pueden ser de diferentes tipos (nominales, numéricas, etc.).

Regresión

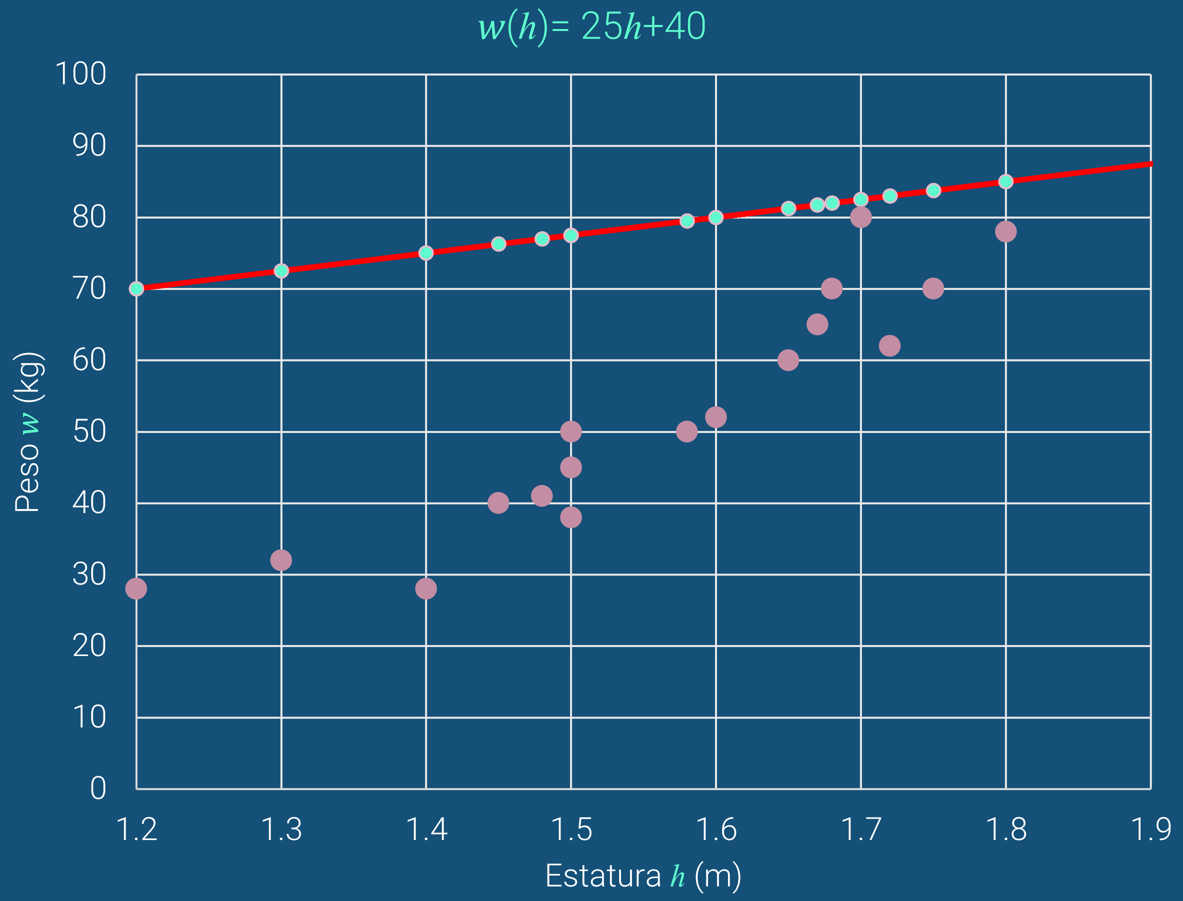


- A mayor estatura, mayor peso.
- La relación parece ser lineal.
- Proponemos como modelo:
  $$w(h) = ah + b$$
- Problema: ¿cómo calculo a y b?

Regresión

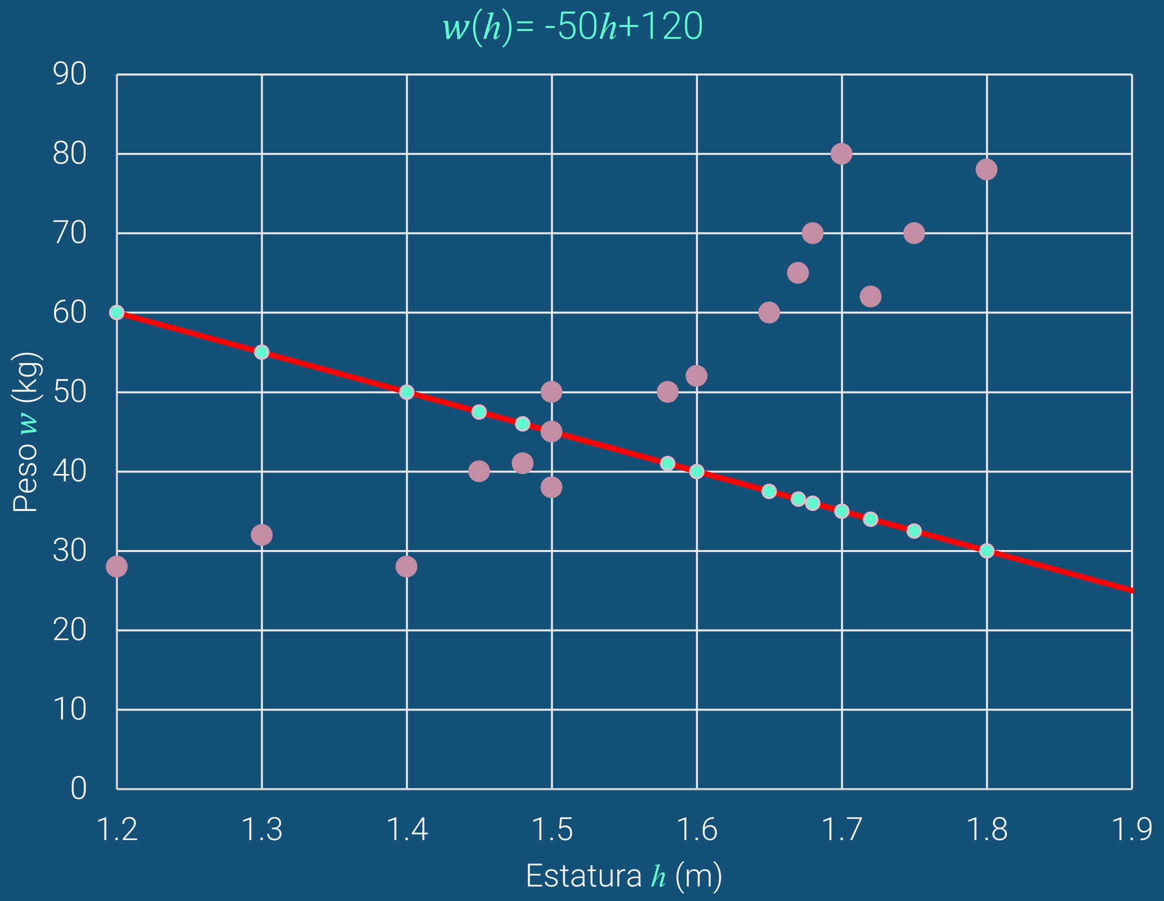


$w(h)$= 40h-7
90
80
70
60
50
40
30
20
10
0
Peso w (kg)
1.2
1.3
1.4
1.5
1.6
1.7
1.8
1.9
Estatura h (m)

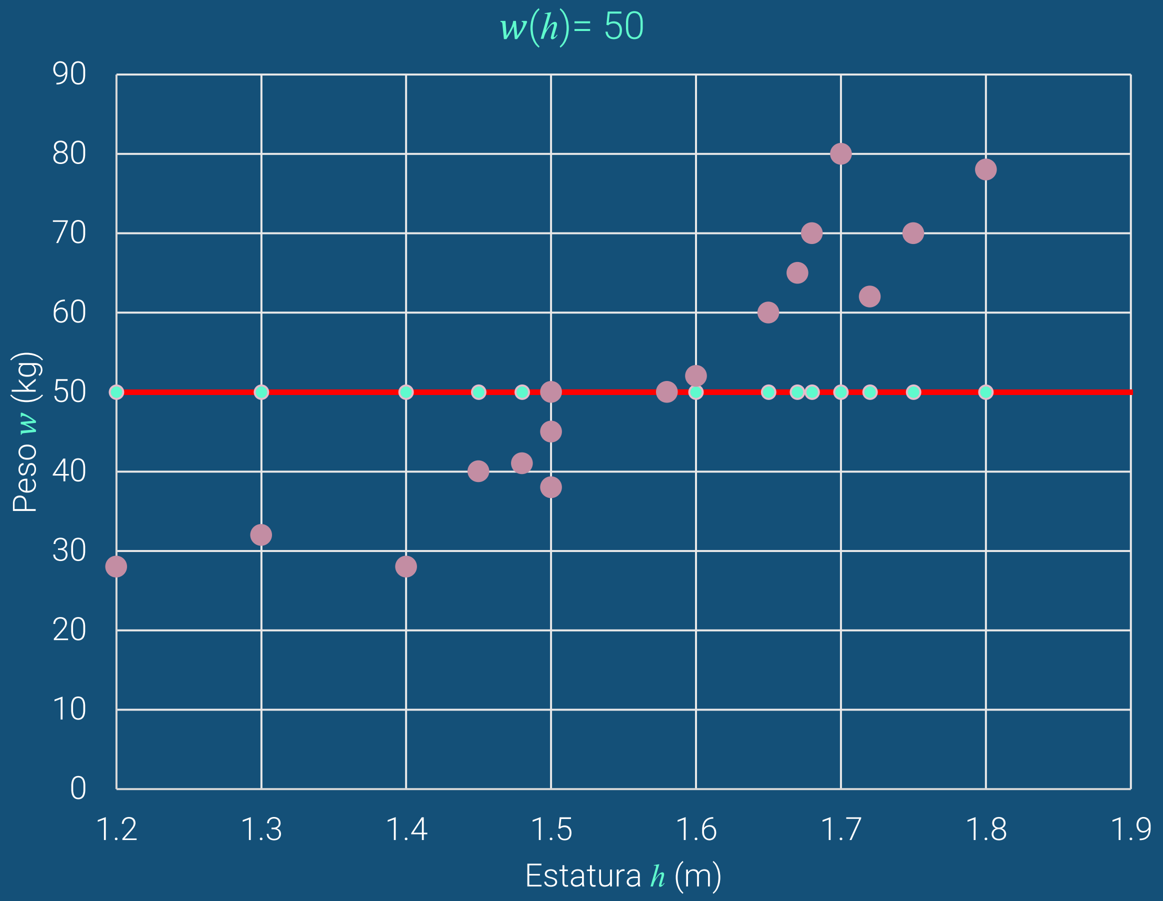


$w(h)$= 25h+40
100
90
80
70
60
50
40
30
20
10
0
Peso w (kg)
1.2
1.3
1.4
1.5
1.6
1.7
1.8
1.9
Estatura h (m)

Regresión



$w(h)$= -50h+120
90
80
70
60
50
40
30
20
10
0
Peso w (kg)
1.2
1.3
1.4
1.5
1.6
1.7
1.8
1.9
Estatura h (m)



$w(h)$= 50
90
80
70
60
50
40
30
20
10
0
Peso w (kg)
1.2
1.3
1.4
1.5
1.6
1.7
1.8
1.9
Estatura h (m)

¿Cómo podemos encontrar el modelo que se ajuste los mejor posible a los datos?

Comenzamos por definir una medida de error que permita comparar modelos y con ello decidir cuál es mejor.

Error cuadrático medio

Sea $f(x; a, b) = ax + b$, y $D = \{(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)\}$ un conjunto de datos con el que vamos a ajustar el modelo (conjunto de entrenamiento). El error cuadrático medio (mean squared error o MSE) del modelo $f(x; a, b)$ está dado por:

$$MSE(D; a, b) = \frac{1}{n}\sum_{i=1}^{n}\left(y_i - f(x_i; a, b)\right)^2$$

Si definimos como el residuo r_i de la observación i a:

$$r_i = y_i - f(x_i; a, b)$$

Entonces:

$$MSE(D; a, b) = \frac{1}{n}\sum_{i=1}^{n} {r_i}^2$$

Una vez que tengamos una medida de error, lo que sigue es calcular a, b, que correspondan al menor error posible.

Es decir, tenemos que resolver el problema:

$$a^*, b^* = \underset{a,b}{\operatorname{argmin}}\Big(MSE(D; a, b)\Big)$$

Optimización de parámetros

- El valor mínimo de $MSE(D; a, b)$ se da cuando el gradiente de esta función respecto a los parámetros a y b es igual a cero. Es decir:

$$\nabla MSE = \begin{bmatrix} \frac{\partial MSE}{\partial a} \\ \frac{\partial MSE}{\partial b} \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

- Derivando:

$$\frac{\partial MSE}{\partial a} = \frac{2}{n} \sum_{i=1}^{n} x_i(-y_i + ax_i + b)$$

$$\frac{\partial MSE}{\partial b} = \frac{2}{n} \sum_{i=1}^{n} (-y_i + ax_i + b)$$

Optimización de parámetros

- Igualando a cero las derivadas parciales y despejando a y b:

$$a = \frac{\sum_{i=1}^{n} x_i y_i - \frac{1}{n}\sum_{i=1}^{n} x_i \sum_{i=1}^{n} y_i}{\sum_{i=1}^{n} x_i^2 - \frac{1}{n}\left(\sum_{i=1}^{n} x_i\right)^2}$$

$$b = \frac{1}{n}\sum_{i=1}^{n} y_i - \frac{a}{n}\sum_{i=1}^{n} x_i$$

Regresión lineal múltiple

Modelo de regresión lineal múltiple

Modelo:

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \cdots + \beta_p x_p$$

Variables predictoras: $x_1, x_2, x_3, \dots, x_p$ (p variables)

Variable dependiente: y

Parámetros del modelo: $\beta_0, \beta_1, \beta_2, \beta_3, \dots, \beta_p$ ($p + 1$ parámetros)

Representación vectorial

Vector de variables predictoras:

$$X = \begin{bmatrix} 1 \\ x_1 \\ x_2 \\ \vdots \\ x_p \end{bmatrix}$$

Parámetros del modelo:

$$\beta = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \\ \vdots \\ \beta_p \end{bmatrix}$$

Modelo:

$$y = X^T \beta$$

Optimización de parámetros

- Para entrenar el modelo, necesitamos un conjunto de datos $D = \{(X_1, y_1), (X_2, y_2), \dots, (X_n, y_n)\}$ con n observaciones.

- Éste lo representaremos por la siguiente matriz X_t y vector Y_t (el subíndice t es de “training”).

$$X_t = \begin{bmatrix} 1 & x_{1,1} & x_{1,2} & \cdots & x_{1,p} \\ 1 & x_{2,1} & x_{2,2} & \dots & x_{2,p} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & x_{n,1} & x_{n,2} & \dots & x_{n,p} \end{bmatrix} \qquad Y_t = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}$$

- Donde $X_i^T = [1 \quad x_{i,1} \quad x_{i,2} \quad x_{i,3} \quad \dots \quad x_{i,p}]$ es el vector de valores de las variables independientes para la observación i.

Optimización de parámetros

Función de error o de costo:

$$MSE(D;\beta) = \frac{1}{n}\sum_{i=1}^{n}\left(y_i - X_i^T\beta\right)^2 = \frac{1}{n}\sum_{i=1}^{n} r_i^2$$

donde el residuo r_i es:

$$r_i = y_i - X_i^T\beta$$

Solución de minimizar $\boldsymbol{MSE(D;\beta)}$:

$$\beta = (X_t^T X_t)^{-1} X_t^T Y_t$$

Regresión lineal múltiple con más de una variable dependiente

Regresión con múltiples variables dependientes

Modelo:

$$
\begin{aligned}
y_1 &= \beta_{0,1} + \beta_{1,1}x_1 + \beta_{2,1}x_2 + \beta_{3,1}x_3 + \cdots + \beta_{p,1}x_p \\
y_2 &= \beta_{0,2} + \beta_{1,2}x_1 + \beta_{2,2}x_2 + \beta_{3,2}x_3 + \cdots + \beta_{p,2}x_p \\
y_3 &= \beta_{0,3} + \beta_{1,3}x_1 + \beta_{2,3}x_2 + \beta_{3,3}x_3 + \cdots + \beta_{p,3}x_p \\
&\vdots \\
y_m &= \beta_{0,m} + \beta_{1,m}x_1 + \beta_{2,m}x_2 + \beta_{3,m}x_3 + \cdots + \beta_{p,m}x_p
\end{aligned}
$$

Variables predictoras: $x_1, x_2, x_3, \dots, x_p$ (p variables)

Variables dependientes: $y_1, y_2, y_3, \dots, y_p$ (m variables)

Parámetros del modelo: $\beta_{i,j}$ ($(p+1) \times m$ parámetros)

Representación vectorial

Vector de variables predictoras:

$$X = \begin{bmatrix} 1 \\ x_1 \\ x_2 \\ \vdots \\ x_p \end{bmatrix}$$

Vector de variables dependientes:

$$Y = \begin{bmatrix} y_1 & y_2 & y_3 & \cdots & y_p \end{bmatrix}$$

Parámetros del modelo:

$$\beta = \begin{bmatrix} \beta_{0,1} & \beta_{0,2} & \beta_{0,3} & \cdots & \beta_{0,m} \\ \beta_{1,1} & \beta_{1,2} & \beta_{1,3} & \cdots & \beta_{1,1} \\ \beta_{2,1} & \beta_{2,2} & \beta_{2,3} & \cdots & \beta_{2,1} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \beta_{p,1} & \beta_{p,2} & \beta_{p,3} & \cdots & \beta_{p,m} \end{bmatrix}$$

Modelo:

$$Y = X^T \beta$$

Optimización de parámetros

- Para entrenar el modelo, necesitamos un conjunto de datos $D = \{(X_1, Y_1), (X_2, Y_2), \dots, (X_n, Y_n)\}$ con n observaciones.

- Éste lo representaremos por las siguientes matrices X_t y Y_t (el subíndice t es de "training").

$$X_t = \begin{bmatrix} 1 & x_{1,1} & x_{1,2} & \cdots & x_{1,p} \\ 1 & x_{2,1} & x_{2,2} & \cdots & x_{2,p} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & x_{n,1} & x_{n,2} & \cdots & x_{n,p} \end{bmatrix} \qquad Y_t = \begin{bmatrix} y_{1,1} & y_{1,2} & \cdots & y_{1,m} \\ y_{2,1} & y_{2,2} & \cdots & y_{2,m} \\ \vdots & \vdots & \ddots & \vdots \\ y_{n,1} & y_{n,2} & \cdots & y_{n,m} \end{bmatrix}$$

- Donde $X_i^T = [1 \quad x_{i,1} \quad x_{i,2} \quad x_{i,3} \quad \dots \quad x_{i,p}]$ es el vector de valores de las variables independientes para la observación i, y $Y_i = [y_{i,1} \quad y_{i,2} \quad y_{i,3} \quad \dots \quad y_{i,m}]$ es el vector de valores de las variables dependientes para la observación i.

Optimización de parámetros

Función de error o de costo:

$$MSE(D;\beta) = \frac{1}{n}\sum_{i=1}^{n}\|Y_i - X_i\beta\|^2 = \frac{1}{n}\sum_{i=1}^{n}\|r_i\|^2 = \frac{1}{n}\sum_{i=1}^{n}\left({r_{i,1}}^2 + {r_{i,2}}^2 + \cdots + {r_{i,m}}^2\right)$$

donde el residuo r_i es:

$$r_i = Y_i - X_i^T\beta = [r_{i,1} \quad r_{i,2} \quad r_{i,3} \quad \cdots \quad r_{i,m}]$$

Solución de minimizar $\boldsymbol{MSE(D;\beta)}$:

$$\beta = (X_t^T X_t)^{-1} X_t^T Y_t$$

Todos los modelos de regresión lineal suponen que las variables son numéricas reales.

En algunos casos, funcionan bien cuando las variables independientes son ordinales codificadas con números enteros.

Bibliografía

- James, G., Witten, D., Hastie, T. & Tibshirani, R. (2023). *An introduction to statistical learning: with applications in Python* (2da ed.). Springer.
    - Capítulo 3

- Hastie, T., Tibshirani, R., & Friedman, J. (2009). *The elements of statistical learning: data mining, inference, and prediction* (2da ed.). Springer.
    - Capítulo 3

- Russell, S. J. & Norvig, P. (2021). *Artificial intelligence: A modern approach (global edition)* (4ta ed.). Person.
    - Capítulo 19