

Regresión Lineal Multivariable

¿Qué pasa cuando queremos hacer una regresión por más de una variable?

Esto es, predecir un valor \hat{y} en base a valores x_1, \dots, x_k . Entonces ahora el modelo se ve:

$$\hat{y} = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_k x_k$$

Si lo escribimos en notación matricial:

$$\hat{y} = \beta \cdot x \quad \text{con} \quad \beta = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \\ \vdots \\ \beta_k \end{bmatrix} \quad \text{y} \quad x = \begin{bmatrix} 1 \\ x_1 \\ x_2 \\ \vdots \\ x_k \end{bmatrix}$$

Así, x es nuestro vector de features. Ahora supongamos un conjunto de n observaciones. Podemos escribir nuestras observaciones en términos de nuestro modelo:

$$\begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} = \begin{bmatrix} x_{11} & x_{12} & \dots & x_{1k} \\ x_{21} & x_{22} & \dots & x_{2k} \\ \vdots & \vdots & \ddots & \vdots \\ x_{n1} & x_{n2} & \dots & x_{nk} \end{bmatrix} \cdot \begin{bmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_k \end{bmatrix} + \begin{bmatrix} e_1 \\ e_2 \\ \vdots \\ e_n \end{bmatrix}$$

Esto es, para una observación i con $x_i = [x_{i1}, x_{i2}, \dots, x_{ik}]$ y con un valor $y = y_i$, sabemos que

$$y_i = \beta \cdot x_i + e_i$$

$$y_i = \hat{y}_i + e_i$$

$$y_i - \hat{y}_i = e_i$$

Nada nuevo hasta ahora: el error e_i es la diferencia entre el valor real y el estimado

Así, nuevamente queremos encontrar los valores β_0, \dots, β_k tal que minimicemos:

$$\sum_{i=1}^n e_i^2 = e^T e$$

Retomemos nuestro modelo

$$\begin{bmatrix} y_1 \\ y_i \\ \vdots \\ y_n \end{bmatrix} = \begin{bmatrix} x_{11} & x_{12} & \dots & x_{1k} \\ x_{21} & x_{22} & \dots & x_{2k} \\ \vdots & \vdots & \ddots & \vdots \\ x_{n1} & x_{n2} & \dots & x_{nk} \end{bmatrix} \times \begin{bmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_k \end{bmatrix} + \begin{bmatrix} e_1 \\ e_2 \\ \vdots \\ e_n \end{bmatrix}$$

↓ Matricialmente

$$Y = X \beta + e \longrightarrow e = Y - X \beta$$

$$\begin{aligned}
 \text{Así } \min e^T e &= (Y - X\beta)^T (Y - X\beta) \\
 &= (Y^T - \beta^T X^T)(Y - X\beta) \\
 &= Y^T Y - \beta^T X^T Y - Y^T X \beta + \beta^T X^T X \beta
 \end{aligned}$$

Notemos que $\beta^T X^T Y = (\beta^T X^T Y)^T = Y^T X \beta$
 porque ambos términos son escalares

$$\text{Así } \min e^T e = Y^T Y - 2\beta^T X^T Y + \beta^T X^T X \beta$$

Haremos lo mismo que para la regresión lineal simple, esto es sacar la derivada parcial respecto a β_i . Aquí usaremos:

$$\frac{\partial \beta^T X^T Y}{\partial \beta} = X^T Y \quad \left. \begin{array}{l} \text{Esta notación nos dice que} \\ \text{queremos un vector que en} \\ \text{la posición } i \text{ tiene el valor:} \end{array} \right\} \frac{\partial \beta^T X^T Y}{\partial \beta_i}$$

Esc vector es $X^T Y$

$$\frac{\partial \beta^T X^T X \beta}{\partial \beta} = 2X^T X \beta$$

$$\text{Así } \frac{\partial e^T e}{\partial \beta} = -2X^T Y + 2X^T X \beta = 0$$

$$\text{Y obtenemos } X^T X \beta = X^T Y$$

Finalmente tenemos:

$$\beta = (X^T X)^{-1} X^T y$$

Pero hay que tener ojo! Para encontrar el mínimo necesitamos que $X^T X$ sea positiva definida, por eso necesitamos que el rango de X sea completo!