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 $x_1, \ldots, x_k$. Entonces ahora el modelo se ve:

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

Si lo escribimos en notación matricial:

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

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

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

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 $\beta_0, \dots, \beta_k$ tal que minimicemos:

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

Retomemos nuestro modelo

$$\begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{bmatrix} = \begin{bmatrix} x_{11} & x_{12} & \dots & x_{1k} \\ x_{21} & x_{22} & \dots & x_{2k} \\ & & \vdots & \\ x_{n1} & x_{n2} & \dots & x_{nk} \end{bmatrix} \times \begin{bmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_k \end{bmatrix} + \begin{bmatrix} e_1 \\ e_2 \\ \vdots \\ e_n \end{bmatrix}$$

$$\Big\downarrow \text{ Matricialmente}$$

$$Y = \underline{X}\beta + e \longrightarrow e = Y - \underline{X}\beta$$

Así $\min e^Te = (Y - X\beta)^T(Y - X\beta)$

$$= (Y^T - \beta^T X^T)(Y - X\beta)$$

$$= Y^TY - \beta^TX^TY - Y^TX\beta + \beta^TX^TX\beta$$

Notemos que $\beta^TX^TY = (\beta^TX^TY)^T = Y^TX\beta$
Porque ambos términos son escalares

Así $\boxed{\min e^Te = Y^TY - 2\beta^TX^TY + \beta^TX^TX\beta}$

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

$$\frac{\delta \beta^TX^TY}{\delta \beta} = X^TY$$

Esta notación nos dice que queremos un vector que en la posición i tenga el valor:

$$\frac{\delta \beta^TX^TY}{\delta \beta_i}$$

Ese vector es X^TY

$$\frac{\delta \beta^TX^TX\beta}{\delta \beta} = 2X^TX\beta$$

Así $\dfrac{\delta e^Te}{\delta \beta} = -2X^TY + 2X^TX\beta = 0$

Y obtenemos $X^TX\beta = X^TY$

Finalmente tenemos:

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

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!