

EL METODO DE ORTOGONALIZACION DE GRAM-SCHMIDT Y LA REGRESION MULTIPLE

Jorge Ortiz P.

Profesor Asistente
Universidad Nacional

Abstract.

A geometric approach to the multiple linear regression allows the use of the Gram-Schmidt orthogonalization process to find the regression equation and sums of squares without inverting a matrix.

Resumen.

Un enfoque geométrico de la regresión lineal múltiple permite usar el proceso de ortogonalización de Gram-Schmidt para encontrar la ecuación de regresión y las sumas de cuadrados sin invertir matrices.

Introducción.

A menudo los estudiantes tienen dificultades cuando se encuentran frente a un problema de regresión lineal múltiple, simplemente porque para su solución deben calcular

la inversa de una matriz de tamaño relativamente grande; este cálculo es muy complejo y sin la ayuda de una calculadora de capacidad apreciable, el problema queda prácticamente fuera de sus manos.

Un enfoque geométrico sencillo del problema permite resolverlo sin necesidad de calcular inversas de matrices y una calculadora con algunas funciones estadísticas es suficiente para la solución numérica.

Problema.

Se tiene una variable Y a explicar con la ayuda de $k+1$ variables explicativas X_0, X_1, \dots, X_k , mediante el modelo lineal

$$Y = b_0 X_0 + b_1 X_1 + \dots + b_k X_k + e \quad (1)$$

donde b_i es el coeficiente de regresión asociado a la variable X_i y e es el error cometido al ajustar la observación de la variable Y al modelo; X_0 es generalmente igual a uno, lo cual hace que b_0 sea una constante independiente. La inclusión de X_0 permite decir que en el caso de error "e" igual a cero, Y es una combinación lineal de las variables X ; en este caso particular,

$$Y = b_0 X_0 + b_1 X_1 + \dots + b_k X_k \quad (2)$$

y se puede decir que Y es una variable explicada completamente por las variables X .

Si el error e es muy grande, el modelo dado no ex-

plica bien el comportamiento de la variable Y; así, cuando se tienen n observaciones (independientes) de las diferentes variables y se quiere describir de la mejor manera posible el comportamiento de la variable Y con la ayuda del modelo (1), lo que se trata de hacer es encontrar los coeficientes b_i que minimicen el error e en forma global para todas las observaciones; de esta manera se encontrará el mejor modelo lineal para explicar el comportamiento de la variable Y con la ayuda de las variables X_0, \dots, X_k .

Notación.

El modelo (1), para cada una de las n observaciones, se escribe así :

$$\begin{aligned} Y_1 &= b_0 X_{10} + b_1 X_{11} + \dots + b_k X_{1k} + e_1 \\ Y_2 &= b_0 X_{20} + b_1 X_{21} + \dots + b_k X_{2k} + e_2 \\ &\dots \dots \dots \\ Y_n &= b_0 X_{n0} + b_1 X_{n1} + \dots + b_k X_{nk} + e_n \end{aligned} \quad (3)$$

Si se llama X_j al vector de observaciones de la variable X_j , es decir :

$$X_j = \begin{bmatrix} X_{1j} \\ X_{2j} \\ \vdots \\ X_{nj} \end{bmatrix}$$

donde X_0 es un vector cuyas componentes son todas iguales

a uno. Si además X es la matriz cuyas columnas son X_0, X_1, \dots, X_k y Y el vector de observaciones de la variable Y , se puede escribir (3) de la siguiente manera :

$$Y = X b + e \quad (4)$$

donde,

$$b = \begin{pmatrix} b_0 \\ b_1 \\ \vdots \\ b_k \end{pmatrix} \quad e = \begin{pmatrix} e_1 \\ e_2 \\ \vdots \\ e_n \end{pmatrix}$$

Para minimizar $e = Y - Xb$ existen varios criterios; nosotros nos ocuparemos del más conocido: el de mínimos cuadrados; este criterio consiste en minimizar con respecto a b la expresión

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

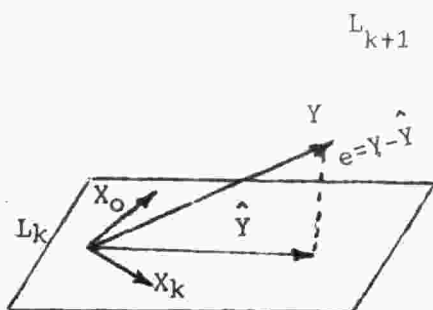
En cualquier libro que trate la regresión múltiple se encuentra que la solución para b está dada por

$$\hat{b} = (X'X)^{-1} X'Y$$

y aquí es clara la necesidad de invertir la matriz $(X'X)$ de dimensión $(k+1, k+1)$.

Interpretación geométrica del problema. Lo dicho en la introducción permite observar que lo que se pretende es en-

contrar la combinación lineal de X_0, X_1, \dots, X_k más próxima de Y ; esto equivale a buscar el vector \hat{Y} generado por X_0, X_1, \dots, X_k , es decir, $\hat{Y} = b_0 X_0 + b_1 X_1 + \dots + b_k X_k$ más próximo de Y . Esto se obtiene cuando Y es la proyección ortogonal de Y sobre el subespacio generado por X_0, X_1, \dots, X_k y así, $e = Y - \hat{Y}$ es ortogonal a ese subespacio. El problema se convierte entonces en buscar \hat{Y} tal que $e = Y - \hat{Y}$ sea ortogonal al subespacio generado por X_0, X_1, \dots, X_k . Esta idea es la que permite utilizar el proceso de ortogonalización de Gram-Schmidt para calcular los coeficientes de regresión lineal.



Método.

Sea L_k el subespacio generado por los vectores X_0, X_1, \dots, X_k y L_{k+1} el subespacio generado por los vectores X_0, X_1, \dots, X_k, Y . Sean U_0, U_1, \dots, U_r los vectores de la base de L_k ($r \leq k+1$) obtenidos por el proceso de Gram-Schmidt, se puede completar una base del subespacio L_{k+1} ; el proceso dice que el vector faltante

para la nueva base está dado por :

$$U_{r+1} = Y - \frac{Y'U_0}{U_0'U_0} U_0 - \frac{Y'U_1}{U_1'U_1} U_1 - \dots - \frac{Y'U_r}{U_r'U_r} U_r \quad (5)$$

y U_{r+1} es ortogonal a todos los vectores de la base de L_k . Además,

$$\hat{Y} = \frac{Y'U_0}{U_0'U_0} U_0 + \frac{Y'U_1}{U_1'U_1} U_1 + \dots + \frac{Y'U_r}{U_r'U_r} U_r \quad (6)$$

\hat{Y} es la proyección de Y sobre el subespacio L_k generado por los vectores X_0, X_1, \dots, X_k . De manera que el error del modelo es precisamente U_{r+1} es decir,

$$e = Y - \hat{Y} = U_{r+1}$$

y el cuadrado de su norma es lo que en regresión se llama la suma residual de cuadrados.

Si U_{r+1} es igual a cero, entonces L_k y L_{k+1} tienen la misma base y Y es explicada completamente por las variables X . La expresión (6) da una descomposición del vector \hat{Y} en términos de una base ortogonal de L_k . Sin embargo, lo que mas interesa es expresar \hat{Y} en términos de los vectores observados X_0, X_1, \dots, X_k . Para ello se hace una descomposición de la base ortogonal en términos de los vectores X .

Sea $P_{A/B}$ la componente de la proyección del vector A sobre el vector B ; es decir,

$$P_{A/B} = \frac{A'B}{B'B}$$

Además, si el vector A se expresa como una combinación lineal de los vectores B_0, B_1, \dots, B_k , entonces $C_{B_i/A}$ representa el coeficiente de B_i en el desarrollo dado para A .

Con estas notaciones el proceso de ortogonalización de Gram-Schmidt se puede escribir de la siguiente manera :

$$U_0 = X_0$$

$$U_1 = X_1 - P_{X_1/U_0} X_0$$

$$\begin{aligned} U_2 &= X_2 - P_{X_2/U_0} U_0 - P_{X_2/U_1} U_1 \\ &= X_2 - P_{X_2/U_1} X_1 - (P_{X_2/U_0} + P_{X_2/U_1} C_{X_0/U_1}) X_0 \end{aligned}$$

$$\begin{aligned} U_3 &= X_3 - P_{X_3/U_0} U_0 - P_{X_3/U_1} U_1 - P_{X_3/U_2} U_2 \\ &= X_3 - P_{X_3/U_2} X_2 - (P_{X_3/U_1} + P_{X_3/U_2} C_{X_1/U_2}) X_1 - \\ &\quad - (P_{X_3/U_0} + P_{X_3/U_1} C_{X_0/U_1} + P_{X_3/U_2} C_{X_0/U_2}) X_0 \end{aligned}$$

Si se tiene en cuenta que : $C_{X_j/U_j} = 1$ para cualquier valor de $j \leq k+1$ se puede demostrar por inducción sobre j que

$$- C_{X_i/U_j} = \sum_{t=i}^{j-1} P_{X_j/U_t} C_{X_i/U_t} \quad \text{para } i \leq j-1 \quad (7)$$

Los coeficientes de regresión de Y en X_0, \dots, X_k están dados entonces por :

$$-C_{X_0/U_{k+1}}, -C_{X_1/U_{k+1}}, \dots, -C_{X_k/U_{k+1}}.$$

Ejemplo práctico.

Los datos de trabajo son los del ejercicio D. del capítulo cuarto del libro de Draper y Smith (ver bibliografía).

X_0	X_1	X_2	Y
1	38	47.5	66
1	41	21.3	43
1	34	36.5	36
1	35	18.0	23
1	31	29.5	22
1	34	14.2	14
1	29	21.0	12
1	32	10.0	7.6

$$1. U_0 = X_0$$

$$P_{X_1/U_0} = 274/8 = 34.25$$

$$2. U_1 = X_1 - P_{X_1/U_0} U_0$$

$$= X_1 - 34.25 X_0$$

$$\begin{aligned}
 3. \quad U_2 &= X_2 - P_{X_2/U_1} X_1 - (P_{X_2/U_0} + P_{X_2/U_1} C_{X_0/U_1}) X_0 \\
 &= X_2 - (94.1/103.5) X_1 - (24.75 - (94.1/103.5)(34.25)) X_0 \\
 &= X_2 - 0.909178744 X_1 - (-6.38937198) X_0
 \end{aligned}$$

$$\begin{aligned}
 4. \quad U_3 &= X_3 - P_{X_3/U_2} X_2 - (P_{X_3/U_1} + P_{X_3/U_2} C_{X_1/U_2}) X_1 \\
 &\quad - (P_{X_3/U_0} + P_{X_3/U_1} C_{X_0/U_1} + P_{X_3/U_2} C_{X_0/U_2}) X_0 \\
 &= X_3 - 1.072682619 X_2 - \{3.776811594 - (1.072682619) \\
 &\quad (9091787)\} X_1 - \{29.75 - (3.776811594)(34.25) - (1.072682619) \\
 &\quad (-6.38937198)\} X_0 \\
 &= X_3 - 1.072682619 X_2 - 2.801551358 X_1 - (-94.5520288) X_0
 \end{aligned}$$

se ha identificado Y con X_3 . U_3 es el error (o vector de residuos) y

$$\hat{Y} = 1.072682619 X_2 + 2.801551358 X_1 - 94.5520288$$

Este es el modelo de regresión lineal buscado.

La norma de U_3 elevada al cuadrado da la suma residual de cuadrados. La suma total se obtiene elevando al cuadrado la norma de $Y - P_{Y/U_0} U_0$. Y la suma de cuadrados de la regresión se obtiene por sustracción. De esta manera se puede formar la tabla de análisis de varianza y calcular el valor de R^2 .

BIBLIOGRAFIA

Draper, N.; Smith, H. Applied Regression Analysis, John Wiley and Sons, New York, 1981.

Scheffé, H; The Analysis of Variance, John Wiley and Sons, New York, 1959.