Análisis Numérico

Tarea 2: Mínimos cuadrados y aplicaciones

Teoría de los mínimos cuadrados

El método de Mínimos Cuadrados, introducido por primera vez en el siglo XIX, se atribuye a los matemáticos Adrien-Marie Legendre y Carl Friedrich Gauss quienes lo desarrollaron en contextos distintos: uno en problemas de ajuste de datos y el otro en predicción astronómica.

Esta es una técnica fundamental en matemáticas y estadística que se utiliza para encontrar la función que mejor se aproxime a un conjunto de datos observados. Su fundamento matemático se basa en minimizar la suma de los cuadrados de las diferencias entre los valores observados ($\mathbf{y}$) y los estimados ($\hat{\mathbf{y}} = A\hat{\beta}$), conocida como la norma cuadrática del residuo $\mathbf{r} = \mathbf{y} - A\hat{\beta}$. En términos de álgebra lineal, se trata de aproximar soluciones a sistemas lineales $\mathbf{y} = A\beta$ que son sobredeterminados, es decir, con más ecuaciones que incógnitas. Y es así como el método encuentra la proyección ortogonal de $\mathbf{y}$ sobre el subespacio columna de A, $Col(A)$, lo cual garantiza que el modelo ajustado minimice la suma de los cuadrados de las componentes del residuo $\mathbf{r}$.

A lo largo de los años, por su simplicidad y efectividad, el método de los mínimos cuadrados es sumamente importante en diversas áreas de la ciencia y la ingeniería.

Perspectiva geométrica

El método de los mínimos cuadrados se basa en el concepto de proyección ortogonal en espacios vectoriales donde se minimiza la distancia perpendicular entre los puntos de datos y el modelo ajustado.

En álgebra lineal, se conoce la proyección ortogonal de un vector sobre un subespacio como el vector más cercano al original que pertenece dicho subespacio. Si un vector $\mathbf{y}$ está en un espacio vectorial $\mathbb{R}^n$ y se desea proyectarlo sobre un subespacio S generado por algún conjunto de vectores, entonces el resultado de esta proyección es un vector $\hat{\mathbf{y}} \in S$ tal que el residuo $\mathbf{r} = \mathbf{r} - \hat{\mathbf{y}}$ es ortogonal a S. Esto asegurará que la proyección minimiza la distancia $||\mathbf{r}||$; en otras palabras, el residuo r tiene la menor magnitud posible. Geométricamente, esto es encontrar el punto más cercano en el subespacio al vector original.

En el caso de la regresión lineal se desea ajustar una línea de la forma $y = \beta_0 + \beta_1 x$ y para ello se puede expresar el problema en términos de álgebra lineal tal y como se explicó anteriormente:

Dado un conjunto de n puntos (x_i, y_i) donde los valores observados y_i se agrupan en un vector $\mathbf{y}$ y la matriz A tiene una columna de unos y una columna con los valores x_i.

El objetivo es encontrar el vector $\hat{\beta} = [\beta_0, \beta_0]^\top$ que minimice la norma del residuo:

$$\mathbf{r} = \mathbf{y} - A\hat{\beta}$$

Esto se traduce en proyectar el vector $\mathbf{y}$ sobre el subespacio generado por las columnas de A. Y así, el resultado $\hat{\mathbf{y}}$ es la proyección ortogonal de $\mathbf{y}$ y el residuo $\mathbf{r}$ es el componente ortogonal a dicho subespacio. En otras palabras, la proyección es ortogonal porque los valores predichos y los valores reales no están correlacionados. Esto se ilustra en la Figura 1, que representa el caso de dos variables independientes (vectores $\mathbf{x_1}$ y $\mathbf{x_2}$) y el vector de datos ($\mathbf{y}$), y muestra que el vector de error ($\mathbf{y_1} - \hat{\mathbf{y}}$) es ortogonal a la estimación del mínimo cuadrado ($\hat{\mathbf{y}}$) que se encuentra en el subespacio definido por las dos variables independientes.

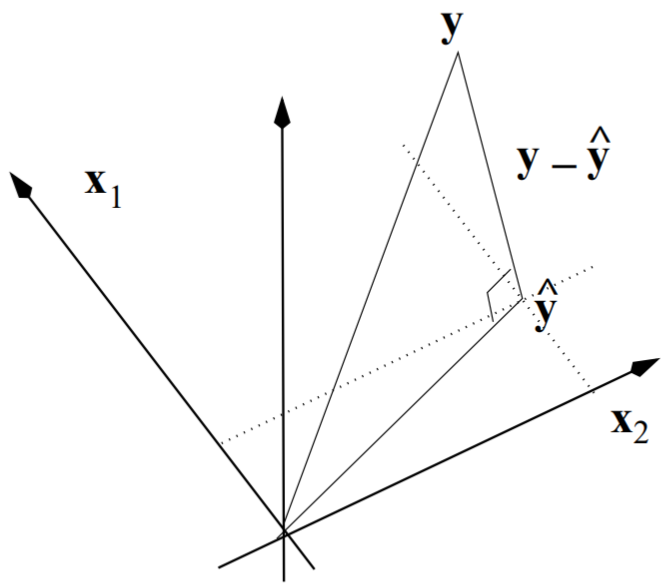

Figura 1: *La estimación de los mínimos cuadrados de los datos es la proyección ortogonal del vector de datos sobre el subespacio de la variable independiente.*

En el caso polinómico, el objetivo es ajustar un polinomio de grado n a un conjunto de datos (x_i, y_i). El polinomio tiene la forma:

$$P(x) = a_0 + a_1 x + a_2 x^2 + \cdots + a_n x^n$$

Cada función $1, x, x^2, \ldots, x^n$ genera un vector en un espacio de dimensión igual al número de observaciones m. Al evaluar dichas funciones en los puntos x_i se obtiene una matriz A cuya columna j-ésima corresponde a la función x^{j-1}. El problema de los mínimos cuadrados consistiría en este caso en encontrar los coeficientes $a = [a_0, a_1, \ldots, a_n]^\top$ que minimicen la distancia entre $\mathbf{y}$ y el subespacio generado por las columnas de A. Análogamente al caso lineal, la solución es una proyección ortogonal y el residuo $\mathbf{r} = \mathbf{y} - \hat{\mathbf{y}}$ es ortogonal al subespacio.

En esencia, los mínimos cuadrados pretende minimizar la distancia perpendicular entre un vector de datos $\mathbf{y}$ y el modelo ajustado $\hat{\mathbf{y}}$. Matemáticamente, implica resolver:

$$\min_{\beta} \|\mathbf{y} - A\hat{\beta}\|^2.$$

donde el término $||\mathbf{y} - A\hat{\beta}\|^2$ representa la suma de los cuadrados de los residuos que son las distancias perpendiculares desde los puntos observados al modelo. La solución se encontraría al resolver las ecuaciones normales $A^\top A\hat{\beta} = A^\top \mathbf{y}$ lo que garantiza que $\hat{\mathbf{y}}$ es la proyección ortogonal de $\mathbf{y}$, cuya derivación y justificación se hará más adelante en la sección **Perspectiva algebraica.**

A continuación, en la Figura 2, se muestra concretamente cómo funciona visualmente el método.

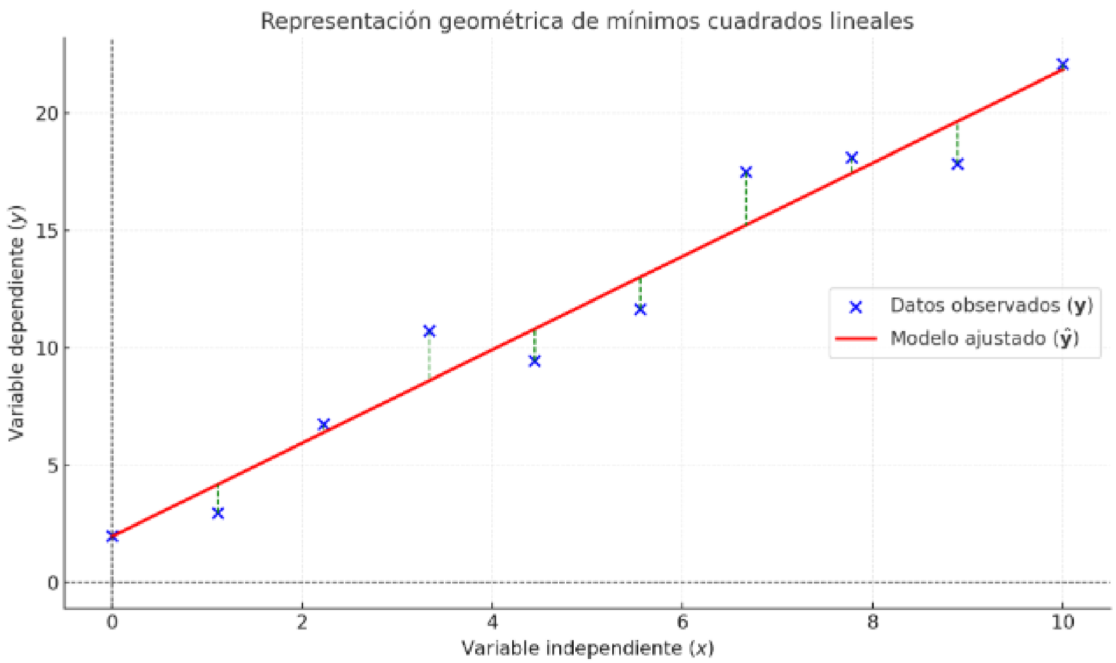

Figura 2: Mínimos cuadrados.

- **Puntos azules.** Representan los datos observados ($\mathbf{y}$)
- **Línea roja.** Modelo ajustado ($\hat{\mathbf{y}}$) que es la proyección ortogonal de los datos sobre el subespacio definido por las columnas de la matriz A.
- **Líneas verdes.** Representan los residuos $\mathbf{r} = \mathbf{y} - \hat{\mathbf{y}}$, es decir, las diferencias entre los valores observados y los ajustados.

Perspectiva algebraica

Deducción general de las ecuaciones normales

Primeramente, veamos la deducción de las ecuaciones normales para cualquier sistema lineal de ecuaciones sobredeterminado donde el modelo ajustado tiene la forma $\mathbf{y} = A\beta$. Como vimos en la sección anterior, en los mínimos cuadrados se pretende resolver:

$$\min_{\beta} \|\mathbf{y} - A\hat{\beta}\|^2.$$

donde:

- $\mathbf{y}$ es el vector de datos observados.
- A es la matriz de diseño, $A \in \mathbb{R}^{mxn}$ cuyas funciones base sería, por ejemplo, $1, x, x^2, \ldots,$ etc.
- $A\hat{\beta}$ es el modelo ajustado que representa la proyección de $\mathbf{y}$ sobre $Col(A)$ y $\hat{\beta} \in \mathbb{R}^n$ es un vector de parámetros a determinar.
- $\mathbf{r} = \mathbf{y} - A\hat{\beta}$ es el residuo que se quiere minimizar.

Si se expande $\|\mathbf{y} - A\hat{\beta}\|^2$ resulta:

$$(\mathbf{y} - A\hat{\beta})^\top(\mathbf{y} - A\hat{\beta}) = \mathbf{r}^\top\mathbf{r}$$

Y al desarrollar resulta:

$$\mathbf{y}^\top\mathbf{y} - 2\hat{\beta}^\top A^\top\mathbf{y} + \hat{\beta}^\top A^\top\hat{\beta}$$

Para encontrar el mínimo, derivamos con respecto a $\hat{\beta}$ y se iguala a cero:

$$\frac{\partial}{\partial\hat{\beta}}[\mathbf{y}^\top\mathbf{y} - 2\hat{\beta}^\top A^\top\mathbf{y} + \hat{\beta}^\top A^\top\hat{\beta}] = 0$$

$$\begin{aligned} -2A^\top\mathbf{y} + 2A^\top A\hat{\beta} &= 0 \\ 2A^\top A\hat{\beta} &= 2A^\top\mathbf{y} \\ \therefore A^\top A\hat{\beta} &= A^\top\mathbf{y} \end{aligned}$$

Estas son las llamadas **ecuaciones normales** donde su solución (o soluciones) $\hat{\beta}$ proporciona los coeficientes del modelo ajustado.

La deducción anterior asegura que $\hat{\mathbf{y}} = A\hat{\beta}$ es la proyección ortogonal de $\mathbf{y}$ sobre el subespacio columna de A, $Col(A)$.

$$\mathbf{r} = \mathbf{y} - A\hat{\beta} \text{ es ortogonal a } Col(A)$$

Algebraicamente, esto significa que: $A^\top\mathbf{r} = A^\top(\mathbf{y} - A\hat{\beta}) = 0$.

Note que hay una conexión evidente con la **perspectiva geométrica** pues el vector residuo $\mathbf{r}$ es perpendicular al subespacio columna de A y la solución $\hat{\beta}$ minimiza la distancia entre $\mathbf{y}$ y $A\hat{\beta}$.

Perspectiva desde la ciencia de datos

Referencias

[1] H. Abdi, *Least Squares*, Encyclopedia for Research Methods for the Social Sciences, pp. 792–795, 2003.

[2] R.L. Burden and J.D. Faires, *Numerical Analysis*, 9th Edition, Brookscole, Boston, 2011.

[3] Strang, G. (2016). *Introduction to Linear Algebra* (5th ed.). Wellesley-Cambridge Press. Capítulos relacionados con proyección ortogonal y mínimos cuadrados.