Ajuste de Datos Mediante Regresión Lineal

El modelo de regresión lineal simple sólo está conformado por dos variables estadísticas llamadas x y y. Considera una única variable independiente y una variable dependiente, asumiendo que la relación entre ambas es lineal. Para la regresión lineal, se asume que x y y se relacionan mediante la relación funcional (siendo β1 y β0 estimadores).

$$y = \beta_0 + \beta_1 X + \varepsilon$$

donde $\beta_0, \beta_1 \in \mathbb{R}$ son constantes desconocidas llamadas coeficientes de regresión. β_1: Se trata del cociente entre la interacción obtenida entre ambas variables y la suma de cuadrados de los valores de la variable dependiente. Este valor corresponde a la pendiente de la recta. Por su parte, β_0 es el resultado de la siguiente ecuación en la que aparecen los valores medios correspondientes a ambas variables y el estimador β_1 obtenido anteriormente. Este valor es la ordenada en el origen.

Dado que los parámetros β_1 y β_0, son constantes desconocidas, estas deben estimarse mediante los datos de la muestra, supóngase que se tienen n datos $\{x_1, x_2\}_{i=1}^{n}$, se estimaran los parámetros utilizando el método de mínimos cuadrados.

Se estiman β_1 y β_0, tal que la suma de los cuadrados de las diferencias entre las observaciones yi y la recta de regresión sea mínima, esto es, buscamos minimizar la función error cuadrático dado por:

$$S(\beta_0, \beta_1) = \sum_{i=1}^{n} (y_1 - \beta_0 - \beta_1 x_i)^2$$

La función de error cuadrático $S(\beta_0, \beta_1)$ alcanza un mínimo en el punto $(\hat{\beta}_0, \widehat{\beta}_1)$ tal que

$$\left.\frac{\partial S}{\partial \beta_0}\right|_{\widehat{\beta}_0, \widehat{\beta}_1} = \left.\frac{\partial S}{\partial \beta_1}\right|_{\widehat{\beta}_0, \widehat{\beta}_1} = 0$$

Entonces derivando respecto a β1 y β0, evaluando en $\widehat{B0}$ y $\widehat{B1}$ e igualando a cero, obtenemos que el siguiente sistema de ecuaciones

$$\begin{cases} -2 \sum_{i=1}^{n} (y_i - \hat{\beta}_0 - \widehat{\beta}_1 x_i) = 0 \\ -2 \sum_{i=1}^{n} (y_i - \hat{\beta}_0 - \widehat{\beta}_1 x_i) x_i = 0 \end{cases}$$

Estas dos ecuaciones son conocidas como ecuaciones normales, la solución de dicho sistema de ecuaciones está dada por:

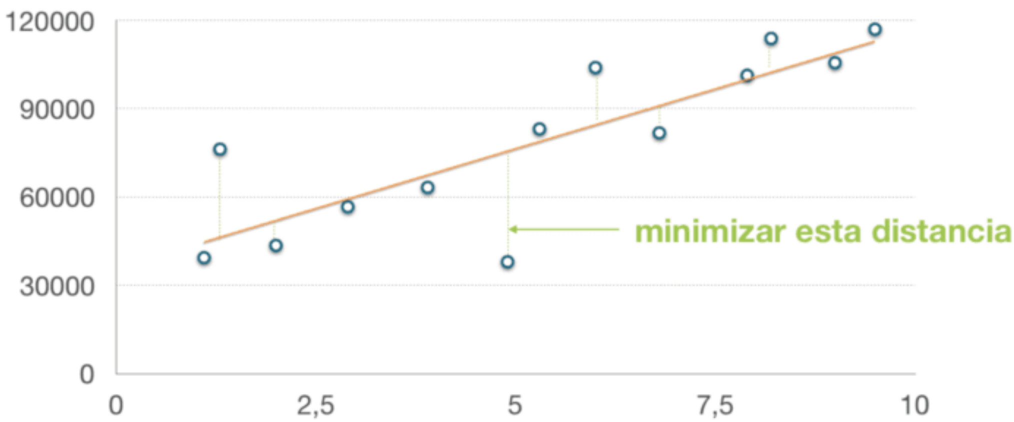
$$\widehat{\beta}_1 = \frac{\sum_{i=1}^{n} x_i y_i - n\bar{x}\bar{y}}{\sum_{i=1}^{n} x_i^2 - \frac{1}{n}(\sum_{i=1}^{n} x_i)^2}$$

$$\widehat{\beta}_0 = \frac{\sum_{i=1}^{n} x_i^2 \sum_{i=1}^{n} y_i - \sum_{i=1}^{n} x_i y_i \sum_{i=1}^{n} x_1}{n \sum_{i=1}^{n} x_i^2 - (\sum_{i=1}^{n} x_i)^2} = \bar{y} - \widehat{\beta}_1 \bar{x}$$

La interpretación del parámetro medio β1 es un incremento en *xi* de una unidad, y_i incrementará en β_1. Luego el modelo ajustado de regresión lineal es:

$$\hat{y} = \hat{B}_0 + \hat{B}_1 x$$

El objetivo con Regresión Lineal Simple es minimizar la distancia vertical entre todos los datos y nuestra línea, por lo tanto, para determinar la mejor línea, debemos minimizar la distancia entre todos los puntos y la distancia de nuestra línea.



Fuente: Aprende IA (s.f.).

Referencias:

Aprende IA. (s.f.). Regresión Lineal – Teoría. Recuperado a partir de: [*https://aprendeia.com/algoritmo-regresion-lineal-simple-machine-learning/*](https://aprendeia.com/algoritmo-regresion-lineal-simple-machine-learning/)