Bootcamp de Machine Learning
Regresión Lineal

Cynthia Elizabeth Castillo Silva @LaMatemaga

Estimadores de parámetros β_0 y β_1

El error cuadrático medio (MSE) está dado por

$$MSE = \frac{1}{n}\sum_{i=1}^{n} e_i^2$$

donde $e_i = Y_i - \hat{Y}_i$, con el estimador de la variable dependiente $\hat{Y}_i = \beta_0 + \beta_1 X_i$. Para poder obtener mejores resultados en nuestro modelo un método recomendado es minimizar el valor de MSE.

Objetivo: Minimizar la siguiente expresión:

$$Q = \sum_{i=1}^{n} e_i^2.$$

Recordemos que el error cuadrático medio está dado por $MSE = \frac{1}{n}Q$.

$$Q = \sum_{i=1}^{n} e_i^2$$
$$Q = \sum_{i=1}^{n} \left(Y_i - \hat{Y}_i\right)^2$$
$$Q = \sum_{i=1}^{n} \left(Y_i - (\beta_0 + \beta_1 X_i)\right)^2$$
$$Q = \sum_{i=1}^{n} \left(Y_i - \beta_0 - \beta_1 X_i\right)^2$$

Observemos que los términos de Q dentro de la sumatoria están elevados al cuadrado, por lo que serán siempre positivos o iguales a cero. Necesitaremos encontrar los valores de β_0 y a β_1 que **minimicen** nuestra función. Para esto necesitaremos derivar respecto a β_0 y a β_1 e igualar a cero.

Para los siguientes pasos necesitaremos conocer las reglas de derivación que usualmente se llevarían dentro de cualquier curso de Cálculo Multivariado (no te preocupes, hay cosas que parecen bastante análogas entre el Univariado y el Multivariado). Hay en este link un par de tablas de referencia rápida. En este documento mencionaremos qué reglas se utilizan en qué pasos.

Para β_0:

$$\frac{\partial Q}{\partial \beta_0} = \sum_{i=1}^{n} (-1)2\left(Y_i - \beta_0 - \beta_1 X_i\right)$$
$$\frac{\partial Q}{\partial \beta_0} = \sum_{i=1}^{n} -2\left(Y_i - \beta_0 - \beta_1 X_i\right)$$

Podemos sacar el valor de la constante -2 afuera de la suma por propiedades de la suma. Entonces nos quedaría

$$\frac{\partial Q}{\partial \beta_0} = -2\sum_{i=1}^{n}(Y_i - \beta_0 - \beta_1 X_i)$$

Para llegar a la expresión anterior utilizamos únicamente la **regla de composición** (o **regla de la cadena**) debido a que la expresión tenía la forma de una función compuesta $g(f(x))$.

Ahora igualamos esta derivada a cero para obtener el valor mínimo para β_0.

$$\begin{aligned}
\frac{\partial Q}{\partial \beta_0} &= 0 \\
-2\sum_{i=1}^{n}(Y_i - \beta_0 - \beta_1 X_i) &= 0 \\
\sum_{i=1}^{n}(Y_i - \beta_0 - \beta_1 X_i) &= 0 \\
\sum_{i=1}^{n} Y_i - \sum_{i=1}^{n}\beta_0 - \sum_{i=1}^{n}\beta_1 X_i &= 0
\end{aligned}$$

Todo lo que no tenga subíndice i podemos sacarlo de la suma, por lo que lo anterior nos quedaría de la siguiente forma:

$$\sum_{i=1}^{n} Y_i - \beta_0\sum_{i=1}^{n} 1 - \beta_1\sum_{i=1}^{n} X_i = 0$$

Recordemos que

$$\sum_{i=1}^{n} 1 = n$$

y que

$$\sum_{i=1}^{n} X_i = n\bar{X} \qquad \text{y} \qquad \sum_{i=1}^{n} Y_i = n\bar{Y},$$

por lo que podemos reescribir la igualdad a cero como

$$\begin{aligned}
n\bar{Y} - n\beta_0 - n\beta_1\bar{X} - &= 0 \\
n\left(\bar{Y} - \beta_0 - \beta_1\bar{X}\right) &= 0 \\
\bar{Y} - \beta_0 - \beta_1\bar{X} &= 0 \\
\bar{Y} - \beta_1\bar{X} &= \beta_0 \\
\beta_0 &= \bar{Y} - \beta_1\bar{X}.
\end{aligned}$$

Para β_1:

$$Q = \sum_{i=1}^{n} (Y_i - \beta_0 - \beta_1 X_i)^2$$
$$\frac{\partial Q}{\partial \beta_1} = (-1)2\sum_{i=1}^{n} X_i\,(Y_i - \beta_0 - \beta_1 X_i)$$
$$\frac{\partial Q}{\partial \beta_1} = -2\sum_{i=1}^{n} X_i\,(\beta_0 + \beta_1 X_i - Y_i)$$

Para llegar a la expresión anterior utilizamos únicamente la **regla de composición** (o **regla de la cadena**) debido a que la expresión tenía la forma de una función compuesta $g(f(x))$. Nos queda distribuir X_i sobre el resto de la expresión dentro de la suma, quedando de la siguiente forma

$$\frac{\partial Q}{\partial \beta_1} = -2\sum_{i=1}^{n} \left(\beta_0 X_i + \beta_1 X_i^2 - X_i Y_i\right)$$

Ahora igualamos esta derivada a cero para obtener el valor mínimo para β_0.

$$\frac{\partial Q}{\partial \beta_1} = 0$$
$$-2\sum_{i=1}^{n} \left(\beta_0 X_i + \beta_1 X_i^2 - X_i Y_i\right) = 0$$
$$\sum_{i=1}^{n} \left(\beta_0 X_i + \beta_1 X_i^2 - X_i Y_i\right) = 0$$
$$\sum_{i=1}^{n} \beta_0 X_i + \sum_{i=1}^{n} \beta_1 X_i^2 - \sum_{i=1}^{n} X_i Y_i = 0$$

Todo lo que no tenga subíndice i podemos sacarlo de la suma, por lo que lo anterior nos quedaría de la siguiente forma:

$$\beta_0 \sum_{i=1}^{n} X_i + \beta_1 \sum_{i=1}^{n} X_i^2 - \sum_{i=1}^{n} X_i Y_i = 0$$

Recordemos que

$$\sum_{i=1}^{n} 1 = n$$

y que

$$\sum_{i=1}^{n} X_i = n\bar{X},$$

por lo que podemos reescribir la igualdad a cero como

$$n\beta_0\bar{X} + \beta_1 \sum_{i=1}^{n} X_i^2 - \sum_{i=1}^{n} X_i Y_i = 0.$$

De lo que obtuvimos anteriormente, sabemos que $\beta_0 = \bar{Y} - \beta_1 \bar{X}$, por lo que podemos reescribir la expresión anterior como

$$\begin{aligned}
n\left(\bar{Y} - \beta_1 \bar{X}\right)\bar{X} + \beta_1 \sum_{i=1}^{n} X_i^2 - \sum_{i=1}^{n} X_i Y_i &= 0 \\
n\bar{Y}\bar{X} - n\beta_1 \bar{X}\bar{X} + \beta_1 \sum_{i=1}^{n} X_i^2 - \sum_{i=1}^{n} X_i Y_i &= 0 \\
n\bar{Y}\bar{X} + \beta_1 \left(\sum_{i=1}^{n} X_i^2 - n\bar{X}\bar{X}\right) - \sum_{i=1}^{n} X_i Y_i &= 0 \\
\beta_1 \left(\sum_{i=1}^{n} X_i^2 - n\bar{X}\bar{X}\right) &= \sum_{i=1}^{n} X_i Y_i - n\bar{Y}\bar{X}
\end{aligned}$$

Tenemos presente que

$$\sum_{i=1}^{n} X_i = n\bar{X} \qquad \text{y} \qquad \sum_{i=1}^{n} Y_i = n\bar{Y},$$

que equivale a decir que

$$\frac{1}{n}\sum_{i=1}^{n} X_i = \bar{X} \qquad \text{y} \qquad \frac{1}{n}\sum_{i=1}^{n} Y_i = \bar{Y},$$

y, reemplazando lo anterior en la expresión de la que queremos obtener β_1, obtenemos

$$\begin{aligned}
\beta_1 \left(\sum_{i=1}^{n} X_i^2 - n\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right)\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right)\right) &= \sum_{i=1}^{n} X_i Y_i - n\bar{Y}\bar{X} \\
\beta_1 \left(\sum_{i=1}^{n} X_i^2 - \frac{1}{n}\sum_{i=1}^{n} X_i \sum_{i=1}^{n} X_i\right) &= \sum_{i=1}^{n} X_i Y_i - n\bar{Y}\bar{X} \\
\beta_1 \left(\sum_{i=1}^{n} X_i^2 - \frac{1}{n}\left(\sum_{i=1}^{n} X_i\right)^2\right) &= \sum_{i=1}^{n} X_i Y_i - n\bar{Y}\bar{X} \\
\beta_1 &= \frac{\displaystyle\sum_{i=1}^{n} X_i Y_i - n\bar{Y}\bar{X}}{\displaystyle\sum_{i=1}^{n} X_i^2 - \frac{1}{n}\left(\sum_{i=1}^{n} X_i\right)^2}
\end{aligned}$$

Rescatando los resultados anteriores, tenemos que

$$\beta_0 = \bar{Y} - \beta_1 \bar{X} \qquad \text{y} \qquad \beta_1 = \frac{\displaystyle\sum_{i=1}^{n} X_i Y_i - n\bar{Y}\bar{X}}{\displaystyle\sum_{i=1}^{n} X_i^2 - \frac{1}{n}\left(\sum_{i=1}^{n} X_i\right)^2}$$

■