UNIVERSITAT DE BARCELONA

Universitat Oberta de Catalunya

Introducción al modelo lineal
Soluciones a los ejercicios opcionales

Francesc Carmona

19 de septiembre de 2020

Ejercicios del libro de Faraway

Los ejercicios 3, 4 y 5 del capítulo 1 son similares a los ejercicios 1 y 2.

Ejercicios del libro de Carmona

Ejercicio 1.1

Hallar las estimaciones de los parámetros en un modelo de regresión lineal simple, minimizando la suma de los cuadrados de los errores:

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

La función S depende de dos incógnitas (β_0, β_1), mientras que las (x_i, y_i) son las observaciones que se consideran conocidas. Entonces para hallar el mínimo de esa función habrá que derivar e igualar a cero:

$$\begin{aligned}
\frac{\partial S}{\partial \beta_0} &= 2\sum_{i=1}^{n}(y_i - \beta_0 - \beta_1 x_i)(-1) \\
\frac{\partial S}{\partial \beta_1} &= 2\sum_{i=1}^{n}(y_i - \beta_0 - \beta_1 x_i)(-x_i)
\end{aligned}$$

y al igualar a cero obtenemos

$$\begin{aligned}
-\sum_{i=1}^{n} y_i + n\beta_0 + \beta_1 \sum_{i=1}^{n} x_i &= 0 \\
-\sum_{i=1}^{n} x_i y_i + \beta_0 \sum_{i=1}^{n} x_i + \beta_1 \sum_{i=1}^{n} x_i^2 &= 0
\end{aligned}$$

que son las llamadas *ecuaciones normales*:

$$\begin{pmatrix} n & \sum_{i=1}^n x_i \\ \sum_{i=1}^n x_i & \sum_{i=1}^n x_i^2 \end{pmatrix} \begin{pmatrix} \beta_0 \\ \beta_1 \end{pmatrix} = \begin{pmatrix} \sum_{i=1}^n y_i \\ \sum_{i=1}^n x_i y_i \end{pmatrix}$$

Para resolver este sistema, dividimos la primera ecuación por n:

$$\beta_0 + \bar{x}\beta_1 = \bar{y} \qquad \Rightarrow \qquad \beta_0 = \bar{y} - \bar{x}\beta_1$$

y sustituimos en la segunda

$$(\bar{y} - \bar{x}\beta_1)\sum_{i=1}^{n} x_i + \beta_1 \sum_{i=1}^{n} x_i^2 = \sum_{i=1}^{n} x_i y_i$$

de donde

$$\beta_1\left(\sum_{i=1}^n x_i^2 - \bar{x}\sum_{i=1}^n x_i\right) = \sum_{i=1}^n x_i y_i - \bar{y}\sum_{i=1}^n x_i \tag{1}$$

Por otra parte

$$\begin{aligned}
\sum_{i=1}^n (x_i-\bar{x})(y_i-\bar{y}) &= \sum_{i=1}^n (x_i y_i - x_i\bar{y} - \bar{x}y_i + \bar{x}\bar{y}) \\
&= \sum_{i=1}^n x_i y_i - \bar{y}\sum_{i=1}^n x_i - \bar{x}\sum_{i=1}^n y_i + n\bar{x}\bar{y} \\
&= \sum_{i=1}^n x_i y_i - \bar{y}\sum_{i=1}^n x_i - n\bar{x}\bar{y} + n\bar{x}\bar{y} \\
&= \sum_{i=1}^n x_i y_i - \bar{y}\sum_{i=1}^n x_i \qquad (2)
\end{aligned}$$

y también

$$\begin{aligned}
\sum_{i=1}^n (x_i-\bar{x})^2 &= \sum_{i=1}^n (x_i^2 - 2x_i\bar{x} + \bar{x}^2) \\
&= \sum_{i=1}^n x_i^2 - 2\bar{x}\sum_{i=1}^n x_i + n\bar{x}^2 \\
&= \sum_{i=1}^n x_i^2 - 2n\bar{x}^2 + n\bar{x}^2 \\
&= \sum_{i=1}^n x_i^2 - n\bar{x}^2 = \sum_{i=1}^n x_i^2 - \bar{x}\sum_{i=1}^n x_i \qquad (3)
\end{aligned}$$

De modo que podemos sustituir (2) y (3) en (1) y tenemos

$$\beta_1\sum_{i=1}^n (x_i-\bar{x})^2 = \sum_{i=1}^n (x_i-\bar{x})(y_i-\bar{y})$$

es decir

$$\hat{\beta}_1 = \frac{\sum_{i=1}^n (x_i-\bar{x})(y_i-\bar{y})}{\sum_{i=1}^n (x_i-\bar{x})^2} = \frac{S_{xy}}{S_{xx}} = \frac{s_{xy}}{s_x^2}$$

donde $s_x^2 = \frac{1}{n}\sum_{i=1}^n (x_i-\bar{x})^2$ y $s_{xy} = \frac{1}{n}\sum_{i=1}^n (x_i-\bar{x})(y_i-\bar{y})$.

Además $\hat{\beta}_0 = \bar{y} - \bar{x}\hat{\beta}_1$ y el sistema queda solucionado.

Hallar una expresión para las predicciones $\hat{y}_i$ y los residuos $e_i = y_i - \hat{y}_i$.

Las predicciones son

$$\hat{y}_i = \hat{\beta}_0 + \hat{\beta}_1 x_i = \bar{y} - \bar{x}\hat{\beta}_1 + \hat{\beta}_1 x_i = \bar{y} + \hat{\beta}_1(x_i - \bar{x})$$

y los residuos

$$e_i = y_i - \hat{y}_i = y_i - \bar{y} - \hat{\beta}_1(x_i - \bar{x})$$

Ejercicio 1.2

Hallar las estimaciones de los parámetros en un modelo de regresión parabólico, minimizando la suma de los cuadrados de los errores:

$$S = \sum_{i=1}^n (y_i - \beta_0 - \beta_1 x_i - \beta_2 x_i^2)^2$$

Del mismo modo que en el ejercicio 1.1 vamos a derivar respecto a cada uno de los parámetros

$$\begin{aligned}
\frac{\partial S}{\partial \beta_0} &= 2\sum_{i=1}^{n}(y_i - \beta_0 - \beta_1 x_i - \beta_2 x_i^2)(-1) \\
\frac{\partial S}{\partial \beta_1} &= 2\sum_{i=1}^{n}(y_i - \beta_0 - \beta_1 x_i - \beta_2 x_i^2)(-x_i) \\
\frac{\partial S}{\partial \beta_2} &= 2\sum_{i=1}^{n}(y_i - \beta_0 - \beta_1 x_i - \beta_2 x_i^2)(-x_i^2)
\end{aligned}$$

y al igualar a cero obtenemos

$$\begin{aligned}
-\sum_{i=1}^{n} y_i + n\beta_0 + \beta_1 \sum_{i=1}^{n} x_i + \beta_2 \sum_{i=1}^{n} x_i^2 &= 0 \\
-\sum_{i=1}^{n} x_i y_i + \beta_0 \sum_{i=1}^{n} x_i + \beta_1 \sum_{i=1}^{n} x_i^2 + \beta_2 \sum_{i=1}^{n} x_i^3 &= 0 \\
-\sum_{i=1}^{n} x_i^2 y_i + \beta_0 \sum_{i=1}^{n} x_i^2 + \beta_1 \sum_{i=1}^{n} x_i^3 + \beta_2 \sum_{i=1}^{n} x_i^4 &= 0
\end{aligned}$$

que son las llamadas *ecuaciones normales*

$$\begin{pmatrix} n & \sum_{i=1}^{n} x_i & \sum_{i=1}^{n} x_i^2 \\ \sum_{i=1}^{n} x_i & \sum_{i=1}^{n} x_i^2 & \sum_{i=1}^{n} x_i^3 \\ \sum_{i=1}^{n} x_i^2 & \sum_{i=1}^{n} x_i^3 & \sum_{i=1}^{n} x_i^4 \end{pmatrix} \begin{pmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \end{pmatrix} = \begin{pmatrix} \sum_{i=1}^{n} y_i \\ \sum_{i=1}^{n} x_i y_i \\ \sum_{i=1}^{n} x_i^2 y_i \end{pmatrix}$$

La solución de este sistema proporciona las estimaciones MC de los parámetros del modelo parabólico. Dado que $\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x} - \hat{\beta}_2(\frac{1}{n}\sum_{i=1}^{n} x_i^2)$, las predicciones son

$$\begin{aligned}
\hat{y}_i &= \hat{\beta}_0 + \hat{\beta}_1 x_i + \hat{\beta}_2 x_i^2 \\
&= \bar{y} - \hat{\beta}_1 \bar{x} - \hat{\beta}_2(\frac{1}{n}\sum_{i=1}^{n} x_i^2) + \hat{\beta}_1 x_i + \hat{\beta}_2 x_i^2 \\
&= \bar{y} + \hat{\beta}_1 (x_i - \bar{x}) + \hat{\beta}_2 (x_i^2 - \frac{1}{n}\sum_{i=1}^{n} x_i^2)
\end{aligned}$$

y los residuos

$$e_i = y_i - \hat{y}_i = y_i - \bar{y} - \hat{\beta}_1 (x_i - \bar{x}) - \hat{\beta}_2 (x_i^2 - \frac{1}{n}\sum_{i=1}^{n} x_i^2)$$