Interpolación y aproximación

Mínimos cuadrados

Joaquin Cavieres

Introducción

Se tienen los siguientes puntos $(1,1), \quad (2,2), \quad (3,2)$ y deseamos encontrar una aproximación mediante $y = mx + b$. Considerando los puntos dados anteriormente entonces las ecuaciones para construir un sistema lineal serían:

$$\begin{aligned} b + m &= 1, \\ b + 2m &= 2, \\ b + 3m &= 2, \end{aligned}$$

que es equivalente a decir:

$$\begin{bmatrix} 1 & 1 \\ 1 & 2 \\ 1 & 3 \end{bmatrix} \begin{bmatrix} b \\ m \end{bmatrix} = \begin{bmatrix} 1 \\ 2 \\ 2 \end{bmatrix} \qquad \text{o expresado en un sistema lineal} \qquad \mathbf{A}\,\mathbf{x} = \mathbf{b}.$$

Sin embargo la recta no pasa por los tres puntos, por tanto, el sistema no tiene solución.

En los modelos estadísticos en la mayoría de las veces nosotros deseamos estimar parámetros x_i basándose en algunas observaciones (las observaciones se asumen que están "contaminadas" por un error o "ruido"). Por ejemplo, si uno desea predecir el promedio de notas de los alumnos de la universidad (PNU) de las solicitudes de estudiantes de primer año, que denotamos por b, basado en promedio de notas de la educación media (PNM), denotado por $a1$, y dos puntajes de prueba de aptitud académica, uno verbal (a_2) y uno cuantitativo (a_3), como parte del proceso de admisión a la universidad. Es una costumbre utilizar un modelo lineal expresado como $b = a_1x_1 + a_2x_2 + a_3x_3$. Esto en forma de mínimos cuadrados se expresaía como:

$$\left\| \mathbf{b} - \sum_{i=1}^{n} x_i \, \mathbf{a}_i \right\|.$$

Lo anterior podemos representarlo con una matriz $\boldsymbol{A}$ con las columnas datas por los vectores independientes $\boldsymbol{a_i}$ y un vector $\boldsymbol{x}$ con elementos x_i. La matriz $\boldsymbol{A}$ es de dimensión $m \times n$ por lo que el largo de los vectores $\boldsymbol{a_i}$ son de largo m. Dado lo anterior entonces quwremos encontrar un $\boldsymbol{x}$

(llamemosle $\hat{\boldsymbol{x}}$) que minimice $\|\mathbf{A}\mathbf{x} - \mathbf{b}\|_2$, así podemos encontrar $\mathbf{A}\hat{\mathbf{x}} = \mathbf{b}$ y el vector de errores cuadrados $\mathbf{e} = \mathbf{b} - \mathbf{A}\hat{\mathbf{x}}$.

Como $\hat{\boldsymbol{x}}$ es una solución al problema de mínimos cuadrados entonces el vector de errores $\mathbf{e} = \mathbf{b} - \mathbf{A}\hat{\mathbf{x}}$ es perpendicular, por tanto este vector $\mathbf{e}$ forma un angulo recto con todos los $\{\mathbf{a}_1, \mathbf{a}_2, \ldots, \mathbf{a}_n\}$, para así tener n ecuaciones con el fin de encontrar $\hat{\boldsymbol{x}}$:

$$\begin{matrix} \mathbf{a}_1^{\mathrm{T}}(\mathbf{b} - \mathbf{A}\hat{\mathbf{x}}) = 0 \\ \vdots \quad \vdots \\ \mathbf{a}_n^{\mathrm{T}}(\mathbf{b} - \mathbf{A}\hat{\mathbf{x}}) = 0 \end{matrix} \qquad \text{o} \qquad \begin{bmatrix} - - - \mathbf{a}_1^{\mathrm{T}} - - - \\ \vdots \\ - - - \mathbf{a}_n^{\mathrm{T}} - - - \end{bmatrix} \begin{bmatrix} \\ \mathbf{b} - \mathbf{A}\hat{\mathbf{x}} \\ \\ \end{bmatrix} = \begin{bmatrix} 0 \\ \vdots \\ 0 \end{bmatrix}$$

en donde las n ecuaciones son exactamente la ecuación normal $\mathbf{A}^{\mathrm{T}}\mathbf{A}\,\hat{\mathbf{x}} = \mathbf{A}^{\mathrm{T}}\,\mathbf{b}$

Otra forma de encontrar los parámetros que minimicen el error es a través del calculo y es el método más utilizado en los modelos estadísticos. Veamos el siguiente ejemplo:

Tenemos un conjunto n de observaciones, con $X = (x_1, x_2, \ldots, x_n)$ y $Y = (y_1, y_2, \ldots, y_n)$. Nosotros asumimos que existe una relación lineal entre ambas y podemos expresarla en la siguiente ecuación (no matricial):

$$y = \beta_0 + \beta_1 x + \epsilon$$

Simulemos un conjunto de datos como ejemplo:

```
set.seed(1000)
n <- 200
x <- rnorm(n, mean = 5, sd=2)
beta_0 <- 2
beta_1 <- 0.5
epsilon <- rnorm(n, mean=0, sd=sqrt(3))
y <- beta_0 + x * beta_1 + epsilon
```

Nosotros estamos interesados en estimar β_0 y β_1 y los denotaremos como $\hat{\beta}_0$ y $\hat{\beta}_1$ y los cuales nos permitiran encontrar una linea lo mas cercana a nuestra variable respuesta Y. Por lo anterior es que nosotros deseamos minimizar los errores (residuales):

$$\begin{aligned} \hat{e} &= y - \hat{y} \\ &= y - \hat{\beta}_0 + \hat{\beta}_1 x \end{aligned}$$

No podemos minimizar drectamente el error $\hat{e}_i$ ya que nuestras predicciones podrían contener valores más altos o más bajos de los reales. Entonces, nos centraremos en minimizar la suma de los errores cuadraticos $RSS = \sum_{i=1}^{n} \hat{e_i}^2$ para evitar el problema anterior.

Minimización con respecto a $\hat{\beta}_0$

$\hat{\beta}_0 = \text{argmin}_{\hat{\beta}_0} \ \sum_{i=1}^{n} \hat{e_i}^2 = \hat{\beta}_0 = \text{argmin}_{\hat{\beta}_0} \ \sum_{i=1}^{n} (y_i - \hat{\beta}_0 + \hat{\beta}_1 x_i)^2$

Aplicamos la derivada parcial a la expresión:

$$\frac{\partial RSS}{\partial \hat{\beta}_0} \sum_{i=1}^{n} (y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i)^2$$

Si aplicamos la regla de la cadena (recordatorio: si hacemos $v = y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i$ y tenemos como función a $f = v^2$, la derivada parcial de v es $2v$, y la de f es -1) entonces:

$$\begin{aligned} &= 2 \cdot -1 \cdot \sum_{i=1}^{n} (y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i) \\ &= -2 \sum_{i=1}^{n} (y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i) - \hat{\beta}_0 + \hat{\beta}_1 x \end{aligned}$$

Igualamos la derivada parcial de $\hat{\beta}_0 = 0$:

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

como $\hat{\beta}_0$ es constante entonces la sumatoria de ese estimado es igual a n veces el mismo, por lo que podemos reescribir lo anterior como:

$$n\hat{\beta}_0 = \sum_{i=1}^{n} y_i - \sum_{i=1}^{n} \hat{\beta}_1 x_i$$

lo que nos da para $\hat{\beta}_0$:

$$\begin{aligned} \hat{\beta}_0 =& \frac{\sum_{i=1}^{n} y_i}{n} - \hat{\beta}_1 \frac{\sum_{i=1}^{n} x_i}{n} \\ \hat{\beta}_0 =& \overline{y} - \hat{\beta}_1 \overline{x} \end{aligned}$$

Minimización con respecto a $\hat{\beta}_1$

Reslver:

$$
\begin{aligned}
\hat{\beta}_1 =& \text{argmin}_{\hat{\beta}_1} \sum_{i=1}^{n} \hat{e_i}^2 \\
=& \text{argmin}_{\hat{\beta}_1} \sum_{i=1}^{n} (y_i - \hat{\beta}_0 + \hat{\beta}_1 x_i)^2
\end{aligned}
$$

La derivada parcial de $\frac{\partial RSS}{\partial \hat{\beta}_1} \sum_{i=1}^{n} (y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i)^2$ via la regla de la cadena es igual

$$
= 2 \cdot x_i \cdot \left(\sum_{i=1}^{n} (y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i) \right)
$$

e igualando a 0:

$$
\begin{aligned}
0 =& 2x \left(\sum_{i=1}^{n} (y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i) \right) \\
=& \sum_{i=1}^{n} (2xy - 2\hat{\beta}_0 x - 2\hat{\beta}_1 x^2) \\
=& \sum_{i=1}^{n} xy - \hat{\beta}_0 \sum_{i=1}^{n} x - \hat{\beta}_1 \sum_{i=1}^{n} x^2
\end{aligned}
$$

reemplazando en $\hat{\beta}_0 = \overline{y} - \hat{\beta}_1 \overline{x}$:

$$
\begin{aligned}
=& \sum_{i=1}^{n} xy - (\overline{y} - \hat{\beta}_1 \overline{x}) \sum_{i=1}^{n} x - \hat{\beta}_1 \sum_{i=1}^{n} x^2 \\
=& \sum_{i=1}^{n} xy - \overline{y} \sum_{i=1}^{n} x + \hat{\beta}_1 \overline{x} \sum_{i=1}^{n} x - \hat{\beta}_1 \sum_{i=1}^{n} x^2
\end{aligned}
$$

factorizando con $\hat{\beta}_1$ en los dos ultimos terminos:

$$
= \sum_{i=1}^{n} xy - \overline{y} \sum_{i=1}^{n} x + \hat{\beta}_1 \left(\overline{x} \sum_{i=1}^{n} x - \sum_{i=1}^{n} x^2 \right)
$$

e igualandolos con la expresión anterior:

$$
-\hat{\beta}_1 \left(\overline{x} \sum_{i=1}^{n} x - \sum_{i=1}^{n} x^2 \right) = \sum_{i=1}^{n} xy - \overline{y} \sum_{i=1}^{n} x
$$

para finalmente dividir en ambos lados por $-\left(\overline{x}\sum_{i=1}^{n} x - \sum_{i=1}^{n} x^2\right)$ nos entrega:

$$\hat{\beta}_1 = \frac{\sum_{i=1}^{n} xy - \overline{y}\sum_{i=1}^{n} x}{\sum_{i=1}^{n} x^2 - \overline{x}\sum_{i=1}^{n} x}$$

Como nota adicional, existe una forma mucho más interpretable para estimar $\hat{\beta}_1$:

$$\frac{SS_{xy}}{SS_x} = \frac{\sum_{i=1}^{n}(x_i - \overline{x})(y_i - \overline{y})}{\sum_{i=1}^{n}(x_i - \overline{x})^2}$$

en donde SS_{xy} es la suma de desviaciones de$\overline{x}$ multipplicado por la suma de las desviaciones de $\overline{y}$ y SS_x es la suma cuadradas de las desviaciones de $\overline{x}$.

Aplicación

```r
# Funcion para encontrar los parametros beta0 y beta 1
# que minimicen la suma de los errores cuadrados
min_cuadrados <- function(x, y){

  sxy <- sum((x - mean(x)) * (y - mean(y)))
  sxx <- sum((x - mean(x))^2)
  beta_1 <- sxy / sxx

  beta_0 <- mean(y) - beta_1 * mean(x)

  return(list('Intercepto'=beta_0,
              'Pendiente'=beta_1,
              'sxy'=sxy,
              'sxx'=sxx))

}

resultados <- min_cuadrados(x, y)
resultados
```

```
## $Intercepto
## [1] 2.186887
##
## $Pendiente
## [1] 0.4568685
##
## $sxy
## [1] 331.6113
##
## $sxx
## [1] 725.8353
```

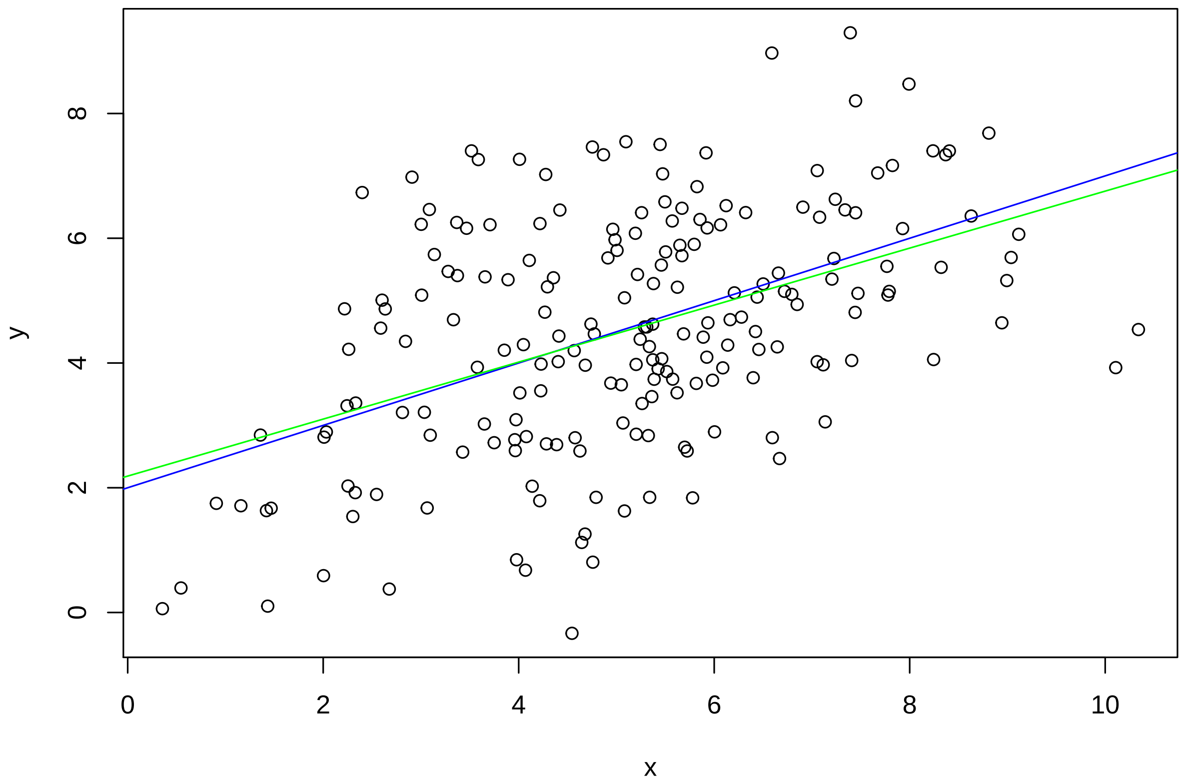
En `R` existe una función para estimar los parámetros mediante mínimos cuadrados llamada `lm()`, así que la vamos a utilizar para comparar los resultados con nuestra función `min_cuadrados()`.

```r
resultados_lm <- lm(y ~ x)
coef(resultados_lm)
```

```
## (Intercept)           x
##   2.1868869   0.4568685
```

Comparemos los resultados con los valores de entrada de los parámetros β_0 y β_1:

```r
plot(x, y)
abline(a = 2, b = 0.5, col='blue')
abline(resultados_lm, col='green')
```



Tarea

Considere a un conjunto de puntos (x_i, y_i) para un $i = 0, 1, ...n$ por la cual no es posible que podamos encontrar una función que pase a través de todos los puntos, sino que queremos encontrar una función $f(x)$ en una forma particular pase los más "cerca" (aproxime) a los puntos (x_i, y_i). Las diferencias (error) entre nuestra $f(x)$ propuesta con los datos reales (y_i) es:

$$e_i = f(x_i) - y_i \quad \text{para} \quad i = 0, 1, 2, \dots n,$$

que son llamados generalmente como los "errores" (o residuales). Podemos encontrar distintos métodos para medir como nuestro $f(x) = y$

- Error máximo: $\|e\|_\infty = \text{máx}_{1 \le i \le n}\{|f(x_i) - y_i|\}$
- Error medio: $\|e\|_1 = \frac{1}{n}\sum_{i=1}^{n} |f(x_i) - y_i|$
- Raíz cuadrada de los errores (o $L^2 - error$): $\|e\|_2 = \left(\frac{1}{n}\sum_{i=1}^{n} |f(x_i) - y_i|^2\right)^{1/2}$

Compare el error máximo, el medio y el error cuadrático medio para la aproximación lineal de $y = f(x) = 3{,}5 - 1{,}3x$ a los puntos de datos $(-1, 5), (0, 4), (1, 2), (2, 1), (3, 0), (4, -2), (5, -3), (6, -4)$.