Práctica 7
Método de los mínimos cuadrados

Índice

1. Ajuste por mínimos cuadrados

El término *mínimos cuadrados* describe el problema muy frecuente de resolver sistemas de ecuaciones lineales "sobredeterminados" o "sobredimensionados", esto es, sistemas lineales *con más ecuaciones que incógnitas*. En tal caso, como normalmente no existe tal solución, en lugar de resolver las ecuaciones de manera exacta se busca sólo una "solución" que haga $\mathsf{A}\vec{x}$ tan próximo a $\vec{b}$ como sea posible.

- Una *solución por mínimos cuadrados* del sistema de ecuaciones lineales $\mathsf{A}\vec{x} = \vec{b}$ es un vector $\vec{x}_M$ en $\mathbb{R}^n$ tal que $\|\vec{b} - \mathsf{A}\vec{x}_M\| \leq \|\vec{b} - \mathsf{A}\vec{x}\|$ para todo $\vec{x} \in \mathbb{R}^n$.

  Si aplicamos el teorema de la aproximación óptima al subespacio $\mathrm{Col}\,\mathsf{A}$ y llamamos $\vec{b}_p$ a la proyección ortogonal de $\vec{b}$ sobre $\mathrm{Col}\,\mathsf{A}$, sabemos que b_p es el vector de $\mathrm{Col}\,\mathsf{A}$ más próximo a b. Entonces habrá un vector $\vec{x}_M$ que será la solución de $\mathsf{A}\vec{x} = \mathsf{b}_p$. Tendremos, pues, que $\mathsf{A}\vec{x}_M = \vec{b}_p$.

  Como también sabemos que $\vec{b} - \vec{b}_p \in (\mathrm{Col}\,\mathsf{A})^\perp$, entonces $\vec{b} - \mathsf{A}\vec{x}_M \in (\mathrm{Col}\,\mathsf{A})^\perp = \mathrm{Nul}\,\mathsf{A}^t$, es decir,
  $$\mathsf{A}^t(\vec{b} - \mathsf{A}\vec{x}_M) = 0,$$
  de donde
  $$\mathsf{A}^t\mathsf{A}\vec{x}_M = \mathsf{A}^t\vec{b}.$$

- La ecuación matricial $\mathsf{A}^t\mathsf{A}\vec{x}_M = \mathsf{A}^t\vec{b}$ se conoce como *sistema de ecuaciones normales* para x_M.

- Si las columnas de A son linealmente independientes, la matriz $\mathsf{A}^t\mathsf{A}$ es invertible y la solución por mínimos cuadrados $\vec{x}_M = (\mathsf{A}^t\mathsf{A})^{-1}\mathsf{A}^t\vec{b}$ será única.
- Si las columnas de A son linealmente independientes y $\mathsf{A} = \mathsf{QR}$, la única solución por mínimos cuadrados viene dada por $\vec{x}_M = \mathsf{R}^{-1}\mathsf{Q}^t\vec{b}$.
- Se llama *error residual* a $\|\vec{b} - \mathsf{A}\vec{x}_M\|$.

Ejemplo 1. Determinamos las soluciones por mínimos cuadrados de $\mathsf{A}\vec{x} = \vec{b}$, siendo

$$\mathsf{A} = \begin{bmatrix} 1 & 3 \\ 1 & -1 \\ 1 & 1 \end{bmatrix}, \qquad \vec{b} = \begin{bmatrix} 5 \\ 1 \\ 0 \end{bmatrix}.$$

Debemos resolver $\mathsf{A}^t\mathsf{A}\vec{x}_M = \mathsf{A}^t\vec{b}$.

$$\mathsf{A}^t\mathsf{A} = \begin{bmatrix} 1 & 1 & 1 \\ 3 & -1 & 1 \end{bmatrix} \begin{bmatrix} 1 & 3 \\ 1 & -1 \\ 1 & 1 \end{bmatrix} = \begin{bmatrix} 3 & 3 \\ 3 & 11 \end{bmatrix}, \qquad \mathsf{A}^t\vec{b} = \begin{bmatrix} 1 & 1 & 1 \\ 3 & -1 & 1 \end{bmatrix} \begin{bmatrix} 5 \\ 1 \\ 0 \end{bmatrix} = \begin{bmatrix} 6 \\ 14 \end{bmatrix}.$$

El sistema

$$\begin{bmatrix} 3 & 3 \\ 3 & 11 \end{bmatrix} \begin{bmatrix} x_{1M} \\ x_{2M} \end{bmatrix} = \begin{bmatrix} 6 \\ 14 \end{bmatrix}$$

tiene como única solución

$$\begin{bmatrix} x_{1M} \\ x_{2M} \end{bmatrix} = \begin{bmatrix} 1 \\ 1 \end{bmatrix}.$$

El error residual será

$$\|\vec{b} - \mathsf{A}\vec{x}_M\| = \left\| \begin{bmatrix} 5 \\ 1 \\ 0 \end{bmatrix} - \begin{bmatrix} 1 & 3 \\ 1 & -1 \\ 1 & 1 \end{bmatrix} \begin{bmatrix} 1 \\ 1 \end{bmatrix} \right\| = \sqrt{6}$$

Con Scilab:

Encontramos la solución por mínimos cuadrados teniendo en cuenta que en Scilab el comando x=A\\b proporciona una solución por mínimos cuadrados en caso de que A no sea una matriz cuadrada. Además, si A es de rango completo por columnas, entonces la solución será única. Del mismo modo, cuando A no es de rango completo por columnas, la solución no es única.

```
-->A=[1 3;1 -1;1 1];b=[5;1;0];
-->Am=A'*A
 Am  =
     3.     3.
     3.     11.
-->bm=A'*b
```

```
 bm  =
    6.
    14.
-->xm=Am\bm
 xm  =
    1.
    1.
```

y puedes comprobar que, directamente, `x=A\b` da la misma solución.

2. Aplicaciones del ajuste por mínimos cuadrados

En ciencias e ingeniería, los experimentos que se realizan producen un conjunto de datos (x_1, y_1), $(x_2, y_2), \ldots,$ (x_n, y_n) con las abscisas diferentes. El problema que se plantea es encontrar una función $y = f(x)$ que relacione los datos lo mejor posible. Buscar esta solución implicará resolver un problema de mínimos cuadrados.

2.1. Ajuste de rectas

La relación más sencilla entre dos variables x e y es la ecuación lineal $y = \beta_0 + \beta_1 x$. A menudo, al representar gráficamente el conjunto de datos experimentales hace el efecto que estos quedan cerca de una recta. Queremos determinar los parámetros β_0 y β_1 que hagan la recta tan «pròxima» a los puntos como sea posible.

La *recta de regresión por mínimos cuadrados* es aquella recta $y = \beta_0 + \beta_1 x$ que minimiza el error residual.

Si los datos estuviesen sobre la recta, los parámetros β_0 y β_1 cumplirian las ecuaciones

$$\beta_0 + \beta_1 x_i = y_i \qquad \text{para } i \in \{1, 2, \ldots, n\}.$$

que, en forma matricial, sería

$$\begin{bmatrix} 1 & x_1 \\ 1 & x_2 \\ \cdots & \cdots \\ 1 & x_n \end{bmatrix} \begin{bmatrix} \beta_0 \\ \beta_1 \end{bmatrix} = \begin{bmatrix} y_1 \\ y_2 \\ \cdots \\ y_n \end{bmatrix}.$$

Llamaremos *matriz de diseño* a

$$\mathsf{X} = \begin{bmatrix} 1 & x_1 \\ 1 & x_2 \\ \cdots & \cdots \\ 1 & x_n \end{bmatrix},$$

vector parámetro a

$$\vec{\beta} = \begin{bmatrix} \beta_0 \\ \beta_1 \end{bmatrix}$$

y *vector observación* a

$$\vec{y} = \begin{bmatrix} y_1 \\ y_2 \\ \dots \\ y_n \end{bmatrix}$$

Se llama *vector residual* a

$$\vec{\varepsilon} = \vec{y} - \mathsf{X}\vec{\beta}.$$

Para determinar el vector parámetro basta con resolver por mínimos cuadrados el sistema $\mathsf{X}\vec{\beta} = \mathsf{y}$, es decir, encontrar la solución que minimiza la norma del vector residual.

Ejemplo 2. Encontraremos la solución de la ecuación $y = \beta_0 + \beta_1 x$ de la recta de mínimos cuadrados que se ajuste a los puntos $(0, 1)$, $(1, 1)$, $(2, 2)$ y $(3, 2)$. Construimos la matriz de diseño X y calculamos $\mathsf{X}^t\mathsf{X}$.

$$\mathsf{X} = \begin{bmatrix} 1 & 0 \\ 1 & 1 \\ 1 & 2 \\ 1 & 3 \end{bmatrix}, \qquad \mathsf{X}^t\mathsf{X} = \begin{bmatrix} 1 & 1 & 1 & 1 \\ 0 & 1 & 2 & 3 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 1 & 1 \\ 1 & 2 \\ 1 & 3 \end{bmatrix} = \begin{bmatrix} 4 & 6 \\ 6 & 14 \end{bmatrix}.$$

Obtenemos

$$\mathsf{X}^t\vec{y} = \begin{bmatrix} 1 & 1 & 1 & 1 \\ 0 & 1 & 2 & 3 \end{bmatrix} \begin{bmatrix} 1 \\ 1 \\ 2 \\ 2 \end{bmatrix} = \begin{bmatrix} 6 \\ 11 \end{bmatrix}$$

Resolvemos

$$\begin{bmatrix} 4 & 6 \\ 6 & 14 \end{bmatrix} \begin{bmatrix} \beta_0 \\ \beta_1 \end{bmatrix} = \begin{bmatrix} 6 \\ 11 \end{bmatrix}, \quad \text{cuya solución es} \quad \begin{bmatrix} \beta_0 \\ \beta_1 \end{bmatrix} = \begin{bmatrix} 9/10 \\ 2/5 \end{bmatrix}$$

Por lo tanto, la recta de ajuste por mínimos cuadrados será

$$y = \frac{9}{10} + \frac{2}{5}x.$$

Con Scilab:

```
-->X=[1 0;1 1;1 2;1 3], y=[1;1;2;2];
-->X1=X'*X,  y1=X'*y
 X1  =
    4.    6.
    6.    14.
 y1  =
    6.
    11.
-->b=X1\y1
 b  =
    0.9
    0.4
```

Calculamos ahora la norma del vector residual $\|\vec{y}-\mathsf{X}\vec{\beta}\|$:

```
-->norm(y-X*b)
 ans  =
    0.4472136
```

2.2. Ajuste de curvas por mínimos cuadrados

Si representamos gráficamente el conjunto de datos y observamos que no están cerca de una recta, buscaremos la curva que mejor los ajuste

$$y = \beta_0 f_0(x) + \beta_1 f_1(x) + \cdots + \beta_k f_k(x),$$

donde f_0, $f_1, \ldots,$ f_n son funciones conocidas y β_0, $\beta_1, \ldots,$ β_n los parámetros que debemos determinar para construir la curva. Procediendo de manera análoga a como lo hemos hecho en la regresión lineal, debemos encontrar un vector parámetro que haga mínima la norma del vector residual. En este caso, el sistema escrito en forma matricial será

$$\begin{bmatrix} f_0(x_1) & f_1(x_1) & \ldots & f_k(x_1) \\ f_0(x_2) & f_1(x_2) & \ldots & f_k(x_1) \\ \cdots & \cdots & \cdots & \cdots \\ \cdots & \cdots & \cdots & \cdots \\ f_0(x_n) & f_1(x_n) & \ldots & f_k(x_n) \end{bmatrix} \begin{bmatrix} \beta_0 \\ \beta_1 \\ \ldots \\ \beta_k \end{bmatrix} = \begin{bmatrix} y_1 \\ y_2 \\ \ldots \\ \ldots \\ y_n \end{bmatrix}.$$

La matriz de este sistema es la matriz de diseño.

Ejemplo 3. Con los datos de la tabla

1	2	3	4	5
1,8	2,7	3,4	3,8	3,9

determinaremos el vector parámetro, el vector residual y la curva de mínimos cuadrados asociada a la función $y = \beta_0 + \beta_1 x + \beta_2 x^2$.

1. Deberemos resolver por el método de mínimos cuadrados el sistema de ecuaciones lineales $\vec{y} = \mathsf{X}\vec{\beta}$:
$$\begin{bmatrix} 1{,}8 \\ 2{,}7 \\ 3{,}4 \\ 3{,}8 \\ 3{,}9 \end{bmatrix} = \begin{bmatrix} 1 & 1 & 1 \\ 1 & 2 & 4 \\ 1 & 3 & 9 \\ 1 & 4 & 16 \\ 1 & 5 & 25 \end{bmatrix} \begin{bmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \end{bmatrix}.$$

2. Calculamos
$$\mathsf{X}^t\mathsf{X} = \begin{bmatrix} 5 & 15 & 55 \\ 15 & 55 & 225 \\ 55 & 225 & 979 \end{bmatrix}, \qquad \mathsf{X}^t\vec{y} = \begin{bmatrix} 15{,}6 \\ 52{,}1 \\ 201{,}5 \end{bmatrix}.$$

3. Resolviendo el sistema (compatible determinado) $\mathsf{X}^t\mathsf{X}\vec{\beta} = \mathsf{X}^t\vec{y}$ obtendremos el vector parámetro que minimiza el error residual

$$\begin{bmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \end{bmatrix} = \begin{bmatrix} 0{,}58 \\ 1{,}34 \\ -0{,}136 \end{bmatrix}.$$

Por lo tanto la función (parábola) que mejor se ajusta a los datos será $y = 0{,}58 + 1{,}34x - 0{,}136x^2$.

Si calculamos $\vec{\varepsilon} = \vec{y} - \mathsf{X}\vec{\beta}$, obtenemos el vector residual

$$\begin{bmatrix} 0{,}114 \\ -0{,}026 \\ 0{,}008 \\ 0{,}014 \\ 0{,}008 \end{bmatrix}.$$

Con Scilab:

Introducimos las matrices X, $\vec{y}$:

```
-->X=[1 1 1;1 2 4;1 3 9;1 4 16;1 5 25], y=[1.8;2.7;3.4;3.8;3.9];
```

y, con la función \ obtenemos la solución por mínimos cuadrados:

```
-->b=X1\y1
 b  =
     0.58
     1.3442857
   - 0.1357143
```

Luego la parábola que mejor se aproxima a los datos tiene los parámetros de b y su error residual será

```
-->ER=norm(y-X*b)
 ER  =
       0.0338062
```

Ejemplo 4. Con los datos del ejemplo anterior, determinaremos los parámetros que ajustan la función $y = \beta_1 \cos(\pi x/3) + \beta_2 \sin(\pi x/3)$.

$$\mathsf{X} = \begin{bmatrix} \cos\frac{\pi}{3} & \sin\frac{\pi}{3} \\ \cos\frac{2\pi}{3} & \sin\frac{2\pi}{3} \\ \cos\frac{3\pi}{3} & \sin\frac{3\pi}{3} \\ \cos\frac{4\pi}{3} & \sin\frac{4\pi}{3} \\ \cos\frac{5\pi}{3} & \sin\frac{5\pi}{3} \end{bmatrix}$$

$$\mathsf{X}^t\mathsf{X} = \begin{bmatrix} 2 & 0 \\ 0 & 3 \end{bmatrix}, \qquad \mathsf{X}^t\vec{y} = \begin{bmatrix} -3{,}8 \\ -2{,}8 \end{bmatrix}$$

Resolvemos el sistema y obtenemos el vector parámetro.

$$\begin{bmatrix} 2 & 0 \\ 0 & 3 \end{bmatrix} \begin{bmatrix} \beta_1 \\ \beta_2 \end{bmatrix} = \begin{bmatrix} -3{,}8 \\ -2{,}8 \end{bmatrix}, \quad \text{cuya solución es} \quad \begin{bmatrix} \beta_1 \\ \beta_2 \end{bmatrix} = \begin{bmatrix} -1{,}9 \\ -0{,}9 \end{bmatrix}.$$

Por lo tanto, la función pedida será $y = -1{,}9\cos(\pi x/3) - 0{,}9\sin(\pi x/3)$, y el vector residual es

$$\vec{\varepsilon} = \vec{y} - \mathsf{X}\vec{\beta} = \begin{bmatrix} 3{,}5 \\ 2{,}5 \\ 1{,}5 \\ 2 \\ 4 \end{bmatrix}.$$

Observamos que el vector residual es menor cuando usamos la primera función para ajustar los datos que cuando usamos la segunda.

Con Scilab:

Introducimos la matriz de diseño X (el vector y ya estaba introducido):

```
-->X=[cos(%pi/3),   sin(%pi/3)
-->   cos(%pi*2/3), sin(%pi*2/3)
-->   cos(%pi*3/3), sin(%pi*3/3)
-->   cos(%pi*4/3), sin(%pi*4/3)
-->   cos(%pi*5/3), sin(%pi*5/3)]
 X  =
    0.5     0.8660254
  - 0.5     0.8660254
  - 1.      1.225D-16
  - 0.5   - 0.8660254
    0.5   - 0.8660254
```

Y podemos obtener los parámetros directamente:

```
-->b=X\y
 b  =
  - 1.9
  - 0.9237604
```

que no ha salido exactamente igual (debido a errores de redondeo con las funciones trigonométricas).

Podemos también repetir con Scilab el proceso hecho a mano para ver si se obtiene el resultado correcto. Primero calculamos $\mathsf{X}^t\mathsf{X}$ y $\mathsf{X}^t\vec{y}$:

```
-->X1=X'*X,y1=X'*y
 X1  =
    2.            2.924D-16
    2.924D-16     3.
```

```
 y1   =
   - 3.8
   - 2.7712813
```

y luego resolvemos el sistema 2×2: $\mathsf{X}_1\vec{b} = \vec{y}_1$

```
-->b1=X1\y1
 b1   =
   - 1.9
   - 0.9237604
```

Como se puede observar, este proceso largo ha obtenido el mismo resultado que el corto (porque $\backslash$ realiza exactamente este proceso siempre que el sistema no es compatible determinado) y sigue conteniendo cierto error de redondeo.

Veamos el error residual que proporciona esta función:

```
-->ER=norm(y-X*b1)
 ER   =
     6.4776539
```

Luego no parece ser una función apropiada a los datos del problema.

Nota: con una recta $y = \beta_0 + \beta_1 x$

```
-->X=[1 1;1 2;1 3;1 4;1 5];
-->b=X\y
 b   =
    1.53
    0.53
-->ER=norm(y-X*b)
 ER   =
     0.5089204
```

Conclusión: por el momento, la curva que mejor se ajusta a los datos es la parábola

$$y = 0{,}58 + 1{,}34x - 0{,}136x^2$$