Mínimos Cuadrados

Mínimos cuadrados es un procedimiento en el que, dado un conjunto de datos (pares ordenados y familia de funciones), se intenta determinar la función continua que *mejor* se aproxime a los datos. En particular, se busca minimizar la suma de cuadrados de las diferencias ordenadas (llamadas residuos) entre los puntos generados por la función y los correspondientes datos.

Problema lineal:

Considere un sistema rectangular de ecuaciones

$$\boldsymbol{A}\boldsymbol{x} = \boldsymbol{b}$$

donde $\boldsymbol{A} \in \mathbb{R}^{m\times n}$, con $n < m$, es una matriz rectangular de m filas y n columnas y $\boldsymbol{b} \in \mathbb{R}^m$.
Este problema, en general, no tiene solución: **sistema sobredeterminado**.

Una alternativa es buscar una solución en el sentido generalizado siguiente:

$$\text{Hallar } \boldsymbol{x} \in \mathbb{R}^n \text{ tal que } \ \|\boldsymbol{b} - \boldsymbol{A}\boldsymbol{x}\|_2 \text{ sea } \textbf{mínima}.$$

Definición. El vector $\boldsymbol{x}$ que minimiza $\|\boldsymbol{b}-\boldsymbol{A}\boldsymbol{x}\|_2$ es la **solución en el sentido de mínimos cuadrados** del sistema rectangular.

Observación. En general:

$$\boldsymbol{A}\boldsymbol{x} \neq \boldsymbol{b}.$$

Este tipo de sistemas rectangulares surgen, por ejemplo, al intentar aproximar una gran cantidad de datos mediante funciones polinomiales.

Ejemplo. Ajuste de polinomios.

Dado un conjunto de puntos

$$(x_1, y_1), \ldots, (x_m, y_m),$$

nos proponemos encontrar el polinomio

$$p(x) = c_0 + c_1 x + \ldots + c_{n-1}x^{n-1},$$

con $n < m$ que está **más cerca** de estos puntos en el sentido que

$$\sum_{i=1}^{m} |p(x_i) - y_i|^2,$$

sea mínima.

Esta suma de cuadrados es el cuadrado de la norma del residuo del sistema rectangular:

$$\begin{pmatrix} 1 & x_1 & \ldots & x_1^{n-1} \\ 1 & x_2 & \ldots & x_2^{n-1} \\ 1 & x_3 & \ldots & x_3^{n-1} \\ \vdots & \vdots & & \vdots \\ 1 & x_m & \ldots & x_m^{n-1} \end{pmatrix} \begin{pmatrix} c_0 \\ c_1 \\ \vdots \\ c_{n-1} \end{pmatrix} = \begin{pmatrix} y_1 \\ y_2 \\ y_3 \\ \vdots \\ y_m \end{pmatrix}.$$

Para encontrar una solución al problema de mínimos cuadrados consideramos el siguiente teorema.

Teorema. *Sean* $\boldsymbol{A} \in \mathbb{R}^{m\times n}(m \geq n)$ *y* $\boldsymbol{b} \in \mathbb{R}^m$*. Un vector* $\boldsymbol{x} \in \mathbb{R}^n$ *minimiza la norma del residuo* $\|\boldsymbol{r}\|_2 = \|\boldsymbol{b} - \boldsymbol{A}\boldsymbol{x}\|_2$ *si y sólo si el residuo* $\boldsymbol{r}$ *es ortogonal a la imagen de* $\boldsymbol{A}$*; esto es si*

$$\boldsymbol{A}^{\mathrm{t}}\boldsymbol{r} = \boldsymbol{0},$$

donde $\boldsymbol{A}^{\mathrm{t}}$ *es la matriz transpuesta de* $\boldsymbol{A}$*.*

Observación:

- $\boldsymbol{x}$ debe satisfacer

$$\boldsymbol{A}^{\mathrm{t}}\boldsymbol{r} = \boldsymbol{0} \quad \Longleftrightarrow \quad \boldsymbol{A}^{\mathrm{t}}(\boldsymbol{b} - \boldsymbol{A}\boldsymbol{x}) = \boldsymbol{0} \quad \Longleftrightarrow \quad \boldsymbol{A}^{\mathrm{t}}\boldsymbol{A}\boldsymbol{x} = \boldsymbol{A}^{\mathrm{t}}\boldsymbol{b}.$$

  Estas últimas ecuaciones reciben el nombre de **ecuaciones normales**.

- En el caso en que $m = n$ y que la matriz $\boldsymbol{A}$ sea una matriz no singular, entonces las ecuaciones normales entregan como solución la solución del sistema lineal $\boldsymbol{A}\boldsymbol{x} = \boldsymbol{b}$.

Las ecuaciones normales tienen solución única si y sólo si todas las columnas de $\boldsymbol{A}$ son l.i.; es decir, si $\operatorname{rango}(\boldsymbol{A}) = n$.

En este caso, además, la matriz $\boldsymbol{A}^{\mathrm{t}}\boldsymbol{A}$ es **simétrica y definida positiva**, de donde, las ecuaciones normales tienen solución única y se pueden utilizar los métodos estudiados para estas matrices, en particular, el **método de Cholesky** o el método iterativo de **Gauss-Seidel**.

Para resolver las ecuaciones normales se puede proceder del siguiente modo:

1. Calcular la matriz $\boldsymbol{A}^{\mathrm{t}}\boldsymbol{A}$ y el vector $\boldsymbol{A}^{\mathrm{t}}\boldsymbol{b}$.
2. Obtener la matriz $\boldsymbol{L}$ de la factorización de Cholesky: $\boldsymbol{A}^{\mathrm{t}}\boldsymbol{A} = \boldsymbol{L}\boldsymbol{L}^{\mathrm{t}}$.
3. Resolver el sistema triangular inferior $\boldsymbol{L}\boldsymbol{y} = \boldsymbol{A}^{\mathrm{t}}\boldsymbol{b}$.
4. Resolver el sistema triangular superior $\boldsymbol{L}^{\mathrm{t}}\boldsymbol{x} = \boldsymbol{y}$.

Inconveniente: El condicionamiento de la matriz $\boldsymbol{A}^{\mathrm{t}}\boldsymbol{A}$ es en general malo, lo que genera gran sensibilidad respecto a errores de redondeo.

Por ejemplo, si $\boldsymbol{A}$ es cuadrada, entonces $\operatorname{cond}_2(\boldsymbol{A}^{\mathrm{t}}\boldsymbol{A}) = \operatorname{cond}_2(\boldsymbol{A})^2$.

Solución: factorización QR. Ortogonalizar las columnas de $\boldsymbol{A}$ mediante, por ejemplo, **Gram-Schmidt**.

Para esto, escribamos:

$$\boldsymbol{A} = \left(\begin{array}{c|c|c|c} \boldsymbol{a}_1 & \boldsymbol{a}_2 & \dots & \boldsymbol{a}_n \end{array}\right) \in \mathbb{R}^{m\times n}$$

donde $\boldsymbol{a}_i \in \mathbb{R}^m$, $i = 1, 2, \ldots, n$, son las columnas de la matriz $\boldsymbol{A}$.

La idea es construir una matriz

$$\boldsymbol{Q} = \left(\begin{array}{c|c|c|c} \boldsymbol{q}_1 & \boldsymbol{q}_2 & \ldots & \boldsymbol{q}_n \end{array} \right) \in \mathbb{R}^{m \times n}$$

y una matriz triangular superior

$$\boldsymbol{R} = \begin{pmatrix} r_{11} & r_{12} & \ldots & r_{1n} \\ 0 & r_{22} & \ldots & r_{2n} \\ \vdots & \ddots & \ddots & \vdots \\ 0 & \ldots & 0 & r_{nn} \end{pmatrix} \in \mathbb{R}^{n \times n},$$

tales que

$$\boldsymbol{A} = \boldsymbol{Q}\boldsymbol{R}.$$

$$\boldsymbol{A} = \boldsymbol{Q}\boldsymbol{R} \quad \Longleftrightarrow \quad \left\{ \begin{array}{rcl} \boldsymbol{a}_1 & = & r_{11}\boldsymbol{q}_1, \\ \boldsymbol{a}_2 & = & r_{12}\boldsymbol{q}_1 + r_{22}\boldsymbol{q}_2, \\ & \vdots & \\ \boldsymbol{a}_n & = & r_{1n}\boldsymbol{q}_1 + r_{2n}\boldsymbol{q}_2 + \ldots + r_{nn}\boldsymbol{q}_n. \end{array} \right.$$

$\boldsymbol{Q}$ y $\boldsymbol{R}$ se pueden construir mediante el proceso de **ortogonalización de Gram-Schmidt**:

$$\boxed{\begin{array}{l} \text{Para } j = 1, \ldots, n: \\ \quad \text{para } i = 1, \ldots, j-1: \\ \quad\quad r_{ij} = \boldsymbol{q}_i^{\mathrm{t}} \boldsymbol{a}_j, \\ \quad r_{jj} = \left\| \boldsymbol{a}_j - \displaystyle\sum_{i=1}^{j-1} r_{ij}\boldsymbol{q}_i \right\|_2, \\ \quad \boldsymbol{q}_j = \dfrac{1}{r_{jj}} \left(\boldsymbol{a}_j - \displaystyle\sum_{i=1}^{j-1} r_{ij}\boldsymbol{q}_i \right). \end{array}}$$

Las columnas de $\boldsymbol{Q}$ son vectores ortonormales:

$$\boldsymbol{q}_i^{\mathrm{t}} \boldsymbol{q}_j = \delta_{ij} = \left\{ \begin{array}{ll} 1 & \text{si } i = j, \\ 0 & \text{si } i \neq j, \end{array} \right.$$

de donde la matriz $\boldsymbol{Q}$ satisface

$$\boldsymbol{Q}^{\mathrm{t}}\boldsymbol{Q} = \begin{pmatrix} \boldsymbol{q}_1^{\mathrm{t}} \\ \hline \boldsymbol{q}_2^{\mathrm{t}} \\ \hline \vdots \\ \hline \boldsymbol{q}_n^{\mathrm{t}} \end{pmatrix} \left(\begin{array}{c|c|c|c} \boldsymbol{q}_1 & \boldsymbol{q}_2 & \ldots & \boldsymbol{q}_n \end{array} \right) = \boldsymbol{I}.$$

Si rango$(\boldsymbol{A}) = n$, entonces el proceso de ortogonalización de Gram-Schmidt genera una matriz $\boldsymbol{R}$ no singular.

Aplicación a la resolución de las ecuaciones normales.

Para resolver el sistema de ecuaciones normales:

$$\boldsymbol{A}^{\mathrm{t}}\boldsymbol{A}\boldsymbol{x} = \boldsymbol{A}^{\mathrm{t}}\boldsymbol{b},$$

$$\text{como} \qquad \begin{cases} \boldsymbol{A} = \boldsymbol{Q}\boldsymbol{R}, \\ \boldsymbol{Q}^{\mathrm{t}}\boldsymbol{Q} = \boldsymbol{I} \quad \text{y} \\ \boldsymbol{R} \text{ es no singular,} \end{cases}$$

entonces:

$$\begin{aligned} \boldsymbol{A}^{\mathrm{t}}\boldsymbol{A}\boldsymbol{x} = \boldsymbol{A}^{\mathrm{t}}\boldsymbol{b} \quad &\Longleftrightarrow \quad \boldsymbol{R}^{\mathrm{t}}\boldsymbol{Q}^{\mathrm{t}}\boldsymbol{Q}\boldsymbol{R}\boldsymbol{x} = \boldsymbol{R}^{\mathrm{t}}\boldsymbol{Q}^{\mathrm{t}}\boldsymbol{b} \\ &\Longleftrightarrow \quad \boldsymbol{R}^{\mathrm{t}}\boldsymbol{R}\boldsymbol{x} = \boldsymbol{R}^{\mathrm{t}}\boldsymbol{Q}^{\mathrm{t}}\boldsymbol{b} \\ &\Longleftrightarrow \quad \boldsymbol{R}\boldsymbol{x} = \boldsymbol{Q}^{\mathrm{t}}\boldsymbol{b}. \end{aligned}$$

Observación. Cuando $\boldsymbol{A} \in \mathbb{R}^{m\times n}$ es una matriz rectangular y $\boldsymbol{b} \in \mathbb{R}^m$, el comando MATLAB

```
x=A\b;
```

devuelve la solución del sistema rectangular en el sentido de mínimos cuadrados.

MATLAB obtiene esta solución mediante el método **QR**, pero las matrices $\boldsymbol{Q}$ y $\boldsymbol{R}$ no se obtienen mediante Gram-Schmidt, sino mediante **transformaciones de Householder** que propagan menos los errores de redondeo.

Ejemplo 1. Un problema de aproximación polinomial.

Considere la siguiente tabla de valores:

x	0.0	0.5	1.0	1.5	2.0	2.5	3.0	3.5	4.0	4.5	5.0
y	10.5000	5.4844	0.0000	-3.6094	-4.5000	-2.9531	0.0000	2.9531	4.5000	3.6094	0.0000

Se pide ajustar estos datos en el sentido de mínimos cuadrados por un polinomio de grado 3.

Solución.

Nuestro problema se reduce a encontrar constantes a, b, c y d para formar el polinomio

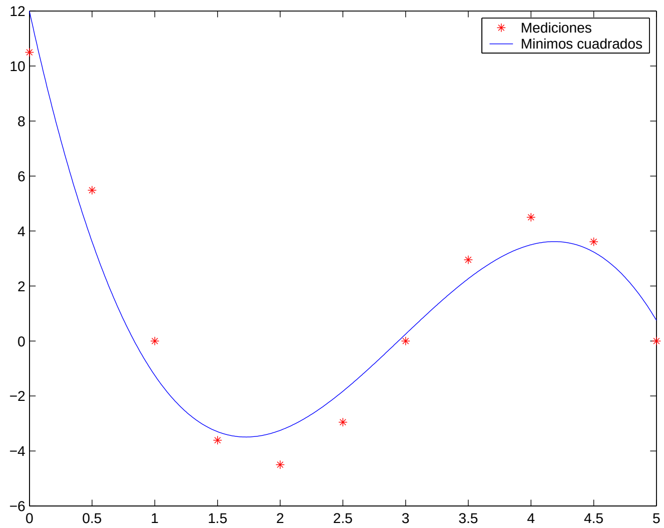
$$p(x) = ax^3 + bx^2 + cx + d.$$

Evaluamos el polinomio p en los diferentes valores x de la tabla, obteniendo así el sistema lineal rectangular con incógnitas a, b, c y d:

$$\underbrace{\begin{pmatrix} 0{,}0000 & 0{,}0000 & 0{,}0000 & 1{,}0000 \\ 0{,}1250 & 0{,}2500 & 0{,}5000 & 1{,}0000 \\ 1{,}0000 & 1{,}0000 & 1{,}0000 & 1{,}0000 \\ 3{,}3750 & 2{,}2500 & 1{,}5000 & 1{,}0000 \\ 8{,}0000 & 4{,}0000 & 2{,}0000 & 1{,}0000 \\ 15{,}6250 & 6{,}2500 & 2{,}5000 & 1{,}0000 \\ 27{,}0000 & 9{,}0000 & 3{,}0000 & 1{,}0000 \\ 42{,}8750 & 12{,}2500 & 3{,}5000 & 1{,}0000 \\ 64{,}0000 & 16{,}0000 & 4{,}0000 & 1{,}0000 \\ 91{,}1250 & 20{,}2500 & 4{,}5000 & 1{,}0000 \\ 125{,}0000 & 25{,}0000 & 5{,}0000 & 1{,}0000 \end{pmatrix}}_{\boldsymbol{A}} \underbrace{\begin{pmatrix} a \\ b \\ c \\ d \end{pmatrix}}_{\boldsymbol{x}} = \underbrace{\begin{pmatrix} 10{,}5000 \\ 5{,}4844 \\ 0{,}0000 \\ -3{,}6094 \\ -4{,}5000 \\ -2{,}9531 \\ 0{,}0000 \\ 2{,}9531 \\ 4{,}5000 \\ 3{,}6094 \\ 0{,}0000 \end{pmatrix}}_{\boldsymbol{b}}$$

Resolviendo el sistema rectangular en el sentido de mínimos cuadrados se obtiene el polinomio:

$$p(x) = -0{,}9583x^3 + 8{,}5x^2 - 20{,}7917x + 12.$$



Linealización de modelos no lineales:

El ajuste por mínimos cuadrados polinomial proporciona una técnica poderosa para ajustar datos. Sin embargo, se basa en el hecho de que la relación entre las variables dependientes e independientes es polinomial. Este no es siempre el caso, y el primer paso en cualquier análisis de mínimos cuadrados debe ser graficar e inspeccionar visualmente los datos para determinar si se aplica un modelo dado.

En el caso de algunos modelos no lineales, se pueden utilzar transformaciones para expresar los datos en una forma que sea compatible con el método de mínimos cuadrados lineal.

Ejemplo 2. Un problema no lineal reducible a lineal.

Considere la siguiente tabla de valores

x	0.0000	0.4000	0.8000	1.2000	1.6000	2.0000
y	3.1437	4.4169	6.0203	8.6512	11.0078	16.2161

Se quiere ajustar una función

$$f(x) = ae^{bx}$$

a estos datos en el sentido de mínimos cuadrados.

Solución.

Tomando logaritmos se transforma en un problema lineal de cuadrados mínimos:

$$z = \ln(y) = \ln(f(x)) = \ln(a) + bx.$$

x	0.0000	0.4000	0.8000	1.2000	1.6000	2.0000
$z = \ln(y)$	1.1454	1.4854	1.7951	2.1577	2.3986	2.7860

$$\underbrace{\begin{pmatrix} 1 & 0{,}0000 \\ 1 & 0{,}4000 \\ 1 & 0{,}8000 \\ 1 & 1{,}2000 \\ 1 & 1{,}6000 \\ 1 & 2{,}0000 \end{pmatrix}}_{\boldsymbol{A}} \underbrace{\begin{pmatrix} \ln(a) \\ b \end{pmatrix}}_{\boldsymbol{x}} = \underbrace{\begin{pmatrix} 1{,}1454 \\ 1{,}4854 \\ 1{,}7951 \\ 2{,}1577 \\ 2{,}3986 \\ 2{,}7860 \end{pmatrix}}_{\boldsymbol{b}}$$

Resolvemos el problema de cuadrados mínimos y obtenemos:

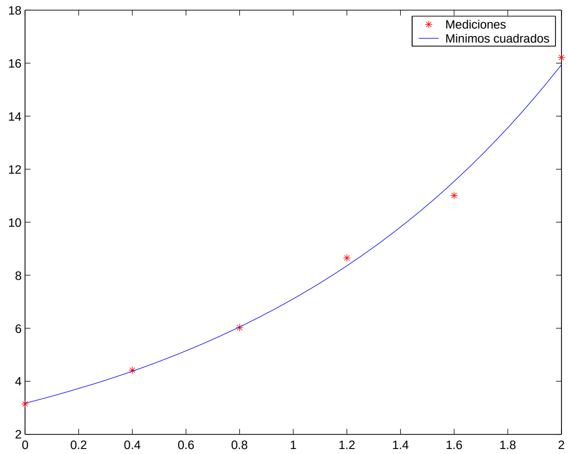
$$\ln(a) = 1{,}1539 \quad \Longrightarrow \quad a = e^{1{,}1539} = 3{,}1705 \qquad \text{y} \qquad b = 0{,}8075.$$

Por lo tanto

$$f(x) = 3{,}1705e^{0{,}8075x}.$$

Otros ejemplos de modelos no lineales reducibles a lineales

- $f(t) = ce^{at-bt^2}$: en este caso, se aplica logaritmo.
- $f(t) = \dfrac{a}{b+t}$: en este caso se toman los recíprocos.
- $f(t) = \dfrac{k_0}{1+ae^{-ct}}$, donde k_0 es una constante conocida: en este caso se toman recíprocos, se resta 1 y después se aplica logaritmo.



Mínimos cuadrados no lineales:

Existen muchos casos en ingeniería y ciencia donde modelos no lineales deben ajustarse a los datos. En el contexto actual, estos modelos se definen como aquellos que tienen una dependencia no lineal de sus parámetros. Por ejemplo

$$y = a_0(1 - e^{a_1 x}) \tag{1}$$

Esta ecuación no se puede manipular para que se ajuste a la forma general de minimos cuadrados lineales.

Al igual que con los mínimos cuadrados lineales, la regresión no lineal se basa en determinar los valores de los parámetros que minimizan la suma de los cuadrados de los residuos. Sin embargo, para el caso no lineal, la solución debe proceder de manera iterativa.

Existen técnicas expresamente diseñadas para la regresión no lineal. Una alternativa es utilizar técnicas de optimización para determinar directamente el ajuste de mínimos cuadrados. Por ejemplo, la ec. (1) se puede expresar como una función objetivo para calcular la suma de los cuadrados

$$f(a_0, a_1) = \sum_{i=1}^{n} [y_i - a_0(1 - e^{a_1 x})]^2$$

Luego se puede usar una rutina de optimización para determinar los valores de a_0 y a_1 que minimizan la función.

Observación. Para resolver el problema de optimización se puede utilizar el comando `fminsearch` de MATLAB.

DMH/VAD.