Problemas de Mínimos Cuadrados

Considere un sistema de ecuaciones

$$Ax = b \tag{1}$$

donde A es una matriz de m filas, n columnas, $b \in \mathbb{R}^m$ y $n < m$.

Este problema no tiene solución (sistema sobredeterminado), a menos que b sea muy particular ($b \in Im(A)$). Una alternativa es buscar una solución en el sentido generalizado siguiente:

$$\begin{gathered} \text{Hallar } x \in \mathbb{R}^n \text{ tal que} \\ \|b - Ax\|_2 \ \textbf{sea mínima}\,. \end{gathered}$$

Ejemplo. Ajuste de polinomios.

Dado un conjunto de puntos $(x_1, y_1), \ldots, (x_m, y_m)$ nos proponemos encontrar el polinomio

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

de grado $n - 1 < m$ que esté más cerca de estos puntos en el sentido que

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

sea mínima.

Vemos que esta suma de cuadrados es la norma del residuo al cuadrado del sistema rectangular

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

Cálculo de la solución

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

$$A^t r = \theta \,,$$

donde A^t es la matriz transpuesta de A.
Consecuencia: x debe satisfacer

$$\begin{aligned} A^t r = \theta \iff & A^t(b - Ax) = 0 \\ \iff & A^t A x = A^t b \,. \end{aligned} \tag{2}$$

Las ecuaciones (2) reciben el nombre de ecuaciones normales.
Observación : En el caso en que $m = n$ y que la matriz A sea una matriz no singular, entonces las ecuaciones normales entregan como solución la solución del sistema lineal $Ax = b$.
Las ecuaciones normales sólo son únicamente solubles si A es de rango completo, es decir, si $rang(A) = n$. En este caso además, la matriz A^tA es simétrica y definida positiva, de donde, las ecuaciones normales son únicamente solubles y se pueden utilizar métodos rápidos (Cholesky, Gauss-Seidel, Gradiente Conjugado).
Pseudoinversa Si A es de rango completo, entonces la solución de las ecuaciones normales está dada por

$$x = (A^tA)^{-1}A^tb \,.$$

A la matriz

$$A^+ := (A^tA)^{-1}A^t \,,$$

se le conoce como matriz pseudoinversa de A, tiene n filas y m columnas.
Para resolver las ecuaciones normales se puede dar el siguiente procedimiento:

1. Calcular la matriz A^tA y el vector A^tb.
2. Calcular la factorización de Cholesky: $A^tA = LL^t$.
3. Resolver el sistema triangular inferior $L^tu = A^tb$.
4. Resolver el sistema triangular superior $Lx = u$.

Problema: El condicionamiento de la matriz A^tA es en general muy malo, lo que genera grandes sensibilidades a errores de redondeo.
Solución: Utilizar un método de factorización QR. Una posibilidad es utilizar Gram-Schmidt. Para esto, escribamos:

$$A = \left[\ a_1 \,|\, a_2 \,|\, \dots \,|\, a_n\ \right] ,$$

donde $a_i \in \mathbb{R}^m$ para $i = 1, 2, \dots, n$, y la idea es construir una matriz

$$Q = \left[\ q_1 \,|\, q_2 \,|\, \dots \,|\, q_n\ \right] ,$$

y una matriz R triangular superior tales que $A = QR$.

Basándonos en el proceso de ortogonalización de Gram-Schmidt Q y R se pueden construir a partir de:

$$\begin{aligned} a_1 &= r_{11}q_1 \,,\; r_{11} \neq 0 \\ a_2 &= r_{12}q_1 + r_{22}q_2 \\ \vdots & \\ a_n &= r_{1n}q_1 + r_{2n}q_2 + \dots + r_{nn}q_n \,, \end{aligned}$$

donde

$$r_{ij} = q_i \cdot a_j \;\; (\text{si } i \neq j) \quad , \quad |r_{jj}| = \|a_j - \sum_{i=1}^{j-1} r_{ij}q_i\|_2 \,.$$

Las columnas de Q satisfacen

$$q_i \cdot q_j = \delta_{ij} = \begin{cases} 1 \text{ si } i = j \\ 0 \text{ si } i \neq j \,, \end{cases}$$

de donde, construyendo las matrices

$$Q = \left[\ q_1 \,|\, q_2 \,|\, \dots \,|\, q_n\ \right] ,\; R = \begin{bmatrix} r_{11} & r_{12} & \dots & r_{1n} \\ 0 & r_{22} & \dots & r_{2n} \\ \vdots & & & \vdots \\ 0 & \dots & 0 & r_{nn} \end{bmatrix} ,$$

se tiene que R es no singular y

$$A = QR \quad \text{y} \quad Q^tQ = \begin{bmatrix} q_1^t \\ \vdots \\ q_n^t \end{bmatrix} \left[\ q_1 \,|\, \dots \,|\, q_n\ \right] = I \,.$$

Aplicación a la resolución de las ecuaciones normales: Queremos resolver

$$A^t A x \;=\; A^t b\,.$$

Vemos que como $A = QR$, $Q^t Q = I$ y R es no singular, entonces

$$\begin{aligned} A^t A x \;=\; A^t b \;&\Longleftrightarrow\; R^t Q^t Q R x \;=\; R^t Q^t b \\ &\Longleftrightarrow\; R^t R x \;=\; R^t Q^t b \\ &\Longleftrightarrow\; R x \;=\; Q^t b\,. \end{aligned}$$

Ejemplo 1. Un problema de aproximación Polinomial.

Considere la siguiente tabla de valores:

x	0	0.5	1.0	1.5	2.0	2.5	3.0	3.5	4.0	4.5	5.0
y	10.5	5.4844	0	-3.6094	-4.5	-2.9531	0	2.9531	4.5	3.6094	0

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

$$p(x) = a\,x^3 + b\,x^2 + c\,x + d.$$

Evaluemos el polinomio p en los diferentes valores de x, obteniendo así el sistema lineal para a, b, c y d:

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

es decir,

$$A\,X = Y.$$

Sistema sobredeterminado. Para resolverlo multiplicamos por A^t obteniendo las ecuaciones normales

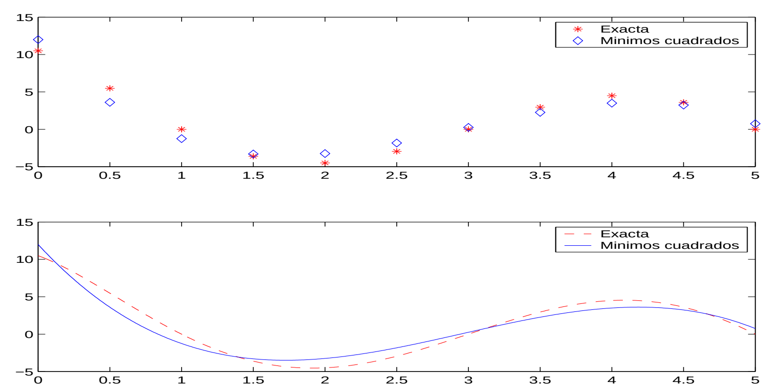
$$\underbrace{10^4 \begin{bmatrix} 3{,}0913 & 0{,}6901 & 0{,}1583 & 0{,}0378 \\ 0{,}6901 & 0{,}1583 & 0{,}0378 & 0{,}0096 \\ 0{,}1583 & 0{,}0378 & 0{,}0096 & 0{,}0027 \\ 0{,}0378 & 0{,}0096 & 0{,}0027 & 0{,}0011 \end{bmatrix}}_{A^tA} \underbrace{\begin{bmatrix} a \\ b \\ c \\ d \end{bmatrix}}_{X} = \underbrace{\begin{bmatrix} 649{,}8809 \\ 138{,}0586 \\ 25{,}5234 \\ 15{,}9844 \end{bmatrix}}_{A^tY}$$

De aquí, el polinomio es:

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

Los datos corresponden al polinomio original

$$q(x) = 0{,}1x^5 - 1{,}5x^4 + 7x^3 - 9x^2 - 7{,}1x + 10{,}5$$

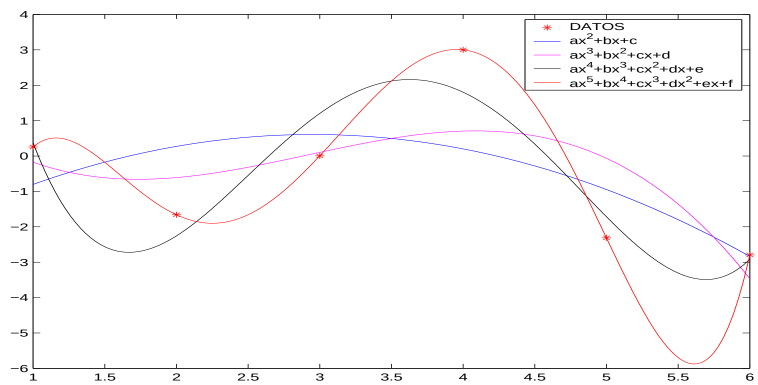


Ejemplo 2.

Considere la siguiente tabla de datos:

x	1	2	3	4	5	6
y	0.2557	-1.6591	0.0056	2.9998	-2.3120	-2.7953

La siguiente figura muestra el gráfico de los polinomios de aproximación en el sentido de los mínimos cuadrados de la tabla por un polinomio de grado 2, 3, 4 y 5.



Ejemplo 3. Un problema nolineal reducible a lineal.

Considere la siguiente tabla de valores

x	0.0000	0.4000	0.8000	1.2000	1.6000	2.0000
y	1.0314	1.0905	1.2608	1.7516	3.1657	7.2405

Se pide ajustar estos datos en el sentido de mínimos cuadrados a una función $f(x) = ae^{bx}$.

x	0.0000	0.4000	0.8000	1.2000	1.6000	2.0000
$z = \ln(y)$	0.0309	0.0867	0.2318	0.5605	1.1524	1.9797

$$z = \ln(a) + bx$$

La matriz

$$A = \begin{bmatrix} 1 & 0 \\ 1 & 0{,}4000 \\ 1 & 0{,}8000 \\ 1 & 1{,}2000 \\ 1 & 1{,}6000 \\ 1 & 2{,}0000 \end{bmatrix}$$

Resolvemos el sistema

$$(A^t A)X = A^t Y$$

Así el polinomio es:

$$p(x) = 0{,}9478x - 0{,}2742$$

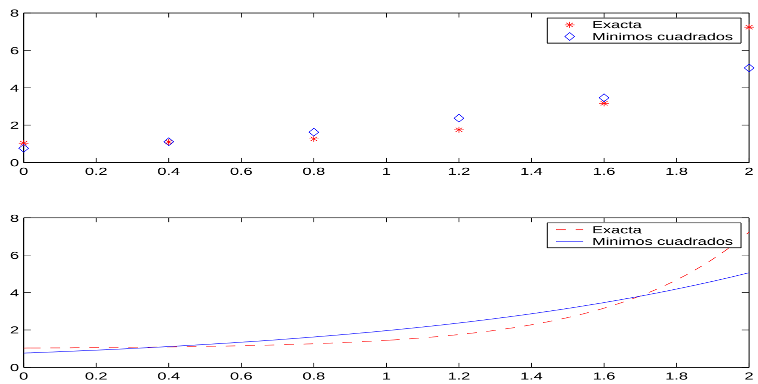
Aplicando exponencial

$$f(x) = 0{,}7602e^{0{,}9478x}$$

La tabla con los valores aproximados antes de aplicar exponencial queda:

x	0.0000	0.4000	0.8000	1.2000	1.6000	2.0000
y	0.7602	1.1107	1.6227	2.3708	3.4638	5.0607

Obs.
En este caso la función original es $f(x) = \dfrac{\pi}{100}e^{\sqrt{7}x} + 1$.



Otros ejemplos de modelos no lineales reducibles a lineales

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