Tema: Fecha: / 1 /

# Mínimos cuadrados

El método de los mínimos cuadrados puede entenderse en dos contextos distintos:

a) Resolver un sistema de ecuaciones

$$Ax = b$$

cuando el sistema es inconsistente.

b) Reemplazar la interpolación cuando el número de puntos es muy grande y se torna ineficaz usar polinomios para interpolar los puntos. Alternativamente, si existen errores al precisar $y_1, \dots, y_n$ no tiene mucho sentido exigir que la función utilizada atraviese todos los puntos.

## Sistemas de ecuaciones inconsistentes

Consideremos el sistema

$$x_1 + x_2 = 2$$

$$x_1 - x_2 = 1$$

$$x_1 + x_2 = 3$$

No existe una solución que satisfaga el sistema, pues la primera y la última ecuación son inconsistentes. Este caso es típico cuando se tienen más ecuaciones que incógnitas. En la práctica este tipo de sistemas surge cuando existen imprecisiones al determinar los coeficientes. Notemos que no es posible usar la eliminación de Gauss para resolver esta clase de sistemas.

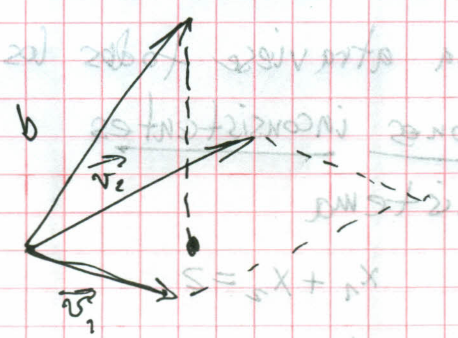
Es posible interpretar la situación desde un punto de vista geométrico. Consideremos la representación matricial

$$\overbrace{\begin{pmatrix} 1 & 1 \\ 1 & -1 \\ 1 & 1 \end{pmatrix}}^{A} \overbrace{\begin{pmatrix} x_1 \\ x_2 \end{pmatrix}}^{x} = \overbrace{\begin{pmatrix} 2 \\ 1 \\ 3 \end{pmatrix}}^{b}$$

que es equivalente a

$$x_1 \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix} + x_2 \begin{pmatrix} 1 \\ -1 \\ 1 \end{pmatrix} = \begin{pmatrix} 2 \\ 1 \\ 3 \end{pmatrix}$$

es decir, que tratamos de escribir b como una combinación lineal de los vectores columna de A. En este caso particular sólo es posible si b está en el mismo plano que los vectores columna, pero no lo está como se ilustra en la figura




El método de mínimos cuadrados consiste en encontrar el punto sobre el plano que se encuentra más cerca a b. Este punto especial, representado por $A\bar{x}$, tiene la característica que su vector residuo $r = b - A\bar{x}$ es perpendicular al plano $Ax$ para todo $x \in \mathbb{R}^n$. Es decir que

$$(b - A\bar{x}) \perp \{Ax \mid x \in \mathbb{R}^n\}$$

ó, de manera equivalente

$$(Ax)^T (b - A\bar{x}) = 0 \quad \text{para todo } x \in \mathbb{R}^n$$

que se puede escribir como

$$x^T A^T (b - A\bar{x}) = 0 \quad \text{para todo } x \in \mathbb{R}^n$$

que implica que $A^T(b-A\bar{x})$ es perpendicular a todo vector $x$ en $\mathbb{R}^n$. Esto sólo puede ocurrir si

$$A^T(b-A\bar{x})=0$$

ó

$$A^TA\bar{x}=A^Tb$$

Ejemplo Resuelva el sistema

$$\begin{pmatrix}1 & 1\\ 1 & -1\\ 1 & 1\end{pmatrix}\begin{pmatrix}x_1\\ x_2\end{pmatrix}=\begin{pmatrix}2\\ 1\\ 3\end{pmatrix}$$

usando el método de mínimos cuadrados R.- $\bar{x}^T=(7/4, 3/4)$.

Una medida de la calidad de la solución es el vector de residuos

$$r=b-A\bar{x}$$

Existen algunas maneras de expresar el tamaño de $r$

a) $\|r\|_2=\sqrt{\sum_{i=1}^{n} r_i^2}$ que es la norma habitual

b) El error cuadrático $r_1^2+\dots+r_n^2=SE$

c) La raíz del error cuadrático medio $RMSE=\sqrt{\frac{1}{n}\sum_{i=1}^{n} r_i^2}$

Los tres están relacionados

$$RMSE=\frac{\sqrt{SE}}{\sqrt{n}}=\frac{\|r\|}{\sqrt{m}}$$

Ajuste de modelos a datos

Si disponemos de un conjunto de m puntos podemos aplicar el método para ajustar un modelo lineal, dado a priori, a nuestro conjunto de datos. El proceso lo podemos sintetizar como

a) Elegir un modelo lineal en los parámetros, como por ejemplo $y=c_1+c_2t$

Tema: Fecha: / /

b) Forzar el modelo a ajustarse a los datos, de manera que se obtenga un sistema del tipo

$$Ax = b$$

c) Resolver el sistema

$$A^T A x = A^T b$$

para encontrar $\bar{x}$.

Ejemplo Encuentre la parábola que mejor se ajusta a los puntos $(-1,1)$, $(0,0)$, $(1,0)$, $(2,-2)$ usando el método de los mínimos cuadrados. R. $y = 0{,}45 - 0{,}65t - 0{,}25t^2$.

$r = (0{,}15; -0{,}45; 0{,}45; -0{,}15)^T$

El número condicionante

El método consiste en resolver

$$A^T A \bar{x} = A^T b$$

¿con qué precisión lo podemos hacer? depende del condicionante de la matriz $A^T A$. Como vimos anteriormente, el condicionante limita superiormente el error al resolver un sistema de ecuaciones. Si el condicionante es grande, puede que el error también lo sea.

Algunos modelos

Es posible ajustar cualquier modelo que sea lineal en los parámetros e incluso adaptar, mediante transformaciones, algunos modelos no lineales.

Por ejemplo, imaginemos que deseamos ajustar el modelo

$$y = c_1 + c_2 \cos 2\pi t + c_3 \operatorname{sen} 2\pi t \qquad (+)$$

a un conjunto de datos. Para esto, basta identificar nuestro modelo ~~se~~ con el usado ~~en~~ en el método de mínimos cuadrados visto anteriormente. $Ax = b$. Para hacer esto podemos definir

$$A = \begin{pmatrix} 1 & \cos 2\pi t_1 & \operatorname{sen} 2\pi t_1 \\ 1 & \cos 2\pi t_2 & \operatorname{sen} 2\pi t_2 \\ \vdots & \vdots & \\ 1 & \cos 2\pi t_n & \operatorname{sen} 2\pi t_n \end{pmatrix}, \quad x = \begin{pmatrix} c_1 \\ c_2 \\ c_3 \end{pmatrix}, \quad b = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix}$$

y no es difícil notar que $Ax = b$ es la evaluación del modelo original (+) al conjunto de datos $(t_1, y_1), (t_2, y_2), \ldots, (t_n, y_n)$.

Por otro lado, modelos no lineales como

$$y = c_1 e^{c_2 t}$$

ó

$$y = c_1 t^{c_2}$$

es posible linealizarlos haciendo algunas transformaciones. Particularmente, podemos aplicar el logaritmo para obtener

$$\ln y = \ln c_1 + c_2 t$$

y

$$\ln y = \ln c_1 + c_2 \ln t.$$

Sin embargo notemos que, a pesar de que podemos linealizar estos modelos específicos, las estimas de algunos de los coeficientes no son los coeficientes en sí, sino funciones de los coeficientes (p.e. $\ln c_1$ en lugar de $c_1$). Es decir que la suma de residuos que se minimiza es una para $y = c_1 e^{c_2 t}$ y otra distinta para $y = \ln c_1 + c_2 t$. Esto puede tener repercusiones en la estimas de $c_1$, en este caso específico.

Para cuantificar la calidad del ajuste podemos hacerlo en ~~el~~ el modelo transformado o en el modelo original. Generalmente se hace esta cuantificación con respecto al modelo original.

Factorización QR

En algunos casos, utilizar las ecuaciones normales de manera directa puede introducir errores muy grandes en el resultado. Esto puede ocurrir cuando

$$\mathrm{cond}(A^TA) >> 1$$

Como alternativa a las ecuaciones normales veremos la denominada factorización QR.

La ortogonalización de Gram-Schmidt es útil para este propósito.

Recordemos que el método de Gram-Schmidt ortogonaliza un conjunto de vectores. Dado un conjunto de m vectores linealmente independientes, el método calcula m vectores mutuamente perpendiculares y unitarios que forman una base de $\mathbb{R}^m$.

Si $A_1, \dots, A_n$ son vectores linealmente independientes en $\mathbb{R}^m$ con $n \leq m$ el método de Gram-Schmidt consiste en determinar

$$y_j = A_j - \sum_{i=1}^{j-1} q_i q_i^T A_j$$

con $\quad q_j = \frac{y_j}{\|y_j\|} \quad$ para $j = 1, \dots, m$

Si definimos

$$r_{jj} = \|y_j\|, \quad r_{ij} = q_i^T A_j$$

entonces podemos escribir

$$A_j = r_{1j} q_1 + \dots + r_{j-1,j} q_{j-1} + r_{jj} q_j = \sum_{i=1}^{j} r_{ij} q_i$$

de manera que

$$(A_1 | \dots | A_n) = (q_1 | \dots | q_n) \begin{pmatrix} r_{11} & r_{12} & \cdots & r_{1n} \\ & r_{22} & \cdots & r_{2n} \\ & & \ddots & \vdots \\ 0 & & & r_{nn} \end{pmatrix}$$

ó $\quad A = QR$, esto se conoce como la factorización QR re ducida.

Ejemplo Encuentre la factorización QR de la matriz $\begin{pmatrix} 1 & -4 \\ 2 & 3 \\ 2 & 2 \end{pmatrix}$

R. $\begin{pmatrix} 1/3 & -14/15 \\ 2/3 & 1/3 \\ 2/3 & 2/15 \end{pmatrix} \begin{pmatrix} 3 & 2 \\ 0 & 5 \end{pmatrix}$.

El algoritmo puede escribirse como

for $j = 1, \ldots, n$

  $y = A_j$

  for $i = 1, \ldots, j-1$

    $r_{ij} = q_i^T A_j$

    $y = y - r_{ij} q_i$

  $r_{jj} = \|y\|$

  $q_j = y / r_{jj}$

y si el algoritmo tiene éxito se aplica la factorización QR completa.

$$(A_1 | \ldots | A_n) = (q_1 | \ldots | q_m) \begin{pmatrix} r_{11} & \cdots & r_{1n} \\ & \ddots & \vdots \\ & & r_{nn} \\ 0 & \cdots & 0 \\ \vdots & & \vdots \\ 0 & & 0 \end{pmatrix}$$

donde $A_1, \ldots, A_n$ son los vectores originales y se han añadido $m-n$ vector a $Q$ para poder tener una base que genere $\mathbb{R}^m$. Notemos que $A$ es $m \times n$, $Q$ es $m \times m$ y $R$ es $m \times n$.

Ejemplo Encuentre la factorización completa de $\begin{pmatrix} 1 & -4 \\ 2 & 3 \\ 2 & 2 \end{pmatrix}$.

R. $\begin{pmatrix} 1/3 & -14/15 & 2/15 \\ 2/3 & 1/3 & 2/3 \\ 2/3 & 2/15 & -11/15 \end{pmatrix} \begin{pmatrix} 3 & 2 \\ 0 & 5 \\ 0 & 0 \end{pmatrix}$

Una de las aplicaciones más importantes de la factorización QR es al problema de mínimos cuadrados.

Si lo que se desea es minimizar $\|Ax-b\|$ entonces podemos reescribirlo como

$$\|Ax-b\| = \|QRx-b\| = \|Rx - Q^T b\|$$

$\Downarrow$

Q es ortogonal

recordando que A es $m \times n$ con $m \geq n$. El hecho de que Q es ortogonal se puede demostrar fácilmente si recordamos que los vectores columna son unitarios y ortogonales entre sí.

Notemos que los residuos son $e = Ax - b$, así que

$$\begin{pmatrix} e_1 \\ \vdots \\ e_n \\ \hline e_{n+1} \\ \vdots \\ e_m \end{pmatrix} = \begin{pmatrix} r_{11} & r_{12} & \cdots & r_{1n} \\ & r_{22} & \cdots & r_{2n} \\ & & \ddots & \vdots \\ & & & r_{nn} \\ \hline 0 & \cdots & \cdots & 0 \\ \vdots & & & \vdots \\ 0 & \cdots & \cdots & 0 \end{pmatrix} \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} - \begin{pmatrix} d_1 \\ \vdots \\ d_n \\ \hline d_{n+1} \\ \vdots \\ d_m \end{pmatrix}$$

donde $d = Q^T b$. Si $r_{ii} \neq 0$ entonces la parte superior de e $(e_1, \ldots, e_n)$ se puede hacer 0 por sustitución sucesiva (recordando el método de la eliminación de Gauss), así que

$$\|e\|^2 = d_{n+1}^2 + \ldots + d_m^2$$

Por tanto, la factorización QR aplicada a los mínimos cuadrados da para un sistema inconsistente

$$Ax = b$$

para luego

$$QRx = b$$

y si se hace $R'$ = matriz superior $n \times n$ de R y $d'$ = n componentes superiores de $d = Q^T b$, entonces la solución de mínimos cuadrados está dada por la solución del sistema

$$R'x = d'.$$

Ejemplo Resolver el sistema $\begin{pmatrix} 1 & -4 \\ 2 & 3 \\ 2 & 2 \end{pmatrix}\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} -3 \\ 15 \\ 9 \end{pmatrix}$ usando el método de mínimos cuadrados y la factorización QR. R.- $\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} 3,8 \\ 1,8 \end{pmatrix}$.

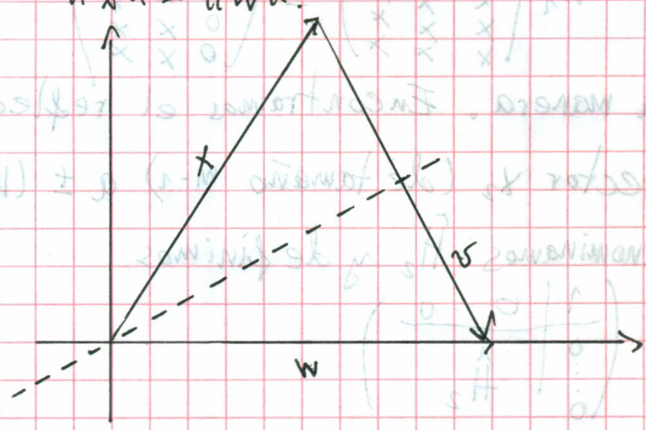
## Los reflectores de Householder

Existe um método alternativo para calcular la factorización QR de una matriz. El método utiliza los reflectores de Householder y requiere menos operaciones que el método de Gram-Schmidt, además de ser más estable.

Un reflector de Householder es una matriz ortogonal que refleja vectores m-dimensionales a través de un plano en m-1 dimensiones. La transformación

$$Hx = w$$

produce un vector reflejado de x a través de un plano. Notemos que $\|x\| = \|w\|$.




* Asumamos que dos vectores cumplen con $\|x\| = \|w\|$. Entonces $w-x$ y $w+x$ son perpendiculares.

Si definimos el vector $v = w - x$ y definimos la matriz

$$P = \frac{vv^T}{v^T v}$$

vemos que $P^2 = P$, $P = P^T$ y que $Pv = v$. Notemos que $Pu$ es la proyección de $u$ sobre $v$.

La matriz
$$H = I - 2P$$
se conoce como el reflector de Householder. Notemos que
$$Hx = w \quad y \quad Hw = x.$$

Ejemplo Para $x = \begin{pmatrix} 3 \\ 4 \end{pmatrix}$ y $w = \begin{pmatrix} 5 \\ 0 \end{pmatrix}$ encuentre un reflector que satisfaga $Hx = w$. R. $H = \begin{pmatrix} 0{,}6 & 0{,}8 \\ 0{,}8 & -0{,}6 \end{pmatrix}$.

Es posible aplicar los reflectores de Householder para encontrar la factorización QR.

Para factorizar A como QR elegimos $x_1$ como la primera columna de A. Si hacemos $w = \pm(\|x_1\|, 0, \dots, 0)$ podemos hacer un reflector de Householder tal que
$$H_1 x_1 = w$$
así que si hacemos $H_1 A$ obtenemos
$$H_1 A = H_1 \begin{pmatrix} x & x & x \\ x & x & x \\ x & x & x \\ x & x & x \end{pmatrix} = \begin{pmatrix} x & x & x \\ 0 & x & x \\ 0 & x & x \\ 0 & x & x \end{pmatrix}$$
y podemos proceder de esta manera. Encontramos el reflector de Householder que mueve el vector $x_2$ (de tamaño $m-1$) a $\pm(\|x_2\|, 0, \dots, 0)$, a dicho reflector lo denominamos $\hat{H}_2$ y definimos
$$H_2 = \left(\begin{array}{c|c} 1 & 0 \cdots 0 \\ \hline 0 & \\ \vdots & \hat{H}_2 \\ 0 & \end{array}\right)$$
de forma que
$$H_2 H_1 A = \begin{pmatrix} x & x & x \\ 0 & x & x \\ 0 & 0 & x \\ 0 & 0 & x \end{pmatrix}$$
Luego, $H_3$ sería
$$H_3 = \begin{pmatrix} 1 & 0 & 0 & 0 & \cdots & 0 \\ 0 & 1 & 0 & 0 & \cdots & 0 \\ & 0 & & & & \\ 0 & \vdots & & \hat{H}_3 & & \\ & 0 & & & & \end{pmatrix}$$
y, una vez determinados los reflectores de Householder, tendría

mos que

$$H_n \cdots H_2 \cdot H_1 A = R$$

donde R es una matriz triangular superior. Si multiplicamos por la izquierda por las inversas de los reflectores de Householder tenemos

$$A = H_1 H_2 \cdots H_n R = QR$$

con $Q = H_1 H_2 \cdots H_n$. Notemos que $H_i = H_i^{-1}$

<u>Ejemplo</u> Encuentre la factorización QR de $\begin{pmatrix} 1 & -4 \\ 2 & 3 \\ 2 & 2 \end{pmatrix}$ usando los reflectores de Householder. R.- $\begin{pmatrix} 1/3 & -14/15 & -2/15 \\ 2/3 & 1/3 & -2/3 \\ 2/3 & 2/15 & 11/15 \end{pmatrix} \begin{pmatrix} 3 & 2 \\ 0 & 5 \\ 0 & 0 \end{pmatrix}$