

3.13 [herramienta] Sean $(\mathbb{R}^n, \langle \cdot, \cdot \rangle)$ y $(\mathbb{R}^m, \langle \cdot, \cdot \rangle)$ los $\mathbb{R}$ -espacios euclídeos canónicos de dimensiones n y m , respectivamente, sea $A \in \mathbb{R}^{m \times n}$ y $b \in \mathbb{R}^m$. Por definición, resolver la ecuación $Ax = b$ por mínimos cuadrados significa determinar el conjunto $\arg \min_{x \in \mathbb{R}^n} \|b - Ax\|$ de todos los $x \in \mathbb{R}^n$ cuyas imágenes por A minimizan la distancia al vector b .

(a) Explicar por qué

$$\arg \min_{x \in \mathbb{R}^n} \|b - Ax\| = \{x \in \mathbb{R}^n : Ax = P_{\text{col}(A)}(b)\}.$$

(b) Observar que $\{x \in \mathbb{R}^n : Ax = P_{\text{col}(A)}(b)\} \neq \emptyset$. Motivo por el cual para cualquier $b \in \mathbb{R}^m$ existe al menos una solución por mínimos cuadrados de la ecuación $Ax = b$.

(c) Verificar que $\text{nul}(A^T) = \text{col}(A)^\perp$.

(d) Utilizando que $b - P_{\text{col}(A)}(b) \perp \text{col}(A)$, concluir que

$$\arg \min_{x \in \mathbb{R}^n} \|b - Ax\| = \{x \in \mathbb{R}^n : A^T Ax = A^T b\}.$$

(e) Verificar que $\text{nul}(A) = \text{fil}(A)^\perp$.

(f) Utilizando que todo $x \in \mathbb{R}^n$ se descompone de manera única en la forma

$$x = x_f + x_h, \text{ con } x_f \in \text{fil}(A), \text{ y } x_h \in \text{nul}(A),$$

deducir que la aplicación $x_f \mapsto Ax_f$ es una biyección de $\text{fil}(A)$ en $\text{col}(A)$.

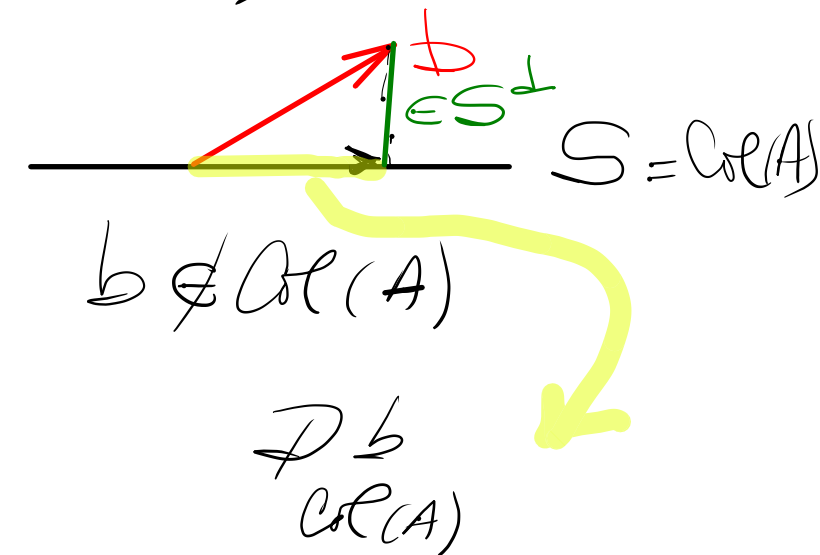
(g) Concluir que para cada $b \in \mathbb{R}^m$ existe un único $x_f(b) \in \text{fil}(A)$ tal que

$$\arg \min_{x \in \mathbb{R}^n} \|b - Ax\| = \{x_f(b) + x_h : x_h \in \text{nul}(A)\}.$$

C. Mínimos
PIC

$Ax \in \text{Col}(A)$
 Ax es una CL de las columnas de A

Si $b \notin \text{Col}(A) \rightarrow$
Sistema incompatible
 $\Rightarrow$ M.C. Mínimos



$$Ax = P_{\text{Col}(A)} b$$

Ax es CL de las columnas de A
Si $\{C_1, C_2, \dots, C_n\}$ es LI $\Rightarrow x$ es único
columnas de A

(c) Verificar que $\text{nul}(A^T) = \text{col}(A)^\perp$.

(d) Utilizando que $b - P_{\text{col}(A)}(b) \perp \text{col}(A)$, concluir que

$$\arg \min_{x \in \mathbb{R}^n} \|b - Ax\| = \{x \in \mathbb{R}^n : A^T Ax = A^T b\}.$$

(e) Verificar que $\text{nul}(A) = \text{fil}(A)^\perp$.

(f) Utilizando que todo $x \in \mathbb{R}^n$ se descompone de manera única en la forma

$$x = x_f + x_h, \text{ con } x_f \in \text{fil}(A), \text{ y } x_h \in \text{nul}(A),$$

deducir que la aplicación $x_f \mapsto Ax_f$ es una biyección de $\text{fil}(A)$ en $\text{col}(A)$.

(g) Concluir que para cada $b \in \mathbb{R}^m$ existe un único $x_f(b) \in \text{fil}(A)$ tal que

$$\arg \min_{x \in \mathbb{R}^n} \|b - Ax\| = \{x_f(b) + x_h : x_h \in \text{nul}(A)\}.$$

$$A = \begin{pmatrix} | & | & & | \\ C_1 & C_2 & \dots & C_n \\ | & | & & | \end{pmatrix}$$

$$C_i \in \mathbb{R}^m$$

$$A \in \mathbb{R}^{m \times n}$$

$$A x_i = \text{es una CL de } C_i \\ \Rightarrow \in \text{Col}(A)$$

$$A^T \in \mathbb{R}^{n \times m} = \begin{pmatrix} -C_1- \\ -C_2- \\ -C_n- \end{pmatrix}$$
$$A^T v = \begin{pmatrix} \langle v, C_1 \rangle \\ \langle v, C_2 \rangle \\ \vdots \\ \langle v, C_n \rangle \end{pmatrix} = 0$$

$$\Rightarrow v \in (\text{Col}(A))^\perp$$

$$\text{de } A \Rightarrow v \in \text{nul}(A^T)$$

$$\text{nul}(A^T) = \left\{ v : A^T v = 0 \right\}$$
$$\downarrow$$
$$\in \mathbb{R}^m$$

$$\Rightarrow \langle v, C_i \rangle = 0 \Rightarrow v \perp C_i$$
$$1 \leq i \leq n \quad \forall C_i$$

$$\text{Si } v \in (\text{Col}(A))^\perp \Rightarrow v \perp \text{a todo } C_i$$

$$\text{nul}(A^T) = (\text{Col}(A))^\perp$$

$$b - A\hat{x} = b - \underset{\text{Col}(A)}{P} b \in (\text{Col}(A))^\perp$$

$$\Rightarrow b - A\hat{x} \in \text{Nul}(A^T) \Rightarrow A^T(b - A\hat{x}) = 0 \Rightarrow$$

$$\boxed{A^T A \hat{x} = A^T b} \quad \text{Este sistema siempre es compatible}$$

$$\text{Nul}(A^T) = (\text{Col}(A))^\perp =$$

$$\text{Nul}(A^T) = (\text{Fil}(A^T))^\perp$$

$$\text{Nul}(A) = (\text{Fil}(A))^\perp$$

(f) Utilizando que todo $x \in \mathbb{R}^n$ se descompone de manera única en la forma

$$x = x_f + x_h, \text{ con } x_f \in \text{fil}(A), \text{ y } x_h \in \text{nul}(A),$$

deducir que la aplicación $x_f \mapsto Ax_f$ es una biyección de $\text{fil}(A)$ en $\text{col}(A)$.

(g) Concluir que para cada $b \in \mathbb{R}^m$ existe un único $x_f(b) \in \text{fil}(A)$ tal que

$$\arg \min_{x \in \mathbb{R}^n} \|b - Ax\| = \{x_f(b) + x_h : x_h \in \text{nul}(A)\}.$$

$$Ax = 0$$

$$\begin{pmatrix} -F_1 - \\ \vdots \\ -F_m - \end{pmatrix} (x) = \begin{pmatrix} \langle F_1, x \rangle \\ \vdots \\ \langle F_m, x \rangle \end{pmatrix} = 0 \quad x \perp F_i \quad 1 \leq i \leq m \quad x \in \text{Nul}(A)$$

$$A \in \mathbb{R}^{m \times n}$$

$$\text{Col}(A) \subseteq \mathbb{R}^m \quad \text{Fil}(A) \subseteq \mathbb{R}^n$$

$$\text{Nul}(A) \subseteq \mathbb{R}^n$$

$$\text{Fil}(A)^\perp = \text{Nul}(A)$$

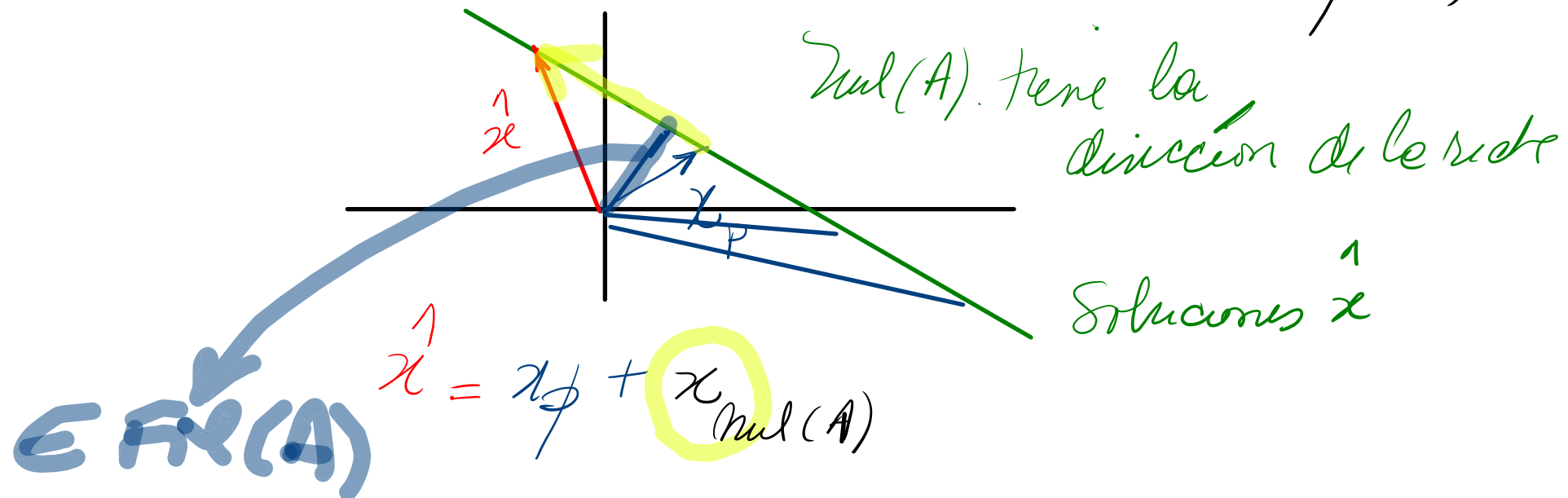
$$A \hat{x} = A (x_{\text{Fil}(A)} + x_{\text{Nul}(A)}) = A x_{\text{Fil}(A)} = \begin{matrix} P \\ \text{Col}(A) \end{matrix} b$$

$$A \hat{x}_1 = \begin{matrix} P \\ \text{Col}(A) \end{matrix} b = A \hat{x}_2 \Rightarrow A(\hat{x}_1 - \hat{x}_2) = \mathbf{0} \Rightarrow$$

$$\hat{x}_1 - \hat{x}_2 \in \text{Nul}(A) \Rightarrow \hat{x}_1 = \hat{x}_2 + x_{\text{Nul}(A)}$$

Ahora bien $\hat{x} \in \mathbb{R}^n$ $\text{Col}(A) \oplus \text{Nul}(A) \Rightarrow$

$$\hat{x} = \hat{x}_{\text{Col}(A)} + x_{\text{Nul}(A)} \text{ es } \acute{u}\text{nica} \Rightarrow \hat{x}_{\text{Fil}(A)} \text{ es } \acute{u}\text{nica}$$



Utilizando el Teorema de Pitágoras observar que de todas las soluciones por mínimos cuadrados de la ecuación $Ax = b$, la solución $x_f(b)$ es la que tiene norma mínima.

(h) Observar que cuando $\dim(\text{col}(A)) = n$ la matriz $A^T A$ es inversible y que por lo tanto, $\hat{x} = (A^T A)^{-1} A^T b$ es la única solución por mínimos cuadrados de la ecuación $Ax = b$. Como $A\hat{x} = P_{\text{col}(A)}(b)$ se deduce que $A(A^T A)^{-1} A^T$ es la matriz en base canónica de la proyección ortogonal sobre $\text{col}(A)$.

$$\hat{x} = \hat{x}_{\text{col}(A)} + \hat{x}_{\text{Nul}(A)}$$

Si la solución es única

$$\hat{x}_{\text{Nul}(A)} = \mathbf{0}$$

$$\hat{x}_{\text{col}(A)} \perp \hat{x}_{\text{Nul}(A)}$$

$\Rightarrow$

$$\hat{x}_f \quad \hat{x}_h$$

$$\|\hat{x}\|^2 = \|\hat{x}_f\|^2 + \|\hat{x}_h\|^2 \geq \|\hat{x}_f\|^2$$

$\Rightarrow \hat{x}_f$ es el de mínima norma

$\dim \text{Col}(A) = n \Rightarrow \{C_1, C_2, \dots, C_n\}$ es L.I. $\Rightarrow \hat{x}$ es único

y además $\text{rg}(A) = n$ Como $A^T A \in \mathbb{R}^{n \times n}$ y

$\text{rango}(A^T A) = \text{rango}(A) \Rightarrow A^T A$ tiene inversa

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

Si tiene inversa

$$(A^T A)^{-1} A^T A \hat{x} = (A^T A)^{-1} A^T b$$

$\Rightarrow$

$$\hat{x} = (A^T A)^{-1} A^T b$$

QTD

dim Col(A) = n

sólo si $\text{rg}(A) = n$
 $\{C_1, \dots, C_n\}$ LI

$$A \hat{x} = \underbrace{A(A^T A)^{-1} A^T}_{[P]_E^E} b = \text{Proy}_{\text{Col}(A)} b$$

Matriz que proyecta sobre $\text{Col}(A)$

$$A = \begin{pmatrix} 1 & 0 \\ 1 & 1 \\ 0 & 1 \end{pmatrix} \quad b = \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}$$

$$A^T A = \begin{pmatrix} 1 & 1 & 0 \\ 0 & 1 & 1 \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 1 & 1 \\ 0 & 1 \end{pmatrix} = \begin{pmatrix} 2 & 1 \\ 1 & 2 \end{pmatrix} \quad \text{rg}(A^T A) = 2$$

$$A^T b = \begin{pmatrix} 1 & 1 & 0 \\ 0 & 1 & 1 \end{pmatrix} \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix} = \begin{pmatrix} 2 \\ 2 \end{pmatrix}$$

tiene inversa

$$A^T A = \begin{pmatrix} 2 & 1 \\ 1 & 2 \end{pmatrix}$$

$$(A^T A)^{-1} = \begin{pmatrix} 2/3 & -1/3 \\ -1/3 & 2/3 \end{pmatrix}$$

$$\hat{x} = (A^T A)^{-1} A^T b = \begin{pmatrix} 2/3 & -1/3 \\ -1/3 & 2/3 \end{pmatrix} \begin{pmatrix} 2 \\ 2 \end{pmatrix} = \begin{pmatrix} 2/3 \\ 2/3 \end{pmatrix} = \hat{x} \text{ única}$$

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

$$\{C_1, C_2\} \subset \mathbb{R}^2$$

$$\dim \text{Col}(A) = 2$$

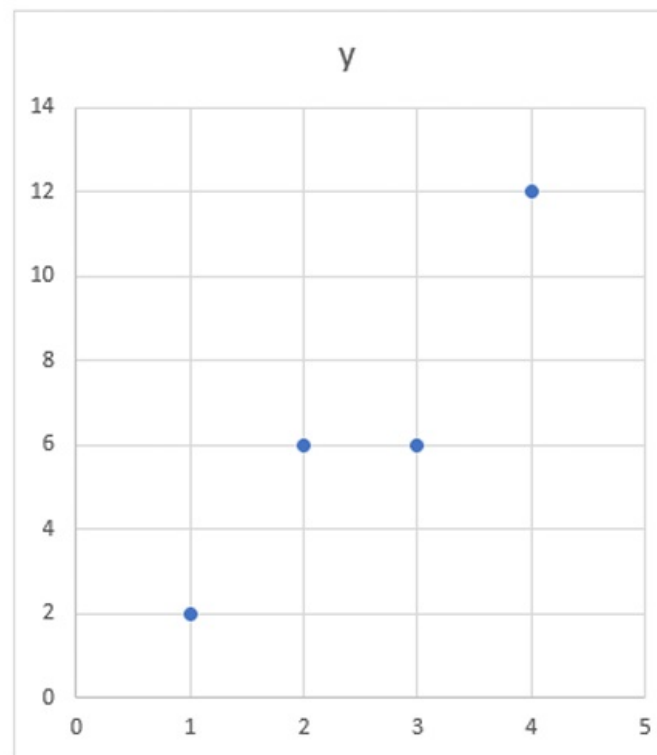
Cuadrados mínimos

Vamos a comenzar con un ejemplo sencillo.

Supongamos que tomamos distintas posiciones de un móvil y_i para distintos valores del tiempo x_i .

tiempo posición

x_i	y_i
1	2
2	6
3	6
4	12



En realidad, en una medición para determinar un modelo matemático que ajuste a cierto fenómeno físico se toman más datos, pero lo hacemos con pocos para no complicar las cuentas.

Supongamos que el móvil realiza un movimiento rectilíneo uniforme.

Entonces intentaremos ajustar con una recta de la forma:

$$y = mx + b$$

Si planteamos para cada par de valores

$$y_i = mx_i + b$$

Obtenemos el siguiente sistema de ecuaciones

$$\begin{cases} 2 = m \cdot 1 + b \\ 6 = m \cdot 2 + b \\ 6 = m \cdot 3 + b \\ 12 = m \cdot 4 + b \end{cases}$$

Matricialmente puede escribirse como:

$$\begin{pmatrix} 1 & 1 \\ 2 & 1 \\ 3 & 1 \\ 4 & 1 \end{pmatrix} \begin{pmatrix} m \\ b \end{pmatrix} = A \begin{pmatrix} m \\ b \end{pmatrix} = \begin{pmatrix} 2 \\ 6 \\ 6 \\ 12 \end{pmatrix} = v$$

A x b

Incompatible

Observemos:

- Las columnas de A forman un conjunto linealmente independiente.
- El sistema es incompatible. Es decir, no hay una recta que pase por los cuatro puntos, el vector v no pertenece al espacio columna de A .
- Desde el punto de vista del álgebra la mejor solución sería el vector $\hat{x} = \begin{pmatrix} m \\ b \end{pmatrix}$ tal que $A\hat{x}$ sea la $P_{\text{Col}(A)}(v)$.
- $\hat{x}$ es la solución por cuadrados mínimos.
- $\hat{x}$ es la solución del sistema $A\hat{x} = P_{\text{Col}(A)}(v)$.

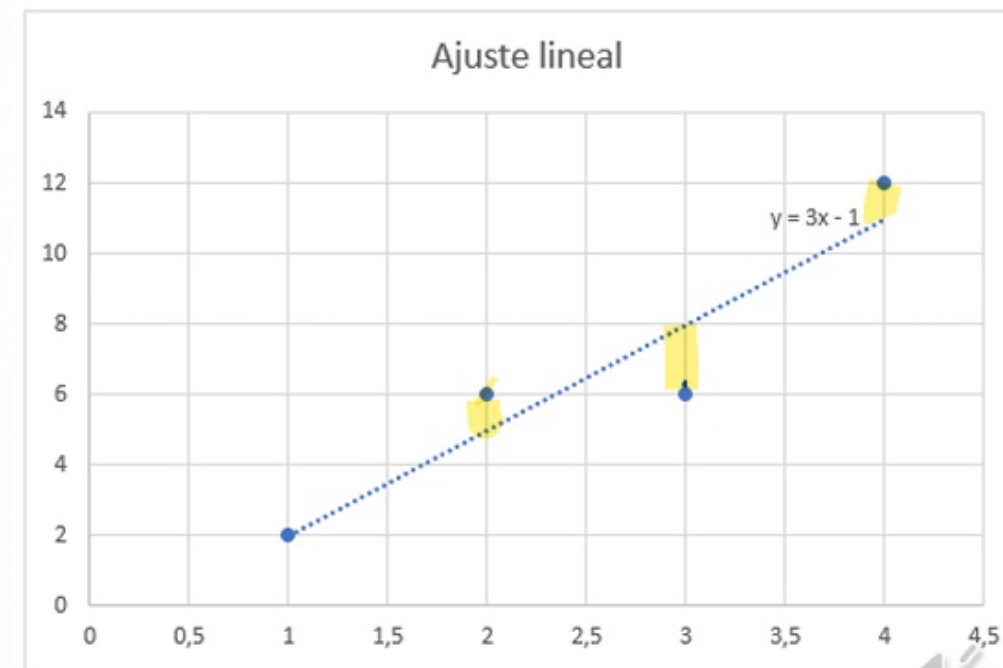
Como vimos en el ejercicio 21 resolvemos planteando
 $A^T A \hat{x} = A^T v$ en nuestro ejemplo

$$\begin{pmatrix} 1 & 2 & 3 & 4 \\ 1 & 1 & 1 & 1 \end{pmatrix} \begin{pmatrix} 1 & 1 \\ 2 & 1 \\ 3 & 1 \\ 4 & 1 \end{pmatrix} \begin{pmatrix} m \\ b \end{pmatrix} = \begin{pmatrix} 1 & 2 & 3 & 4 \\ 1 & 1 & 1 & 1 \end{pmatrix} \begin{pmatrix} 2 \\ 6 \\ 6 \\ 12 \end{pmatrix}$$

$$\begin{pmatrix} 30 & 10 \\ 10 & 4 \end{pmatrix} \begin{pmatrix} m \\ b \end{pmatrix} = \begin{pmatrix} 80 \\ 26 \end{pmatrix}$$

Cuya solución es $\hat{x} = \begin{pmatrix} 3 \\ -1 \end{pmatrix}$ es decir que la recta
 que mejor ajusta es:

$$y = 3x - 1$$



Veamos cual es el error

$$\hat{y}_i = 3x_i - 1$$

x_i	y_i	$\hat{y}_i$	$\varepsilon_i = \hat{y}_i - y_i$
1	2	2	0
2	6	5	-1
3	6	8	2
4	12	11	-1

$$\varepsilon = \begin{pmatrix} 0 \\ -1 \\ 2 \\ -1 \end{pmatrix} \quad y \|\varepsilon\| = \sqrt{6}$$

Veamos ahora lo siguiente:

$$y_i + \varepsilon_i = mx_i + b = \hat{y}_i$$

La idea es minimizar el error, pero si sumamos los errores vemos que pueden compensarse (eso justamente pasa en este ejemplo), entonces podríamos pensar en minimizar la suma de los módulos que es equivalente a minimizar la suma de los módulos al cuadrado. Es decir, lo que vamos a minimizar es la norma al cuadrado del vector ε .

$$\sum_{i=1}^n \varepsilon_i^2 = \sum_{i=1}^n (\hat{y}_i - y_i)^2 = \sum_{i=1}^n (mx_i + b - y_i)^2 = f(m, b)$$

Obtenemos una función en dos variables. Tenemos que encontrar m y b que minimicen la f .

Como $f \geq 0$ el punto donde $\nabla f = 0$ nos proporcionará un mínimo.

$$f_m = 2 \sum_{i=1}^n (mx_i + b - y_i)x_i = 0$$

$$\sum_{i=1}^n (mx_i^2 + bx_i - y_i x_i) = 0$$

$$m \sum_{i=1}^n x_i^2 + b \sum_{i=1}^n x_i = \sum_{i=1}^n x_i y_i \quad (1)$$

$$f_b = 2 \sum_{i=1}^n (mx_i + b - y_i) = 0$$

$$m \sum_{i=1}^n x_i + nb = \sum_{i=1}^n y_i \quad (2)$$

(1) y (2) se pueden escribir matricialmente de la siguiente manera:

$$\begin{pmatrix} \sum_{i=1}^n x_i^2 & \sum_{i=1}^n x_i \\ \sum_{i=1}^n x_i & n \end{pmatrix} \begin{pmatrix} m \\ b \end{pmatrix} = \begin{pmatrix} \sum_{i=1}^n x_i y_i \\ \sum_{i=1}^n y_i \end{pmatrix}$$

Si tomamos los datos del ejemplo nos queda:

$\begin{pmatrix} 30 & 10 \\ 10 & 4 \end{pmatrix} \begin{pmatrix} m \\ b \end{pmatrix} = \begin{pmatrix} 80 \\ 26 \end{pmatrix}$ *el mismo sistema compatible que obtuvimos haciendo el análisis desde el álgebra.*

En realidad, se trata de minimizar el error cuadrático medio

$$\frac{1}{n} \sum_{i=1}^n \varepsilon_i^2 = \frac{6}{4} = \frac{3}{2}$$

que equivale a minimizar el error cuadrático. En nuestro ejemplo sería $\frac{\sqrt{6}}{4}$.

El objetivo del investigador es hallar $y = p_m(x)$ de modo de minimizar:

$$\frac{1}{n} \sum_{i=1}^n \varepsilon_i^2 = \sum_{i=1}^n (\hat{y}_i - y_i)^2 = \frac{1}{n} \sum_{i=1}^n (p_m(x_i) - y_i)^2$$

Volviendo a nuestro ejemplo pensemos que ahora el móvil se mueve con movimiento uniformemente acelerado:

Es decir $y = a + bx + cx^2$ donde a es la posición inicial, b es la velocidad inicial y c es la mitad de la aceleración.

Entonces intentaremos ajustar con una función cuadrática.

Si planteamos para cada par de valores

$$y_i = c + bx_i + ax_i^2$$

Obtenemos el siguiente sistema de ecuaciones

$$\begin{cases} 2 = c + b \cdot 1 + a \cdot 1 \\ 6 = c + b \cdot 2 + a \cdot 4 \\ 6 = c + b \cdot 3 + a \cdot 9 \\ 12 = c + b \cdot 4 + a \cdot 16 \end{cases} \quad \text{que matricialmente puede escribirse como:}$$

$$\begin{pmatrix} 1 & 1 & 1 \\ 1 & 2 & 4 \\ 1 & 3 & 9 \\ 1 & 4 & 16 \end{pmatrix} \begin{pmatrix} c \\ b \\ a \end{pmatrix} = A \begin{pmatrix} c \\ b \\ a \end{pmatrix} = \begin{pmatrix} 2 \\ 6 \\ 6 \\ 12 \end{pmatrix} = v \quad \text{este sistema es incompatible,}$$

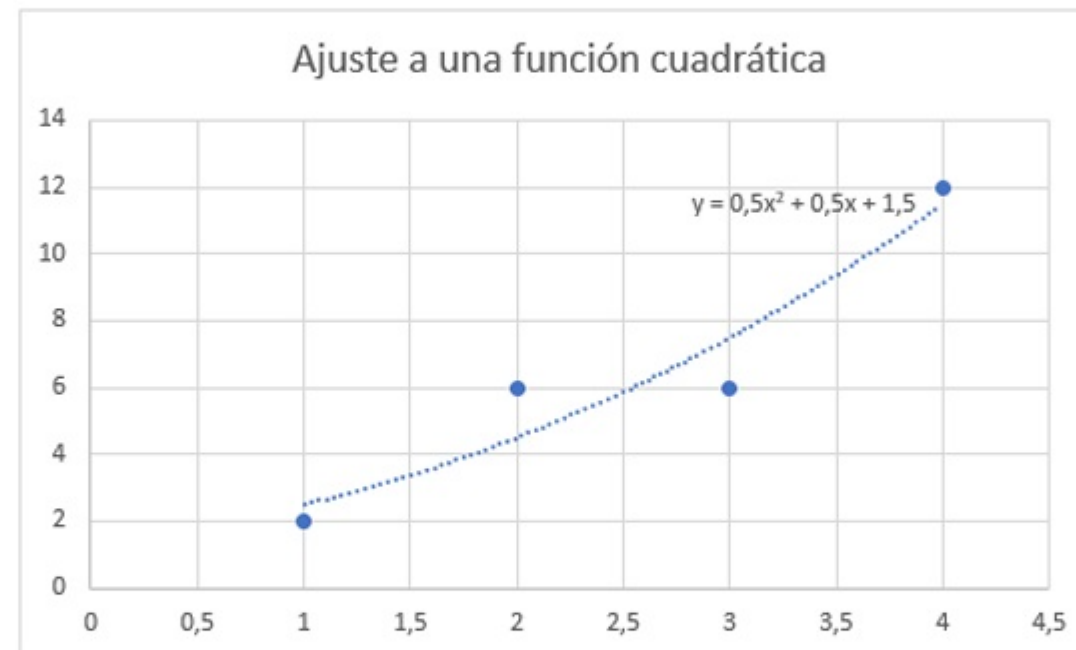
resolvemos entonces el sistema compatible $A^T A \hat{x} = A^T v$

$$\begin{pmatrix} 1 & 1 & 1 & 1 \\ 1 & 2 & 3 & 4 \\ 1 & 4 & 9 & 16 \end{pmatrix} \begin{pmatrix} 1 & 1 & 1 \\ 1 & 2 & 4 \\ 1 & 3 & 9 \\ 1 & 4 & 16 \end{pmatrix} \begin{pmatrix} c \\ b \\ a \end{pmatrix} = A^T A \begin{pmatrix} c \\ b \\ a \end{pmatrix} = A^T \begin{pmatrix} 2 \\ 6 \\ 6 \\ 12 \end{pmatrix} = \begin{pmatrix} 1 & 1 & 1 & 1 \\ 1 & 2 & 3 & 4 \\ 1 & 4 & 9 & 16 \end{pmatrix} \begin{pmatrix} 2 \\ 6 \\ 6 \\ 12 \end{pmatrix}$$

$$\begin{pmatrix} 4 & 10 & 30 \\ 10 & 30 & 100 \\ 30 & 100 & 354 \end{pmatrix} \begin{pmatrix} c \\ b \\ a \end{pmatrix} = \begin{pmatrix} 26 \\ 80 \\ 272 \end{pmatrix}$$

La solución del sistema compatible determinado es

$$\begin{pmatrix} c \\ b \\ a \end{pmatrix} = \begin{pmatrix} 1,5 \\ 0,5 \\ 0,5 \end{pmatrix}$$



Igual a como hicimos con el ajuste lineal podríamos haber minimizado el error cuadrático medio $\frac{1}{n} \sum_{i=1}^n \varepsilon_i^2$

Veamos cual es el error

$$\hat{y}_i = 0,5x_i^2 + 0,5x_i + 1,5$$

$$\varepsilon = \begin{pmatrix} 0,5 \\ -1,5 \\ 1,5 \\ -0,5 \end{pmatrix} \quad \|\varepsilon\| = \sqrt{5}$$

x_i	y_i	$\hat{y}_i$	$\varepsilon_i = \hat{y}_i - y_i$
1	2	2,5	0,5
2	6	4,5	-1,5
3	6	7,5	1,5
4	12	11,5	-0,5

Y el error cuadrático medio resulta:

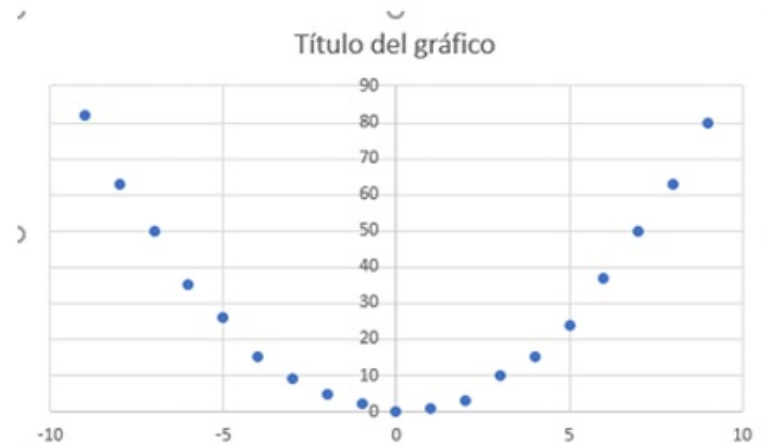
$$\frac{1}{n} \sum_{i=1}^n \varepsilon_i^2 = \frac{5}{4}$$

Este error resulta menor al que obtuvimos con la aproximación lineal.

Conclusiones:

- Cuando se realiza un experimento el n es mucho mayor que el grado del polinomio que ajusta los datos. En nuestro ejemplo esto no sucede por eso es lógico que el error cuadrático disminuya al aproximar con una función cuadrática ya que si aumentamos un grado más la función contiene a todos los puntos.
- Cuando se tienen los datos y el gráfico de dispersión se puede “intuir” el comportamiento de y como $f(x)$.

Por ejemplo, a nadie se le ocurriría hacer un ajuste lineal para esta nube de puntos.



- Muchas veces se sabe cuál debe ser el comportamiento y el cálculo del error sirve para ver si los datos son confiables o no.
- Otras veces se tienen modelos probables y el modelo que provee el menor error cuadrático medio será el elegido.