Propiedades de matrices simétricas definidas positivas:

Propiedades de matrices simétricas definidas positivas:

Sea $\boldsymbol{A} = [a_{ij}] \in \mathbb{R}^{n \times n}$.

La matriz $\boldsymbol{A}$ es simétrica si $\boldsymbol{A} = \boldsymbol{A}^\top$.

La matriz $\boldsymbol{A}$ es definida positiva si para todo $\boldsymbol{x} \neq \boldsymbol{0}$ se tiene que

$$\boldsymbol{x}^\top \boldsymbol{A} \boldsymbol{x} > 0.$$

Notación: con $\boldsymbol{A} > 0$ indicamos que la matriz es definida positiva.

Usamos *s.d.p.* para indicar que una matriz es simétrica y definida positiva.

Decimos que $\boldsymbol{H}$ es una submatriz principal de $\boldsymbol{A}$ si es una submatriz cuadrada formada con las entradas alrededor de la diagonal principal:

$$\boldsymbol{H} = \boldsymbol{A}(j:k,\ j:k)$$

Propiedades de matrices simétricas definidas positivas:

Proposición

Las siguientes proposiciones son equivalentes:

1. Sea $\boldsymbol{X}$ no singular. $\boldsymbol{A}$ es s.d.p. si y sólo si $\boldsymbol{X}^\top \boldsymbol{A}\boldsymbol{X}$ es s.d.p.
2. Si $\boldsymbol{A}$ es s.d.p. y $\boldsymbol{H}$ es cualquier submatriz principal de $\boldsymbol{A}$, entonces $\boldsymbol{H}$ es s.d.p.
3. $\boldsymbol{A}$ es s.d.p. si y sólo si $\boldsymbol{A}$ es simétrica y todos sus eigenvalores son positivos.
4. Si $\boldsymbol{A}$ es s.d.p., entonces $a_{ii} > 0$ y $\max_{ij} |a_{ij}| = \max_i a_{ii} > 0$.
5. $\boldsymbol{A}$ es s.d.p. si y sólo si existe una única matriz triangular inferior no singular $\boldsymbol{L}$, con entradas positivas en la diagonal, tal que $\boldsymbol{A} = \boldsymbol{L}\boldsymbol{L}^\top$.

Propiedades de matrices simétricas definidas positivas:

Lema 4. Toda submatriz principal de A es positiva definida.

Demostración. Sea $S \supset \{1, \ldots, n\}$, $S \neq \varnothing$ y sea A_S la submatriz al elmininar las filas y columnas de A listadas en S^C, por lo que A_S es una submatriz principal. Sea $x \in \mathbb{C}^n$ no cero tal que

$$x^i = \begin{cases} z & i \in S,\ z \in \mathbb{C} \\ 0 & i \notin S \end{cases}$$

y sea $x_S \in \mathbb{C}^{n-|S|}$ el vector obtenido a partir de eliminar las entradas $i \notin S$ de x. Notemos que:

$$x_S^* A_S x_S = x^* A x > 0$$

Como x arbitrario y no cero, concluimos que A_S es positiva definida. □

(Demostración escrita por alumno E.A.M)

Teoría de la aproximación Métodos de Mínimos cuadrados

MAT-251

Dr. Alonso Ramírez Manzanares
CIMAT A.C.
e-mail: alram@cimat.mx
web: http://www.cimat.mx/~alram/met_num/

Dr. Salvador Botello Rionda
CIMAT A.C.
e-mail: joaquin@cimat.mx

Teoría de la aproximación

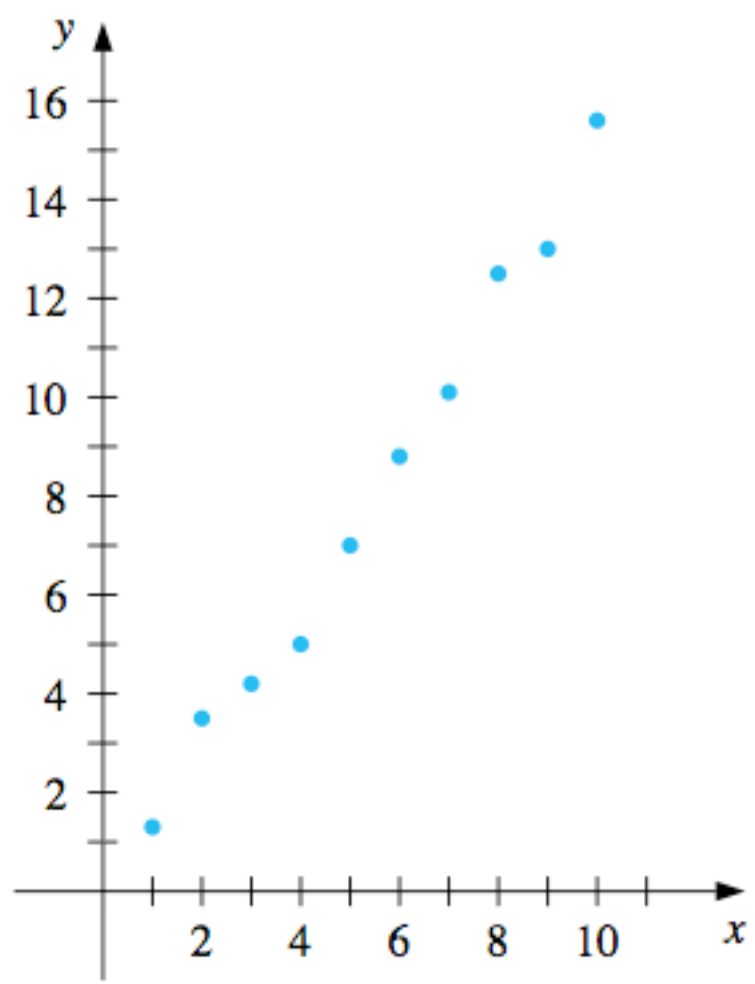
- Dos problemas fundamentales
    - Dada una función de manera explícita, queremos aproximarla por otra función mas simple (quizá mas fácil de derivar) con la cual podamos trabajar. Por ejemplo podemos querer usar polinomios en un intervalo para representarla.
    - **Dada una serie de puntos (observaciones de una fenómeno) queremos ajustar una función simple (para lo cual definimos una clase de función a utilizar) que sea óptima en la adaptación de los datos y que podamos usar para representarlos.**
- Veamos con cuidado esta segunda parte...

Aproximación discreta por mínimos cuadrados

- Considere el problema de estimar valores de una función en puntos no tabulados en la siguiente tabla de observaciones experimentales de un fenómeno.

Table 8.1

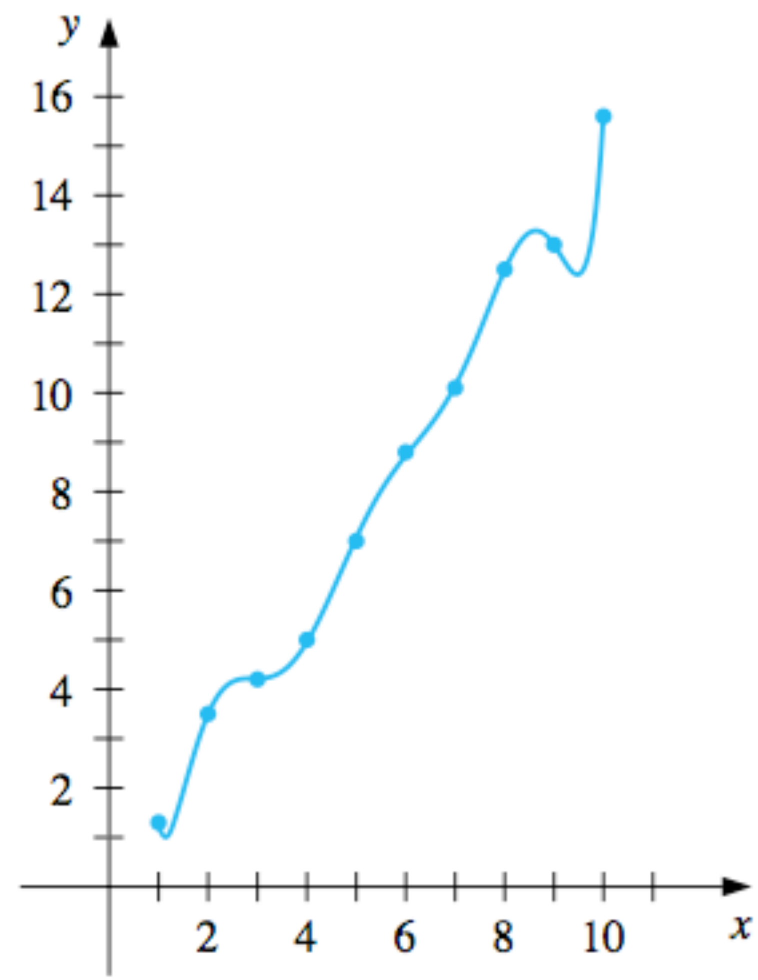
x_i	y_i	x_i	y_i
1	1.3	6	8.8
2	3.5	7	10.1
3	4.2	8	12.5
4	5.0	9	13.0
5	7.0	10	15.6



Viendo los datos, es probable que exista una relación lineal en los datos, pero es claro que ninguna linea va a ajustarse a los datos de manera exacta

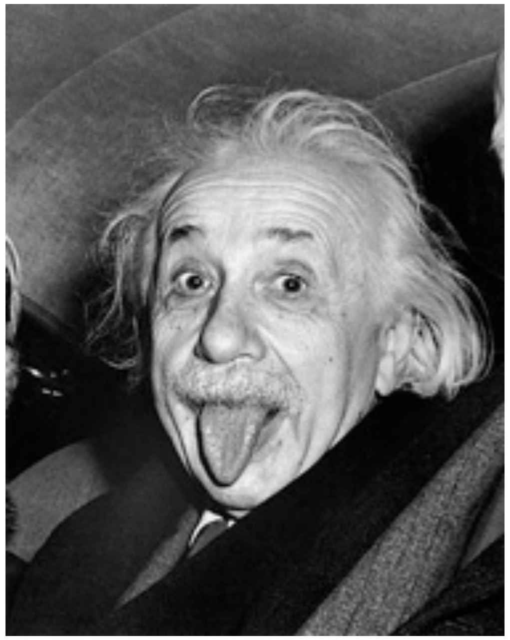
Aproximación discreta por mínimos cuadrados

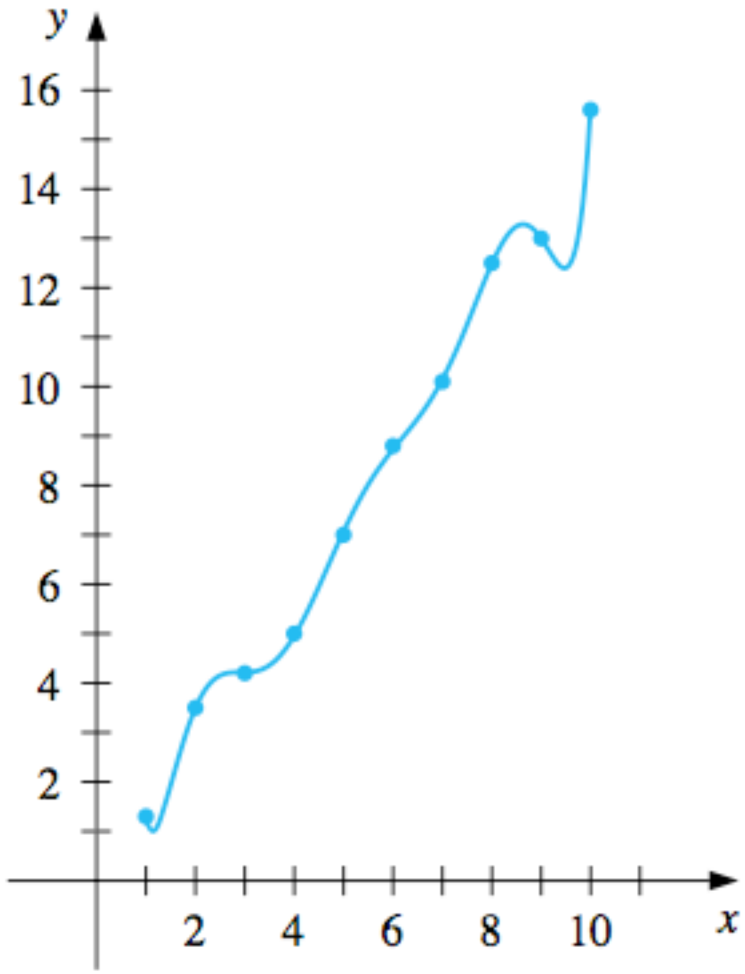
- Podríamos entonces ajustar una función más compleja que la lineal para representar los datos con error muy pequeño, pero esto no siempre es lo más correcto.



Aproximación discreta por mínimos cuadrados

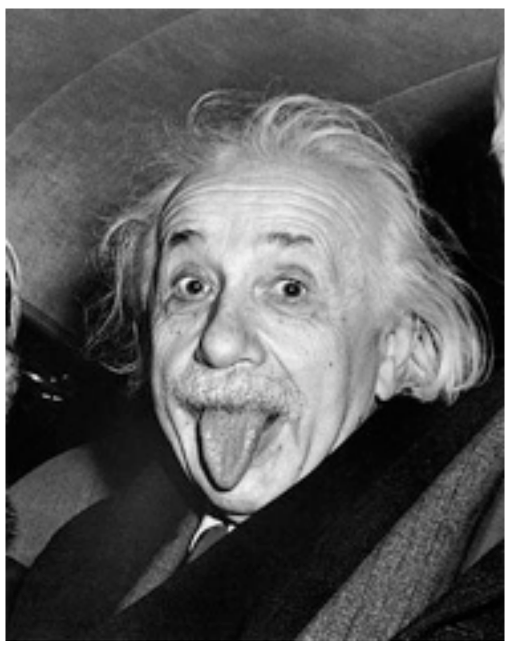
- Podríamos entonces ajustar una función más compleja que la lineal para representar los datos con error muy pequeño, pero esto no siempre es lo más correcto.





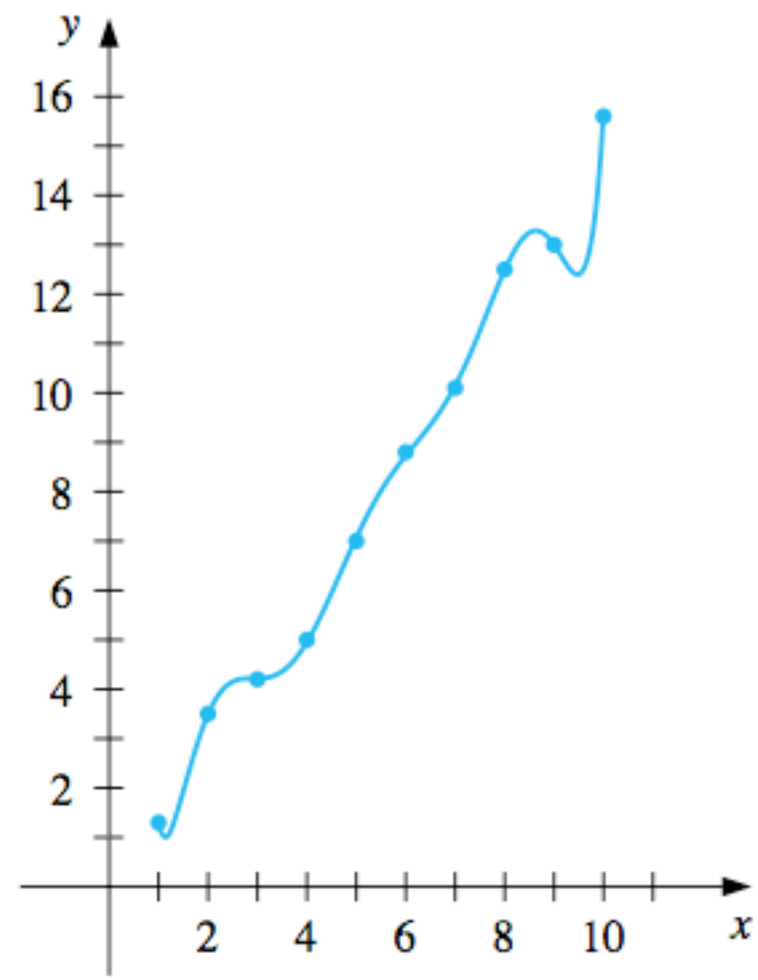
Aproximación discreta por mínimos cuadrados

- Podríamos entonces ajustar una función más compleja que la lineal para representar los datos con error muy pequeño, pero esto no siempre es lo más correcto.



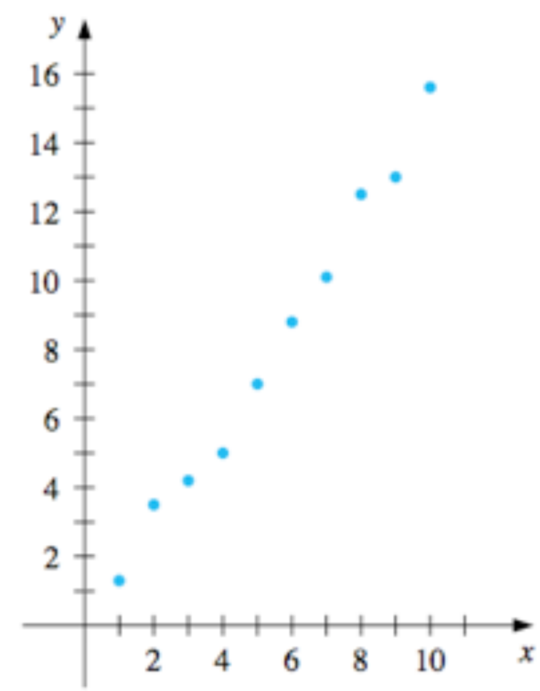
Albert Einstein: “A model should be as simple as possible, but no simpler”

“a model” = “a scientific theory” = “everything”

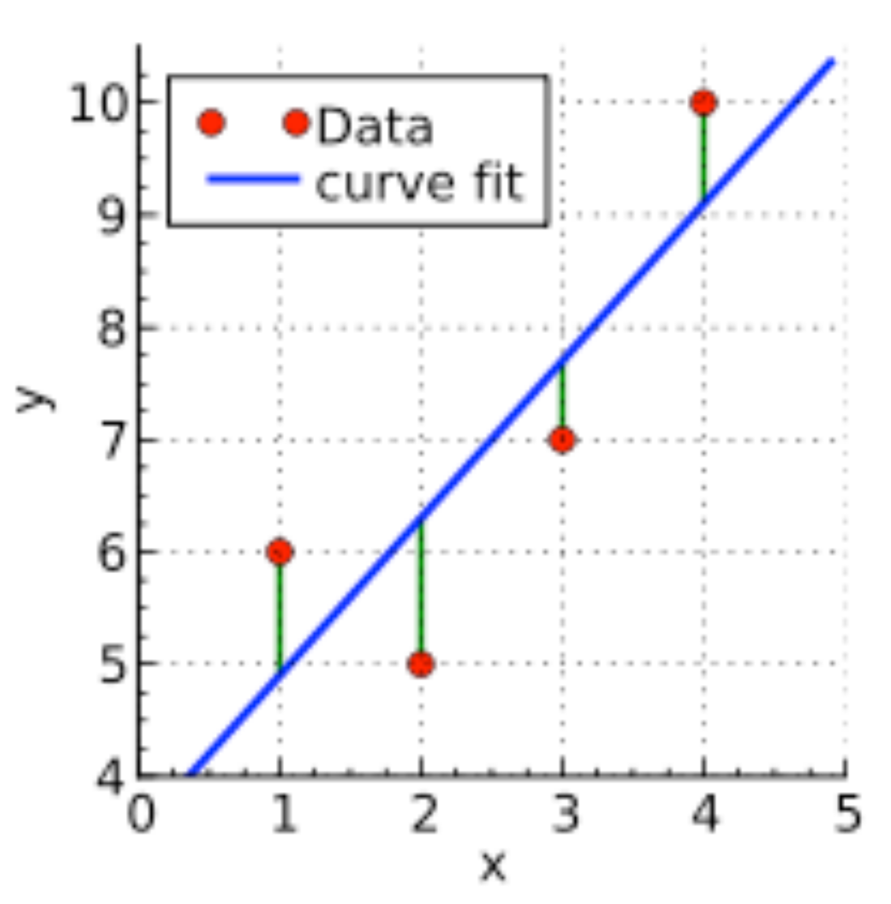


Aproximación discreta por mínimos cuadrados (LS)

- Entonces nosotros, haciéndole caso a Don Einstein vamos a usar un modelo simple para representar a los datos, usaremos un modelo lineal en este ejemplo, por lo tanto suponemos que el modelo de los datos es

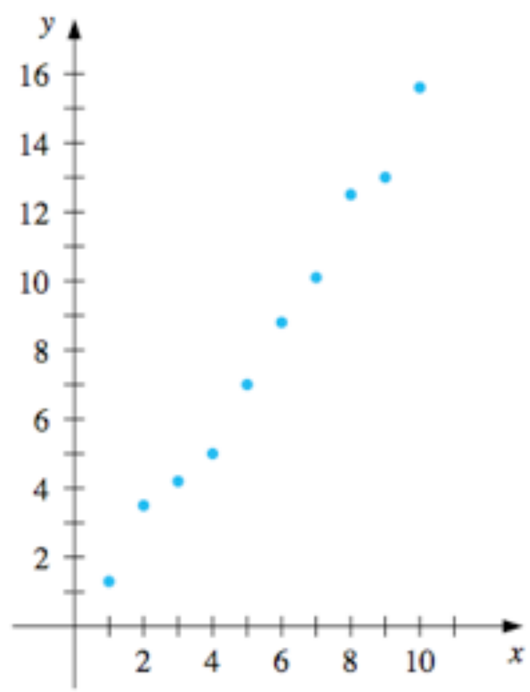


  - $y_i = a_1 x_i + a_0 + \varepsilon$



Forma matricial del sistema mínimos cuadrados lineales

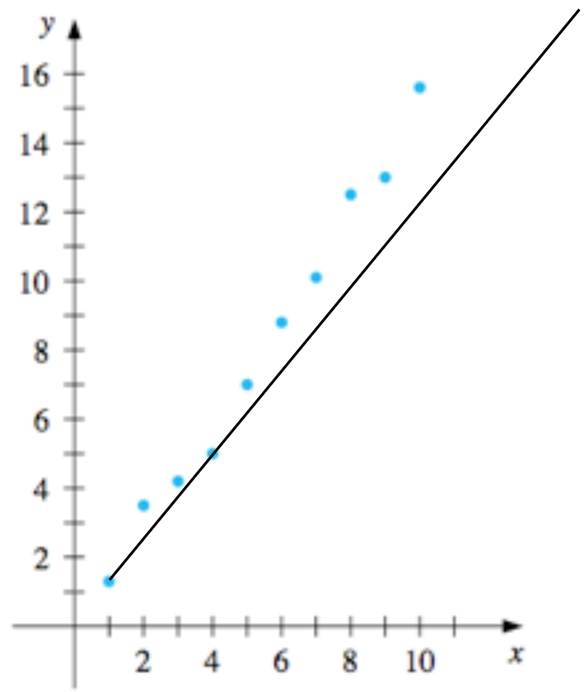
- Nótese que estrictamente tenemos un sistema ***sobre-determinado Ax*** = ***b***



- Con número de ecuaciones generalmente mucho mayor que el número de incógnitas
    - $y_1 = a_1 x_1 + a_0$
    - ...
    - $y_i = a_1 x_i + a_0$
    - ...
    - $y_n = a_1 x_n + a_0$
- Entonces, la matriz ***A*** del sistema tiene dimensiones *m* x 2 con *m* >> 2.

Forma matricial del sistema mínimos cuadrados lineales

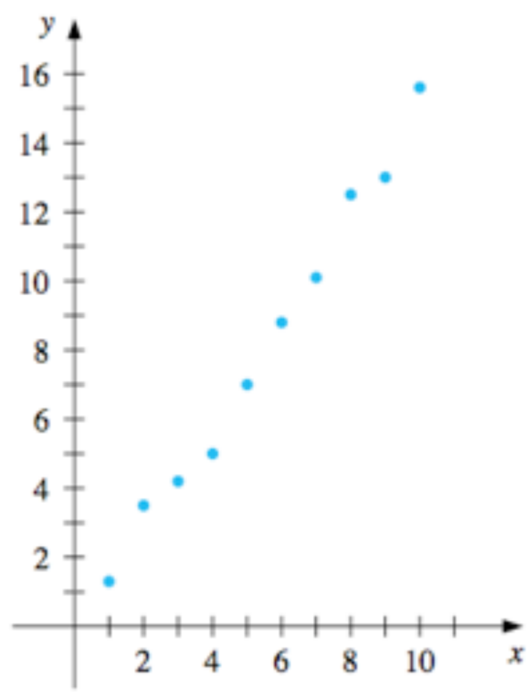
- Nótese que estrictamente tenemos un sistema ***sobre-determinado Ax*** = ***b***



- Con número de ecuaciones generalmente mucho mayor que el número de incógnitas
    - $y_1 = a_1 x_1 + a_0$
    - ...
    - $y_i = a_1 x_i + a_0$
    - ...
    - $y_n = a_1 x_n + a_0$
- Entonces, la matriz ***A*** del sistema tiene dimensiones *m* x 2 con *m* >> 2.

Forma matricial del sistema mínimos cuadrados lineales

- Nótese que estrictamente tenemos un sistema ***sobre-determinado Ax*** = ***b***

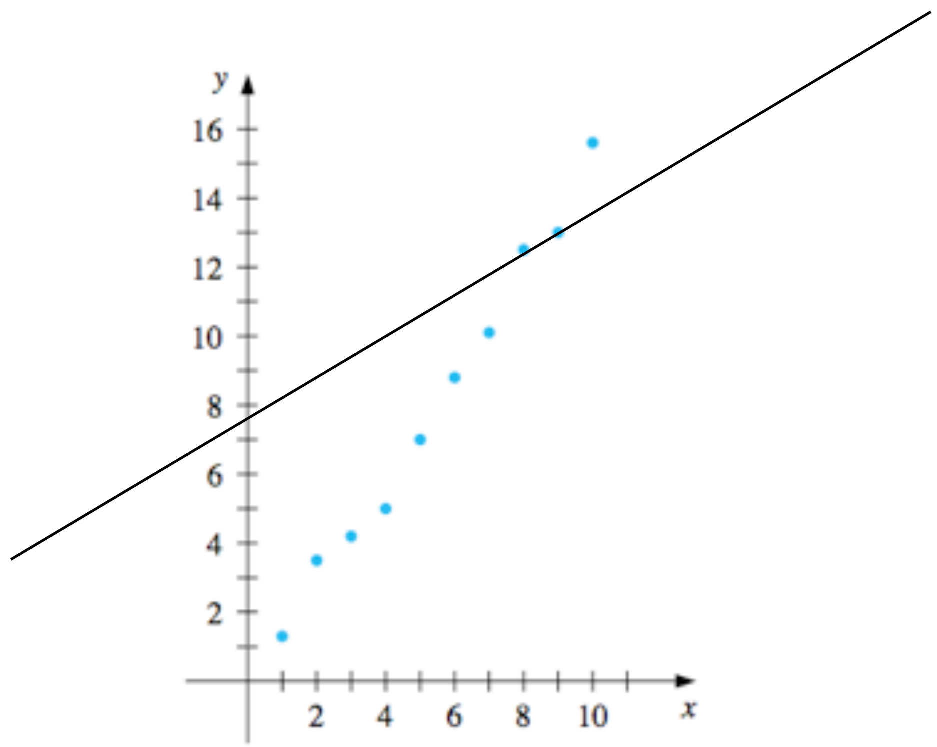


- Con número de ecuaciones generalmente mucho mayor que el número de incógnitas
  - $y_1 = a_1 x_1 + a_0$
  - ...
  - $y_i = a_1 x_i + a_0$
  - ...
  - $y_n = a_1 x_n + a_0$

- Entonces, la matriz ***A*** del sistema tiene dimensiones *m* x 2 con *m* >> 2.

Forma matricial del sistema mínimos cuadrados lineales

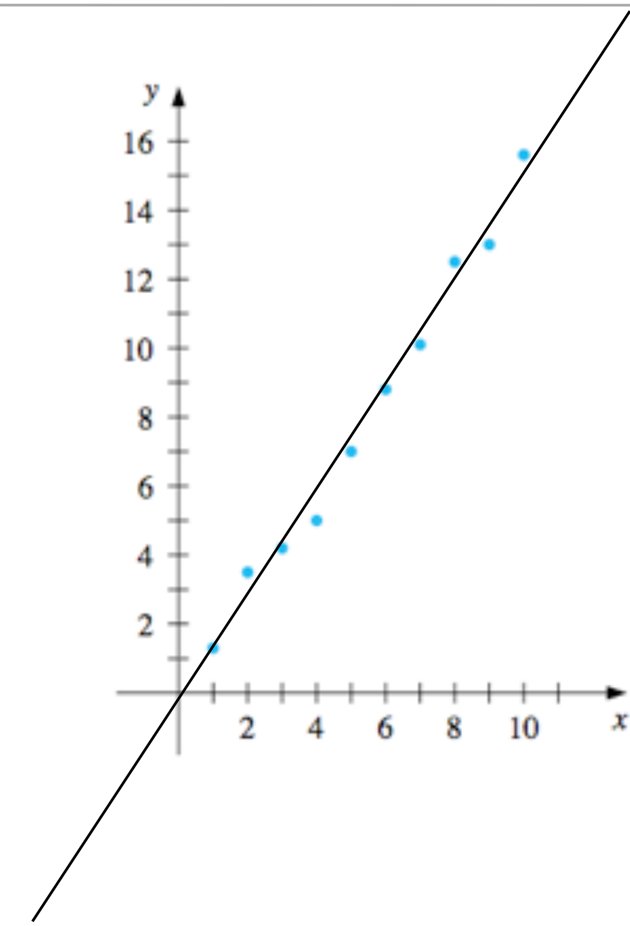
- Nótese que estrictamente tenemos un sistema ***sobre-determinado Ax*** = ***b***



- Con número de ecuaciones generalmente mucho mayor que el número de incógnitas
    - $y_1 = a_1 x_1 + a_0$
    - ...
    - $y_i = a_1 x_i + a_0$
    - ...
    - $y_n = a_1 x_n + a_0$
- Entonces, la matriz ***A*** del sistema tiene dimensiones m x 2 con $m >> 2$.

Aproximación discreta por mínimos cuadrados (LS)

- Por lo tanto, si queremos ajustar una sola línea podemos plantear las siguientes dos funciones a optimizar:



$$E_1(a_0, a_1) = \sum_{i=1}^{10} |y_i - (a_1 x_i + a_0)|\,.$$

Esta no nos gusta porque la derivada de | | en cero no está definida.

$$E_2(a_0, a_1) = \sum_{i=1}^{m} (y_i - (a_1 x_i + a_0))^2$$

Esta si nos gusta, y mucho (por lo menos en esta clase), y se llama error total de mínimos cuadrados

Aproximación discreta por mínimos cuadrados (LS)

- Para optimizar esta función de $\Re^2 \rightarrow \Re$

$$E_2(a_0, a_1) = \sum_{i=1}^{m} (y_i - (a_1 x_i + a_0))^2$$

  - Que no es otra cosa min ($\|\boldsymbol{Ax}-\boldsymbol{b}\|_2$)^2

- Calculamos el gradiente dado por las componentes

Aproximación discreta por mínimos cuadrados (LS)

- Para optimizar esta función de $\Re^2 \rightarrow \Re$

$$E_2(a_0, a_1) = \sum_{i=1}^{m} (y_i - (a_1 x_i + a_0))^2$$

  - Que no es otra cosa min ($||\boldsymbol{Ax}-\boldsymbol{b}||_2$)^2

- Calculamos el gradiente dado por las componentes

$$0 = \frac{\partial}{\partial a_0} \sum_{i=1}^{m} (y_i - (a_1 x_i + a_0))^2 = 2 \sum_{i=1}^{m} (y_i - a_1 x_i - a_0)(-1)$$

Aproximación discreta por mínimos cuadrados (LS)

- Para optimizar esta función de $\Re^2 \rightarrow \Re$

$$E_2(a_0, a_1) = \sum_{i=1}^{m} (y_i - (a_1 x_i + a_0))^2$$

  - Que no es otra cosa min ($||\boldsymbol{Ax}-\boldsymbol{b}||_2$)^2

- Calculamos el gradiente dado por las componentes

$$0 = \frac{\partial}{\partial a_0} \sum_{i=1}^{m} (y_i - (a_1 x_i + a_0))^2 = 2\sum_{i=1}^{m} (y_i - a_1 x_i - a_0)(-1)$$

$$0 = \frac{\partial}{\partial a_1} \sum_{i=1}^{m} (y_i - (a_1 x_i + a_0))^2 = 2\sum_{i=1}^{m} (y_i - a_1 x_i - a_0)(-x_i)$$

Aproximación discreta por mínimos cuadrados (LS)

- De las 2 derivadas parciales anteriores se forma el conjunto de ***ecuaciones normales***

$$a_0 \sum_{i=1}^{m} x_i + a_1 \sum_{i=1}^{m} x_i^2 = \sum_{i=1}^{m} x_i y_i$$

$$a_0 \cdot m + a_1 \sum_{i=1}^{m} x_i = \sum_{i=1}^{m} y_i.$$

- Este SEL de 2x2 nos lleva a la siguiente solución:

Aproximación discreta por mínimos cuadrados (LS)

- De las 2 derivadas parciales anteriores se forma el conjunto de ***ecuaciones normales***

$$a_0 \sum_{i=1}^{m} x_i + a_1 \sum_{i=1}^{m} x_i^2 = \sum_{i=1}^{m} x_i y_i$$

$$a_0 \cdot m + a_1 \sum_{i=1}^{m} x_i = \sum_{i=1}^{m} y_i.$$

- Este SEL de 2x2 nos lleva a la siguiente solución:

$$a_0 = \frac{\left(\sum_{i=1}^{m} x_i^2\right)\left(\sum_{i=1}^{m} y_i\right) - \left(\sum_{i=1}^{m} x_i y_i\right)\left(\sum_{i=1}^{m} x_i\right)}{m\left(\sum_{i=1}^{m} x_i^2\right) - \left(\sum_{i=1}^{m} x_i\right)^2}$$

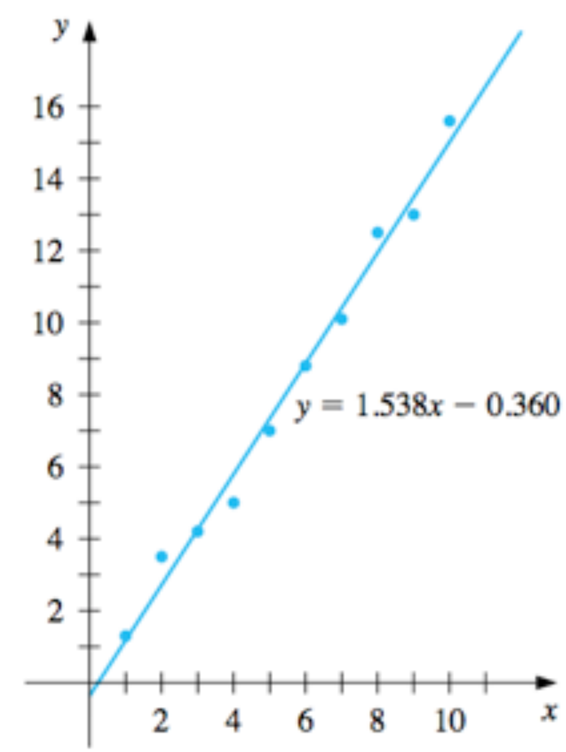
$$a_1 = \frac{m\left(\sum_{i=1}^{m} x_i y_i\right) - \left(\sum_{i=1}^{m} x_i\right)\left(\sum_{i=1}^{m} y_i\right)}{m\left(\sum_{i=1}^{m} x_i^2\right) - \left(\sum_{i=1}^{m} x_i\right)^2}.$$

Ejemplo de Linear LS (mínimos cuadrados lineales)

x_i	y_i	x_i^2	$x_i y_i$	$P(x_i) = 1.538x_i - 0.360$
1	1.3	1	1.3	1.18
2	3.5	4	7.0	2.72
3	4.2	9	12.6	4.25
4	5.0	16	20.0	5.79
5	7.0	25	35.0	7.33
6	8.8	36	52.8	8.87
7	10.1	49	70.7	10.41
8	12.5	64	100.0	11.94
9	13.0	81	117.0	13.48
10	15.6	100	156.0	15.02
55	81.0	385	572.4	$E = \sum_{i=1}^{10}(y_i - P(x_i))^2 \approx 2.34$

$$a_0 = \frac{385(81) - 55(572.4)}{10(385) - (55)^2} = -0.360$$

$$a_1 = \frac{10(572.4) - 55(81)}{10(385) - (55)^2} = 1.538.$$



En general los mínimos cuadrados para polinomios

- Si queremos ajustar el polinomio con grado n < m-1 :

$$P_n(x) = a_n x^n + a_{n-1} x^{n-1} + \cdots + a_1 x + a_0$$

- Minimizamos

$$E_2 = \sum_{i=1}^{m} (y_i - P_n(x_i))^2.$$

En general los mínimos cuadrados para polinomios

$$E_2 = \sum_{i=1}^{m}(y_i - P_n(x_i))^2$$

$$= \sum_{i=1}^{m} y_i^2 - 2\sum_{i=1}^{m} P_n(x_i)y_i + \sum_{i=1}^{m}(P_n(x_i))^2$$

$$= \sum_{i=1}^{m} y_i^2 - 2\sum_{i=1}^{m}\left(\sum_{j=0}^{n} a_j x_i^j\right) y_i + \sum_{i=1}^{m}\left(\sum_{j=0}^{n} a_j x_i^j\right)^2$$

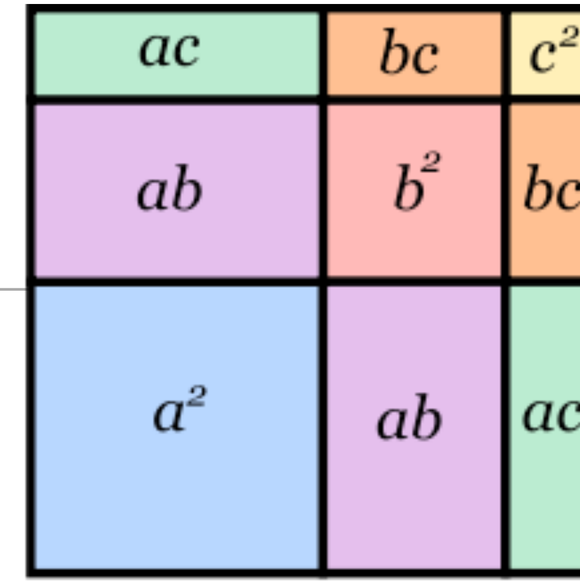
Podemos escribir

$$\left(\sum_{j=0}^{n} a_j x_i^j\right)^2 = \sum_{j=0}^{n}\sum_{k=0}^{n} a_j a_k x_i^{j+k}$$

Por ejemplo, para un polinomio de grado 1 al cuadrado:

$$\begin{aligned}\sum_{j=0}^{1}\sum_{k=0}^{1} a_j a_k x_i^{j+k} &= (a_0 a_0 x_i^0 + a_0 a_1 x_i^1) + (a_1 a_0 x_i^1 + a_1 a_1 x_i^2) \\ &= a_0^2 + 2a_0 a_1 x_i + a_1^2 x_i^2 = (a_0 + a_1 x_i)^2 \\ &= \left(\sum_{j=0}^{1} a_j x_i^j\right)^2\end{aligned}$$

Productos Notables



$$(a+b+c)^2 = a^2 + b^2 + c^2 + 2(ab+bc+ca)$$

Podemos escribir

$$\left(\sum_{j=0}^{n} a_j x_i^j\right)^2 = \sum_{j=0}^{n}\sum_{k=0}^{n} a_j a_k x_i^{j+k}$$

Por ejemplo, para un polinomio de grado 1 al cuadrado:

$$\begin{aligned}\sum_{j=0}^{1}\sum_{k=0}^{1} a_j a_k x_i^{j+k} &= (a_0 a_0 x_i^0 + a_0 a_1 x_i^1) + (a_1 a_0 x_i^1 + a_1 a_1 x_i^2)\\ &= a_0^2 + 2a_0 a_1 x_i + a_1^2 x_i^2 = (a_0 + a_1 x_i)^2\\ &= \left(\sum_{j=0}^{1} a_j x_i^j\right)^2\end{aligned}$$

En general los mínimos cuadrados para polinomios

$$E_2 = \sum_{i=1}^{m} (y_i - P_n(x_i))^2$$

$$= \sum_{i=1}^{m} y_i^2 - 2\sum_{i=1}^{m} P_n(x_i) y_i + \sum_{i=1}^{m} (P_n(x_i))^2$$

$$= \sum_{i=1}^{m} y_i^2 - 2\sum_{i=1}^{m} \left(\sum_{j=0}^{n} a_j x_i^j \right) y_i + \sum_{i=1}^{m} \left(\sum_{j=0}^{n} a_j x_i^j \right)^2$$

$$= \sum_{i=1}^{m} y_i^2 - 2\sum_{j=0}^{n} a_j \left(\sum_{i=1}^{m} y_i x_i^j \right) + \sum_{j=0}^{n} \sum_{k=0}^{n} a_j a_k \left(\sum_{i=1}^{m} x_i^{j+k} \right)$$

En general los mínimos cuadrados para polinomios

- Dada la forma de la función

$$E_2 = \sum_{i=1}^{m} y_i^2 - 2\sum_{j=0}^{n} a_j \left(\sum_{i=1}^{m} y_i x_i^j\right) + \sum_{j=0}^{n}\sum_{k=0}^{n} a_j a_k \left(\sum_{i=1}^{m} x_i^{j+k}\right)$$

- Las derivadas parciales son

$$0 = \frac{\partial E_2}{\partial a_j} = -2\sum_{i=1}^{m} y_i x_i^j + 2\sum_{k=0}^{n} a_k \sum_{i=1}^{m} x_i^{j+k}.$$

En general los mínimos cuadrados para polinomios

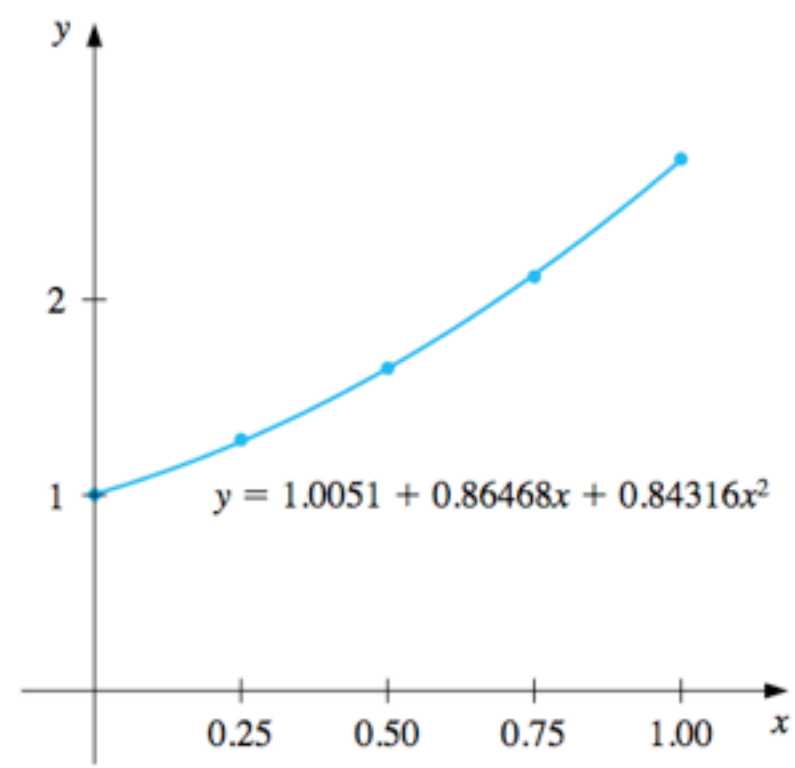
- Quedando el conjunto de ecuaciones normales y por lo tanto la matriz de mínimos cuadrados como:

$$
\begin{aligned}
a_0 \sum_{i=1}^{m} x_i^0 + a_1 \sum_{i=1}^{m} x_i^1 + a_2 \sum_{i=1}^{m} x_i^2 + \cdots + a_n \sum_{i=1}^{m} x_i^n &= \sum_{i=1}^{m} y_i x_i^0, \\
a_0 \sum_{i=1}^{m} x_i^1 + a_1 \sum_{i=1}^{m} x_i^2 + a_2 \sum_{i=1}^{m} x_i^3 + \cdots + a_n \sum_{i=1}^{m} x_i^{n+1} &= \sum_{i=1}^{m} y_i x_i^1, \\
&\vdots \\
a_0 \sum_{i=1}^{m} x_i^n + a_1 \sum_{i=1}^{m} x_i^{n+1} + a_2 \sum_{i=1}^{m} x_i^{n+2} + \cdots + a_n \sum_{i=1}^{m} x_i^{2n} &= \sum_{i=1}^{m} y_i x_i^n.
\end{aligned}
$$

Ejemplo de mínimos cuadrados para polinomios

- Dados los datos, ajustar un polinomio de grado 2

i	1	2	3	4	5
x_i	0	0.25	0.50	0.75	1.00
y_i	1.0000	1.2840	1.6487	2.1170	2.7183
$P(x_i)$	1.0051	1.2740	1.6482	2.1279	2.7129
$y_i - P(x_i)$	−0.0051	0.0100	0.0004	−0.0109	0.0054



Las ecuaciones normales

$$
\begin{aligned}
5a_0 &+ 2.5a_1 + 1.875a_2 = 8.7680,\\
2.5a_0 &+ 1.875a_1 + 1.5625a_2 = 5.4514,\\
1.875a_0 &+ 1.5625a_1 + 1.3828a_2 = 4.4015.
\end{aligned}
$$

$$a_0 = 1.0051, \quad a_1 = 0.86468, \quad \text{and} \quad a_2 = 0.84316.$$

$$\sum_{i=1}^{5} (y_i - P(x_i))^2 = 2.76 \times 10^{-4},$$

Casos no lineales que se pueden linealizar

$$0 = \frac{\partial E}{\partial b} = 2\sum_{i=1}^{m}(y_i - be^{ax_i})(-e^{ax_i})$$

$$0 = \frac{\partial E}{\partial a} = 2\sum_{i=1}^{m}(y_i - be^{ax_i})(-bx_i e^{ax_i})$$

Casos no lineales que se pueden linealizar

- Dado el modelo de los datos $y = be^{ax}$ o bien la forma $y = b\,x^a$

$$0 = \frac{\partial E}{\partial b} = 2\sum_{i=1}^{m}(y_i - be^{ax_i})(-e^{ax_i})$$

$$0 = \frac{\partial E}{\partial a} = 2\sum_{i=1}^{m}(y_i - be^{ax_i})(-bx_i e^{ax_i})$$

Casos no lineales que se pueden linealizar

- Dado el modelo de los datos $y = be^{ax}$ o bien la forma $y = b\, x^a$

- Las ecuaciones normales asociadas al primer modelo son

$$0 = \frac{\partial E}{\partial b} = 2\sum_{i=1}^{m}(y_i - be^{ax_i})(-e^{ax_i})$$

$$0 = \frac{\partial E}{\partial a} = 2\sum_{i=1}^{m}(y_i - be^{ax_i})(-bx_i e^{ax_i})$$

Casos no lineales que se pueden linealizar

- Dado el modelo de los datos $y = be^{ax}$ o bien la forma $y = b\,x^{a}$

- Las ecuaciones normales asociadas al primer modelo son

$$0 = \frac{\partial E}{\partial b} = 2\sum_{i=1}^{m}(y_i - be^{ax_i})(-e^{ax_i})$$

$$0 = \frac{\partial E}{\partial a} = 2\sum_{i=1}^{m}(y_i - be^{ax_i})(-bx_ie^{ax_i})$$

Casos no lineales que se pueden linealizar

- Dado el modelo de los datos $y = be^{ax}$ o bien la forma $y = b\, x^a$
- Las ecuaciones normales asociadas al primer modelo son

$$0 = \frac{\partial E}{\partial b} = 2\sum_{i=1}^{m}(y_i - be^{ax_i})(-e^{ax_i})$$

$$0 = \frac{\partial E}{\partial a} = 2\sum_{i=1}^{m}(y_i - be^{ax_i})(-bx_ie^{ax_i})$$

Casos no lineales que se pueden linealizar

- Dado el modelo de los datos $y = be^{ax}$ o bien la forma $y = b\,x^{a}$

- Las ecuaciones normales asociadas al primer modelo son

$$0 = \frac{\partial E}{\partial b} = 2\sum_{i=1}^{m}(y_i - be^{ax_i})(-e^{ax_i})$$

$$0 = \frac{\partial E}{\partial a} = 2\sum_{i=1}^{m}(y_i - be^{ax_i})(-bx_ie^{ax_i})$$

- Las cuales son no lineales, pero podemos operar

Casos no lineales que se pueden linealizar

- Dado el modelo de los datos $y = be^{ax}$ o bien la forma $y = b\,x^{a}$

- Las ecuaciones normales asociadas al primer modelo son

$$0 = \frac{\partial E}{\partial b} = 2\sum_{i=1}^{m}(y_i - be^{ax_i})(-e^{ax_i})$$

$$0 = \frac{\partial E}{\partial a} = 2\sum_{i=1}^{m}(y_i - be^{ax_i})(-bx_i e^{ax_i})$$

- Las cuales son no lineales, pero podemos operar

  - *ln y = ln b + a x*

Casos no lineales que se pueden linealizar

- Dado el modelo de los datos *y = be^ax^* o bien la forma *y = b x^a^*

- Las ecuaciones normales asociadas al primer modelo son

$$0 = \frac{\partial E}{\partial b} = 2\sum_{i=1}^{m}(y_i - be^{ax_i})(-e^{ax_i})$$

$$0 = \frac{\partial E}{\partial a} = 2\sum_{i=1}^{m}(y_i - be^{ax_i})(-bx_i e^{ax_i})$$

- Las cuales son no lineales, pero podemos operar

    - *ln y = ln b + a x*

- y estimar las incógnitas *ln b* y *a,* lo cual nos lleva a un problema lineal. **Pero, nótese que esta no es la solución de mínimos cuadrados del problema original. A veces conviene resolver el problema con métodos no lineales.**

Temas adicionales

Información sobre la forma matricial de LS

- Esta notación es muy común en la literatura, sobre todo usada en estadística
- El residuo del sistema ***X*** *β*= *y*, **X** tiene dimensiones *m* x *n* $r_i = y_i - \sum_{j=1}^{n} X_{ij}\beta_j.$
- la suma de residuos
- la derivada parcial total
- la derivada parcial del r_i
- substituyendo

Información sobre la forma matricial de LS

- Esta notación es muy común en la literatura, sobre todo usada en estadística
- El residuo del sistema $\boldsymbol{X}\beta = y$, $\mathbf{X}$ tiene dimensiones $m \times n$ $$r_i = y_i - \sum_{j=1}^{n} X_{ij}\beta_j.$$
- la suma de residuos $$S = \sum_{i=1}^{m} r_i^2.$$
- la derivada parcial total
- la derivada parcial del r_i
- substituyendo

Información sobre la forma matricial de LS

- Esta notación es muy común en la literatura, sobre todo usada en estadística
- El residuo del sistema $\boldsymbol{X}\beta = y$, **X** tiene dimensiones m x n $$r_i = y_i - \sum_{j=1}^{n} X_{ij}\beta_j.$$
- la suma de residuos $$S = \sum_{i=1}^{m} r_i^2.$$
- la derivada parcial total $$\frac{\partial S}{\partial \beta_j} = 2\sum_{i=1}^{m} r_i \frac{\partial r_i}{\partial \beta_j} \ (j = 1, 2, \ldots, n).$$
- la derivada parcial del r_i
- substituyendo

Información sobre la forma matricial de LS

- Esta notación es muy común en la literatura, sobre todo usada en estadística
- El residuo del sistema $\boldsymbol{X}\beta = y$, **X** tiene dimensiones m x n $$r_i = y_i - \sum_{j=1}^{n} X_{ij}\beta_j.$$
- la suma de residuos $$S = \sum_{i=1}^{m} r_i^2.$$
- la derivada parcial total $$\frac{\partial S}{\partial \beta_j} = 2\sum_{i=1}^{m} r_i \frac{\partial r_i}{\partial \beta_j} \ (j = 1, 2, \ldots, n).$$
- la derivada parcial del r_i $$\frac{\partial r_i}{\partial \beta_j} = -X_{ij}.$$
- substituyendo

Información sobre la forma matricial de LS

- Esta notación es muy común en la literatura, sobre todo usada en estadística
- El residuo del sistema ***X*** *β= y*, **X** tiene dimensiones *m* x *n* $r_i = y_i - \sum_{j=1}^{n} X_{ij}\beta_j.$
- la suma de residuos $$S = \sum_{i=1}^{m} r_i^2.$$
- la derivada parcial total $$\frac{\partial S}{\partial \beta_j} = 2\sum_{i=1}^{m} r_i \frac{\partial r_i}{\partial \beta_j} \ (j = 1, 2, \ldots, n).$$
- la derivada parcial del r_i $$\frac{\partial r_i}{\partial \beta_j} = -X_{ij}.$$
- substituyendo $$\frac{\partial S}{\partial \beta_j} = 2\sum_{i=1}^{m} \left(y_i - \sum_{k=1}^{n} X_{ik}\beta_k\right)(-X_{ij}) \ (j = 1, 2, \ldots, n).$$

Información sobre la forma matricial de LS

- con lo anterior: $$\frac{\partial S}{\partial \beta_j} = 2\sum_{i=1}^{m}\left(y_i - \sum_{k=1}^{n} X_{ik}\beta_k\right)(-X_{ij}) \ (j = 1, 2, \ldots, n).$$
- el vector β gorro es el optimal
- Separando términos dependientes e independientes
- La forma matricial del sistema original ***X*** *β= y*

Información sobre la forma matricial de LS

- con lo anterior:

$$\frac{\partial S}{\partial \beta_j} = 2\sum_{i=1}^{m}\left(y_i - \sum_{k=1}^{n} X_{ik}\beta_k\right)(-X_{ij})\ (j = 1, 2, \ldots, n).$$

- el vector β gorro es el optimal

$$2\sum_{i=1}^{m}\left(y_i - \sum_{k=1}^{n} X_{ik}\hat{\beta}_k\right)(-X_{ij}) = 0\ (j = 1, 2, \ldots, n).$$

- Separando términos dependientes e independientes

- La forma matricial del sistema original $\boldsymbol{X}\,\beta = y$

Información sobre la forma matricial de LS

- con lo anterior:

$$\frac{\partial S}{\partial \beta_j} = 2\sum_{i=1}^{m}\left(y_i - \sum_{k=1}^{n} X_{ik}\beta_k\right)(-X_{ij})\ (j = 1, 2, \dots, n).$$

- el vector β gorro es el optimal

$$2\sum_{i=1}^{m}\left(y_i - \sum_{k=1}^{n} X_{ik}\hat{\beta}_k\right)(-X_{ij}) = 0\ (j = 1, 2, \dots, n).$$

- Separando términos dependientes e independientes

$$\sum_{i=1}^{m}\sum_{k=1}^{n} X_{ij}X_{ik}\hat{\beta}_k = \sum_{i=1}^{m} X_{ij}y_i\ (j = 1, 2, \dots, n).$$

- La forma matricial del sistema original ***X*** *β*= *y*

Información sobre la forma matricial de LS

- con lo anterior: $$\frac{\partial S}{\partial \beta_j} = 2\sum_{i=1}^{m}\left(y_i - \sum_{k=1}^{n} X_{ik}\beta_k\right)(-X_{ij})\ (j = 1, 2, \dots, n).$$
- el vector β gorro es el optimal $$2\sum_{i=1}^{m}\left(y_i - \sum_{k=1}^{n} X_{ik}\hat{\beta}_k\right)(-X_{ij}) = 0\ (j = 1, 2, \dots, n).$$
- Separando términos dependientes e independientes $$\sum_{i=1}^{m}\sum_{k=1}^{n} X_{ij}X_{ik}\hat{\beta}_k = \sum_{i=1}^{m} X_{ij}y_i\ (j = 1, 2, \dots, n).$$
- La forma matricial del sistema original ***X*** *β= y* $$(\mathbf{X}^\top \mathbf{X})\hat{\boldsymbol{\beta}} = \mathbf{X}^\top \mathbf{y}$$

Información sobre la forma matricial de LS

- La forma matricial del sistema original ***X*** *β= y* $$(\mathbf{X}^\top\mathbf{X})\hat{\boldsymbol{\beta}} = \mathbf{X}^\top\mathbf{y}$$
  - donde **X** es rectangular con dimensiones *m* x *n* con *m>>n*
- *De tal forma que la solución al sistema esta dada por*

$$\hat{\boldsymbol{\beta}} = (\mathbf{X}^\top\mathbf{X})^{-1}\mathbf{X}^\top\mathbf{y} = \mathbf{X}^{+}\mathbf{y}$$

- *Se puede demostrar que la matriz simetrica* $\boldsymbol{X}^T\boldsymbol{X}$ *es no singular si todos los* x_i *son distintos y n < m-1.*
- *Si la matriz* $\boldsymbol{X}^T\boldsymbol{X}$ *es definida positiva (ya era simétrica) las ecuaciones normales se pueden resolver por factorización de Cholesky.*

Información sobre la forma matricial de LS

- La forma matricial del sistema original ***X*** *β= y* $(\mathbf{X}^\top\mathbf{X})\hat{\boldsymbol{\beta}} = \mathbf{X}^\top\mathbf{y}$
  - donde **X** es rectangular con dimensiones *m* x *n* con *m>>n*
- *De tal forma que la solución al sistema esta dada por*

$$\hat{\boldsymbol{\beta}} = (\mathbf{X}^\top\mathbf{X})^{-1}\mathbf{X}^\top\mathbf{y} = \mathbf{X}^+\mathbf{y}$$

Esta matriz de dimensiones *nxn* es la pseudo inversa de Moore-Penrose

- *Se puede demostrar que la matriz simetrica* $\boldsymbol{X}^T\boldsymbol{X}$ *es no singular si todos los* x_i *son distintos y* $n < m-1$.
- *Si la matriz* $\boldsymbol{X}^T\boldsymbol{X}$ *es definida positiva (ya era simétrica) las ecuaciones normales se pueden resolver por factorización de Cholesky.*

Aproximación para funciones continuas

- Queremos encontrar una polinomio que se parezca a *f(x)* continua en [a,b]

$$E(a_0, a_1, \ldots, a_n) = \int_a^b (f(x) - P_n(x))^2\, dx = \int_a^b \left(f(x) - \sum_{k=0}^{n} a_k x^k \right)^2 dx.$$