

Algoritmos matemáticos para: sistemas de ecuaciones lineales, inversión de matrices y mínimos cuadrados

Jose Aguilar

Inversión de matrices

Definición (Inversa de una matriz): Sea \mathbf{A} una matriz $n \times n$. Una matriz \mathbf{C} de $n \times n$ es una **inversa** de \mathbf{A} si $\mathbf{CA} = \mathbf{AC} = \mathbf{I}$.

Para la matriz $\mathbf{A} = \begin{pmatrix} -4 & 9 \\ 1 & -2 \end{pmatrix}$ no es difícil verificar que la matriz $\mathbf{C} = \begin{pmatrix} 2 & 9 \\ 1 & 4 \end{pmatrix}$

satisface que

$$\begin{pmatrix} 2 & 9 \\ 1 & 4 \end{pmatrix} \begin{pmatrix} -4 & 9 \\ 1 & -2 \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$$

Se dice entonces que la matriz \mathbf{C} es una **inversa** de la matriz \mathbf{A} . Esto se define enseguida:

Teorema: Sea \mathbf{A} una matriz $n \times n$ con inversa \mathbf{C} tal que $\mathbf{CA} = \mathbf{AC} = \mathbf{I}$. Si \mathbf{D} es otra matriz $n \times n$ tal que $\mathbf{AD} = \mathbf{I}$, entonces $\mathbf{C} = \mathbf{D}$.

Demostración: Como la multiplicación de matrices es asociativa, se tiene que

$\mathbf{C}(\mathbf{AD}) = (\mathbf{CA})\mathbf{D}$, de donde, como $\mathbf{AD} = \mathbf{I}$ y $\mathbf{CA} = \mathbf{I}$, se tiene que

$\mathbf{C}(\mathbf{AD}) = \mathbf{CI} = \mathbf{C}$ y $(\mathbf{CA})\mathbf{D} = \mathbf{ID} = \mathbf{D}$, por tanto, $\mathbf{C} = \mathbf{D}$.

Inversión de matrices

Se denotará la inversa de una matriz A , cuando exista, como A^{-1} . Entonces $A A^{-1} = A^{-1} A = I$. Nótese que no se debe expresar A^{-1} como $1/A$.

Definición: Una matriz cuadrada que tiene inversa se llama **invertible**. Una matriz cuadrada que no tiene inversa se llama **singular**.

Teorema: La matriz $A = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$

es invertible si $ad - bc \neq 0$, en cuyo caso la inversa está dada por la fórmula

$$A^{-1} = \frac{1}{ad - bc} \begin{pmatrix} d & -b \\ -c & a \end{pmatrix} = \begin{pmatrix} \frac{d}{ad - bc} & \frac{-b}{ad - bc} \\ \frac{-c}{ad - bc} & \frac{a}{ad - bc} \end{pmatrix}$$

Teorema: Sean A y B matrices invertibles $n \times n$. Entonces:

a) AB es invertible

b) $(AB)^{-1} = B^{-1} A^{-1}$

inversión de matrices

Propiedades de la inversión de matrices

La matriz inversa, si existe, es única

$$A^{-1} \cdot A = A \cdot A^{-1} = I$$

$$(A \cdot B)^{-1} = B^{-1} \cdot A^{-1}$$

$$(A^{-1})^{-1} = A$$

$$(kA)^{-1} = (1/k) \cdot A^{-1}$$

$$(A^t)^{-1} = (A^{-1})^t$$

inversión de matrices

Observación:

Podemos encontrar matrices que cumplen $A \cdot B = I$, pero que $B \cdot A \neq I$, en tal caso, podemos decir que A es la inversa de B "por la izquierda" o que B es la inversa de A "por la derecha".

Hay varios métodos para calcular la matriz inversa de una matriz dada:

- Por el método de Gauss-Jordan
- Usando determinantes
- Directamente

inversión de matrices

Cálculo Directo de la Matriz Inversa

Dada la matriz $A = \begin{pmatrix} 2 & -1 \\ 1 & 1 \end{pmatrix}$ buscamos una matriz que cumpla $A \cdot A^{-1} = I$, es decir

$$\begin{pmatrix} 2 & -1 \\ 1 & 1 \end{pmatrix} \cdot \begin{pmatrix} a & b \\ c & d \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$$

Para ello planteamos el sistema de ecuaciones:


$$\left. \begin{array}{l} 2a - c = 1 \\ 2b - d = 0 \\ a + c = 0 \\ b + d = 1 \end{array} \right\} \Rightarrow a = -c \Rightarrow \left. \begin{array}{l} -2c - c = 1 \\ d = 2b \\ a = -c \\ b + 2b = 1 \end{array} \right\} \Rightarrow \left. \begin{array}{l} a = \frac{1}{3} \\ c = -\frac{1}{3} \end{array} \right\} \begin{array}{l} b = \frac{1}{3} \\ d = \frac{2}{3} \end{array} \Rightarrow A^{-1} = \begin{pmatrix} \frac{1}{3} & \frac{1}{3} \\ -\frac{1}{3} & \frac{2}{3} \end{pmatrix}$$

La matriz que se ha calculado realmente sería la inversa por la "derecha", pero es fácil comprobar que también cumple $A^{-1} \cdot A = I$, con lo cual es realmente la inversa de A .

Inversión de matrices

Cálculo de las inversas:

Sea $\mathbf{A}=\{a_{ij}\}$ una matriz $n \times n$. Para hallar \mathbf{A}^{-1} si es que existe, se debe encontrar una matriz $\mathbf{X}=\{x_{ij}\}$ $n \times n$ tal que $\mathbf{AX}=\mathbf{I}$, esto es, tal que

$$\begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nn} \end{pmatrix} \begin{pmatrix} \text{blue } x_{11} & \text{red } x_{12} & \cdots & \text{purple } x_{1n} \\ \text{blue } x_{21} & \text{red } x_{22} & \cdots & \text{purple } x_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ \text{blue } x_{n1} & \text{red } x_{n2} & \cdots & \text{purple } x_{nn} \end{pmatrix} = \begin{pmatrix} \text{blue } 1 & \text{red } 0 & \cdots & \text{purple } 0 \\ \text{blue } 0 & \text{red } 1 & \cdots & \text{purple } 0 \\ \vdots & \vdots & \ddots & \vdots \\ \text{blue } 0 & \text{red } 0 & \cdots & \text{purple } 1 \end{pmatrix}$$


Esto es un sistema de ecuaciones con n vectores de incógnitas, y entonces es posible aplicar el Método de Gauss-Jordan para encontrar la inversa de \mathbf{A} . La idea es transformar, por medio de operaciones elementales por filas, la matriz aumentada del sistema (\mathbf{A}, \mathbf{I}) a un sistema $(\mathbf{I}, \mathbf{A}^{-1})$

$$\mathbf{A}^{-1} (\mathbf{A}, \mathbf{I}) \Leftrightarrow (\mathbf{A}^{-1} \mathbf{A}, \mathbf{A}^{-1} \mathbf{I}) \Leftrightarrow (\mathbf{I}, \mathbf{A}^{-1})$$

Sistema de ecuaciones lineales

$$a_{11}x_1 + a_{12}x_2 + \dots + a_{1n}x_n = b_1$$

.

.

.

$$a_{n1}x_1 + a_{n2}x_2 + \dots + a_{nn}x_n = b_n$$

SISTEMA DE ECUACIONES LINEALES

sistema de ecuaciones lineales, por ejemplo

$$2x_1 - 3x_2 = 7$$

$$3x_1 - x_2 = 2,$$

tiene asociado una matriz **A** correspondiente a las incógnitas, y un vector **b** correspondiente a los términos independientes, es decir,

$$A = \begin{pmatrix} 2 & -3 \\ 3 & -1 \end{pmatrix} \quad b = \begin{pmatrix} 7 \\ 2 \end{pmatrix}$$

Si ahora se escriben las incógnitas como un vector se puede denotar el sistema de ecuaciones lineales como **$Ax=b$** , es decir

$$x = \begin{pmatrix} x_1 \\ x_2 \end{pmatrix}$$

$$\begin{pmatrix} 2 & -3 \\ 3 & -1 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} 7 \\ 2 \end{pmatrix}$$

Esta última ecuación sugiere la noción de multiplicación de una matriz **A** por un vector columna **x**.

Métodos directos e iterativos

DIRECTOS

- $\mathbf{Ax} = \mathbf{b}$
- $\mathbf{x} = \mathbf{A}^{-1} \mathbf{b}$
- Tamaño pequeño

ITERATIVOS

- $\mathbf{x} = \mathbf{Cx} + \mathbf{d}$
- $\mathbf{x}^{(k+1)} = \mathbf{Cx}^{(k)} + \mathbf{d}$
- Tamaño grande

$$\mathbf{x} = \mathbf{A}^{-1} \mathbf{b}$$

$$\lim_{k \rightarrow \infty} \mathbf{x}^{(k)} = \mathbf{Cx} + \mathbf{d}$$

Sistema de ecuaciones lineales

$$2x_1 + 3x_2 = 2$$

$$x_1 + x_2 = 3$$



$$x_2 = 2/3 - (2/3)x_1$$

$$x_2 = 3 - x_1$$

$$3 - x_1 = 2/3 - (2/3)x_1$$



$$3 - 2/3 = x_1 - (2/3)x_1$$

$$2.333 = 0.333x_1$$



$$x_1 = 7$$

Método de Jacobi

$$A = L + D + U$$

$$x^{(k+1)} = D^{-1}(b - (L + U)x^{(k)})$$

La ecuación $A x = b$ se transforma en $(D - L - U) x = b$

$$x = D^{-1}(L + U) x + D^{-1}b$$

- $U = \text{triang. sup}; L = \text{triang. Inf.}$
- $D = \text{diag}(A);$

Algoritmo Método de Jacobi

función Jacobi (A,)

// x^0 es una aproximación inicial a la solución//

K=0

Mientras no convergencia

para i=1 **hasta** n

y=0

para j=1 **hasta** n

si $j \neq i$ **entonces**

$$y = y + a_{ij} x_j^k$$

$$x_i^{k+1} = (b_i - y) / a_{ii}$$

K=K+1

Método de Gauss-Seidel

$$\left. \begin{aligned}
 x_1^{(k+1)} &= (b_1 - a_{12}x_2^{(k)} - a_{13}x_3^{(k)} \dots - a_{1n}x_n^{(k)}) / a_{11} \\
 x_2^{(k+1)} &= (b_2 - a_{21}x_1^{(k+1)} - a_{23}x_3^{(k)} \dots - a_{2n}x_n^{(k)}) / a_{22} \\
 x_3^{(k+1)} &= (b_3 - a_{31}x_1^{(k+1)} - a_{32}x_2^{(k+1)} \dots - a_{3n}x_n^{(k)}) / a_{33} \\
 &\vdots \\
 x_n^{(k+1)} &= (b_n - a_{n1}x_1^{(k+1)} - a_{n2}x_2^{(k+1)} \dots - a_{n,n-1}x_{n-1}^{(k+1)}) / a_{nn}
 \end{aligned} \right\}$$

Modelo matricial

$$\mathbf{A} = \mathbf{L} + \mathbf{D} + \mathbf{U}$$

$$(\mathbf{L} + \mathbf{D})\mathbf{x}^{(k+1)} = \mathbf{b} - \mathbf{U}\mathbf{x}^{(k)}$$

$$\mathbf{x}^{(k+1)} = (\mathbf{L} + \mathbf{D})^{-1}(\mathbf{b} - \mathbf{U}\mathbf{x}^{(k)})$$

- $\mathbf{D} = \text{diag}(\mathbf{A})$
- $\mathbf{U} = \text{triang. Sup}; \mathbf{L} = \text{triang. inf}$

Algoritmo Método de Gauss-Seidel

función Gaussi (A, x, b)

// x^0 es una aproximación inicial a la solución//

para k=1 **hasta** convergencia

para i=1 **hasta** n

y=0

para j=1 **hasta** n

si $j \neq i$ **entonces**

$$y = y + a_{ij} x_j$$

$$x_i = (b_i - y) / a_{ii}$$

APROXIMACIÓN DE MÍNIMOS CUADRADOS

Modelado de datos

- El modelado de datos se puede expresar de la siguiente forma:
- Dadas:
 - Una colección finita de datos (x_i, y_i)
 - Una forma funcional $y = f(x)$
- Hallar los parámetros de la función
 - que mejor representen la relación entre los datos

Ejemplo

Para comprender datos experimentales, deseamos determinar una recta o una curva que “encaje” o “se ajuste” más (o describa mejor) estos datos

Imaginemos la siguiente tabla con los pasados de un curso en semestres pasados.

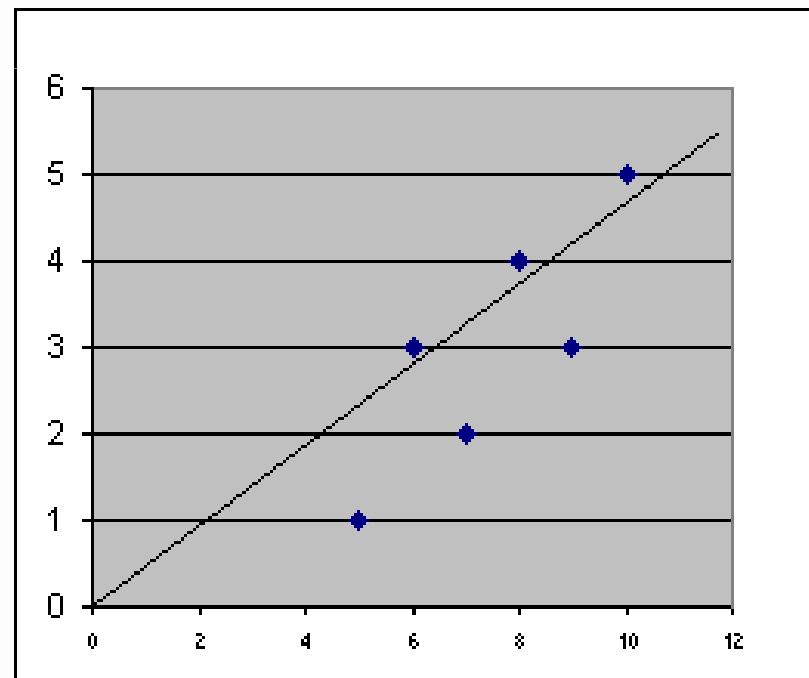
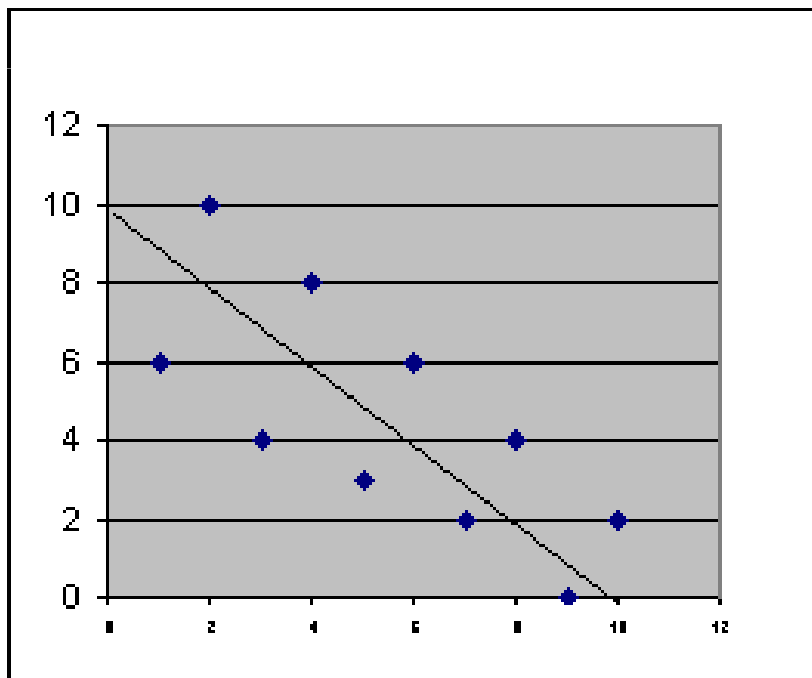
Curso	1	2	3	4	5	6
Porcentaje de pasados	0.70	0.75	0.80	0.80	0.85	0.80

Si quisiéramos trazar una recta que acerque a los puntos en la tabla hay muchas opciones. Sin embargo, hay una que se ajusta mejor a estos datos, bajo cierto criterio.

Caso anterior es $y = 0.13333 + 0.05 x$

GRÁFICOS DE DISPERSIÓN / RECTA DE REGRESIÓN

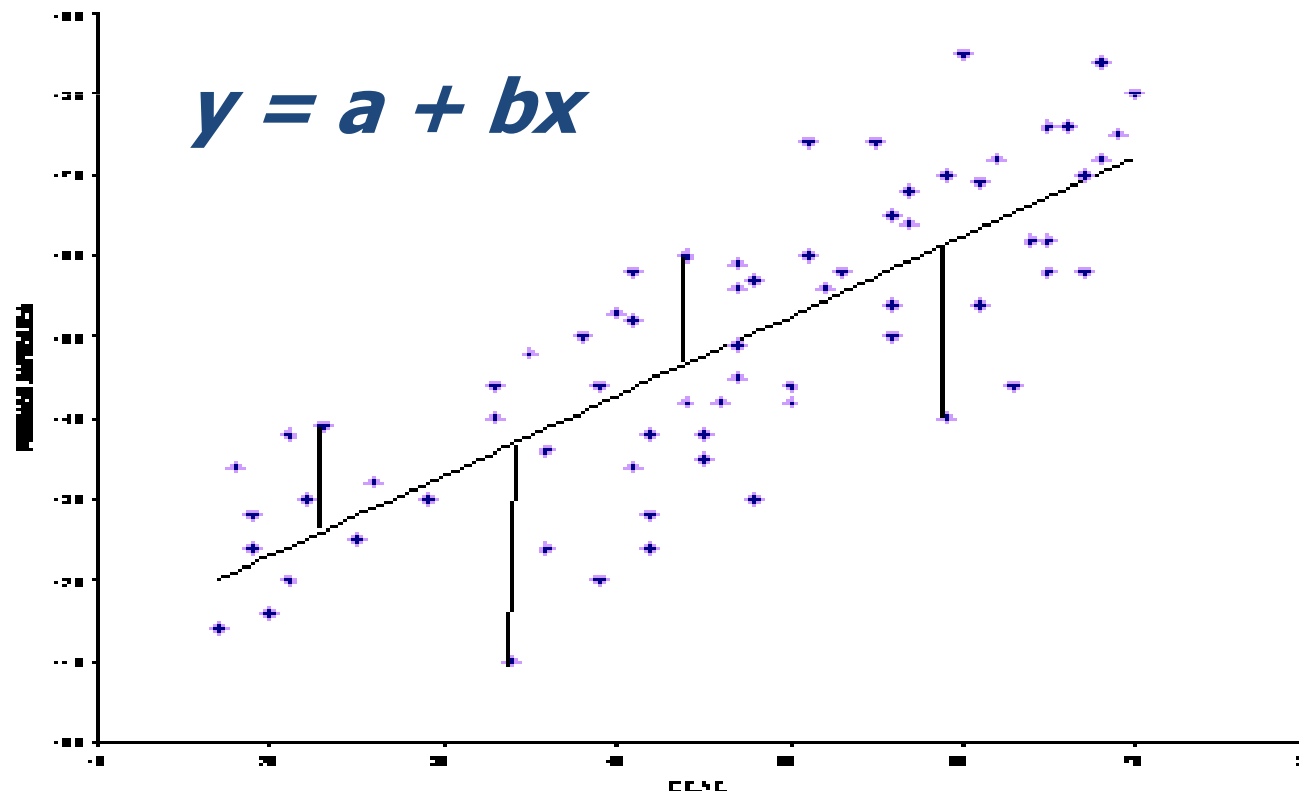
La relación entre dos variables métricas puede ser representada mediante la línea de mejor ajuste a los datos. Esta recta se le denomina recta de regresión, y puede ser negativa o positiva, la primera con tendencia decreciente y la segunda creciente.



GRÁFICOS DE DISPERSIÓN / RECTA DE REGRESIÓN

Para el cálculo de la recta de regresión se aplica el método de mínimos cuadrados entre dos variables.

Esta línea es la que hace mínima la suma de los cuadrados de los residuos.

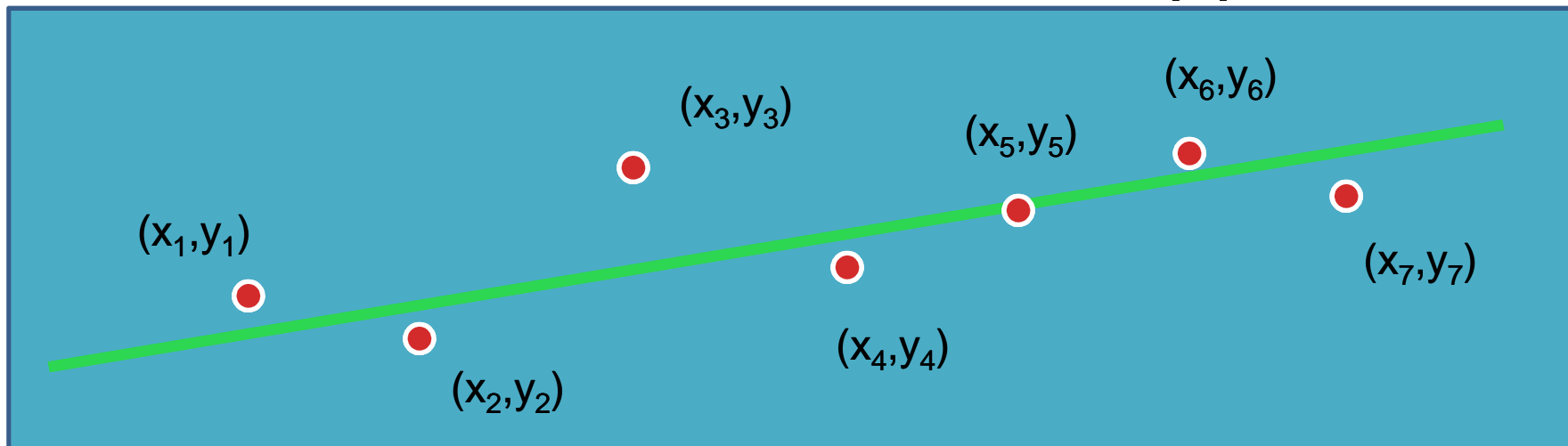


Modelado de datos

- Se busca minimizar unos residuos

$$\varepsilon_i(k) = y_i(k) - f(x_i)$$

Un ejemplo



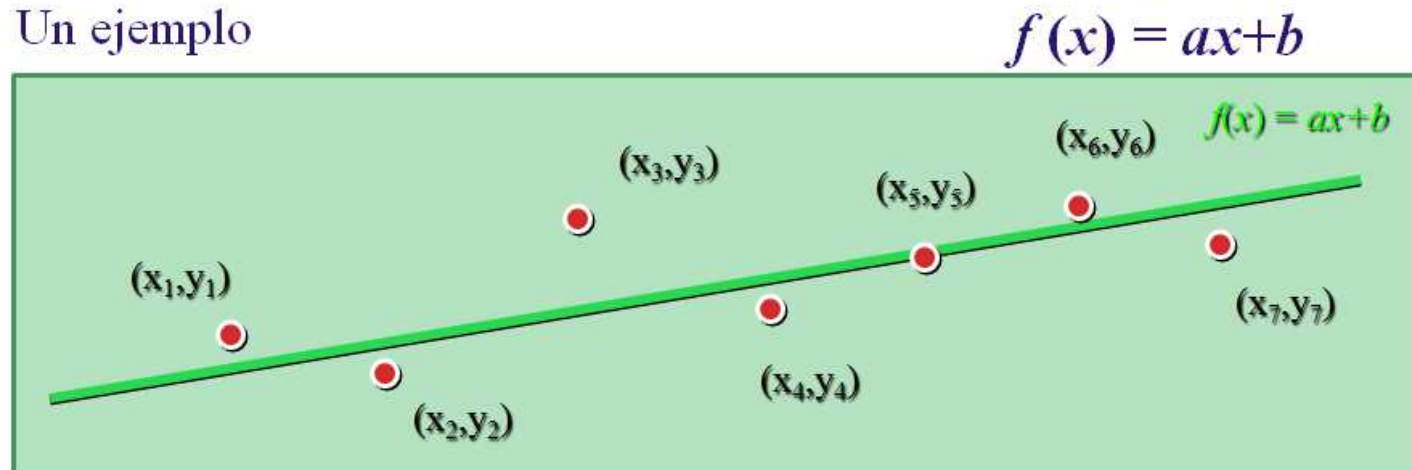
Criterio de los minimos cuadrados

- Formulacion del ajuste por Minimos cuadrados:

$$J(a,b) = \sum_{k=1}^N \varepsilon^2(k)$$

$$\varepsilon(k) = y(k) - f(x(k))$$

Un ejemplo



donde N es el numero de datos entrada-salida dado

Llamemos a “u” perturbación o error, siendo la diferencia que hay entre el valor observado de la variable exógena (y) y el valor estimado que obtendremos a través de la recta de regresión \hat{y}_i .

$$\hat{y}_i = a + bx_i$$

La metodología para la obtención de la recta será hacer **MÍNIMA** la **suma** de los **CUADRADOS** de las perturbaciones. ¿Por qué se elevan al cuadrado?

$$u_i^2 = (y_i - \hat{y}_i)^2 \longrightarrow \sum_{i=1}^n u_i^2 = \sum_{i=1}^n (y_i - \hat{y}_i)^2$$

$$\min_{q,p} \left[\sum_{i=1}^n u_i^2 = \sum_{i=1}^n (y_i - \hat{y}_i)^2 = \sum_{i=1}^n [y_i - \hat{a} + b x_i]^2 \right]$$

Un problema de optimizacion

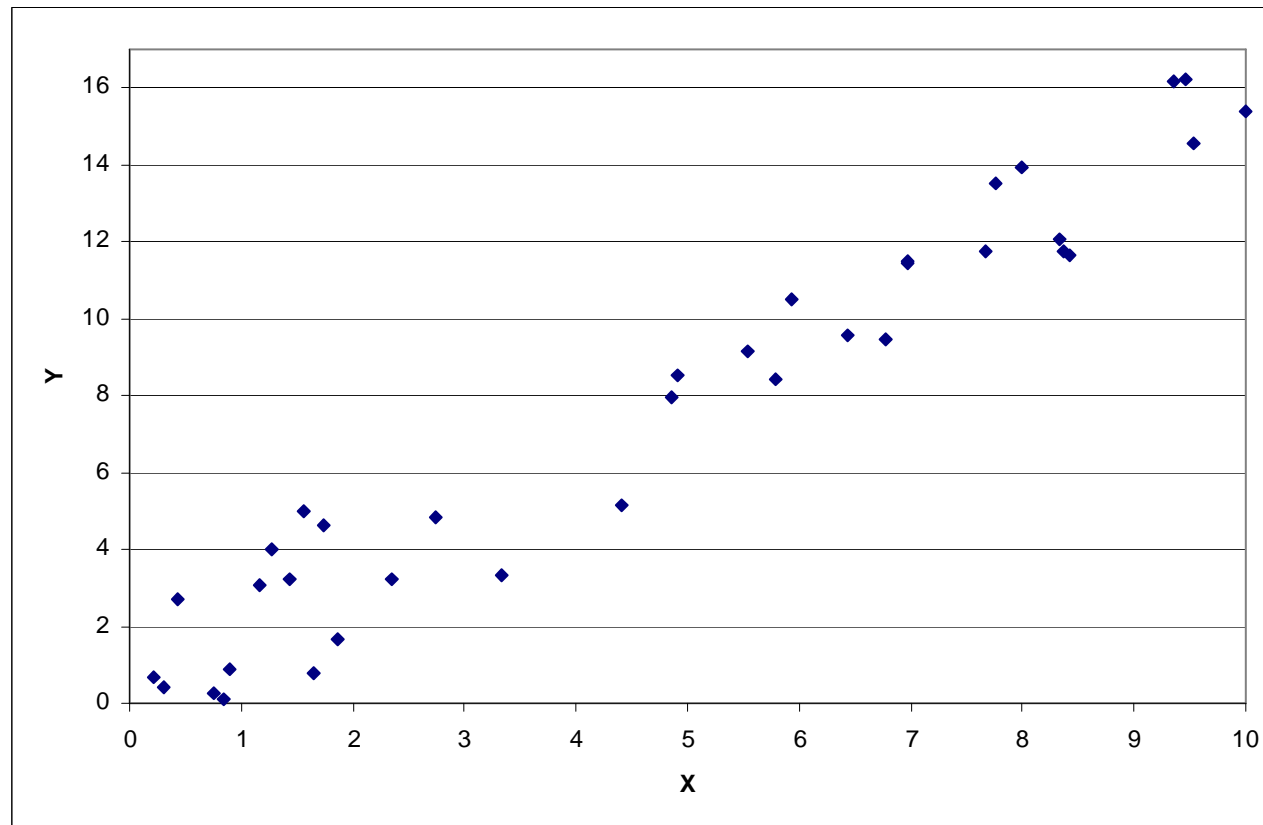
- Aproximaciones computacionales:
 - Algoritmos numéricos generales para la minimización de una función
 - Basados en el gradiente; algoritmos numéricos generales para hallar raíces; algoritmos que aprovechan la forma de la función
 - Algoritmos con una aproximación basada en la inteligencia artificial: algoritmos genéticos
- **Solución analítica:** mínimos cuadrados lineal

La aproximacion de funciones

- Al realizar la aproximacion de una funcion, sólo están disponibles un número finito de muestras

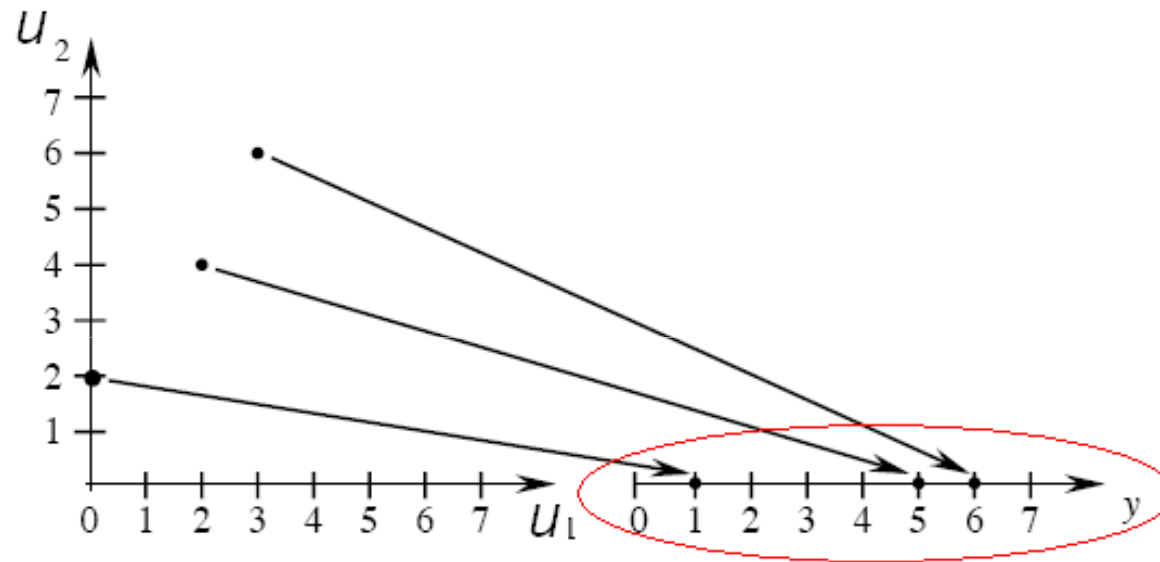
$$Z^N = \left\{ (u(1), \hat{y}(1)), \dots, (u(k), \hat{y}(k)), \dots, (u(N), \hat{y}(N)) \right\}$$

Ejemplo: una entrada, una salida



¿Como podemos modelar el proceso que genera estos datos?

Ejemplo: dos entradas, una salida



$$Z^T = \left[\begin{bmatrix} 0 \\ 2 \end{bmatrix}, 1 \quad \begin{bmatrix} 2 \\ 4 \end{bmatrix}, 5 \quad \begin{bmatrix} 3 \\ 6 \end{bmatrix}, 6 \right]$$

Modelos lineales vs . No lineales

- Es común asumir que $f(u)$ pertenece a una familia de funciones que comparten la misma *estructura* y difieren por los valores tomados por ciertos *parámetros* θ .

$$y = f(u, \theta)$$

El modelo lineal

- Un modelo lineal asume que la función es lineal respecto a los *parámetros* θ

$$f(u, \theta) = \theta_1 f_1(u) + \theta_2 f_2(u) + \cdots + \theta_q f_q(u)$$

Aquí, la linealidad se refiere a “con respecto a los parámetros”

Modelos no-lineales

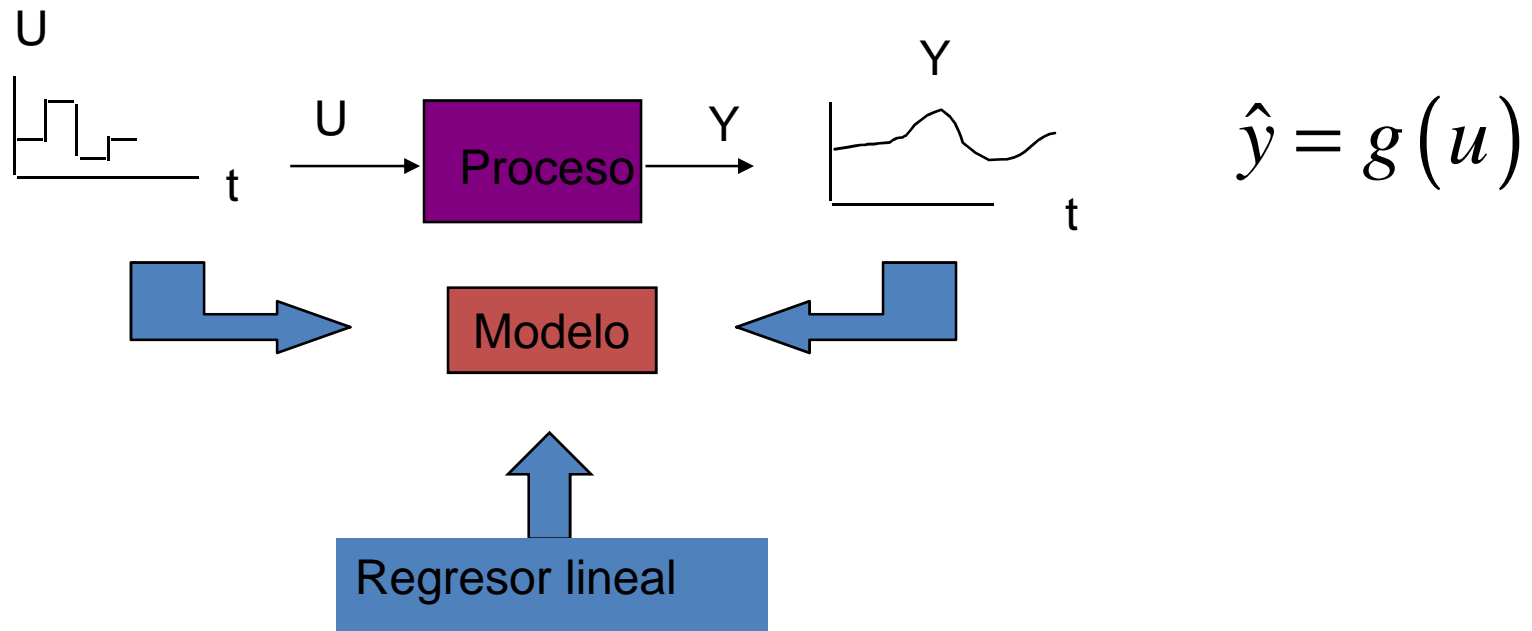
- En los modelos no-lineales la función es no-lineal respecto a los *parámetros* θ

$$f(u, \theta) = \exp(-u\theta)$$

Estimación de Mínimos Cuadrados Lineal

Dada una colección finita de observaciones

$$Z^N = \{u(0), y(0), u(1), y(1), \dots, u(N), y(N)\}$$



Mínimos Cuadrados

- El método de regresión de mínimos cuadrados consiste en encontrar la curva o función que mejor se ajuste a una serie de puntos (X_i, Y_i) , obtenidos generalmente a partir de un experimento.
- La estrategia consiste en minimizar las diferencias entre la función y los datos observados.
- El caso o ejemplo más sencillo es el ajuste de una función lineal a la serie de puntos.

Regresión lineal

Se asume que la relación entrada-salida puede ser descrita por una estructura de **regresor lineal**

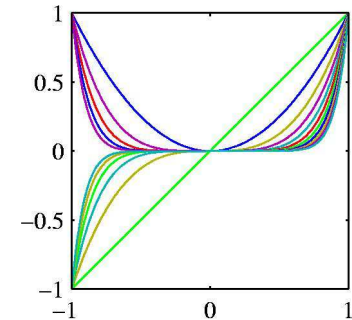
$$f(u, \theta) = \theta_1 f_1(u) + \theta_2 f_2(u) + \cdots + \theta_q f_q(u) + \varepsilon$$

$f(u, \theta)$ es denominada la *funcion de ajuste*. Las $f_i(u)$ son denominadas las *funciones base*

Algunas funciones base

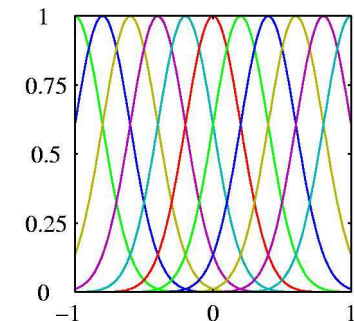
- Funciones polinomiales

$$f_j(u) = u^j$$



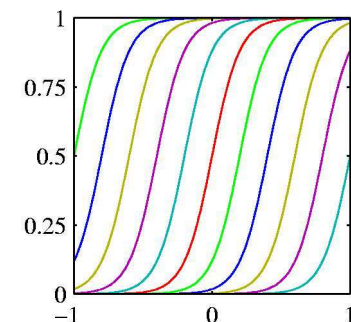
- Funciones base Gaussianas

$$f_j(u) = \exp\left(-\frac{(u - \mu_j)^2}{2\sigma^2}\right)$$



- Funciones base Sigmoidales

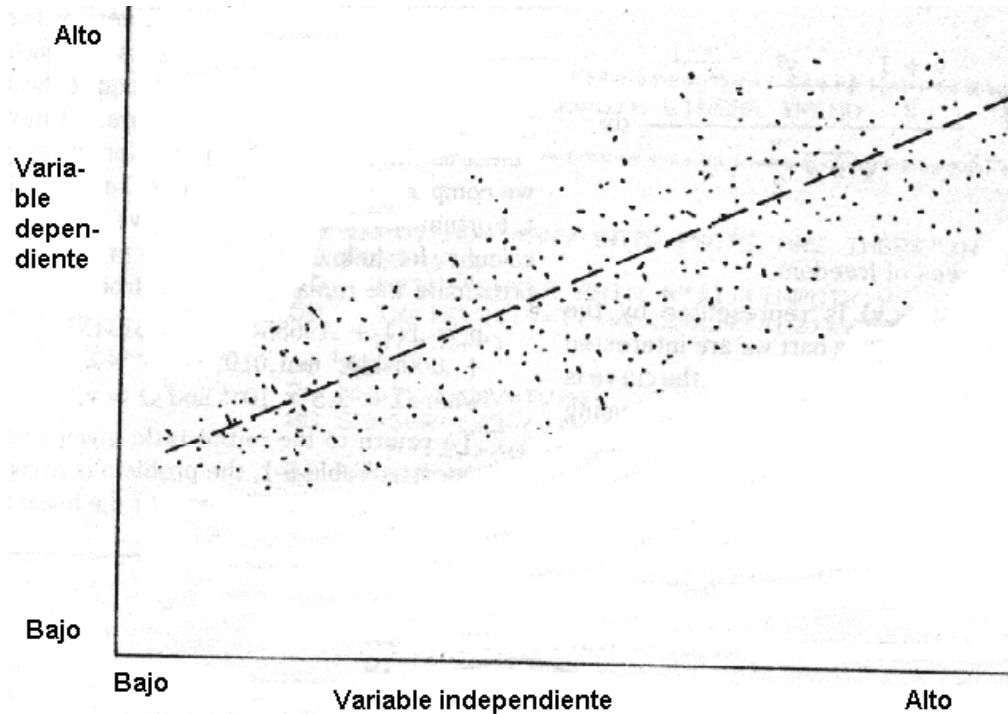
$$f_j(u) = \frac{1}{1 + \exp(-a)}$$



- Fourier \rightarrow wavelets

Regresión Lineal

$$Y = A + BX + e$$



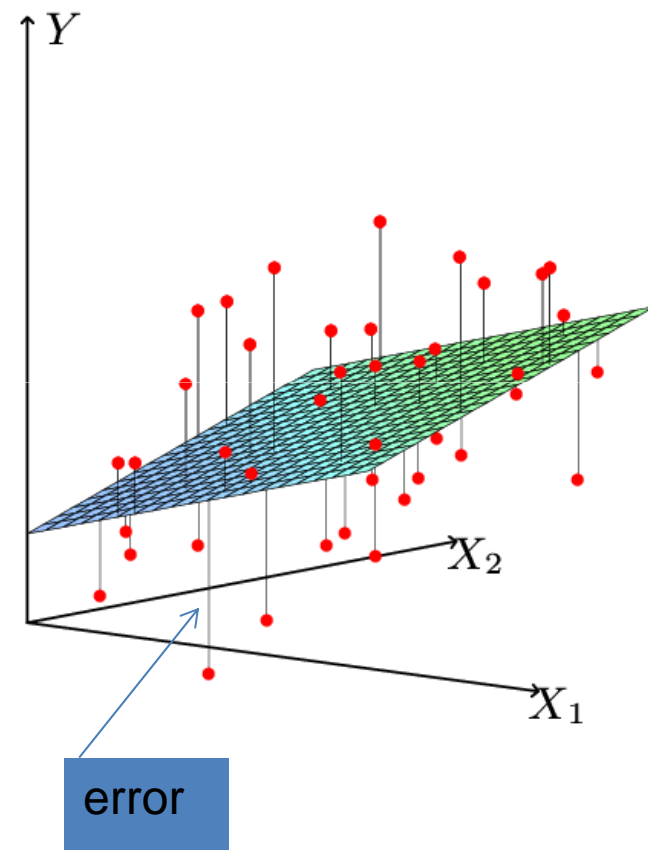
e representa las diferencias entre el modelo lineal y las observaciones

La gráfica muestra el ajuste de la nube de puntos a una línea recta

Como los datos X , Y son conocidos, el objetivo es entonces encontrar los mejores valores para los coeficientes A , B , tal que $e \rightarrow 0$.

Los errores cometidos

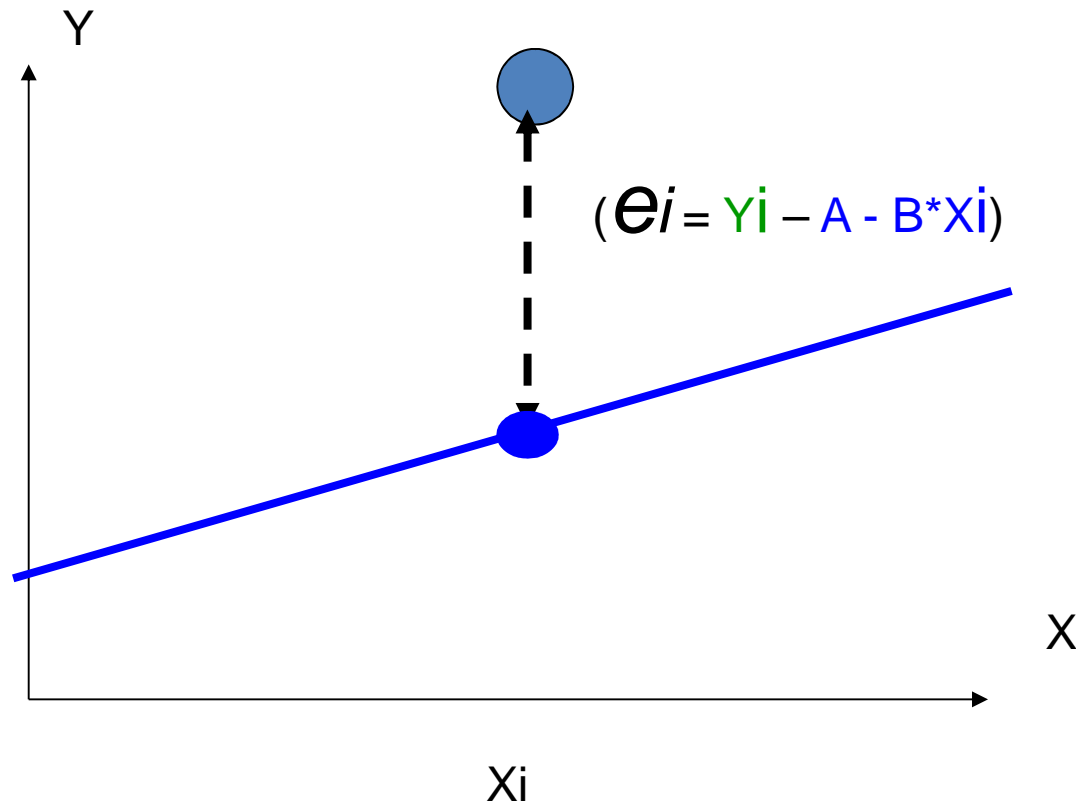
- Dados unos datos y el modelo lineal, deseamos calcular los “mejores” parámetros.
- Queremos minimizar los errores.



Error en la aproximación

El objetivo es minimizar el error cometido con la aproximación

Éste se representa como la distancia entre el valor real y el aproximado



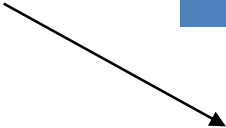
Los residuos

- El ajuste de minimos cuadrados halla el vector de parametros ϑ tal que se minimiza

$$J = \frac{1}{N} \sum_{k=1}^N \varepsilon^2(k)$$

residuos

= errores


$$\varepsilon(k) = \hat{y}(k) - f(u(k), \theta)$$

Criterio para el mejor ajuste

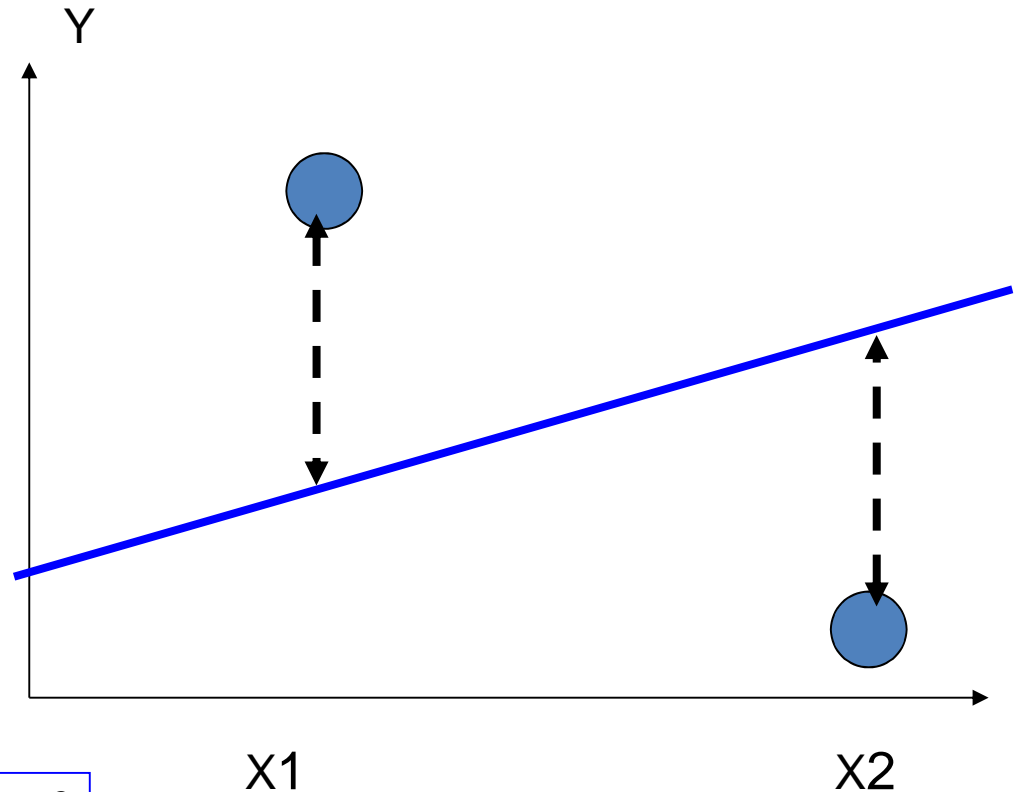
- Como se tiene una serie de n puntos (X_i, Y_i) ($i=1, \dots, n$), la acumulación de los errores será:

$$\sum_{i=1}^n e_i = \sum_{i=1}^n (Y_i - A - BX_i)$$

- Para que los valores de error positivos y negativos no se cancelen entre sí, éstos se deben elevar al cuadrado

Cancelación de errores...

- Para este ejemplo de dos puntos, los errores e_1 y e_2 se cancelan.
- La suma de los errores = 0



$$\sum_{i=1}^n e_i^2 = \sum_{i=1}^n (Y_i - A - BX_i)^2$$

Regresión Lineal

- Sea:

$$S = \sum_{i=1}^n e^2_i = \sum_{i=1}^n (Y_i - A - BX_i)^2$$

- Como el objetivo es encontrar A y B , tal que S sea mínimo, para esto se deriva S parcialmente con respecto a A y B respectivamente y se igualan a cero

Sistema de ecuaciones para encontrar A y B

- Las derivadas parciales de S con respecto a A y a B se hacen cero, así:

$$\frac{\partial S}{\partial A} = \sum 2(Y - A - BX)(-1) = 0 \quad (1)$$

$$\frac{\partial S}{\partial B} = \sum 2(Y - A - BX)(-X) = 0 \quad (2)$$

Regresión Lineal - Fórmulas

En resumen las fórmulas para calcular los coeficientes **A** y **B** de una función lineal de regresión con sólo dos tipos de variables **X** y **Y** son:

$$B = \frac{\sum_{i=1}^N (X_i - \bar{X})(Y_i - \bar{Y})}{\sum_{i=1}^N (X_i - \bar{X})^2} \quad ; \quad \bar{X} = \frac{1}{N} \sum_{i=1}^N X_i \quad ; \quad \bar{Y} = \frac{1}{N} \sum_{i=1}^N Y_i$$

$$A = \bar{Y} - B \bar{X}$$

Regresión Lineal - Algoritmo

Entrada: Número de datos n , datos (x,y)

$sumx, sumy, sumxy, sumx2 = 0$

$i=0$

Mientras $i \leq n-1$

$sumx = sumx + x(i)$

$sumy = sumy + y(i)$

$sumx2 = sumx2 + (x(i) * x(i))$

$sumxy = sumxy + (x(i) * y(i))$

$i = i + 1$

$Denominador = sumx * sumy - n * sumx2$

$a = (sumx * sumy - n * sumxy) / Denominador$

$b = (sumx * sumxy - sumx2 * sumy) / Denominador$

Imprimir a y b

Regresión No Lineal

Hay ocasiones en las cuales la relación existente entre X y Y no es lineal, sin embargo ésta puede ser descrita por algún otro tipo de función. EJ:

$$\textit{Potencia} : Y = AX^B$$

$$\textit{Exponencial} : Y = Ae^{BX}$$

$$\textit{Logarítmica} : Y = A + B\text{Log}(X); Y = B\text{Log}(X) + \text{Log}(A)$$

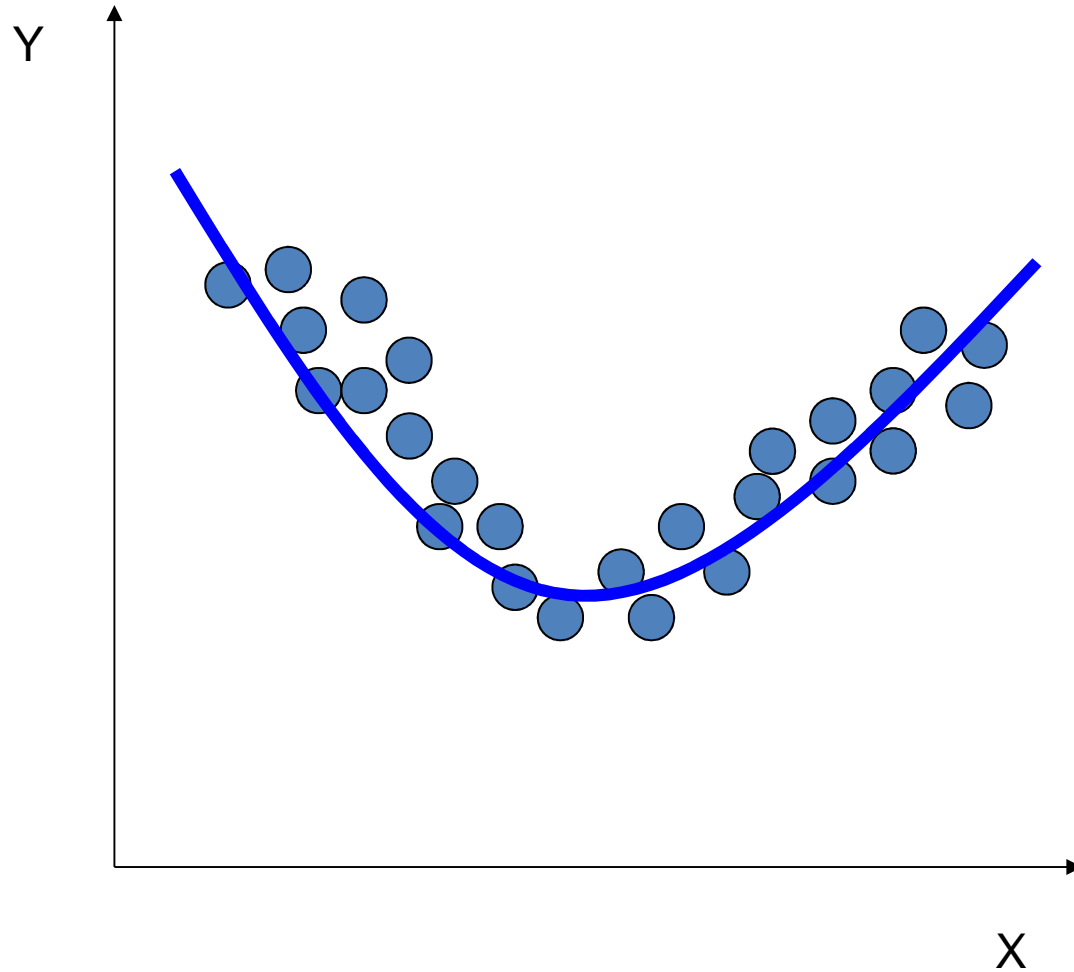
$$\text{Log}_e(Y) = BX + \text{Log}_e(A)$$

$$\textit{Polinómica} : Y = A_0 + A_1X + A_2X^2 + \dots + A_NX^N$$

$$\textit{Parabólica} : Y = A + BX + CX^2$$

3. Regresión No Lineal

- Relación no-lineal entre las variables X y Y .
- Posiblemente parabólica???



Regresión no-lineal: Potencia:

$$Y = AX^B$$

$$B = \frac{\sum_{i=1}^N (\text{Log}(X_i) * \text{Log}(Y_i)) - \frac{\left(\sum_{i=1}^N \text{Log}(X_i)\right)\left(\sum_{i=1}^N \text{Log}(Y_i)\right)}{N}}{\sum_{i=1}^N \text{Log}(X_i)^2 - \frac{\left(\sum_{i=1}^N \text{Log}(X_i)\right)^2}{N}};$$

$$A = \exp\left(\frac{\sum_{i=1}^N \text{Log}(Y_i)}{N} - B \frac{\sum_{i=1}^N \text{Log}(X_i)}{N}\right)$$

Regresión no-lineal: Exponencial:

$$Y = Ae^{BX}$$

$$B = \frac{\sum_{i=1}^N (X_i \text{Log}(Y_i)) - \frac{1}{N} \left(\sum_{i=1}^N X_i \right) \left(\sum_{i=1}^N \text{Log}(Y_i) \right)}{\sum_{i=1}^N X_i^2 - \frac{1}{N} \left(\sum_{i=1}^N X_i \right)^2} ;$$

$$A = \exp \left[\frac{\sum_{i=1}^N \text{Log}(Y_i)}{N} - B \frac{\sum_{i=1}^N X_i}{N} \right]$$