Matemática Superior Aplicada

Regresión Lineal

Prof.: Dr. Alejandro S. M. Santa Cruz

J.T.P.: Dr. Juan Ignacio Manassaldi

J.T.P.: Ing. Amalia Rueda

Pasos a Seguir

Experimentos - Medición

Recopilación y procesamiento de datos

Modelo matemático

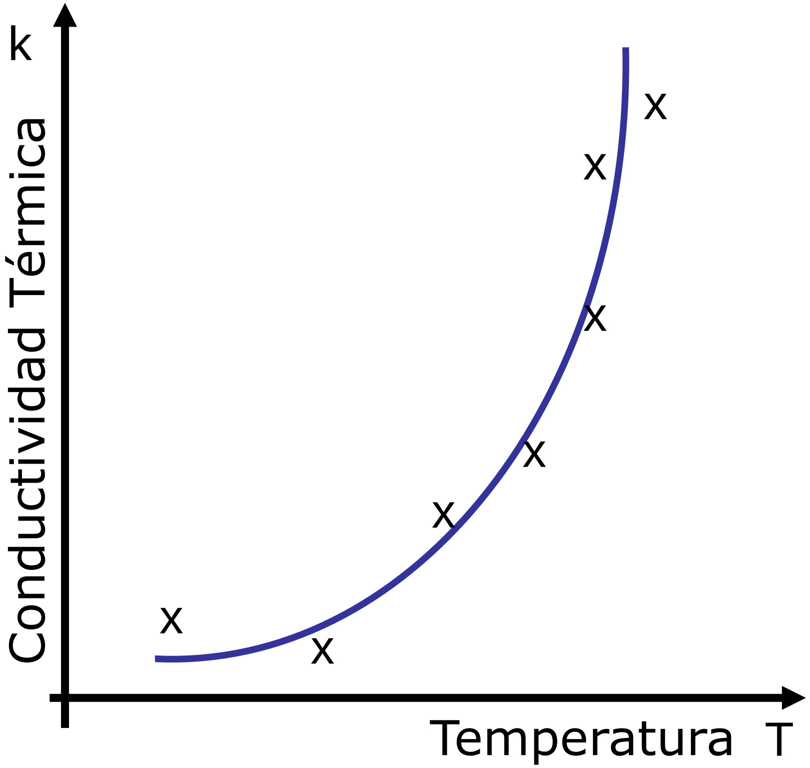
$$E = mc^2$$

$$y = a + bc$$

$$P = \frac{RT}{v-b} - \frac{a}{v^2}$$

Estimación de los parámetros del modelo

Regresión Lineal



T	k
T_1	k_1
T_2	k_2
$\vdots$	$\vdots$
T_m	k_m

$$k = c + bT + aT^2$$

El análisis de regresión es una técnica estadística para investigar la relación funcional entre dos o más variables, ajustando algún modelo matemático.

Regresión Lineal

T	k
T_1	k_1
T_2	k_2
$\vdots$	$\vdots$
T_m	k_m

Datos

Por ejemplo:

T [ºC]	k [W/(m K)]
5	0.5694
15	0.5882
25	0.6106
35	0.6240
45	0.6345
55	0.6434
65	0.6569
75	0.6683
85	0.6741
95	0.6776

$$\mathrm{k} = c + b\mathrm{T} + a\mathrm{T}^2$$

Modelo

Regresión Lineal

T	k
T_1	k_1
T_2	k_2
$\vdots$	$\vdots$
T_m	k_m

Datos

$$k_1 = c + bT_1 + aT_1^{\,2}$$
$$k_2 = c + bT_2 + aT_2^{\,2}$$
$$k_3 = c + bT_3 + aT_3^{\,2}$$
$$k_4 = c + bT_4 + aT_4^{\,2}$$
$$\vdots \quad \vdots \qquad \qquad \vdots$$
$$k_m = c + bT_m + aT_m^{\,2}$$

residuo

m ecuaciones

m + 3 incógnitas

$$\mathrm{k} = c + b\mathrm{T} + a\mathrm{T}^2$$

Modelo

Regresión Lineal

$$\begin{aligned}
r_1 + k_1 &= c + bT_1 + aT_1^{\,2} \\
r_2 + k_2 &= c + bT_2 + aT_2^{\,2} \\
r_3 + k_3 &= c + bT_3 + aT_3^{\,2} \\
r_4 + k_4 &= c + bT_4 + aT_4^{\,2} \\
\vdots \quad \vdots \quad &\vdots \qquad\qquad \vdots \\
r_m + k_m &= c + bT_m + aT_m^{\,2}
\end{aligned}$$

$$\begin{aligned}
r_1 &= c + bT_1 + aT_1^{\,2} - k_1 \\
r_2 &= c + bT_2 + aT_2^{\,2} - k_2 \\
r_3 &= c + bT_3 + aT_3^{\,2} - k_3 \\
r_4 &= c + bT_4 + aT_4^{\,2} - k_3 \\
\vdots \quad &\vdots \qquad\qquad \vdots \\
r_m &= c + bT_m + aT_m^{\,2} - k_m
\end{aligned}$$

Regresión Lineal

$$
\begin{aligned}
r_1 &= c + bT_1 + aT_1^{\,2} - k_1 \\
r_2 &= c + bT_2 + aT_2^{\,2} - k_2 \\
r_3 &= c + bT_3 + aT_3^{\,2} - k_3 \\
r_4 &= c + bT_4 + aT_4^{\,2} - k_3 \\
&\vdots \\
r_m &= c + bT_m + aT_m^{\,2} - k_m
\end{aligned}
$$

$$\underline{r} = \underline{\underline{A}}\,\underline{x} - \underline{y}$$

Parámetros de Modelización

$$\underline{x} = \begin{pmatrix} c \\ b \\ a \end{pmatrix}$$

$$\underline{\underline{A}} = \begin{pmatrix} 1 & T_1 & T_1^{\,2} \\ 1 & T_2 & T_2^{\,2} \\ \vdots & \vdots & \vdots \\ 1 & T_m & T_m^{\,2} \end{pmatrix}$$

Matriz de las Funciones de Modelización

$$\begin{pmatrix} c + bT_1 + aT_1^{\,2} \\ c + bT_2 + aT_2^{\,2} \\ \vdots \\ c + bT_m + aT_m^{\,2} \end{pmatrix}$$

Regresión Lineal

$$r_1 = c + bT_1 + aT_1^{\,2} - k_1$$

$$r_2 = c + bT_2 + aT_2^{\,2} - k_2$$

$$\vdots \quad \vdots \qquad\qquad \vdots$$

$$r_m = c + bT_m + aT_m^{\,2} - k_m$$

$$\underline{r} = \underline{\underline{A}}\,\underline{x} - \underline{y} \Rightarrow$$ Residuo

$$\left\|\underline{r}\right\|^2 = \left\|\underline{\underline{A}}\,\underline{x} - \underline{y}\right\|^2 \Rightarrow$$ El objetivo es minimizar la norma del residuo al cuadrado

Regresión Lineal

Matemática Superior Aplicada

UTN

$$r_1 = c + bT_1 + aT_1^{\,2} - k_1$$

$$r_2 = c + bT_2 + aT_2^{\,2} - k_2$$

$$\vdots \quad \vdots \qquad\qquad \vdots$$

$$r_m = c + bT_m + aT_m^{\,2} - k_m$$

$$\left\|\underline{r}\right\|^2 = \left\|\underline{\underline{A}}\,\underline{x} - \underline{y}\right\|^2$$

$$\underline{r} = \underline{\underline{A}}\,\underline{x} - \underline{y}$$

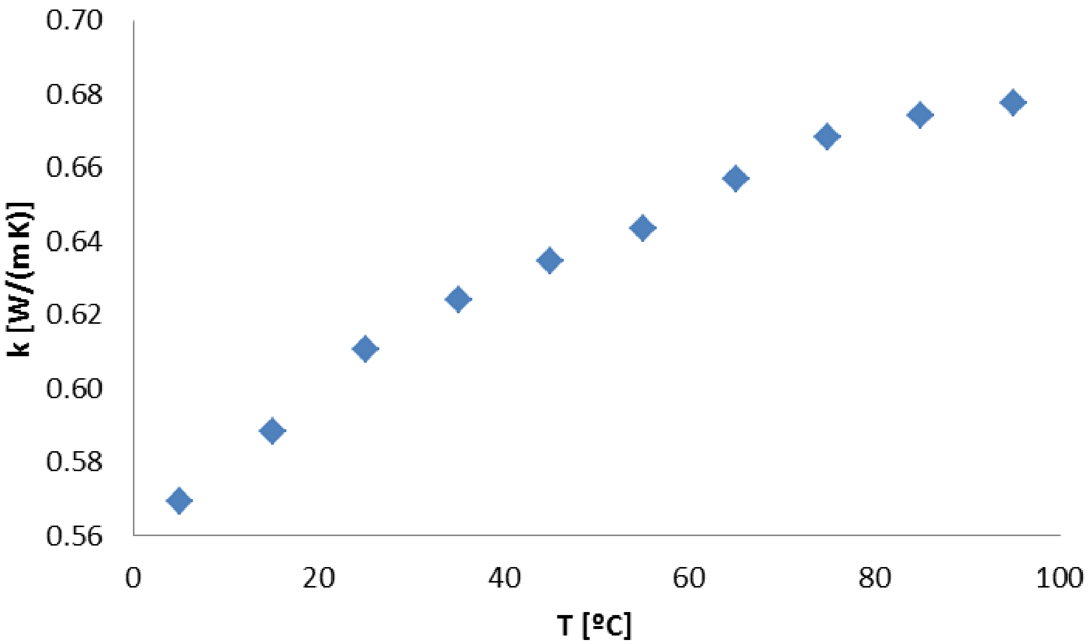
Residuo

$$\min_{\|x\|\neq 0} \left\|\underline{r}\right\|^2 = \min_{\|x\|\neq 0} \left\|\underline{\underline{A}}\,\underline{x} - \underline{y}\right\|^2$$

Problema de mínimos cuadrados

- A cada valor de *x* le corresponde un valor de la norma al cuadrado del residuo.
- El problema de mínimos cuadrados encuentra el valor de x para el cual la norma al cuadrado del vector residuo resulta lo mas pequeña posible.

Ejemplo: k vs T (Agua líquida 1 atm)



T [ºC]	k [W/(m K)]
5	0.5694
15	0.5882
25	0.6106
35	0.6240
45	0.6345
55	0.6434
65	0.6569
75	0.6683
85	0.6741
95	0.6776

$$\mathrm{k} = c + b\mathrm{T} + a\mathrm{T}^2$$

$$\boxed{\underline{r} = \underline{\underline{A}}\,\underline{x} - \underline{y} \quad \text{Residuo}}$$

Ejemplo: Problema de mínimos cuadrados

$$\min_{\|x\|\neq 0} \|\underline{r}\|^2 = \min_{\|x\|\neq 0} \left\|\underline{\underline{A}}\,\underline{x} - \underline{y}\right\|^2$$

$$\underline{\underline{A}} = \begin{bmatrix} 1 & 5 & 5^2 \\ 1 & 15 & 15^2 \\ 1 & 25 & 25^2 \\ 1 & 35 & 35^2 \\ 1 & 45 & 45^2 \\ 1 & 55 & 55^2 \\ 1 & 65 & 65^2 \\ 1 & 75 & 75^2 \\ 1 & 85 & 85^2 \\ 1 & 95 & 95^2 \end{bmatrix} \quad \underline{x} = \begin{bmatrix} c \\ b \\ a \end{bmatrix} \quad \underline{y} = \begin{bmatrix} 0.5694 \\ 0.5882 \\ 0.6106 \\ 0.6240 \\ 0.6345 \\ 0.6434 \\ 0.6569 \\ 0.6683 \\ 0.6741 \\ 0.6776 \end{bmatrix}$$

Ecuaciones Normales

$$\min_{\|x\|\neq 0} \|\underline{r}\|^2 = \min_{\|x\|\neq 0} \|\underline{\underline{A}}\,\underline{x} - \underline{y}\|^2$$

$$\underline{\underline{A}} = \begin{bmatrix} 1 & 5 & 5^2 \\ 1 & 15 & 15^2 \\ 1 & 25 & 25^2 \\ 1 & 35 & 35^2 \\ 1 & 45 & 45^2 \\ 1 & 55 & 55^2 \\ 1 & 65 & 65^2 \\ 1 & 75 & 75^2 \\ 1 & 85 & 85^2 \\ 1 & 95 & 95^2 \end{bmatrix} \quad \underline{x} = \begin{bmatrix} c \\ b \\ a \end{bmatrix} \quad \underline{y} = \begin{bmatrix} 0.5694 \\ 0.5882 \\ 0.6106 \\ 0.6240 \\ 0.6345 \\ 0.6434 \\ 0.6569 \\ 0.6683 \\ 0.6741 \\ 0.6776 \end{bmatrix}$$

$$\underbrace{\underline{\underline{A}}^T\,\underline{\underline{A}}}_{\underline{\underline{ata}}}\,\underline{x} = \underbrace{\underline{\underline{A}}^T\,\underline{y}}_{\underline{aty}}$$

$$\underline{\underline{ata}}\,\underline{x} = \underline{aty}$$

La solución a este SEAL minimiza la norma dos al cuadrado del residuo. (teoría)

Ecuaciones Normales

$$\underset{=}{A}^T \underset{=}{A} \, \underline{x} = \underset{=}{A}^T \underline{y}$$

$\underset{=}{ata}$ $\qquad$ $\underline{aty}$

$$\underset{=}{A} = \begin{bmatrix} 1 & 5 & 5^2 \\ 1 & 15 & 15^2 \\ 1 & 25 & 25^2 \\ 1 & 35 & 35^2 \\ 1 & 45 & 45^2 \\ 1 & 55 & 55^2 \\ 1 & 65 & 65^2 \\ 1 & 75 & 75^2 \\ 1 & 85 & 85^2 \\ 1 & 95 & 95^2 \end{bmatrix} \qquad \underline{y} = \begin{bmatrix} 0.5694 \\ 0.5882 \\ 0.6106 \\ 0.6240 \\ 0.6345 \\ 0.6434 \\ 0.6569 \\ 0.6683 \\ 0.6741 \\ 0.6776 \end{bmatrix}$$

$$\underbrace{\begin{bmatrix} 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 5 & 15 & 25 & 35 & 45 & 55 & 65 & 75 & 85 & 95 \\ 5^2 & 15^2 & 25^2 & 35^2 & 45^2 & 55^2 & 65^2 & 75^2 & 85^2 & 95^2 \end{bmatrix}}_{\underset{=}{A}^T}$$

$$\underbrace{\begin{bmatrix} 10 & 500 & 33250 \\ 500 & 33250 & 2487500 \\ 33250 & 2487500 & 198336250 \end{bmatrix}}_{\underset{=}{ata}} \qquad \underbrace{\begin{bmatrix} 6.347 \\ 327.206 \\ 22044.055 \end{bmatrix}}_{\underline{aty}}$$

Ecuaciones Normales

$$\underline{\underline{ata}}\,\underline{x} = \underline{aty}$$

$$\begin{bmatrix} 10 & 500 & 33250 \\ 500 & 33250 & 2487500 \\ 33250 & 2487500 & 198336250 \end{bmatrix} \begin{bmatrix} c \\ b \\ a \end{bmatrix} = \begin{bmatrix} 6.347 \\ 327.206 \\ 22044.055 \end{bmatrix}$$

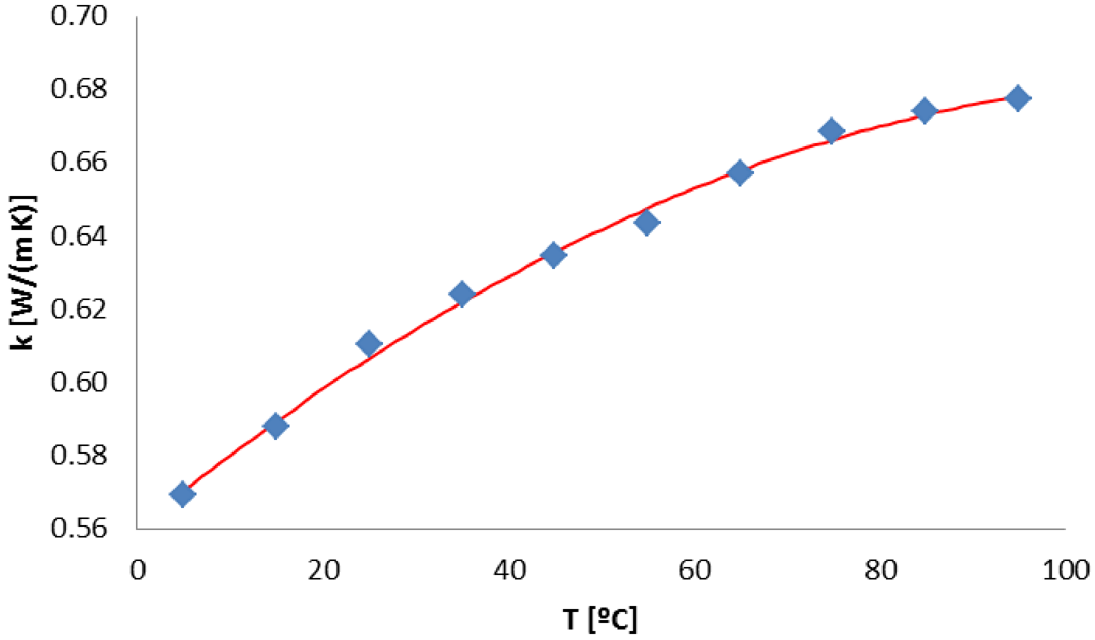
3x3

¡Ya podemos resolver este sistema!

Solución del ejemplo

```
//Ecuaciones Normales
T=[5 15 25 35 45 55 65 75 85 95]';
k=[0.5694 0.5882 0.6106 0.6240 0.6345 0.6434 0.6569 0.6683 0.6741 0.6776]';
A=[ones(10,1)  T  T.^2];
y=k;
ata=A'*A;
aty=A'*y;
x=ata\aty;
c=x(1)
b=x(2)
a=x(3)
```

Solución del ejemplo



T [ºC]	k [W/(m K)]
5	0.5694
15	0.5882
25	0.6106
35	0.6240
45	0.6345
55	0.6434
65	0.6569
75	0.6683
85	0.6741
95	0.6776

$$k = 0.56059 + 0.002053T - 8.58333e - 6T^2$$

Solución del ejemplo

```
--> r=A*x-y
 r  =


   0.00124
   0.0012533
  -0.00405
  -0.00207
   0.0010933
   0.00414
   0.00087
  -0.0020167
  -0.00102
   0.00056



--> norm(r)^2
 ans  =


   0.0000483
```

Descomposición QR

$\underline{\underline{A}} = \underline{\underline{Q}}\,\underline{\underline{R}}$ ➜ Algoritmo de descomposición QR

$$\underbrace{n\left\{\overbrace{\begin{matrix} a_{11} & \cdots & a_{1p} \\ a_{21} & \cdots & a_{2p} \\ \vdots & \ddots & \vdots \\ \vdots & \ddots & \vdots \\ a_{n1} & \cdots & a_{np} \end{matrix}}^{p}\right.}_{\underline{\underline{A}}(n\times p)} = \overbrace{\underbrace{\begin{bmatrix} q_{11} & \cdots & q_{1n} \\ \vdots & \ddots & \vdots \\ q_{n1} & \cdots & q_{nn} \end{bmatrix}}_{\underline{\underline{Q}}(n\times n)}}^{\text{Ortogonal}} \underbrace{\overbrace{\begin{bmatrix} r_{11} & r_{12} & \cdots & r_{1\,p-1} & r_{1\,p} \\ 0 & r_{22} & \cdots & r_{2\,p-1} & r_{2\,p} \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & r_{p-1\,p-1} & r_{p-1\,p} \\ 0 & 0 & \cdots & 0 & r_{pp} \\ 0 & 0 & \cdots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}}^{\text{Triangular Superior}}}_{\underline{\underline{R}}(n\times p)}$$

Descomposición QR

$$\underline{\underline{A}} = \underline{\underline{Q}}\,\underline{\underline{R}}$$

Ortogonal

$$\underline{\underline{A}}\,\underline{x} = \underline{y} \rightarrow \underline{\underline{Q}}\,\underline{\underline{R}}\,\underline{x} = \underline{y} \rightarrow \boxed{\underline{\underline{Q}}^t\,\underline{\underline{Q}}}\,\underline{\underline{R}}\,\underline{x} = \underline{\underline{Q}}^t\,\underline{y}$$

$$\underline{\underline{R}}\,\underline{x} = \underline{\underline{Q}}^t\,\underline{y}$$

Debemos resolver este sistema

La solución a este SEAL también minimiza la norma dos al cuadrado del residuo.
(<u>teoría</u>)

Descomposición QR

$$\underline{\underline{R}}\,\underline{x} = \underline{\underline{Q}}^t\,\underline{y}$$

Resolvemos el sistema con los elementos no nulos de la matriz R. Observamos que se puede aplicar sustitución hacia atrás.

$$\underbrace{\begin{bmatrix} r_{11} & r_{12} & \cdots & r_{1\,p-1} & r_{1\,p} \\ 0 & r_{22} & \cdots & r_{2\,p-1} & r_{2\,p} \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & r_{p-1\,p-1} & r_{p-1\,p} \\ 0 & 0 & \cdots & 0 & r_{pp} \\ 0 & 0 & \cdots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}}_{\underline{\underline{R}}(n\times p)} \underbrace{\begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_{p-1} \\ x_p \end{bmatrix}}_{\underline{x}} = \underbrace{\begin{bmatrix} qty_1 \\ qty_2 \\ \vdots \\ qty_{p-1} \\ qty_p \\ qty_{p+1} \\ \vdots \\ qty_n \end{bmatrix}}_{\underline{\underline{Q}}'\underline{y}}$$

¡Podemos calcular el Residuo!

Descomposición QR

$$\underbrace{\begin{bmatrix} qty_1 \\ qty_2 \\ \vdots \\ qty_{p-1} \\ qty_p \\ \boxed{\begin{matrix} qty_{p+1} \\ \vdots \\ qty_N \end{matrix}} \end{bmatrix}}_{\underline{\underline{Q}}'\underline{y}}$$

$$R = \|Ax - y\|^2 = \left\| \begin{pmatrix} qty_{p+1} & qty_{p+2} & \cdots & qty_N \end{pmatrix} \right\|^2$$

Estos elementos se utilizan para obtener el valor del residuo sin tener que resolver el sistema.

Descomposición QR

```
[n p]=size(A);
[Q,R]=qr(A)
qty=Q'*y
x=s_atras(R(1:p,:),qty(1:p))
Residuo=norm(qty(p+1:n))^2
```

$$\underline{\underline{A}} = \underline{\underline{U}}\,\underline{\underline{\Sigma}}\,\underline{\underline{V}}^t$$

Solo con elementos en la diagonal principal

$$N\left\{\underbrace{\overbrace{\begin{matrix} a_{11} & \cdots & a_{1p}\\ a_{21} & \cdots & a_{2p}\\ \vdots & \ddots & \vdots\\ \vdots & \ddots & \vdots\\ a_{n1} & \cdots & a_{np}\end{matrix}}^{p}}_{\underline{\underline{A}}(n\times p)}\right. = \underbrace{\begin{bmatrix} u_{11} & \cdots & u_{1n}\\ \vdots & \ddots & \vdots\\ u_{n1} & \cdots & u_{nn}\end{bmatrix}}_{\underline{\underline{U}}(N\times N)} \underbrace{\begin{bmatrix} d_{11} & 0 & \cdots & 0\\ 0 & d_{22} & \cdots & 0\\ \vdots & \vdots & \ddots & \vdots\\ 0 & 0 & \cdots & d_{pp}\\ 0 & 0 & \cdots & 0\\ \vdots & \vdots & \vdots & \vdots\\ 0 & 0 & \cdots & 0\end{bmatrix}}_{\underline{\underline{\Sigma}}(n\times p)} \underbrace{\begin{bmatrix} v_{11} & \cdots & v_{1p}\\ \vdots & \ddots & \vdots\\ v_{p1} & \cdots & v_{pp}\end{bmatrix}}_{\underline{\underline{V}}^T(p\times p)}$$

Ortogonales

Descomposición en Valores Singulares

$$\underline{\underline{A}} = \underline{\underline{U}}\,\underline{\underline{\Sigma}}\,\underline{\underline{V}}\,'$$

$$\underline{\underline{A}}\,\underline{x} = \underline{y} \quad \Rightarrow \quad \underline{\underline{U}}\,\underline{\underline{\Sigma}}\,\underline{\underline{V}}\,'\underline{x} = \underline{y} \quad \Rightarrow \quad \underline{\underline{U}}\,'\underline{\underline{U}}\,\underline{\underline{\Sigma}}\,\underline{\underline{V}}\,'\underline{x} = \underline{\underline{U}}\,'\underline{y}$$

Ortogonal

$$\underline{\underline{\Sigma}}\,\underbrace{\underline{\underline{V}}\,'\underline{x}}_{\underline{z}} = \underbrace{\underline{\underline{U}}\,'\underline{y}}_{\underline{d}} \quad \Rightarrow \quad \underline{\underline{\Sigma}}\,\underline{z} = \underline{d} \quad \Rightarrow \quad \underline{z}$$

Obtengo el vector z de manera directa

$$\underline{z} = \underline{\underline{V}}\,'\underline{x} \quad \Rightarrow \quad \underline{x} = \underline{\underline{V}}\,\underline{z}$$

La solución de esta secuencia también minimiza la norma dos al cuadrado del residuo. (teoría)

```
//Descomposición en Valores Singulares
[n p]=size(A);
[U,S,V]=svd(A)
d=U'*y
z=d(1:p)./diag(S)
x=V*z
```