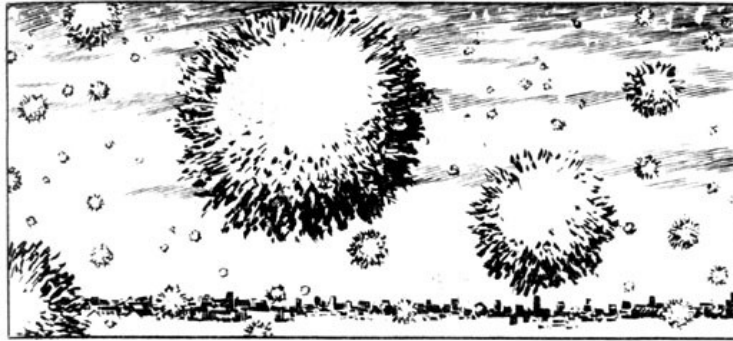

Álgebra II (Curso 23)
Segundo cuatrimestre, 2021
NOTAS EN LA EMERGENCIA SANITARIA:
BORRADORES PARA LA CLASE DEL 25 DE OCTUBRE
Sebastian GRYNBERG



*El único héroe válido es el héroe “en grupo”,
nunca el héroe individual, el héroe solo.*

H. G. OESTERHELD

ÍNDICE

1. Introducción	2
1.1. Geometría de los subespacios fundamentales de una matriz	2
1.2. Ecuaciones lineales	3
2. Mínimos cuadrados	4
2.1. Introducción al problema	4
2.2. Soluciones por mínimos cuadrados	5
2.3. La solución de norma mínima	6
2.4. Matrices de rango máximo	8
3. Ajuste de datos	9
3.1. Regresión polinomial	10
3.2. Recta de regresión	12

1. INTRODUCCIÓN

En todo lo que sigue $A \in \mathbb{R}^{m \times n}$ será una matriz de m filas por n columnas con coeficientes reales. Los espacios \mathbb{R}^n y \mathbb{R}^m serán considerados con la estructura geométrica provista por sus productos internos canónicos.

1.1. Geometría de los subespacios fundamentales de una matriz.

Repaso. En la Guía 1 se trataron los problemas relacionados con la resolución de la ecuación lineal $Ax = b$. El análisis de la ecuación y de su conjunto de soluciones se realiza mediante los *cuatro espacios fundamentales de la matriz A*:

$$\begin{aligned} \text{col}(A) &= \{Ax : x \in \mathbb{R}^n\} \subset \mathbb{R}^m, \\ \text{nul}(A) &= \{x \in \mathbb{R}^n : Ax = 0\} \subset \mathbb{R}^n, \\ \text{col}(A^T) &= \{A^T y : y \in \mathbb{R}^m\} \subset \mathbb{R}^n, \\ \text{nul}(A^T) &= \{y \in \mathbb{R}^m : A^T y = 0\} \subset \mathbb{R}^m. \end{aligned}$$

Estos cuatro subespacios permiten narrar la historia completa del sistema lineal $Ax = b$. Dos están en \mathbb{R}^n , el $\text{nul}(A)$ y el $\text{col}(A^T)$, y los otros dos en \mathbb{R}^m , el $\text{col}(A)$ y el $\text{nul}(A^T)$.

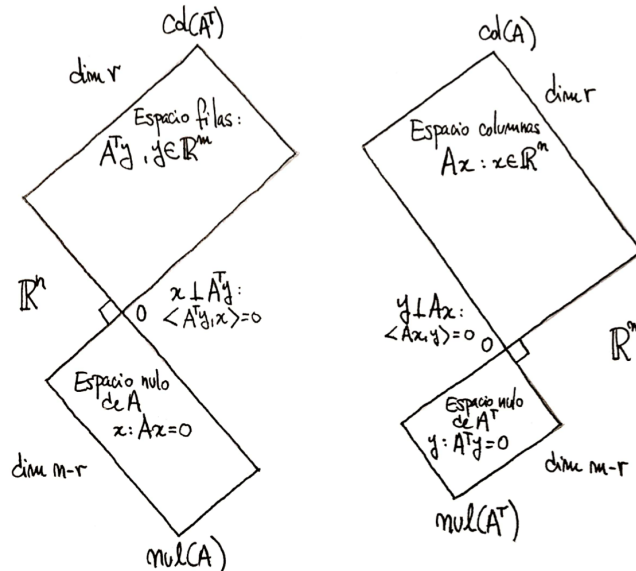


FIGURA 1. Dimensiones y relaciones de ortogonalidad entre espacios fundamentales de una matriz $A \in \mathbb{R}^{m \times n}$ de rango $\dim(\text{col}(A)) = r$.

Lema 1.1. Vale que $\text{nul}(A) = \text{col}(A^T)^\perp$. En consecuencia,

$$(1) \quad \mathbb{R}^n = \text{col}(A^T) \oplus \text{nul}(A).$$

Demostración. Una cadena de equivalencias con base en las definiciones

$$\begin{aligned}
 x \in \text{nul}(A) &\iff Ax = 0 \\
 &\iff y^T Ax = 0 \text{ para todo } y \in \mathbb{R}^m \\
 &\iff (A^T y)^T x = 0 \text{ para todo } y \in \mathbb{R}^m \\
 &\iff \langle x, A^T y \rangle = 0 \text{ para todo } y \in \mathbb{R}^m \\
 &\iff x \perp A^T y \text{ para todo } y \in \mathbb{R}^m \\
 &\iff x \in \text{col}(A^T)^\perp
 \end{aligned}$$

□

Nota Bene. Nótese que intercambiando papeles entre m y n y entre A y A^T se obtienen las relaciones $\text{nul}(A^T) = \text{col}(A)^\perp$ y $\mathbb{R}^m = \text{col}(A) \oplus \text{nul}(A^T)$.

Corolario 1.2. $\dim(\text{col}(A)) = \dim(\text{col}(A^T))$.

1.2. Ecuaciones lineales.

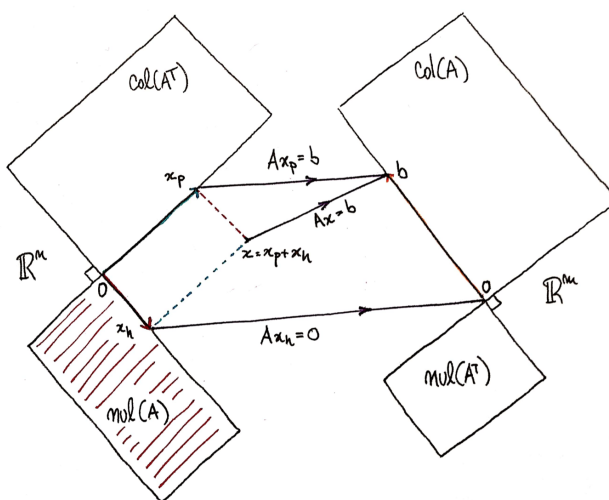


FIGURA 2. La acción de A sobre \mathbb{R}^n : el espacio filas de A se transforma en el espacio columnas, el nulo de A en el vector cero. Dado $b \in \text{col}(A)$, el sistema $Ax = b$ tiene solución. Hay una única solución particular $x_p \in \text{col}(A^T)$. La solución general es $x = x_p + x_h$, donde $x_h \in \text{nul}(A)$. La particularidad de x_p es que $x_p \perp \text{nul}(A)$.

Enfoque. Considerar \mathcal{B} una base de \mathbb{R}^n compuesta por la unión de una base $\mathcal{B}_1 = \{x_1, \dots, x_r\}$ de $\text{col}(A^T)$ y una base $\mathcal{B}_0 = \{x_{r+1}, \dots, x_n\}$ de $\text{nul}(A)$. El conjunto $\{Ax_1, \dots, Ax_r\}$ es una base de $\text{col}(A)$. Si $b \in \text{col}(A)$ existe un único $x_p \in \text{col}(A^T)$ tal que $Ax_p = b$.

La introducción del error $\varepsilon = y - [\varphi_j(x_i)]a$ y el criterio para determinar los coeficientes $a \in \mathbb{R}^n$, que minimicen el error cuadrático $\|\varepsilon\|^2 = \|y - [\varphi_j(x_i)]a\|^2$, permiten tratar y resolver el problema para cualquier conjunto de datos $y \in \mathbb{R}^m$.

2.2. Soluciones por mínimos cuadrados.

Definición 2.1. Sean $A \in \mathbb{R}^{m \times n}$ y $b \in \mathbb{R}^m$. Las soluciones por mínimos cuadrados de la ecuación $Ax = b$ son aquellos vectores $x \in \mathbb{R}^n$ que minimizan el error cuadrático $\|b - Ax\|^2$. En otras palabras, resolver la ecuación $Ax = b$ por mínimos cuadrados significa determinar el conjunto

$$\operatorname{argmin}_{x \in \mathbb{R}^n} \|b - Ax\| = \{x \in \mathbb{R}^n : \|b - Ax\| \leq \|b - A\xi\| \text{ para todo } \xi \in \mathbb{R}^n\},$$

de todos los $x \in \mathbb{R}^n$ cuyas imágenes por A minimizan la distancia al vector b .

Nota Bene. Nótese que para $b \in \operatorname{col}(A)$ vale que $\min_{x \in \mathbb{R}^n} \|b - Ax\| = 0$ y en este caso $\operatorname{arg} \min_{x \in \mathbb{R}^n} \|b - Ax\| = \{x \in \mathbb{R}^n : Ax = b\}$, las soluciones por cuadrados mínimos coinciden con las soluciones de la ecuación $Ax = b$.

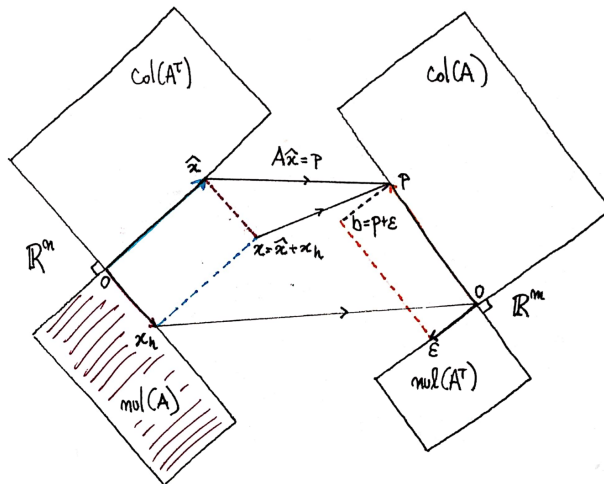


FIGURA 3. Una solución de mínimos cuadrados de $Ax = b$ es cualquier vector $x \in \mathbb{R}^n$ que minimiza $\|b - Ax\|^2$, se obtienen resolviendo las ecuaciones normales $A^T Ax = A^T b$. La solución de mínimos cuadrados de norma mínima \hat{x} pertenece al subespacio $\operatorname{col}(A^T) = \operatorname{nul}(A)^\perp$, porque $\|\hat{x} + x_h\|^2 = \|\hat{x}\|^2 + \|x_h\|^2 \geq \|\hat{x}\|^2$.

Enfoque. Como Ax pertenece al espacio columnas, elegimos el punto $p \in \operatorname{col}(A)$ más cercano a b . Este punto es la proyección ortogonal de b sobre $\operatorname{col}(A)$, $p = P_{\operatorname{col}(A)}(b)$. Tenemos así que $\operatorname{arg} \min_{x \in \mathbb{R}^n} \|b - Ax\| = \{x \in \mathbb{R}^n : Ax = p\} \neq \emptyset$. Como el error $\varepsilon = b - p \in \operatorname{col}(A)^\perp = \operatorname{nul}(A^T)$ tenemos que

$$x \in \operatorname{arg} \min_{x \in \mathbb{R}^n} \|b - Ax\| \iff A^T(b - Ax) = 0 \iff A^T Ax = A^T b.$$

Lema 2.2. Las soluciones por mínimos cuadrados de la ecuación $Ax = b$ son las soluciones de la ecuación

$$A^T Ax = A^T b.$$

Ejemplo 2.3. Hallar las soluciones por mínimos cuadrados de la ecuación

$$\underbrace{\begin{bmatrix} 1 & 1 & 2 & 2 \\ 1 & 2 & 2 & 3 \\ 2 & 3 & 4 & 5 \\ 1 & 3 & 2 & 4 \\ 0 & 1 & 0 & 1 \end{bmatrix}}_A \underbrace{\begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix}}_x = \underbrace{\begin{bmatrix} 2 \\ 3 \\ 5 \\ 3 \\ 2 \end{bmatrix}}_b$$

y calcular el error cuadrático cometido.

Resolución. Planteamos la ecuación normal $A^T Ax = A^T b$

$$\underbrace{\begin{bmatrix} 7 & 12 & 14 & 19 \\ 12 & 24 & 24 & 36 \\ 14 & 24 & 28 & 38 \\ 19 & 36 & 38 & 55 \end{bmatrix}}_{A^T A} \underbrace{\begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix}}_x = \underbrace{\begin{bmatrix} 18 \\ 34 \\ 36 \\ 52 \end{bmatrix}}_{A^T b}$$

y la resolvemos utilizando el método de Gauss-Jordan

$$\begin{bmatrix} 1 & 0 & 2 & 1 \\ 0 & 1 & 0 & 1 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix} = \begin{bmatrix} 1 \\ \frac{11}{12} \\ 0 \\ 0 \end{bmatrix}.$$

Obtenemos que $x_1 = 1 - 2x_3 - x_4$ y $x_2 = \frac{11}{12} - x_4$. Por lo tanto, el conjunto de todas las soluciones por mínimos cuadrados de la ecuación $Ax = b$ es

$$\operatorname{argmin}_{x \in \mathbb{R}^4} \|b - Ax\| = \underbrace{\begin{bmatrix} 1 & \frac{11}{12} & 0 & 0 \end{bmatrix}^T}_{x_p} + \operatorname{gen} \left\{ \underbrace{\begin{bmatrix} -2 & 0 & 1 & 0 \end{bmatrix}^T, \begin{bmatrix} -1 & -1 & 0 & 1 \end{bmatrix}^T}_{\operatorname{nul}(A)} \right\}.$$

El error cuadrático que se comete es

$$\min_{x \in \mathbb{R}^4} \|b - Ax\|^2 = \|b - Ax_p\|^2,$$

donde x_p es una solución particular por mínimos cuadrados de $Ax = b$. De

$$b - Ax_p = \begin{bmatrix} 2 \\ 3 \\ 5 \\ 3 \\ 2 \end{bmatrix} - \begin{bmatrix} 1 & 1 & 2 & 2 \\ 1 & 2 & 2 & 3 \\ 2 & 3 & 4 & 5 \\ 1 & 3 & 2 & 4 \\ 0 & 1 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 \\ \frac{11}{12} \\ 0 \\ 0 \end{bmatrix} = \begin{bmatrix} 2 \\ 3 \\ 5 \\ 3 \\ 2 \end{bmatrix} - \frac{1}{12} \begin{bmatrix} 23 \\ 34 \\ 57 \\ 45 \\ 11 \end{bmatrix} = \frac{1}{12} \begin{bmatrix} 1 \\ 2 \\ 3 \\ -9 \\ 13 \end{bmatrix},$$

resulta que $\|b - Ax_p\|^2 = \frac{11}{6}$.

2.3. La solución de norma mínima.

En la sección anterior mostramos que

$$(4) \quad \operatorname{argmin}_{x \in \mathbb{R}^n} \|b - Ax\| = \{x \in \mathbb{R}^n : Ax = P_{\operatorname{col}(A)}(b)\}.$$

Las soluciones de la ecuación $Ax = P_{\text{col}(A)}(b)$ son de la forma $x = x_p + x_h$, donde x_p es una solución particular y $x_h \in \text{nul}(A)$. Por otra parte, hemos visto que si el rango de A es r y $\mathcal{B}_1 = \{x_1, \dots, x_r\}$ es una base de $\text{col}(A^T)$, entonces $\{Ax_1, Ax_2, \dots, Ax_r\}$ es una base de $\text{col}(A)$. Por lo tanto, la ecuación $Ax = P_{\text{col}(A)}(b)$ admite una única solución $\hat{x} \in \text{col}(A^T)$, y su vector de coordenadas en base \mathcal{B}_1 satisface la ecuación

$$[Ax_1 \quad Ax_2 \quad \cdots \quad Ax_r] [\hat{x}]^{\mathcal{B}_1} = P_{\text{col}(A)}(b).$$

Como $\text{col}(A^T) \perp \text{nul}(A)$ tenemos que

$$\|\hat{x} + x_h\|^2 = \|\hat{x}\|^2 + \|x_h\|^2 \geq \|\hat{x}\|^2.$$

Esto significa que de todas las soluciones por mínimos cuadrados de la ecuación $Ax = b$ la que tiene la norma mínima es \hat{x} .

Ejemplo 2.4 (Continuación). Hallar la solución por mínimos cuadrados de norma mínima de la ecuación $Ax = b$ del Ejemplo 2.3.

Resolución. Seguimos al pie de la letra el desarrollo del argumento anterior.

Paso 1. Hallamos una base \mathcal{B}_1 de $\text{col}(A^T)$ utilizando el algoritmo espacio filas. De la matriz escalonada por filas reducida

$$E_A = \begin{bmatrix} 1 & 0 & 2 & 1 \\ 0 & 1 & 0 & 1 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}$$

se obtiene que $\mathcal{B}_1 = \{x_1, x_2\}$, donde $x_1 = [1 \ 0 \ 2 \ 1]^T$ y $x_2 = [0 \ 1 \ 0 \ 1]^T$, es una base de $\text{col}(A^T)$

Paso 2. Planteamos la ecuación $[Ax_1 \quad Ax_2] [\hat{x}]^{\mathcal{B}_1} = P_{\text{col}(A)}(b)$ y la resolvemos:

$$\underbrace{\begin{bmatrix} 1 & 1 & 2 & 2 \\ 1 & 2 & 2 & 3 \\ 2 & 3 & 4 & 5 \\ 1 & 3 & 2 & 4 \\ 0 & 1 & 0 & 1 \end{bmatrix}}_A \underbrace{\begin{bmatrix} 1 & 0 \\ 0 & 1 \\ 2 & 0 \\ 1 & 1 \end{bmatrix}}_{[x_1 \ x_2]} \underbrace{\begin{bmatrix} \xi_1 \\ \xi_2 \end{bmatrix}}_{[\hat{x}]^{\mathcal{B}_1}} = \frac{1}{12} \underbrace{\begin{bmatrix} 23 \\ 34 \\ 57 \\ 45 \\ 11 \end{bmatrix}}_{P_{\text{col}(A)}(b)} \iff \underbrace{\begin{bmatrix} 7 & 3 \\ 8 & 5 \\ 15 & 8 \\ 9 & 7 \\ 1 & 2 \end{bmatrix}}_{[Ax_1 \ Ax_2]} \underbrace{\begin{bmatrix} \xi_1 \\ \xi_2 \end{bmatrix}}_{[\hat{x}]^{\mathcal{B}_1}} = \frac{1}{12} \underbrace{\begin{bmatrix} 23 \\ 34 \\ 57 \\ 45 \\ 11 \end{bmatrix}}_{P_{\text{col}(A)}(b)} \\ \iff \underbrace{\begin{bmatrix} \xi_1 \\ \xi_2 \end{bmatrix}}_{[\hat{x}]^{\mathcal{B}_1}} = \frac{1}{132} \begin{bmatrix} 13 \\ 54 \end{bmatrix}$$

Paso 3. Utilizamos la inversa del isomorfismo de coordenadas para reconstruir \hat{x} .

$$[\hat{x}]^{\mathcal{B}_1} = \frac{1}{132} \begin{bmatrix} 13 \\ 54 \end{bmatrix} \iff \hat{x} = \underbrace{\begin{bmatrix} 1 & 0 \\ 0 & 1 \\ 2 & 0 \\ 1 & 1 \end{bmatrix}}_{[x_1 \ x_2]} \frac{1}{132} \begin{bmatrix} 13 \\ 54 \end{bmatrix} = \frac{1}{132} \begin{bmatrix} 13 \\ 54 \\ 26 \\ 67 \end{bmatrix}$$

Conclusión. La solución por mínimos cuadrados de norma mínima de la ecuación $Ax = b$ es el vector $\hat{x} = [\frac{13}{132} \ \frac{54}{132} \ \frac{26}{132} \ \frac{67}{132}]^T$ y su norma vale $\|\hat{x}\| = \frac{\sqrt{8250}}{132} \approx 0.688$.

2.4. Matrices de rango máximo.

Definición 2.5. Sean $m \geq n$ y $A \in \mathbb{R}^{m \times n}$. Decimos que A es de rango máximo cuando su rango es n . En otras palabras, A es de rango máximo si y solamente si

$$\text{rango}(A) = \dim(\text{col}(A)) = n.$$

Nota Bene. Nótese que decir que A es de rango máximo es lo mismo que decir que $\text{nul}(A) = \{0\}$.

Lema 2.6. Sean $m \geq n$ y $A \in \mathbb{R}^{m \times n}$. Vale que $\text{nul}(A) = \text{nul}(A^T A)$. En particular, $A^T A$ es inversible si y solamente si A es de rango máximo.

Demostración. Como $\text{nul}(A) \subseteq \text{nul}(A^T A)$ basta verificar que $\text{nul}(A^T A) \subseteq \text{nul}(A)$: si $A^T A x = 0$, entonces $\|Ax\|^2 = 0$, de donde sigue que $Ax = 0$. \square

Corolario 2.7. Si A es de rango de máximo la única solución por mínimos cuadrados de la ecuación $Ax = b$ es $\hat{x} = (A^T A)^{-1} A^T b$.

Demostración. Inmediata de los Lemas 2.2 y 2.6. \square

Nomenclatura. Si $A \in \mathbb{R}^{m \times n}$ es de rango de máximo, la matriz $A^\dagger \in \mathbb{R}^{n \times m}$ definida por

$$A^\dagger := (A^T A)^{-1} A^T$$

se denomina la *seudoinversa de Moore-Penrose* de A .

Nota Bene. Nótese que si A es de rango máximo, entonces

$$A^\dagger A = [I_{\mathbb{R}^{n \times n}}] \quad \text{y} \quad AA^\dagger = [P_{\text{col}(A)}].$$

Ejemplo 2.8. Hallar la solución por mínimos cuadrados de la ecuación

$$\underbrace{\begin{bmatrix} 1 & 0 \\ 1 & 1 \\ 1 & 2 \end{bmatrix}}_A \underbrace{\begin{bmatrix} x_1 \\ x_2 \end{bmatrix}}_x = \underbrace{\begin{bmatrix} 6 \\ 0 \\ 0 \end{bmatrix}}_b.$$

Resolución. Como la matriz A es de rango máximo la solución deseada es $\hat{x} = A^\dagger b$.

Paso 1. Para construir la pseudoinversa A^\dagger seguimos los pasos al pie de la letra:

$$\begin{aligned} A^\dagger &= (A^T A)^{-1} A^T \\ &= \begin{bmatrix} 3 & 3 \\ 3 & 5 \end{bmatrix}^{-1} \begin{bmatrix} 1 & 1 & 1 \\ 0 & 1 & 2 \end{bmatrix} \\ &= \frac{1}{6} \begin{bmatrix} 5 & -3 \\ -3 & 3 \end{bmatrix} \begin{bmatrix} 1 & 1 & 1 \\ 0 & 1 & 2 \end{bmatrix} \\ &= \frac{1}{6} \begin{bmatrix} 5 & 2 & -1 \\ -3 & 0 & 3 \end{bmatrix}. \end{aligned}$$

Paso 2. Hallamos la solución por mínimos cuadrados de la ecuación $Ax = b$.

$$\hat{x} = A^\dagger b = \frac{1}{6} \begin{bmatrix} 5 & 2 & -1 \\ -3 & 0 & 3 \end{bmatrix} \begin{bmatrix} 6 \\ 0 \\ 0 \end{bmatrix} = \begin{bmatrix} 5 \\ -3 \end{bmatrix}.$$

La distancia de $A\hat{x}$ al vector b vale

$$\|b - A\hat{x}\| = \left\| \begin{bmatrix} 6 & 0 & 0 \end{bmatrix}^T - \begin{bmatrix} 5 & 2 & -1 \end{bmatrix}^T \right\| = \sqrt{6}.$$

Ejemplo 2.9 (Variante). Hallar la matriz de la proyección ortogonal de \mathbb{R}^3 sobre el subespacio $\mathbb{S} = \text{gen} \left\{ \begin{bmatrix} 1 & 1 & 1 \end{bmatrix}^T, \begin{bmatrix} 0 & 1 & 2 \end{bmatrix}^T \right\}$.

Resolución. Como $\dim(\mathbb{S}) = 2$, construimos una matriz $A \in \mathbb{R}^{3 \times 2}$ de rango máximo cuyas columnas sean una base de \mathbb{S} . De esa manera garantizamos que $\mathbb{S} = \text{col}(A)$ y recurrimos a la relación $AA^\dagger = [P_{\text{col}(A)}]$ para encontrar la matriz deseada.

Obtenemos que

$$[P_{\mathbb{S}}] = AA^\dagger = \begin{bmatrix} 1 & 0 \\ 1 & 1 \\ 1 & 2 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 1 & 1 \\ 1 & 2 \end{bmatrix}^\dagger = \frac{1}{6} \begin{bmatrix} 1 & 0 \\ 1 & 1 \\ 1 & 2 \end{bmatrix} \begin{bmatrix} 5 & 2 & -1 \\ -3 & 0 & 3 \end{bmatrix} = \frac{1}{6} \begin{bmatrix} 5 & 2 & -1 \\ 2 & 2 & 2 \\ -1 & 2 & 5 \end{bmatrix}.$$

3. AJUSTE DE DATOS

Volvamos al problema presentado en la sección 2.1. Disponemos de una tabla que presenta m parejas de datos experimentales

$$(5) \quad \begin{array}{c|cccc} x & x_1 & x_2 & \cdots & x_m \\ \hline y & y_1 & y_2 & \cdots & y_m \end{array}$$

donde $x_i \neq x_j$ para $i \neq j$, y proponemos ajustarlos con un modelo lineal basado en un conjunto de funciones $\{\varphi_j : j \in \mathbb{I}_n\}$. Esto es, se propone una función

$$\varphi = \sum_{j=1}^n a_j \varphi_j,$$

con $a_1, a_2, \dots, a_n \in \mathbb{R}$ elegidos de manera tal que minimicen la suma de los cuadrados de los errores $\varepsilon_i = y_i - \varphi(x_i)$, $i \in \mathbb{I}_m$. Esto significa que los parámetros a_1, a_2, \dots, a_n se determinan mediante las soluciones por mínimos cuadrados del sistema de ecuaciones

$$\underbrace{\begin{bmatrix} \varphi_1(x_1) & \varphi_2(x_1) & \cdots & \varphi_n(x_1) \\ \varphi_1(x_2) & \varphi_2(x_2) & \cdots & \varphi_n(x_2) \\ \vdots & \vdots & & \vdots \\ \varphi_1(x_m) & \varphi_2(x_m) & \cdots & \varphi_n(x_m) \end{bmatrix}}_{[\varphi_j(x_i)]} \underbrace{\begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_n \end{bmatrix}}_a = \underbrace{\begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_m \end{bmatrix}}_y.$$

Este último problema se resuelve hallando las soluciones $a \in \mathbb{R}^n$ de la ecuación

$$[\varphi_j(x_i)]^T [\varphi_j(x_i)] a = [\varphi_j(x_i)]^T y.$$

Si la matriz $[\varphi_j(x_i)]$ es de rango máximo, este problema tiene solución única determinada por

$$\hat{a} = [\varphi_j(x_i)]^\dagger y.$$

3.1. Regresión polinomial.

Para ajustar los datos de la tabla (5) se fija $n < m$ y se propone un polinomio de la forma

$$p_n = a_0 + a_1x + \cdots + a_nx^n,$$

con $a_0, a_1, \dots, a_n \in \mathbb{R}$ elegidos de manera tal que minimicen la suma de los cuadrados de los errores $\varepsilon_i = y_i - p_n(x_i)$, $i \in \mathbb{I}_m$. Los coeficientes del polinomio se determinan resolviendo por mínimos cuadrados el sistema de ecuaciones

$$\underbrace{\begin{bmatrix} 1 & x_1 & \cdots & x_1^n \\ 1 & x_2 & \cdots & x_2^n \\ \vdots & \vdots & & \vdots \\ 1 & x_m & \cdots & x_m^n \end{bmatrix}}_{V_n(x_1, x_2, \dots, x_m)} \underbrace{\begin{bmatrix} a_0 \\ a_1 \\ \vdots \\ a_n \end{bmatrix}}_{[p_n]^{\mathcal{E}_n}} = \underbrace{\begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_m \end{bmatrix}}_y.$$

Nota Bene. Nótese que las columnas de la matriz $V_n(x_1, x_2, \dots, x_m)$ son linealmente independientes porque coinciden con las primeras $n+1$ columnas de la matriz de cambio de coordenadas, $M_{\mathcal{E}_{m-1}}^{\mathcal{B}_x}$, de la base canónica $\mathcal{E}_{m-1} = \{1, x, \dots, x^{m-1}\}$ de $\mathbb{R}_{m-1}[x]$ en la base de polinomios interpoladores de Lagrange correspondiente al conjunto de abscisas $\mathcal{X} = \{x_1, x_2, \dots, x_m\}$,

$$\mathcal{B}_x = \left\{ \prod_{k \in \mathbb{I}_m: k \neq i} \frac{x - x_k}{x_i - x_k} : i \in \mathbb{I}_m \right\}.$$

Como la matriz $V_n(x_1, x_2, \dots, x_m)$ es de rango máximo se concluye que

$$(6) \quad [p_n]^{\mathcal{E}_n} = V_n(x_1, x_2, \dots, x_m)^\dagger y,$$

donde $y = [y_1 \ y_2 \ \cdots \ y_m]^T$.

Ejemplo 3.1. Un investigador recolecta información sobre las proporciones p_i de una sustancia en un compuesto a distintas temperaturas T_i (en cientos de grados Celsius). Los datos recolectados se presentan en la siguiente tabla

T	1	1.2	1.4	1.6	1.8
p	0.45	0.54	0.62	0.75	0.92

- Hallar la recta $p = a + bT$ que mejor se ajusta a esos datos.
- Hallar la parábola $p = a + bT + cT^2$ que mejor se ajusta a esos datos.
- Determinar cuál de esos dos modelos se ajusta mejor a esos datos.

Resolución.

(a) Para hallar la recta $p = a + bT$ que mejor se ajusta a esos datos tenemos que resolver por mínimos cuadrados el sistema de ecuaciones $a + bT_i = p_i$, $i \in \mathbb{I}_5$:

$$\underbrace{\begin{bmatrix} 1 & 1 \\ 1 & 1.2 \\ 1 & 1.4 \\ 1 & 1.6 \\ 1 & 1.8 \end{bmatrix}}_{V_2} \begin{bmatrix} a \\ b \end{bmatrix} = \begin{bmatrix} 0.45 \\ 0.54 \\ 0.62 \\ 0.75 \\ 0.92 \end{bmatrix}$$

Como la matriz $V_2 \in \mathbb{R}^{5 \times 2}$ es de rango máximo y su pseudoinversa es

$$V_2^\dagger = \frac{1}{10} \begin{bmatrix} 16 & 9 & 2 & -5 & -12 \\ -10 & -5 & 0 & 5 & 10 \end{bmatrix},$$

tenemos que

$$\begin{bmatrix} a \\ b \end{bmatrix} = \frac{1}{10} \begin{bmatrix} 16 & 9 & 2 & -5 & -12 \\ -10 & -5 & 0 & 5 & 10 \end{bmatrix} \begin{bmatrix} 0.45 \\ 0.54 \\ 0.62 \\ 0.75 \\ 0.92 \end{bmatrix} = \frac{1}{1000} \begin{bmatrix} -149 \\ 575 \end{bmatrix}$$

Por lo tanto, la recta que mejor ajusta los datos observados es

$$p = -\frac{149}{1000} + \frac{575}{1000}T.$$

(b) Para hallar la parábola $p = a + bT + cT^2$ que mejor se ajusta a esos datos tenemos que resolver por mínimos cuadrados el sistema de ecuaciones $a + bT_i + cT_i^2 = p_i$, $i \in \mathbb{I}_5$:

$$\underbrace{\begin{bmatrix} 1 & 1 & 1^2 \\ 1 & 1.2 & 1.2^2 \\ 1 & 1.4 & 1.4^2 \\ 1 & 1.6 & 1.6^2 \\ 1 & 1.8 & 1.8^2 \end{bmatrix}}_{V_3} \begin{bmatrix} a \\ b \\ c \end{bmatrix} = \begin{bmatrix} 0.45 \\ 0.54 \\ 0.62 \\ 0.75 \\ 0.92 \end{bmatrix}$$

Como la matriz $V_3 \in \mathbb{R}^{5 \times 3}$ es de rango máximo y su pseudoinversa es

$$V_3^\dagger = \frac{1}{70} \begin{bmatrix} 582 & -172 & -456 & -270 & 386 \\ -770 & 315 & 700 & 385 & -630 \\ 250 & -125 & -250 & -125 & 250 \end{bmatrix},$$

tenemos que

$$\begin{bmatrix} a \\ b \\ c \end{bmatrix} = \frac{1}{70} \begin{bmatrix} 582 & -172 & -456 & -270 & 386 \\ -770 & 315 & 700 & 385 & -630 \\ 250 & -125 & -250 & -125 & 250 \end{bmatrix} \begin{bmatrix} 0.45 \\ 0.54 \\ 0.62 \\ 0.75 \\ 0.92 \end{bmatrix} = \frac{1}{7000} \begin{bmatrix} 3892 \\ -3325 \\ 2625 \end{bmatrix}$$

Por lo tanto, la parábola que mejor ajusta los datos observados es

$$p = -\frac{3892}{7000} + \frac{-3325}{7000}T + \frac{2625}{7000}T^2.$$

(c) Para determinar cuál de los dos modelos se ajusta mejor a esos datos tenemos que calcular sus errores cuadráticos.

$$\varepsilon_1 = \frac{1}{1000} \begin{bmatrix} 450 \\ 540 \\ 620 \\ 750 \\ 920 \end{bmatrix} - \frac{1}{1000} \begin{bmatrix} 1 & 1 \\ 1 & 1.2 \\ 1 & 1.4 \\ 1 & 1.6 \\ 1 & 1.8 \end{bmatrix} \begin{bmatrix} -149 \\ 575 \end{bmatrix} = \frac{1}{1000} \begin{bmatrix} 24 \\ -1 \\ -36 \\ -21 \\ 34 \end{bmatrix},$$

$$\varepsilon_2 = \frac{1}{1000} \begin{bmatrix} 450 \\ 540 \\ 620 \\ 750 \\ 920 \end{bmatrix} - \frac{1}{7000} \begin{bmatrix} 1 & 1 & 1^2 \\ 1 & 1.2 & 1.2^2 \\ 1 & 1.4 & 1.4^2 \\ 1 & 1.6 & 1.6^2 \\ 1 & 1.8 & 1.8^2 \end{bmatrix} \begin{bmatrix} 3892 \\ -3325 \\ 2625 \end{bmatrix} = \frac{1}{1000} \begin{bmatrix} -6 \\ 14 \\ -6 \\ -6 \\ 4 \end{bmatrix},$$

Como $\|\varepsilon_1\|^2 = \frac{3470}{10^6}$ y $\|\varepsilon_2\|^2 = \frac{320}{10^6}$, se concluye que el ajuste cuadrático es mejor que el ajuste lineal.

3.2. Recta de regresión.

Para ajustar los datos de la tabla (5) se propone una recta de la forma $y = a + bx$. Los valores de a y b son la solución del sistema

$$\begin{bmatrix} 1^T \\ x^T \end{bmatrix} \begin{bmatrix} 1 & x \end{bmatrix} \begin{bmatrix} a \\ b \end{bmatrix} = \begin{bmatrix} 1^T \\ x^T \end{bmatrix} y$$

donde para simplificar la escritura abusamos de la notación y ponemos

$$1 = [1 \ 1 \ \cdots \ 1]^T, \quad x = [x_1 \ x_2 \ \cdots \ x_m]^T, \quad y = [y_1 \ y_2 \ \cdots \ y_m]^T.$$

La solución es

$$\begin{aligned} \begin{bmatrix} a \\ b \end{bmatrix} &= \begin{bmatrix} m & \sum_{i=1}^m x_i \\ \sum_{i=1}^m x_i & \sum_{i=1}^m x_i^2 \end{bmatrix}^{-1} \begin{bmatrix} \sum_{i=1}^m y_i \\ \sum_{i=1}^m x_i y_i \end{bmatrix} \\ &= \begin{bmatrix} 1 & \bar{x} \\ \bar{x} & \frac{1}{m} \sum_{i=1}^m x_i^2 \end{bmatrix}^{-1} \begin{bmatrix} \bar{y} \\ \frac{1}{m} \sum_{i=1}^m x_i y_i \end{bmatrix} \\ &= \frac{1}{\frac{1}{m} \sum_{i=1}^m x_i^2 - \bar{x}^2} \begin{bmatrix} \frac{1}{m} \sum_{i=1}^m x_i^2 & -\bar{x} \\ -\bar{x} & 1 \end{bmatrix} \begin{bmatrix} \bar{y} \\ \frac{1}{m} \sum_{i=1}^m x_i y_i \end{bmatrix}. \end{aligned}$$

De aquí que

$$b = \frac{\frac{1}{m} \sum_{i=1}^m x_i y_i - \bar{x} \bar{y}}{\frac{1}{m} \sum_{i=1}^m x_i^2 - \bar{x}^2}, \quad a = \bar{y} - b \bar{x}.$$

Esto es así porque

$$\begin{aligned} a &= \frac{(\frac{1}{m} \sum_{i=1}^m x_i^2) \bar{y} - \bar{x} \frac{1}{m} \sum_{i=1}^m x_i y_i}{\frac{1}{m} \sum_{i=1}^m x_i^2 - \bar{x}^2} \\ &= \frac{(\frac{1}{m} \sum_{i=1}^m x_i^2 - \bar{x}^2) \bar{y} - (\frac{1}{m} \sum_{i=1}^m x_i y_i - \bar{x} \bar{y}) \bar{x}}{\frac{1}{m} \sum_{i=1}^m x_i^2 - \bar{x}^2} \\ &= \bar{y} - b \bar{x}. \end{aligned}$$

Por lo tanto, la recta de regresión tiene la forma

$$y = \bar{y} + \left(\frac{\overline{xy} - \bar{x}\bar{y}}{\overline{x^2} - \bar{x}^2} \right) (x - \bar{x}).$$

Ejemplo 3.2. La *ley de enfriamiento de Newton* establece que un objeto cambia de temperatura T a un ritmo proporcional a la diferencia entre T y la temperatura ambiente. Supongamos que la temperatura ambiente es de 70°C , entonces

$$\frac{dT}{dt} = -\lambda(T - 70),$$

donde λ es un parámetro positivo. La solución general de esta ecuación diferencial es de la forma

$$T(t) = ae^{-\lambda t} + 70,$$

con $a \in \mathbb{R}$.

Se mide la temperatura en instantes sucesivos y se desea hallar un modelo para la temperatura T en función del tiempo t .

Por ejemplo, veamos qué ocurre cuando los resultados de las mediciones son los que se muestran en la siguiente tabla

t	0	1	2	3	4
T	100	90	85	83	82

El modelo tiene la forma

$$T(t) = ae^{-\lambda t} + 70,$$

donde los parámetros desconocidos son λ y a . Mediante el cambio de variable $y = \ln(T - 70)$ el problema se transforma en un problema de regresión lineal, porque

$$\ln(T - 70) = \ln(a) - \lambda t.$$

t	0	1	2	3	4
$\ln(T - 70)$	$\ln(30)$	$\ln(20)$	$\ln(15)$	$\ln(13)$	$\ln(12)$

Para hallar los valores de los parámetros que producen el modelo más ajustado a los datos observados podemos resolver las ecuaciones normales $A^T A \xi = A^T b$, donde

$$A = \begin{bmatrix} 1 & 0 \\ 1 & 1 \\ 1 & 2 \\ 1 & 3 \\ 1 & 4 \end{bmatrix}, \quad b = \begin{bmatrix} \ln(30) \\ \ln(20) \\ \ln(15) \\ \ln(13) \\ \ln(12) \end{bmatrix}, \quad \xi = \begin{bmatrix} \ln(a) \\ -\lambda \end{bmatrix}$$

Calculamos $A^T A$ y $(A^T A)^{-1}$

$$(7) \quad A^T A = \begin{bmatrix} 5 & 10 \\ 10 & 30 \end{bmatrix}, \quad (A^T A)^{-1} = \frac{1}{50} \begin{bmatrix} 30 & -10 \\ -10 & 5 \end{bmatrix}$$

Calculamos $A^T b$

$$(8) \quad A^T b \approx \begin{bmatrix} 14.15 \\ 26.05 \end{bmatrix}$$

y resolvemos la ecuación normal $A^T A \xi = A^T b$ multiplicando por la inversa de $A^T A$:

$$\xi = \begin{bmatrix} 3.284 \\ -0.2263 \end{bmatrix}$$

Tenemos así que $\ln(a) \approx 3.284 \iff a \approx e^{3.284} \approx 26.68$ y $-\lambda = -0.2263$. Obtengamos así que

$$(9) \quad T(t) = 26.68e^{-0.2263t} + 70.$$

Este resultado nos permite producir buenas estimaciones de la temperatura $T(t)$ para otros instantes de tiempo. Si la ley de enfriamiento de Newton es un modelo preciso y los datos se recopilaron cuidadosamente, nuestro modelo nos debería permitir hacer predicciones precisas sobre los valores que ira adoptando la temperatura a medida que transcurra el tiempo. \square