Tema 2

Espacio Vectorial Euclídeo. Mínimos cuadrados

En el tema anterior estudiamos el Espacio Vectorial $\mathbb{R}^n$, donde definimos las operaciones de suma de vectores y producto de un vector por escalares. En este tema introducimos una nueva operación: *producto escalar* entre vectores. De esta nueva operación se derivan los conceptos de módulo (longitud) de un vector y ángulo que forman dos vectores, y en consecuencia, los conceptos de distancia y ortogonalidad geométrica como se intuye en $\mathbb{R}^2$ y $\mathbb{R}^3$.

Definiremos con esto la ortogonalidad entre subespacios y como resultado principal veremos que $\mathbb{R}^n$ es suma directa de cualquier subespacio vectorial y su ortogonal. Como consecuencia, podremos calcular la distancia mínima de un vector a un subespacio F, y el vector de F donde se alcanza dicho mínimo: la proyección ortogonal sobre F.

Todo el estudio realizado sobre el espacio vectorial euclídeo lo utilizaremos para construir soluciones aproximadas, óptimas en el sentido de lo que denominaremos *mínimos cuadrados*, de sistemas de ecuaciones lineales incompatibles.

2.1. Producto escalar. Norma y distancia euclídea

Definición 1. *Sea V un espacio vectorial sobre $\mathbb{R}$. Un* **producto escalar** *o* **interno** *sobre V es una aplicación tal que a cada par de vectores $\vec{x}$ e $\vec{y}$ de V le asigna un número real, que denotaremos por $\vec{x}\cdot\vec{y}$, de manera que para cualesquiera $\vec{x}$, $\vec{y}$, $\vec{z} \in V$, y cualquier $k \in \mathbb{R}$ se satisfacen las siguientes propiedades:*

1. $\vec{x}\cdot\vec{x} \geq 0 \qquad \vec{x}\cdot\vec{x} = 0 \Leftrightarrow \vec{x} = \vec{0}$
2. $\vec{x}\cdot\vec{y} = \vec{y}\cdot\vec{x}$
3. $\vec{x}\cdot(\vec{y}+\vec{z}) = \vec{x}\cdot\vec{y} + \vec{x}\cdot\vec{z}$
4. $(k\vec{x})\cdot\vec{y} = k(\vec{x}\cdot\vec{y})$

Un espacio vectorial en el que se ha definido un producto escalar se llama **espacio vectorial euclídeo**. Todo espacio vectorial de dimensión finita puede dotarse de un producto escalar y convertirse por tanto en espacio vectorial euclídeo.

Ejemplo 2. *Cuando $V = \mathbb{R}^n$, podemos definir la siguiente operación entre vectores (comprobar que se trata de un producto escalar): si $\vec{x}, \vec{y} \in V$, con $\vec{x} = (x_1, ..., x_n)$, $\vec{y} = (y_1, ..., y_n)$, entonces*

$$\vec{x} \cdot \vec{y} = x_1 y_1 + ... + x_n y_n.$$

Observemos que si escribimos los vectores $\vec{x}$ e $\vec{y}$ como matrices columna, $x = \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix}$, $y = \begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix}$, se tiene que $\vec{x} \cdot \vec{y} = x^t y = (x_1 \dots x_n) \begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix}$, donde recordemos que para una matriz A, su matriz transpuesta A^t es la resultante de cambiar filas por columnas.

Definición 3. *Sea V un espacio vectorial donde se ha definido un producto escalar, llamamos* **norma** *o* **módulo** *de un vector $\vec{x} \in V$ al siguiente número real*

$$\|\vec{x}\| = +\sqrt{\vec{x} \cdot \vec{x}}$$

A los vectores $\vec{x} \in V$ que satisfagan $\|\vec{x}\| = 1$ se les llamará vectores **unitarios**.

A partir de la definición y de las propiedades del producto escalar es fácil comprobar las siguientes propiedades de la norma: para cualesquiera $\vec{x}, \vec{y} \in V$ y $k \in \mathbb{R}$ se satisface:

1. $\|\vec{x}\| = 0$ si y sólo si $\vec{x} = \vec{0}$
2. $\|k\vec{x}\| = |k| \cdot \|\vec{x}\|$
3. **(Desigualdad triangular)** $\|\vec{x} + \vec{y}\| \leq \|\vec{x}\| + \|\vec{y}\|$
4. **(Desigualdad de Cauchy-Schwarz)** $|\vec{x} \cdot \vec{y}| \leq \|x\| \, \|y\|$, y la igualdad se da si y sólo si $\vec{x}$ e $\vec{y}$ son proporcionales.
5. **(Ley del paralelogramo)** $\|\vec{x} + \vec{y}\|^2 + \|\vec{x} - \vec{y}\|^2 = 2\left(\|\vec{x}\|^2 + \|\vec{y}\|^2\right)$.

A partir de la norma, vamos de definir una distancia en $\mathbb{R}^n$, la **distancia euclídea**: Dados $\vec{x}, \ \vec{y} \ \in V$, llamaremos distancia entre $\vec{x}$ e $\vec{y}$, y la denotamos por $\mathrm{d}(\vec{x}, \vec{y})$, al número positivo

$$\mathrm{d}(\vec{x}, \vec{y}) = \|\vec{x} - \vec{y}\| \qquad \Big(\text{ por tanto } \quad \|\vec{x}\| = d(\vec{x}, \vec{0}) \Big).$$

A partir de las propiedades de la norma se deduce que para cualesquiera $\vec{x}, \ \vec{y}, \ \vec{z} \ \in \ V$, la distancia satisface:

1. $\mathrm{d}(\vec{x}, \vec{x}) = 0$
2. $\mathrm{d}(\vec{x}, \vec{y}) = \mathrm{d}(\vec{y}, \vec{x})$ (simetría)
3. $\mathrm{d}(\vec{x} + \vec{z}, \vec{y} + \vec{z}) = \mathrm{d}(\vec{x}, \vec{y})$ (invarianza por traslación)
4. $\mathrm{d}(\vec{x}, \vec{y}) \leq \mathrm{d}(\vec{x}, \vec{z}) + \mathrm{d}(\vec{z}, \vec{y})$ (desigualdad triangular).

Todo espacio vectorial que tenga definida una distancia recibe el nombre de **espacio métrico**.

Para finalizar esta sección definimos la distancia de un vector $\vec{x}$ a un subespacio F como

$$\mathrm{d}(\vec{x}, F) = \min\{\mathrm{d}(\vec{x}, \vec{y});\ \ \vec{y} \in F\} = \min\{\|\vec{x} - \vec{y}\|;\ \ \vec{y} \in F\}.$$

En la siguiente sección encontraremos el vector de F donde se alcanza el mínimo anterior, y calcularemos, por tanto, la distancia de un vector a un subespacio.

2.2. Ortogonalidad. Proyecciones ortogonales sobre subespacios

Es bien conocido que para hallar la mínima distancia desde un punto a un plano en el espacio tridimensional, hay que proyectar perpendicularmente dicho punto sobre el plano. Extenderemos por analogía dicha noción a un espacio vectorial euclídeo: la distancia de un vector a un subespacio vectorial se obtiene *proyectando ortogonalmente* el vector sobre dicho subespacio. Daremos sentido a esta expresión:

En esta sección definiremos el ortogonal de un subespacio vectorial F. Veremos que cualquier espacio vectorial se puede descomponer como suma directa de un subespacio y su ortogonal (que denotaremos por $F^{\perp}$), esto es, $V = F \oplus F^{\perp}$. Por tanto, F y $F^{\perp}$ son complementarios. Esta descomposición nos permitirá definir la proyección ortogonal de un vector sobre un subespacio.

Definición 4. *Dos vectores $\vec{x}, \vec{y} \in V$ son* **ortogonales** *o* **perpendiculares***, ($\vec{x} \perp \vec{y}$), si su producto escalar el cero. Es decir,*

$$\vec{x} \perp \vec{y} \implies \vec{x} \cdot \vec{y} = 0.$$

Nota: En $\mathbb{R}^2$ y $\mathbb{R}^3$, el producto escalar de dos vectores es el producto de los módulos por el coseno del ángulo que forman:

$$\vec{x} \cdot \vec{y} = \|x\|\,\|y\| \cos\widehat{\vec{x}\vec{y}}.$$

Con esta definición podemos dar la siguiente versión del teorema de Pitágoras

Teorema 5 (Pitágoras.). *Sea V un espacio vectorial euclídeo. Si $\vec{x}$ e $\vec{y}$ son vectores ortogonales en V, entonces*

$$\|\vec{x} + \vec{y}\|^2 = \|\vec{x}\|^2 + \|\vec{y}\|^2.$$

En general, si $\vec{x}_1, \vec{x}_2, \ldots, \vec{x}_r$ son ortogonales dos a dos, entonces

$$\|\vec{x}_1 + \vec{x}_2 + \cdots + \vec{x}_r\|^2 = \|\vec{x}_1\|^2 + \|\vec{x}_2\|^2 + \cdots + \|\vec{x}_r\|^2.$$

Definición 6. *Un conjunto de vectores no nulos $\{\vec{x}_1, \vec{x}_2, \ldots, \vec{x}_n\}$ forman un* **sistema ortonormal** *si*

(a) son ortogonales dos a dos, es decir, $\vec{x}_i \cdot \vec{x}_j = 0, \qquad \forall i \neq j$,

(b) los vectores son unitarios, esto es, $\|\vec{x}_i\| = 1 \qquad \forall i.$

Ejercicios:

1. Demostrar que en un espacio vectorial euclídeo, todo conjunto ortonormal de vectores es linealmente independiente.

2. Comprobar que los vectores de la base canónica de $\mathbb{R}^n$, $B = \{\vec{e}_1, \vec{e}_2, ..., \vec{e}_n\}$, forman un sistema ortonormal.

Sea F un subespacio de un espacio vectorial euclídeo V. Se llama **ortogonal** de F , y se escribe $F^{\perp}$, al conjunto de los vectores que son ortogonales a todos los vectores de F, esto es,

$$F^{\perp} \; = \; \{\vec{x} \; \in \; V \; / \; \vec{x} \cdot \vec{y} \; = \; 0, \;\; \forall \; \vec{y} \; \in \; F\}.$$

Ejercicio: Comprueba que $F^{\perp}$ es un subespacio vectorial.

Ejemplo 7. *Sea* $F =< (1, 1, 1, 0), (0, 1, -1, 0) > \subset \mathbb{R}^4$. *Vamos a calcular las ecuaciones implícitas del subespacio* $F^{\perp}$. *Por definición,*

$$F^{\perp} \; = \; \{\vec{x} \; \in \; \mathbb{R}^4 \; / \; \vec{y} \cdot \vec{x} \; = \; 0, \;\; \forall \; \vec{y} \; \in \; F\}.$$

Por las propiedades de linealidad de los subespacios vectoriales, basta con calcular los vectores que son ortogonales a los de una base de F. *En nuestro caso, tomamos*

$$B_F = \{(1, 1, 1, 0), (0, 1, -1, 0)\} = \{\vec{v}_1, \vec{v}_2\}.$$

Un vector $\vec{x} = (x_1, x_2, x_3, x_4)$ *pertenecerá a* $F^{\perp}$ *si* $\vec{v}_1 \cdot \vec{x} = 0$ *y* $\vec{v}_2 \cdot \vec{x} = 0$. *Escribiendo los vectores en forma matricial, se tendrá que verificar*

$$(1\;1\;1\;0)\begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{pmatrix} = 0, \qquad (0\;1\;-1\;0)\begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{pmatrix} = 0;$$

o lo que es lo mismo,

$$\begin{pmatrix} 1 & 1 & 1 & 0 \\ 0 & 1 & -1 & 0 \end{pmatrix}\begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$$

$$\begin{pmatrix} x_1 + x_2 + x_3 \\ x_2 - x_3 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix},$$

con lo que encontramos las ecuaciones implícitas de $F^{\perp}$

$$F^{\perp} = \left\{\vec{x} = (x_1, x_2, x_3, x_4) \in \mathbb{R}^4 \mid \; x_1 + x_2 + x_3 = 0, \;\; x_2 - x_3 = 0\right\}.$$

Resolviendo el sistema homogéneo obtendríamos las ecuaciones paramétricas de $F^{\perp}$.

Recíprocamente, dado un subespacio vectorial en forma de ecuaciones implícitas, podemos obtener una base del ortogonal tan sólo fijándonos en los coeficientes de las ecuaciones. Por *ejemplo, para hallar* $(F^\perp)^\perp$, *el ortogonal de* $F^\perp$, *nos fijamos en los coeficientes de sus ecuaciones implícitas:*

$$\begin{array}{lll} Ec,1: & x_1+x_2+x_3=0 & \textit{Coeficientes de las incógnitas} \longrightarrow \quad (1,1,1,0) \\ Ec,2: & x_2-x_3=0 & \textit{Coeficientes de las incógnitas} \longrightarrow \quad (0,1,-1,0). \end{array}$$

Entonces, $\left(F^\perp\right)^\perp =< (1,1,1,0),(0,1,-1,0) >$. *Observamos que* $(F^\perp)^\perp = F$. *Esta es una de las propiedades que enunciamos en el siguiente teorema:*

Teorema 8. *Sean* F *y* G *subespacios vectoriales de* V. *Se verifican las siguientes propiedades:*

(a) $F^\perp$ *es un subespacio de* V *.*

(b) $(F+G)^\perp \; = \; F^\perp \cap G^\perp$ *.*

(c) $F \; = \; (F^\perp)^\perp$*.*

(d) Si $F \subset G$*, entonces,* $G^\perp \subset \; F^\perp$*.*

(e) $(F \cap G)^\perp \; = \; F^\perp + G^\perp$*.*

Veamos ahora que un subespacio y su ortogonal son complementarios. Esto nos permite escribir un espacio vectorial V como suma directa de un subespacio y su ortogonal. Como dijimos al principio de esta sección, esta descomposición nos permitirá definir la proyección ortogonal de un vector sobre un subespacio.

Teorema 9. *Sean* V *un espacio vectorial euclídeo, y* F *un subespacio de* V, *entonces* $V \; = F \oplus F^\perp$, *es decir:*

$$F \cap F^\perp \; = \; \{\vec{0}\} \qquad V \; = \; F + F^\perp.$$

Ejercicio: Comprobar que $\mathbb{R}^4 \; = \; F \oplus F^\perp$, con F es subespacio vectorial dado en el Ejemplo 7.

Como consecuencia del teorema anterior, cualquier vector $\vec{x} \in V$ se puede descomponer como suma de dos vectores, $\vec{x} \; = \; \vec{v}_1 + \vec{v}_2$, con $\vec{v}_1 \; \in \; F$ y $\vec{v}_2 \; \in \; F^\perp$.

Definición 10. *Sean* V *un espacio vectorial euclídeo,* F *un subespacio de* V *y* $\vec{x} \in V$. *Si escribimos* $\vec{x} \; = \; \vec{v}_1 + \vec{v}_2$, *con* $\vec{v}_1 \; \in \; F$, $\vec{v}_2 \; \in \; F^\perp$, *la* **proyección ortogonal de** $\vec{x}$ **sobre** F *(resp.* **sobre** $F^\perp$*) es el vector* $\vec{v}_1 \in F$, *(resp. el vector* $\vec{v}_2$*).*

Escribiremos $\forall\, \vec{x} \in V$, $\vec{x} \; = \; \vec{v}_1 + \vec{v}_2$, con $\vec{v}_1 \; \in \; F$, $\vec{v}_2 \; \in \; F^\perp$

$$\vec{v}_1 = \mathrm{P}_F(\vec{x}) \qquad \vec{v}_2 = \mathrm{P}_{F^\perp}(\vec{x}).$$

Observemos entonces que el vector proyección ortogonal de $\vec{x}$ sobre F, $\mathrm{P}_F(\vec{x})$, es aquel tal que $\vec{x} - \mathrm{P}_F(\vec{x})$ es ortogonal a todos los vectores de F.

La proyección ortogonal de un vector sobre un subespacio nos permite calcular la distancia del vector a dicho subespacio:

Teorema 11 (Teorema de la mejor aproximación). *El vector* $P_F(\vec{x}) \in F$ *es el más cercano al vector* $\vec{x}$, *esto es,*

$$\mathrm{d}(\vec{x},F) = \text{mín}\,\{\|\vec{x}-\vec{y}\|;\;\; \vec{y} \;\in\; F\} = \|\vec{x} - P_F(\vec{x})\|$$

por tanto, $\qquad \|\vec{x} - P_F(\vec{x})\| \leq \|\vec{x} - \vec{y}\| \qquad \forall\, \vec{y} \; \in \; F.$

Para concluir esta sección, vamos a definir tres subespacios vectoriales que se obtienen de una matriz:

Sea $A \in \mathcal{M}_{m\times n}(\mathbb{R})$. Se definen los siguientes **espacios fundamentales de** A:

- **Espacio fila**: es el subespacio vectorial de $\mathbb{R}^n$ engendrado por los vectores cuyas coordenadas coinciden con los elementos de las filas:

$$\mathcal{F}(A) =< \vec{f}_1,\ \vec{f}_2,\ \dots,\ \vec{f}_m > .$$

- **Espacio columna**: es el subespacio vectorial de $\mathbb{R}^m$ engendrado por los vectores cuyas coordenadas coinciden con los elementos de las columnas:

$$\mathcal{R}(A) =< \vec{c}_1,\ \vec{c}_2,\ \dots,\ \vec{c}_n > .$$

  Nótese que $\mathcal{R}(A) = \{x_1\vec{c}_1 + x_2\vec{c}_2 + \cdots + x_n\vec{c}_n,\ \forall (x_1,\dots,\ x_n) \in \mathbb{R}^n\}$, escrito en forma matricial sería

$$\mathcal{R}(A) = \{ \begin{pmatrix} \vdots & \vdots & & \vdots \\ c_1 & c_2 & \cdots & c_n \\ \vdots & \vdots & & \vdots \end{pmatrix} \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix},\ \forall (x_1,\dots,\ x_n) \in \mathbb{R}^n\} \ = \ \{A\vec{x}\ /\ \forall \vec{x} \in \mathbb{R}^n\}.$$

- **Espacio nulo**: es el subespacio vectorial de $\mathbb{R}^n$ definido por

$$\mathcal{N}(A) = \{\vec{x} \in \mathbb{R}^n\ /\ A\vec{x} = \vec{0}\}.$$

Se satisfacen las siguientes condiciones de ortogonalidad entre los subespacios fundamentales de una matriz.

1. $\mathcal{N}(A) \ = \ (\mathcal{F}(A))^{\perp}$, por tanto, $\mathcal{F}(A) \ = \ (\mathcal{N}(A))^{\perp}$.
2. $\mathcal{N}(A^t) \ = \ (\mathcal{R}(A))^{\perp}$, por tanto, $\mathcal{R}(A) \ = \ (\mathcal{N}(A^t))^{\perp}$.

Observemos que en el ejemplo 7, para obtener $F^{\perp}$ lo que se hizo fue hallar el espacio nulo de la matriz $A = \begin{pmatrix} 1 & 1 & 1 & 0 \\ 0 & 1 & -1 & 0 \end{pmatrix}$, cuyas filas son los vectores de una base de F. Por tanto, $F^{\perp} = \mathcal{N}(A)$.

2.3. Aproximación por mínimos cuadrados. Aplicaciones

Para el desarrollo de esta sección consideraremos los vectores de $\mathbb{R}^m$ como matrices columnas de m componentes, por tanto, a partir de ahora no escribiremos con notación vectorial ($\vec{x}$). Sea A una matriz real de m filas y n columnas, y b un vector columna de m componentes. Si $m > n$, el sistema $Ax = b$ será, en general, incompatible: hay más ecuaciones que incógnitas. Cuando el sistema sea incompatible, no existe solución exacta, pero buscaremos una "solución aproximada"x_0: la que hace más pequeña la distancia entre Ax y b; es decir, la que minimiza el error $E^2 = \|Ax - b\|^2$, que recibe el nombre de **desviación cuadrática**.

Diremos por tanto que x_0 es una **solución en el sentido de los mínimos cuadrados** para el sistema $Ax = b$ si se verifica:

$$\|Ax_0 - b\| \leq \|Ax - b\| \quad \forall\ x \in \mathbb{R}^n.$$

Por tanto, nuestro problema se reduce a encontrar x_0 tal que Ax_0 sea el que más se aproxime a b entre todos los elementos de la forma Ax, $x \in \mathbb{R}^n$, esto es, buscamos x_0 tal que

$$\|Ax_0 - b\| \leq \|y - b\| \quad \forall\ \ y = Ax, \quad x \in \mathbb{R}^n. \tag{2.1}$$

Recordemos que el espacio columna de A es $\mathcal{R}(A) = \{Ax\ /\ \forall x \in \mathbb{R}^n\}$, por tanto, podemos escribir (2.1) como sigue

$$\|Ax_0 - b\| \leq \|y - b\| \qquad y \in \mathcal{R}(A).$$

De esta manera, tenemos que encontrar Ax_0, el elemento de $\mathcal{R}(A)$ que más se aproxima a b. Por el teorema de la mejor aproximación, se tiene que $Ax_0 = \mathrm{P}_{\mathcal{R}(A)}(b)$, la proyección ortogonal de b sobre $\mathcal{R}(A)$. En consecuencia, se deduce que el vector $Ax_0 - b \in \mathcal{R}(A)^{\perp}$.

Como $Ax_0 - b \in \mathcal{R}(A)^{\perp}$, y todos los elementos de $\mathcal{R}(A)$ son de la forma Ay ,$\forall\ y \in \mathbb{R}^n$, entonces,

$$Ay \cdot (Ax_0 - b) = 0 \ \ \forall\ y \in \mathbb{R}^n$$

Teniendo en cuenta que el producto escalar en $\mathbb{R}^m$ se expresa matricialmente de la forma $x \cdot y = x^t y$ obtenemos la expresión:

$$(Ay)^t(Ax_0 - b) = 0$$

de donde, sabiendo que $(Ay)^t = y^t A^t$, llegamos a que

$$y^t(A^t Ax_0 - A^t b) = 0.$$

Al ser esta expresión cierta para todo $y \in \mathbb{R}^n$, es claro que el segundo factor vale cero, es decir,

$$A^t Ax_0 = A^t b.$$

Estas ecuaciones, que siempre tienen solución, se llaman **Ecuaciones Normales de Gauss**. Resumimos el proceso en el siguiente resultado:

Teorema 12. *Sean A una matriz real de orden $m \times n$ y $b \in \mathbb{R}^m$. Las siguientes condiciones son equivalentes:*

(a) x_0 es una solución en el sentido de los mínimos cuadrados para el sistema $Ax = b$.

(b) $Ax_0 = P_{\mathcal{R}(A)}(b)$ (Ax_0 es la proyección de b sobre el espacio columna).

(c) x_0 es una solución de las ecuaciones normales de Gauss $A^t Ax_0 = A^t b$.

Como hemos visto, el sistema de ecuaciones normales de Gauss es siempre compatible. Pueden ocurrir dos casos:

1. $A^tAx_0 = A^tb$ es compatible determinado (esto es, cuando A^tA es invertible). Entonces, existe una única solución ($x_0 = (A^tA)^{-1}A^tb$), que es la solución única en el sentido de los mínimos cuadrados.

2. $A^tAx_0 = A^tb$ es compatible indeterminado (esto es, cuando A^tA NO es invertible). Entonces, existen infinitas soluciones en el sentido de los mínimos cuadrados. En este caso, elegimos aquella solución que tiene norma mínima, a la que llamaremos solución óptima. Es decir:

   x_0 es una **solución óptima** en el sentido de los mínimos cuadrados para el sistema $Ax = b$ si es una solución en mínimos cuadrados y $\|x_0\| \leq \|x\|$ para cualquier otra solución en mínimos cuadrados x.

En el siguiente Teorema veremos que la solución óptima así definida, es única y pertenece al espacio fila de la matriz A. Como consecuencia, de las infinitas soluciones del sistema de ecuaciones normales de Gauss, la solución óptima será la que pertenece a $\mathcal{F}(A)$.

Teorema 13. *Sea A una matriz real $m \times n$ y $b \in \mathbb{R}^m$. Entonces las siguientes condiciones son equivalentes:*

(a) x_0 es una solución óptima en mínimos cuadrados para $Ax = b$.

(b) $Ax_0 = P_{\mathcal{R}(A)}b$ y $x_0 \in \mathcal{F}(A)$ (o $\mathcal{R}(A^t)$).

Además, en ese caso, x_0 es la única solución óptima.

Vamos a demostrar el teorema utilizando alguno de los resultados principales de esta sección. (Por tanto, la comprensión de esta demostración implica la asimilación de los conceptos utilizados).

Demostración:

$(a) \Longrightarrow (b)$. Supongamos que x_0 es la solución óptima, esto es,

- Es solución en el sentido de los mínimos cuadrados y
- $\|x_0\| \leq \|x\|$ para cualquier otra solución x en el sentido de los mínimos cuadrados.

Consideramos el espacio nulo de A, $\mathcal{N}(A)$. Descomponemos el vector x_0:

$$x_0 = x + y \qquad x \in \mathcal{N}(A), \quad y \in (\mathcal{N}(A))^\perp.$$

Veamos que y también es solución en el sentido de los mínimos cuadrados:

Como $x \in \mathcal{N}(A)$, se tendrá que $Ax = 0$. Por tanto, si multiplicamos a la izquierda en la igualdad anterior por A, obtenemos:

$$Ax_0 = Ax + Ay = Ay.$$

Por el Teorema 12, se tiene que $Ax_0 = P_{\mathcal{R}(A)}b$, luego de la igualdad anterior se obtiene que $Ay = P_{\mathcal{R}(A)}b$. Por tanto y también es solución en el sentido de los mínimos cuadrados.

Como $x \in \mathcal{N}(A)$ y $y \in (\mathcal{N}(A)))^\perp$, se tiene que $x \perp y$. Por el teorema de Pitágoras:

$$\|x_0\|^2 = \|x + y\|^2 = \|x\|^2 + \|y\|^2 \geq \|y\|^2.$$

Pero x_0 era la solución óptima, por tanto, $\|x_0\| \leq \|y\|$. Deducimos entonces que $x_0 = y \in (\mathcal{N}(A))^\perp = \mathcal{F}(A)$.

Ejercicio: Demostrar el recíproco.

Aplicación: Regresión

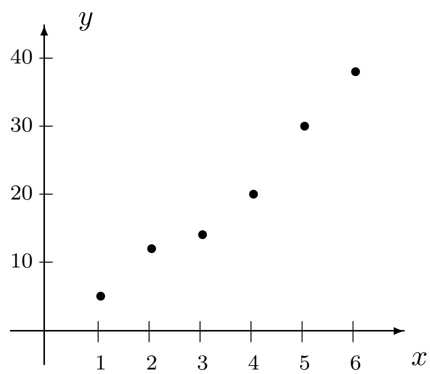
Existen muchos fenómenos de las ciencias experimentales en los que intervienen dos variables, x e y, de las que se sabe que están relacionadas, pero no se conoce cuál es la función $y = f(x)$ que expresa su dependencia. Nuestro problema consiste en determinar la mejor expresión matemática o línea geométrica para expresar los valores de dos variables (x,y) observadas.

Consideramos los pares de valores (x_i, y_i) observados en un fenómeno experimental (por ejemplo, x_i es la carga aplicada sobre una cierta estructura e y_i es la deformación que se produce; y_i es la distancia a un satélite que va camino de Marte y x_i el tiempo; x_i los costes de producción en un proceso económico e y_i los volúmenes producidos con los precios y las ganancias). Al conjunto de pares (x_i, y_i), $i = 1, 2, \ldots, m$ se les denomina **nube de puntos**.

Ejemplo 14. *Supongamos que hemos observado las alturas (en centímetros) de las plantas de soja en función de las semanas que hace que han germinado, y hemos obtenido los siguientes datos:*

x_i *(semanas)*	*1*	*2*	*3*	*4*	*5*	*6*
y_i *(cm)*	*5*	*12*	*14*	*20*	*30*	*38*

Podemos representar gráficamente los pares (x_i, y_i)*, obteniendo la nube de puntos de nuestro experimento:*



Si se admite que las variables x e y tienen una relación lineal, la curva que más se adapta a los valores observados (x_i, y_i) es una recta (como en el ejemplo anterior): su ecuación sería $y = ax + b$. Está claro que como la función $y = ax + b$ es sólo una aproximación, en general, la recta no parará por los puntos (x_i, y_i), por tanto la igualdad $y_i = ax_i + b$ no se va a cumplir para ciertas parejas de datos, cualesquiera que sean a y b.

El sistema a resolver es:

$$\begin{aligned} ax_1 + b &= y_1 \\ ax_2 + b &= y_2 \\ &\vdots \\ ax_m + b &= y_m \end{aligned}$$

La **regresión lineal** consiste en calcular a_0 y b_0 de manera que la recta $y = a_0x + b_0$ es la que mejor se ajusta a la nube de puntos (x_i, y_i) en el sentido de los mínimos cuadrados. Para ello, buscamos la solución del sistema anterior en el sentido de los mínimos cuadrados. Obtendremos dos valores a_0 y b_0, y con esto, la recta $y = a_0x + b_0$, que denominaremos **recta de regresión de y sobre x**.

Ejemplo 15. *Calculemos la recta de regresión en el ejemplo 14. Buscamos valores a, b solución en el sentido de los mínimos cuadrados del sistema*

$$\left\{\begin{array}{rl} a+b & =5 \\ 2a+b & =12 \\ 3a+b & =14 \\ 4a+b & =20 \\ 5a+b & =30 \\ 6a+b & =38 \end{array}\right. \quad \begin{array}{c} \textit{matricialmente} \\ \longrightarrow \\ A\begin{pmatrix} a \\ b \end{pmatrix} = m \end{array} \quad \begin{pmatrix} 1 & 1 \\ 2 & 1 \\ 3 & 1 \\ 4 & 1 \\ 5 & 1 \\ 6 & 1 \end{pmatrix} \begin{pmatrix} a \\ b \end{pmatrix} = \begin{pmatrix} 5 \\ 12 \\ 14 \\ 20 \\ 30 \\ 38 \end{pmatrix}.$$

Calculamos el sistema de ecuaciones normales de Gauss $A^tA\begin{pmatrix} a \\ b \end{pmatrix} = A^tm$*:*

$$\begin{pmatrix} 91 & 21 \\ 21 & 6 \end{pmatrix}\begin{pmatrix} a \\ b \end{pmatrix} = \begin{pmatrix} 529 \\ 119 \end{pmatrix}.$$

Es un sistema compatible determinado, por lo que se obtienen los valores $a_0 = \frac{45}{7}$, $b_0 = -\frac{8}{3}$. Así, la recta de regresión resultante es

$$y = \frac{45}{7}x - \frac{8}{3}.$$

Observaciones.

- Supongamos que tenemos un experimento en el que la curva que mejor se adapta a la nube de puntos es una parábola. En este caso, aplicamos el procedimiento anterior para buscar los coeficientes a, b y c, del polinomio

  $$y = ax^2 + bx + c.$$

  En general, puede aplicarse el procedimiento al problema de la regresión polinómica, $y = a_0 + a_1x + \cdots + a_nx^n$ (en este caso la matriz que se obtiene es de orden $m \times (n+1)$).

- Algunos ajustes no lineales pueden reducirse a ajustes lineales. Por ejemplo, $y = ae^{bx}$ se reduce a un ajuste lineal tomando logaritmos: $\log(y) = \log(a) + bx$.

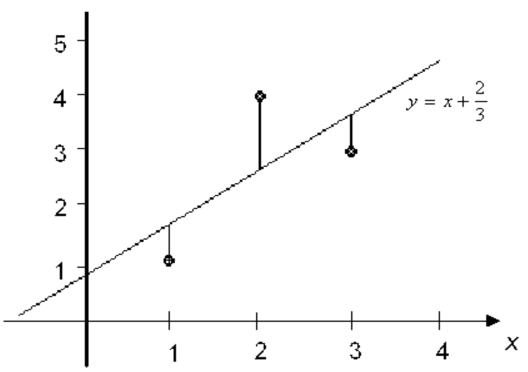
- Se pueden resolver también ajustes con más variables, como por ejemplo: $z = ax+by+c$.

Análogamente se puede buscar la **recta de regresión de x sobre y** de la forma $x = \alpha y + \beta$. En general, ambas rectas no tienen por qué ser iguales. En la recta de regresión $y = ax + b$ las distancias se miden proyectando el punto verticalmente sobre la recta, mientras que en la recta $x = \alpha y + \beta$ las distancias se miden proyectando horizontalmente:

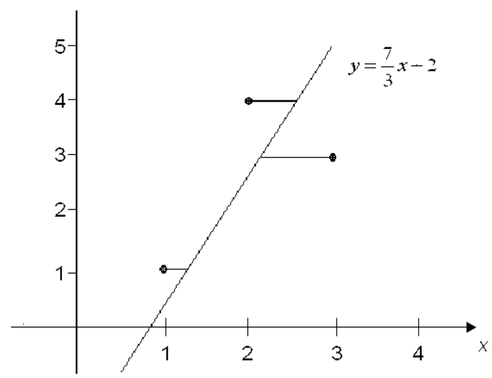
Ejemplo 16. *Para los datos:* $(1, 1), (2, 4), (3, 3)$*, se obtienen las siguientes rectas de regresión:*

- *Recta de regresión de y sobre x:* $y = x + 2/3$.

- *Recta de regresión de x sobre y:* $x = \frac{3}{7}y + \frac{6}{7}$*, de donde, despejando y,* $y = \frac{7}{3}x - 2$.

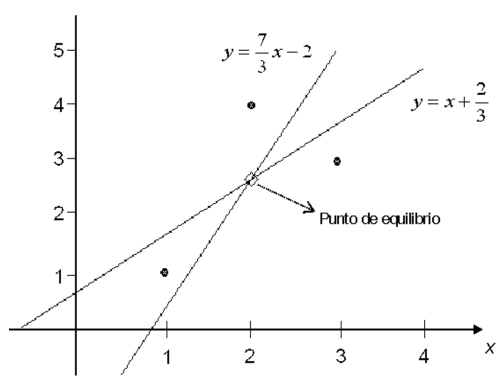


Recta de regresión de x sobre y

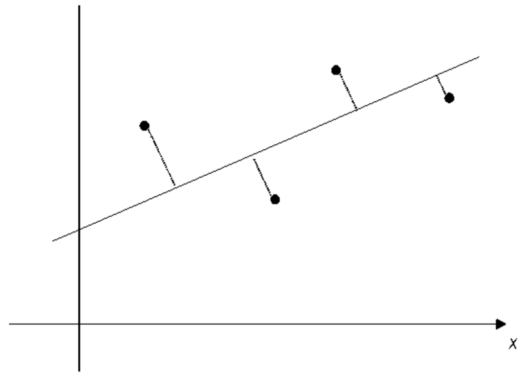


Recta de regresión de y sobre x

El punto donde se cortan ambas restas se denomina **punto de equilibrio***, y sus coordenadas vienen dadas por la media de los datos. En este ejemplo, el punto de equilibrio de los tres datos es: (media de x_i, media de y_i)=*$\left(\frac{1+2+3}{3}, \frac{1+4+3}{3}\right) = (2, \frac{8}{3})$.



Por último, para aproximar los puntos por una recta, también podríamos considerar la recta r que hace mínima la suma de los cuadrados de las distancias de cada punto a la recta: $\sum_i d((x_i, y_i), r)^2$. Estas distancias se miden proyectando el punto perpendicularmente sobre la recta:



Recta de regresión generalizada