Comentarios sobre el ejercicio 19

Enunciado:

Se considera el espacio euclídeo $(R_2[x], <\cdot,\cdot>)$ *y un producto interno dado por la fórmula:* $< p, q >= \int_0^\infty p(x)q(x)e^{-x}dx.$

Se pide calcular $\text{mín}_{a_1 a_2 \in R} \int_0^\infty [1-(a_1 x + a_2 x^2)]^2\, dx$

Nota: Recordamos que $\int_0^\infty f(x)dx = \lim_{a\to\infty} \int_0^a f(x)dx$ *para* f *continua en reales positivos, siempre que exista efectivamente ese límite.*

Vamos a ver cómo relacionar el producto interno con lo pedido.

Primero consideremos $S = gen\{x, x^2\}$ *y el* $p(x) = 1$

Recordamos la definición de norma inducida adaptada a este producto interno:

$< p, p >= \int_0^{\infty} p(x)p(x)e^{-x}dx = \|p(x)\|^2$ *no es difícil entonces comprender que en* $\int_0^{\infty}[1 - (a_1x + a_2x^2)]^2\, dx = \|1 - (a_1x + a_2x^2)\|^2$ *y como se nos pide que tal resultado sea el mínimo, podemos entonces pensar que* $a_1x + a_2x^2 = P_S(1)$.

Para calcular tal proyección necesitamos una ***base ortogonal*** *del subespacio* S.

Para eso podemos tomar el conjunto $\{x, h(x)\}$ *donde* $h(x) = \alpha x + \beta x^2 \perp x$

El planteo lleva a que $\{x, -3x + x^2\}$ *efectivamente es un conjunto ortogonal de vectores de* S.

Hemos construido una base ortogonal…. tenemos todos los elementos para ir a plantear la fórmula de la proyección:

$$P_S(1) = \underbrace{\frac{<1,x>}{\|x\|^2}}_{\gamma} x + \underbrace{\frac{<1,-3x+x^2>}{\|-3x+x^2\|^2}}_{\delta} (-3x+x^2)$$

Queda por desarrollar la última expresión para deducir quiénes son los coeficientes $a_1 a_2$.

Las cuentas están planteadas……queda trabajo pendiente.

Continuamos con el ejercicio 3 20

Sea $(V,<.,.>)$ *un espacio euclídeo y sea* $T \in L(V)$ *una proyección* $(T^2 = T)$ *con dimensión de* $V < \infty$

Son equivalentes las siguientes afirmaciones.

a) T *es una proyección ortogonal*
b) $\forall\, x, y \in V :< T(x), y >=< x, T(y) >$
c) $\forall\, x \in V: \|T(x)\| \leq \|x\|$

Recordamos (Ej. 2_ 24_27)

Si $T \in L(V)$ *es tal que* $T^2 = T \Rightarrow T$ *es la proyección de* V *sobre* $Im(T)$ *en la dirección de* $Nu(T)$ *y vale que* $V = Nu(T) \oplus Im(T)$

Como en el espacio vectorial V *se ha definido algún producto interno* $<\cdot,\cdot>$ *podemos pensar entonces que* $T(v) = v$ *si* $v \in Im(T)$ *y* $T(v) = 0_V$ *si* $v \in Im(T)^{\perp}$ *y así decimos que* T *es una proyección ortogonal.*

Ahora …… a justificar $a) \Rightarrow b)$

Para eso tomamos en el espacio vectorial V *a cada vector par de vectores* x, y *definidos de modo único como:*

$x = x_1 + x_2$ *con* $x_1 \in Im(T)$ *y* $x_2 \in Im(T)^{\perp}$

$y = y_1 + y_2$ *con* $y_1 \in Im(T)$ *y* $y_2 \in Im(T)^{\perp}$.

Evaluamos entonces

$$< \boldsymbol{T(x), y} > = < T\,(x_1 + x_2),\ y_1 + y_2 > = < T(x_1) + T(x_2), y_1 + y_2 >$$

$$= < x_1, y_1 + y_2 > = < x_1, y_1 > + < x_1\,, y_2 > = < \boldsymbol{x_1, y_1} >$$

Por otro lado:

$$< \boldsymbol{x, T(y)} > = < x_1 + x_2, T(y_1 + y_2) > = < x_1 + x_2, T(y_1) + T(y_2) > =$$

$$= < x_1 + x_2, y_1 > \ = < x_1, y_1 > + < x_2\,, y_1 > = < \boldsymbol{x_1, y_1} >$$

Conclusión: $<\boldsymbol{T}(\boldsymbol{x}),\boldsymbol{y}>=<\boldsymbol{x},\boldsymbol{T}(\boldsymbol{y})>$

$\boxed{T \text{ es una proyección ortogonal} \Rightarrow \forall\, \mathrm{x},\mathrm{y} \in \mathrm{V} :< \mathrm{T}(\mathrm{x}),\mathrm{y} >=< \mathrm{x},\mathrm{T}(\mathrm{y}) >}$

Seguimos ahora con $b \Rightarrow a$

Recordemos que siempre tendremos como hipótesis principal que $T^2 = T$ *y como consecuencia que* $\forall\, x \in V : T[T(x) - x] = 0_V$ *y* $T(x) - x \in Nu(T)$

En este caso necesitamos probar que:

$\forall\, x,y \in V :< T(x),y >=< x,T(y) > \Rightarrow T$ *es una proyección ortogonal.*

$Como\, T^2 = T$ resulta que $V = Nu(T) \oplus Im(T)$ y también $V = Nu(T)^{\perp} \oplus Nu(T)$ usando el teorema de la dimensión:

$$dim(V) = dim\, Nu(T) + dim\, Im(T)$$

y también $dim(V) = \dim Nu(T) + \dim Nu(T)^{\perp}$.

De acá se deduce que: $dim\, Im(T) = dim\, Nu(T)^{\perp}$

¿Por qué nos importa?

Porque nuestro objetivo ahora es probar que $Im\,(T) = Nu(T)^{\perp}$

Dado que ambos tienen igual dimensión, alcanza con probar que

$$Im(T) \subseteq Nu(T)^{\perp}$$

Tomemos un $y \in Im(T)$ *y* $w \in Nu(T)$ *y calculemos*

$$< y,w > = < T(v),w > \underset{< T(x),y > = < x,T(y) >}{=} < v,T(w) > = < v,0_V > = 0$$

Habiendo probado que $Nu(T) = Im(T)^{\perp}$ *entonces* T *es una proyección ortogonal*

$$\boxed{\forall\, x,y \in V : < T(x),y > = < x,T(y) > \Rightarrow T \text{ es una proyección ortogonal}}$$

Vamos con $b \Rightarrow c$

Antes, veamos algunas observaciones:

Como $\forall\, x,y \in V : < T(x),y > = < x,T(y) >$ *en particular se cumple si* $x = y$

Con esto entonces inferimos dos resultados:

- $< x,T(x) > = < T(x),x > = \overline{< x,T(x) >} \Rightarrow < x,T(x) > \in R$
- $\|T(x)\|^2 = < T(x),T(x) > \underset{< T(x),y > = < x,T(y) >}{=} < x,T\big(T(x)\big) > \underset{T = T^2}{=} < x,T(x) >$

Vamos con la prueba

$$0 \leq \|T(x)-x\|^2 = < T(x)-x, T(x)-x > = \|T(x)\|^2 + \|x\|^2 - < \underbrace{x, T(x)}_{\in R} > - < T(x), x > =$$

$$= \|T(x)\|^2 + \|x\|^2 - 2 < \underbrace{x, T(x)}_{\|T(x)\|^2} > = \|T(x)\|^2 + \|x\|^2 - 2\|T(x)\|^2$$

$$= -\|T(x)\|^2 + \|x\|^2$$

Conclusión: $\|T(x)\|^2 \leq \|x\|^2$ *y ahora sí, es inmediato que* $\|T(x)\| \leq \|x\|$

$$\boxed{\forall\, x, y \in V : < T(x), y > = < x, T(y) > \;\Rightarrow \forall\, x \in V: \|T(x)\| \leq \|x\|}$$

Hasta ahora hemos probado que $a) \Leftrightarrow b)$ *y* $b) \Rightarrow c)$

Podemos probar que $c) \Rightarrow a)$

En este caso además de la hipótesis de $T^2 = T$ *y que se cumple lo enunciado en* $c)$ *, no olvidamos que* $\dim(V) < \infty$.

De nuevo como $T^2 = T$

$V = Nu(T) \oplus Im(T)$ *y por el teorema de la dimensión*

$dim(V) = dim\, Nu(T) + dim\, Im(T)$

Además $V = Nu(T)^{\perp} \oplus Nu(T)$ *y* $dim(V) = dim\, Nu(T)^{\perp} + dim\, Nu(T)$ *.*

De acá se deduce que: $dim\, Im(T) = dim\, Nu(T)^{\perp}$

Nos importa porque nuestro objetivo ahora es probar que $Im(T) = Nu(T)^{\perp}$ *sabiendo que* $\|T(x)\| \leq \|x\| \forall x \in V$

Sólo necesitamos probar que $Im(T) \subseteq Nu(T)^{\perp}$

Tomemos $x \in Im(T) : x = x_1 + x_2$ *con* $x_1 \in Nu(T)$ *y* $x_2 \in Nu(T)^{\perp}$ *y debemos concluir que* $x_1 = 0_V$

Ahora $T(x) = T(x_1) + T(x_2) = T(x_2)$

También $T(x) = x = T(x_2)$ *ya que T es una proyección*

Siendo $x_1 \perp x_2$ *usamos el Teorema de Pitágoras* $\|x\|^2 = \|x_1\|^2 + \|x_2\|^2$

Recién ahora usamos $\|T(x)\| \leq \|x\|$ *en particular con* $\|T(x_2)\| \leq \|x_2\|$

y así $\|x_1\|^2 + \|x_2\|^2 = \|x\|^2 = \|T(x_2)\|^2 \leq \|x_2\|^2 \Rightarrow \|x_1\|^2 = 0 \Rightarrow x_1 = 0_V$

Como hemos probado que efectivamente $Im(T) \subseteq Nu(T)^{\perp}$ *y ambos con igual dimensión, concluimos entonces* $Im(T) = Nu(T)^{\perp}$, T *es una proyección ortogonal.*

$$\boxed{\forall x \in V: \|T(x)\| \leq \|x\| \Rightarrow T \text{ es una proyección ortogonal}}$$

Habiendo probado que $c) \Rightarrow a)$ *y* $a) \Rightarrow b)$ *sale también* $c) \Rightarrow b)$

Mínimos Cuadrados

Ejercicio 3_21

$(R^n, <\cdot,\cdot>)$ *con* $< \boldsymbol{x}, \boldsymbol{y} >= \boldsymbol{x}^T \boldsymbol{y}$

Sea S *subespacio de* R^n *donde* $\{v_1, v_{2,} v_3, \dots\ v_m\}$ *es una base de S.*

Sea una matriz $A \in R^{nxm}$ *:* $A = (v_1\ \ v_2\ \ v_3\ \ \dots\ \ \ v_m\ )$.

Sabemos que la proyección es una transformación lineal que en este contexto verifica:

$P: R^n \rightarrow R^n \ /\ P_S\ (v) = v\ si\ v \in S = Col(A)\ y\ P_S\ (v) = 0_{R^n}\ si\ v \in S^{\perp} = Col(A)^{\perp}.$

$$\boxed{P_S\ (v) \in Col(A)\ y\ v - P_S(v) \in Col(A)^{\perp}}$$

Como la $P_S\ (v)$ *es un vector* $\in Col(A)$ *debe existir*

$$\tilde{x} = \begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ \vdots \\ x_m \end{pmatrix} : A \begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ \vdots \\ x_m \end{pmatrix} = P_S\ (v).$$

Por otro lado (Ej_3_11) se ha demostrado que $Nul(A)^{\perp} = fil(A)$ *de modo que*

$Nul(A^T)^{\perp} = fil(A^T) = Col(A)$ *y también* $Nul\ (A^T) = Col(A)^{\perp}$

Por propiedad de la proyección $v - A \begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ \vdots \\ x_m \end{pmatrix} \in Col(A)^{\perp}$

y así ...

$$A^T(v - A \begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ \vdots \\ x_m \end{pmatrix}) = 0_{R^m} \Rightarrow A^T A \begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ \vdots \\ x_m \end{pmatrix} = A^T v \ con\ A^T A \in R^{mxm}$$

Siempre trabajando con el ***producto interno*** *canónico en* R^n *se puede demostrar que para cualquier matriz* $A \in R^{nxm}$ *es* $\boldsymbol{Nul\,(A\,) = Nul(A^T A)}$.

Prueba:

a) $Nul\,(A\,) \subset Nul(A^T A)$

$x \in Nul(A) \Rightarrow Ax = 0_{R_n} \Rightarrow A^T(Ax) = A^T.0_{R_n} \Rightarrow (A^T A)x = 0_{R_m} \Rightarrow x \in Nul(A^T A)$

b) $Nul(A^T A) \subseteq Nul(A)$

$x \in Nul(A^T A) \Rightarrow A^T Ax = 0_{R_m} \Rightarrow x^T(A^T Ax) = x^T.0_{R_m} \Rightarrow (x^T A^T)(Ax) = 0 \Rightarrow$
$(Ax)^T(Ax) = 0 \Rightarrow < Ax, Ax >= 0 \Rightarrow Ax = 0_{R^n} \Rightarrow x \in Nul(A)$

¿Por qué es útil esta propiedad en este contexto?

Porque con la misma se puede probar que $rg(A) = rg(A^T A)$.

Basta escribir el teorema de la dimensión

$dim\,Nul(A) + rg(A) = m \;y\; dim\,Nul(A^T A) + rg(A^T A) = m$

De esta manera volviendo al problema $A^T A \begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ \vdots \\ x_m \end{pmatrix} = A^T v$ *, dado que* A *tiene*

$rg(A) = m$ *se deduce que* $rg\,(A^T A) = m$ *y ya que* $A^T A \in R^{mxm}$ *la matriz* $A^T A$

tiene inversa y finalmente se puede escribir $\begin{pmatrix} \boldsymbol{x_1} \\ \boldsymbol{x_2} \\ \boldsymbol{x_3} \\ \vdots \\ \boldsymbol{x_m} \end{pmatrix} = (\boldsymbol{A^T A})^{-1} \boldsymbol{A^T v}.$

A la matriz $(A^T A)^{-1} A^T$ *se la anota como* $A^{\#}$

Es sencillo probar que $A^{\#} A = (A^T A)^{-1} A^T A = I_{mxm}$

Y también $(AA^{\#})^2 = (AA^{\#})(AA^{\#}) = A(A^{\#} A)A^{\#} = AA^{\#}.$

Más aún: $P_S\,(v) = AA^{\#}(v) = A(A^T A)^{-1} A^T v = A\tilde{x}$

De manera que efectivamente $AA^{\#}$ *es la matriz de tal proyección en la base canónica.*

Por las mismas propiedades de la proyección:

$d(v,S)^2 = \|v - P_S(v)\|^2$ *es el* ***error cuadrático*** *y es por esta razón que el vector indicado como*

$$\tilde{x} = \begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ \vdots \\ x_m \end{pmatrix}$$ *se conoce como la* ***solución por cuadrados mínimos*** *de* $Ax = v$

Comentario

Si las columnas de la matriz A no forman un conjunto linealmente independiente la solución por cuadrados mínimos de la ecuación $Ax = v$ *se encuentra resolviendo la ecuación*

$$A^T A \begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ \vdots \\ x_m \end{pmatrix} = A^T v$$

Estas ecuaciones llevan a un sistema compatible indeterminado y cuyas soluciones serán de la forma $\tilde{x} = x_p + x_H$ *con* $x_H \in Nul(A^T A) = Nul(A)$.

Ilustramos con ejemplos de cálculos sencillos:

1)Dado el sistema $Ax = v$ *con* $A = \begin{pmatrix} 1 & 0 \\ 0 & 1 \\ 1 & 1 \end{pmatrix}$ *y* $v = \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}$.

Se pide encontrar la solución por cuadrados mínimos del problema y calcular el error cuadrático.

Las columnas de A *forman un conjunto LI de manera que:*

$$\tilde{x} = (A^T A)^{-1} A^T \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix} = \left(\begin{pmatrix} 1 & 0 & 1 \\ 0 & 1 & 1 \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 0 & 1 \\ 1 & 1 \end{pmatrix} \right)^{-1} \cdot \begin{pmatrix} 1 & 0 & 1 \\ 0 & 1 & 1 \end{pmatrix} \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix} = \begin{pmatrix} \frac{1}{3} \\ \frac{1}{3} \end{pmatrix}$$

$$\textit{El error cuadrático viene dado por} \left\| \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix} - A \begin{pmatrix} \frac{1}{3} \\ \frac{1}{3} \end{pmatrix} \right\|^2 = \left\| \begin{pmatrix} \frac{-1}{3} \\ \frac{-1}{3} \\ \frac{1}{3} \end{pmatrix} \right\|^2 = \frac{1}{3}$$

$$\textit{Ya que: } P_S\,(v) = A \begin{pmatrix} \frac{1}{3} \\ \frac{1}{3} \end{pmatrix} = \begin{pmatrix} \frac{1}{3} \\ \frac{1}{3} \\ \frac{2}{3} \end{pmatrix}$$

2) En este caso tomaremos una matriz $A = \begin{pmatrix} 1 & 0 & 1 \\ 0 & 1 & 1 \\ 1 & 1 & 2 \end{pmatrix}$ *con* $v = \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}$

Planteando nuevamente $A^T A\, \tilde{x} = A^T \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}$

$\begin{pmatrix} 2 & 1 & 3 \\ 1 & 2 & 3 \\ 3 & 3 & 6 \end{pmatrix} \tilde{x} = \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}$ *tomando en este caso* $\begin{pmatrix} a \\ b \\ c \end{pmatrix}$ *y resolviendo el sistema de*

$$\tilde{x} = \begin{pmatrix} a \\ b \\ c \end{pmatrix} = \begin{pmatrix} \frac{1}{3} - c \\ \frac{1}{3} - c \\ c \end{pmatrix} = \begin{pmatrix} \frac{1}{3} \\ \frac{1}{3} \\ 0 \end{pmatrix} + c \begin{pmatrix} -1 \\ -1 \\ 1 \end{pmatrix} \; c \in R : x_p = \begin{pmatrix} \frac{1}{3} \\ \frac{1}{3} \\ 0 \end{pmatrix}, x_H = c \begin{pmatrix} -1 \\ -1 \\ 1 \end{pmatrix}$$

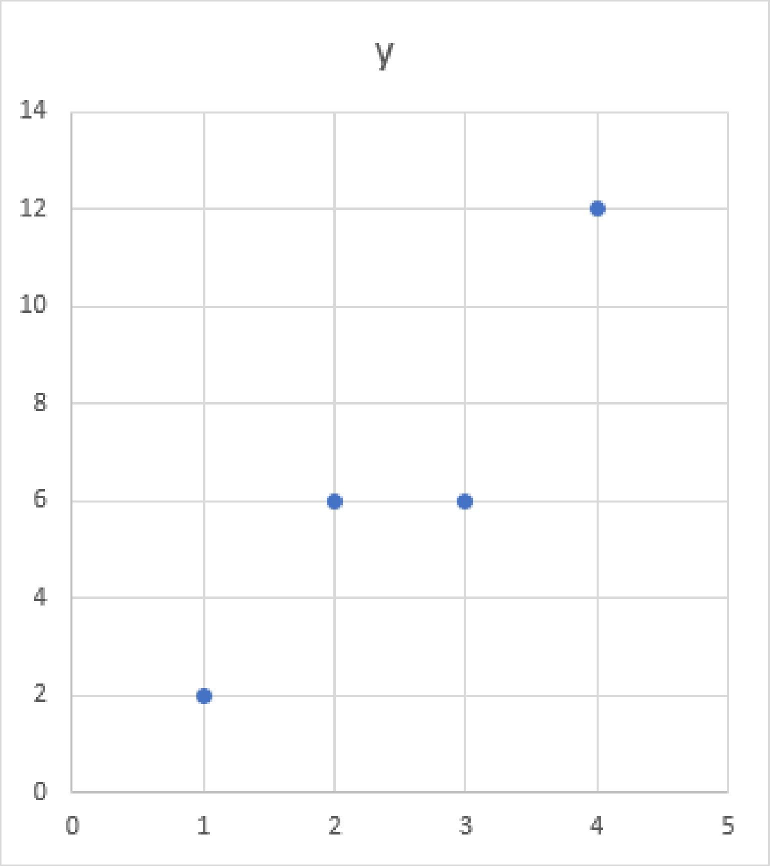
Y nuevamente: $A \begin{pmatrix} \frac{1}{3} \\ \frac{1}{3} \\ 0 \end{pmatrix} = \begin{pmatrix} 1 & 0 & 1 \\ 0 & 1 & 1 \\ 1 & 1 & 2 \end{pmatrix} \begin{pmatrix} \frac{1}{3} \\ \frac{1}{3} \\ 0 \end{pmatrix} = \begin{pmatrix} \frac{1}{3} \\ \frac{1}{3} \\ \frac{2}{3} \end{pmatrix}$ *Con error cuadrático* $\left\| \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix} - A \begin{pmatrix} \frac{1}{3} \\ \frac{1}{3} \end{pmatrix} \right\|^2 = \left\| \begin{pmatrix} \frac{-1}{3} \\ \frac{-1}{3} \\ \frac{1}{3} \end{pmatrix} \right\|^2 = \frac{1}{3}$

Cuadrados mínimos

Vamos a comenzar con un ejemplo sencillo.

Supongamos que tomamos distintas posiciones de un móvil y_i *para distintos valores del tiempo* x_i.

x_i	y_i
1	2
2	6
3	6
4	12



En realidad, en una medición para determinar un modelo matemático que ajuste a cierto fenómeno físico se toman más datos, pero lo hacemos con pocos para no complicar las cuentas.

Supongamos que el móvil realiza un movimiento rectilíneo uniforme.

Entonces intentaremos ajustar con una recta de la forma:

$$y = mx + b$$

Si planteamos para cada par de valores

$$y_i = mx_i + b$$

Obtenemos el siguiente sistema de ecuaciones

$$\begin{cases} 2 = m.1 + b \\ 6 = m.2 + b \\ 6 = m.3 + b \\ 12 = m.4 + b \end{cases}$$

Matricialmente puede escribirse como:

$$\begin{pmatrix} 1 & 1 \\ 2 & 1 \\ 3 & 1 \\ 4 & 1 \end{pmatrix}\begin{pmatrix} m \\ b \end{pmatrix} = A\begin{pmatrix} m \\ b \end{pmatrix} = \begin{pmatrix} 2 \\ 6 \\ 6 \\ 12 \end{pmatrix} = v$$

Observemos:

- *Las columnas de* A *forman un conjunto linealmente independiente.*
- *El sistema es incompatible. Es decir, no hay una recta que pase por los cuatro puntos, el vector* v *no pertenece al espacio columna de* A.
- *Desde el punto de vista del álgebra la mejor solución sería el vector* $\hat{x} = \begin{pmatrix} m \\ b \end{pmatrix}$ *tal que* $A\hat{x}$ *sea la* $P_{Col(A)}(v)$.
- $\hat{x}$ *es la solución por cuadrados mínimos.*
  $\hat{x}$ *es la solución del sistema* $A\hat{x} = P_{Col(A)}(v)$.

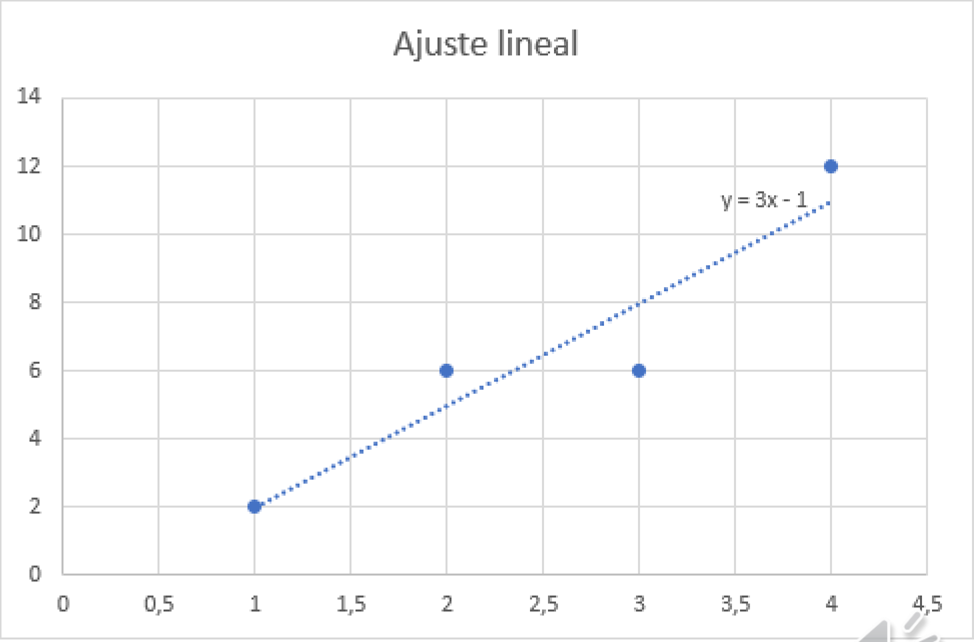
Como vimos en el ejercicio 21 resolvemos planteando
$A^T A\hat{x} = A^T v$ *en nuestro ejemplo*

$$\begin{pmatrix} 1 & 2 & 3 & 4 \\ 1 & 1 & 1 & 1 \end{pmatrix}\begin{pmatrix} 1 & 1 \\ 2 & 1 \\ 3 & 1 \\ 4 & 1 \end{pmatrix}\begin{pmatrix} m \\ b \end{pmatrix} = \begin{pmatrix} 1 & 2 & 3 & 4 \\ 1 & 1 & 1 & 1 \end{pmatrix}\begin{pmatrix} 2 \\ 6 \\ 6 \\ 12 \end{pmatrix}$$

$$\begin{pmatrix} 30 & 10 \\ 10 & 4 \end{pmatrix}\begin{pmatrix} m \\ b \end{pmatrix} = \begin{pmatrix} 80 \\ 26 \end{pmatrix}$$

Cuya solución es $\hat{x} = \begin{pmatrix} 3 \\ -1 \end{pmatrix}$ *es decir que la recta que mejor ajusta es:*

$$y = 3x - 1$$



Veamos cual es el error

$$\hat{y}_i = 3x_i - 1$$

x_i	y_i	$\hat{y}_i$	$\varepsilon_i = \hat{y}_i - y_i$
1	2	2	0
2	6	5	-1
3	6	8	2
4	12	11	-1

$$\varepsilon = \begin{pmatrix} 0 \\ -1 \\ 2 \\ -1 \end{pmatrix} y \|\varepsilon\| = \sqrt{6}$$

Veamos ahora lo siguiente:

$$y_i + \varepsilon_i = mx_i + b = \hat{y}_i$$

La idea es minimizar el error, pero si sumamos los errores vemos que pueden compensarse (eso justamente pasa en este ejemplo), entonces podríamos pensar en minimizar la suma de los módulos que es equivalente a minimizar la suma de los módulos al cuadrado. Es decir, lo que vamos a minimizar es la norma al cuadrado del vector ε.

$$\sum_{i=1}^{n} {\varepsilon_i}^2 = \sum_{i=1}^{n} (\widehat{y_i} - y_i)^2 = \sum_{i=1}^{n} (mx_i + b - y_i)^2 = f(m, b)$$

Obtenemos una función en dos variables. Tenemos que encontrar m *y* b *que minimicen la* f*.*

Como $f \geq 0$ *el punto donde el* $\nabla f = 0$ *nos proporcionará un mínimo.*

$$f_m = 2\sum_{i=1}^{n}(mx_i + b - y_i)x_i = 0$$

$$\sum_{i=1}^{n} ({mx_i}^2 + bx_i - y_i x_i) = 0$$

$$m\sum_{i=1}^{n} x_i^2 + b\sum_{i=1}^{n} x_i = \sum_{i=1}^{n} x_i y_i \quad (1)$$

$$f_b = 2\sum_{i=1}^{n}(mx_i + b - y_i) = 0$$

$$m\sum_{i=1}^{n} x_i + nb = \sum_{i=1}^{n} y_i \quad (2)$$

(1) *y* (2) *se pueden escribir matricialmente de la siguiente manera:*

$$\begin{pmatrix} \sum_{i=1}^{n} {x_i}^2 & \sum_{i=1}^{n} x_i \\ \sum_{i=1}^{n} x_i & n \end{pmatrix} \begin{pmatrix} m \\ b \end{pmatrix} = \begin{pmatrix} \sum_{i=1}^{n} x_i y_i \\ \sum_{i=1}^{n} y_i \end{pmatrix}$$

Si tomamos los datos del ejemplo nos queda:

$\begin{pmatrix} 30 & 10 \\ 10 & 4 \end{pmatrix}\begin{pmatrix} m \\ b \end{pmatrix} = \begin{pmatrix} 80 \\ 26 \end{pmatrix}$ *el mismo sistema compatible que obtuvimos haciendo el análisis desde el álgebra.*

En realidad, se trata de minimizar el error cuadrático medio

$$\frac{1}{n}\sum_{i=1}^{n} {\varepsilon_i}^2$$

que equivale a minimizar el error cuadrático. En nuestro ejemplo sería $\frac{\sqrt{6}}{4}$*.*

El objetivo del investigador es hallar $y = p_m(x)$ *de modo de minimizar:*

$$\frac{1}{n}\sum_{i=1}^{n} {\varepsilon_i}^2 = \sum_{i=1}^{n} (\widehat{y_i} - y_i)^2 = \frac{1}{n}\sum_{i=1}^{n} (p_m(x_i) - y_i)^2$$

Volviendo a nuestro ejemplo pensemos que ahora el móvil se mueve con movimiento uniformemente acelerado:

Es decir $y = a + bx + cx^2$ *donde* a *es la posición inicial,* b *es la velocidad inicial y* c *es la mitad de la aceleración.*

Entonces intentaremos ajustar con una función cuadrática.

Si planteamos para cada par de valores

$$y_i = c + bx_i + a{x_i}^2$$

Obtenemos el siguiente sistema de ecuaciones

$$\begin{cases} 2 = c + b.1 + a.1 \\ 6 = c + b.2 + a.4 \\ 6 = c + b.3 + a.9 \\ 12 = c + b.4 + a.16 \end{cases}$$

que matricialmente puede escribirse como:

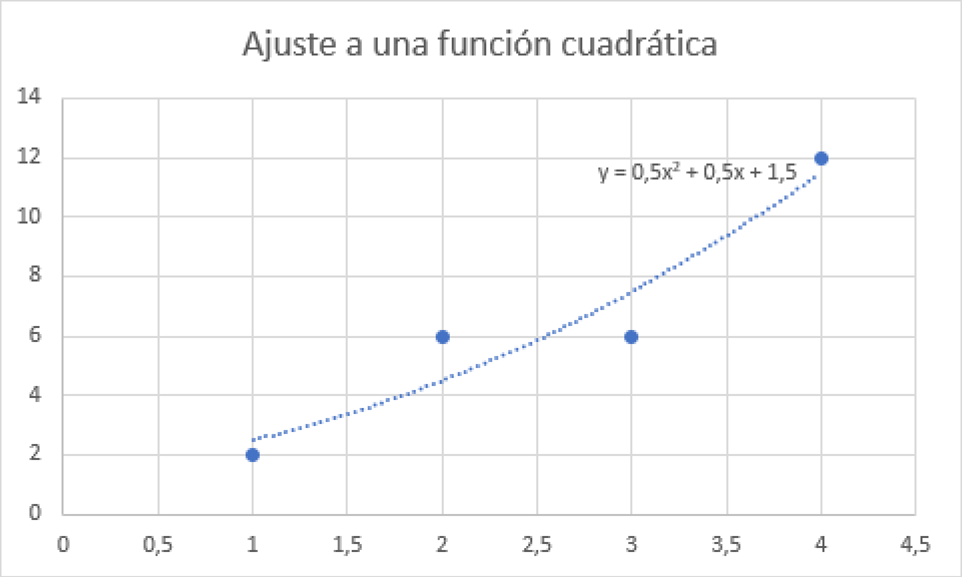
$$\begin{pmatrix} 1 & 1 & 1 \\ 1 & 2 & 4 \\ 1 & 3 & 9 \\ 1 & 4 & 16 \end{pmatrix}\begin{pmatrix} c \\ b \\ a \end{pmatrix} = A\begin{pmatrix} c \\ b \\ a \end{pmatrix} = \begin{pmatrix} 2 \\ 6 \\ 6 \\ 12 \end{pmatrix} = v$$ *este sistema es incompatible,*

resolvemos entonces el sistema compatible $A^T A\hat{x} = A^T v$

$$\begin{pmatrix} 1 & 1 & 1 & 1 \\ 1 & 2 & 3 & 4 \\ 1 & 4 & 9 & 16 \end{pmatrix}\begin{pmatrix} 1 & 1 & 1 \\ 1 & 2 & 4 \\ 1 & 3 & 9 \\ 1 & 4 & 16 \end{pmatrix}\begin{pmatrix} c \\ b \\ a \end{pmatrix} = A^T A\begin{pmatrix} c \\ b \\ a \end{pmatrix} = A^T\begin{pmatrix} 2 \\ 6 \\ 6 \\ 12 \end{pmatrix} = \begin{pmatrix} 1 & 1 & 1 & 1 \\ 1 & 2 & 3 & 4 \\ 1 & 4 & 9 & 16 \end{pmatrix}\begin{pmatrix} 2 \\ 6 \\ 6 \\ 12 \end{pmatrix}$$

$$\begin{pmatrix} 4 & 10 & 30 \\ 10 & 30 & 100 \\ 30 & 100 & 354 \end{pmatrix}\begin{pmatrix} c \\ b \\ a \end{pmatrix} = \begin{pmatrix} 26 \\ 80 \\ 272 \end{pmatrix}$$

La solución del sistema compatible determinado es $\begin{pmatrix} c \\ b \\ a \end{pmatrix} = \begin{pmatrix} 1,5 \\ 0,5 \\ 0,5 \end{pmatrix}$



Igual a como hicimos con el ajuste lineal podríamos haber minimizado el error cuadrático medio $\frac{1}{n}\sum_{i=1}^{n} \varepsilon_i{}^2$

Veamos cual es el error

$$\hat{y}_i = 0{,}5{x_i}^2 + 0{,}5x_i + 1{,}5$$

x_i	y_i	$\hat{y}_i$	$\varepsilon_i = \hat{y}_i - y_i$
1	2	2,5	0,5
2	6	4,5	-1,5
3	6	7,5	1,5
4	12	11,5	-0,5

$$\varepsilon = \begin{pmatrix} 0{,}5 \\ -1{,}5 \\ 1{,}5 \\ -0{,}5 \end{pmatrix} y \, \|\varepsilon\| = \sqrt{5}$$

Y el error cuadrático medio resulta:

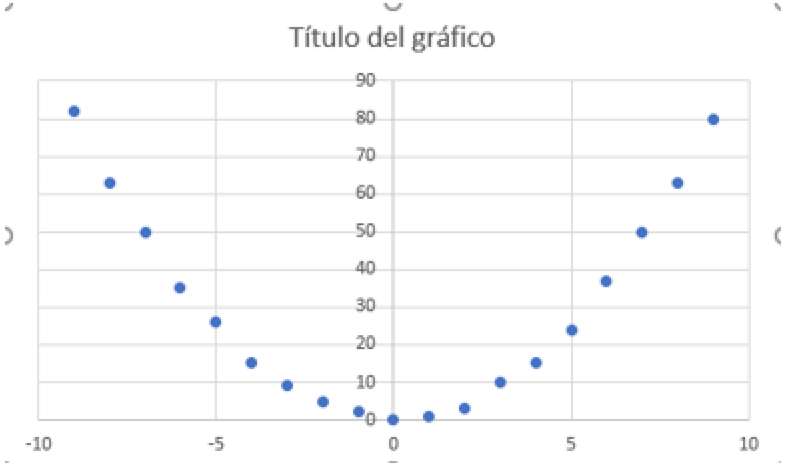
$$\frac{1}{n}\sum_{i=1}^{n} {\varepsilon_i}^2 = \frac{\sqrt{5}}{4}$$

Este error resulta menor al que obtuvimos con la aproximación lineal.

Conclusiones:

- *Cuando se realiza un experimento el n es mucho mayor que el grado del polinomio que ajusta los datos. En nuestro ejemplo esto no sucede por eso es lógico que el error cuadrático disminuya al aproximar con una función cuadrática ya que si aumentamos un grado más la función contiene a todos los puntos.*
- *Cuando se tienen los datos y el gráfico de dispersión se puede "intuir" el comportamiento de y como $f(x)$.*

Por ejemplo, a nadie se le ocurriría hacer un ajuste lineal para esta nube de puntos.



- *Muchas veces se sabe cuál debe ser el comportamiento y el cálculo del error sirve para ver si los datos son confiables o no.*

- *Otras veces se tienen modelos probables y el modelo que provee el menor error cuadrático medio será el elegido.*

Veamos ahora el ejercicio 23 de la práctica

23. Se consideran n datos experimentales $\begin{bmatrix} x_1 \\ y_1 \end{bmatrix}, \begin{bmatrix} x_2 \\ y_2 \end{bmatrix} \ldots, \begin{bmatrix} x_n \\ y_n \end{bmatrix}$, donde $x_i \neq x_j$ para $i \neq j$, y se propone un modelo polinomial para explicar el comportamiento de los mismos. Esto es, se postula la existencia de un polinomio p_m de grado m tal que

$$p_m(x_i) = y_i + \varepsilon_i$$

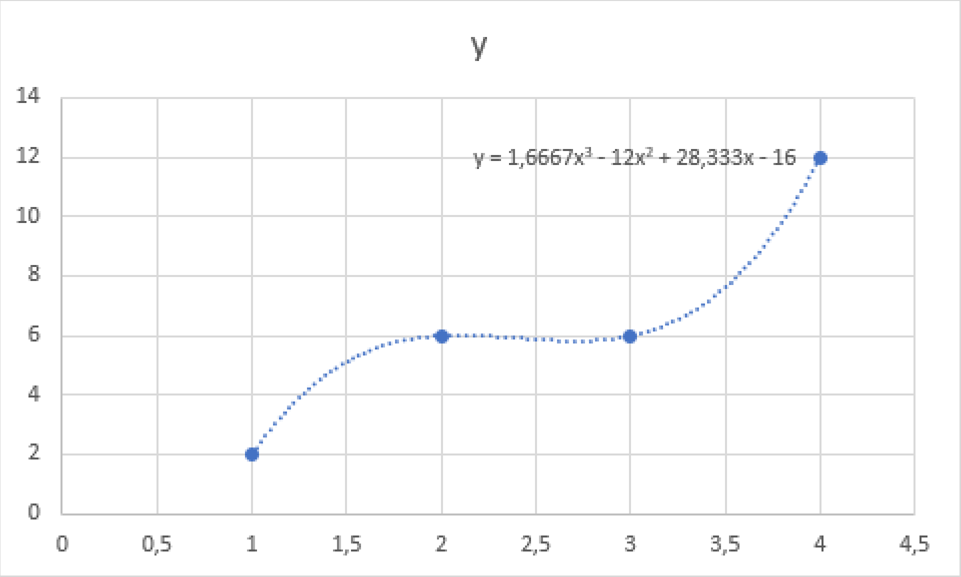
Donde $p_m(x) = \sum_{i=0}^{m} a_i x^i$

Si planteamos para cada x_i, $p_m(x_i) = y_i$ *veremos que en general el sistema resulta incompatible.*

¿Cuándo podríamos asegurar que el sistema es compatible?

En nuestro ejemplo ¿cuál debería ser el grado del polinomio para que el error resultara el vector nulo?

Exacto… debería ser de grado 3



En general si tenemos n datos el ajuste con un polinomio de grado n-1 es exacto.

Sigamos con el ejercicio…

(**a**) Escribir $p_m(x) = \sum_{i=0}^{m} a_i x^i$ y comprobar que el problema de minimizar el error cuadrático medio es equivalente al de hallar la proyección ortogonal del vector $y = \begin{bmatrix} y_1 & y_2 & \cdots & y_n \end{bmatrix}^T$ sobre el espacio columna de la matriz

$$V_m(x_1, x_2, \ldots, x_n) := \begin{bmatrix} 1 & x_1 & x_1^2 & \cdots & x_1^m \\ 1 & x_2 & x_2^2 & \cdots & x_2^m \\ \vdots & \vdots & \vdots & & \vdots \\ 1 & x_n & x_n^2 & \cdots & x_n^m \end{bmatrix} \in \mathbb{R}^{n\times(m+1)}$$

$$p_m(x) = \sum_{i=0}^{m} a_i x^i$$

Planteamos que $p_m(x_i) = y_i$ *y obtenemos en general el siguiente sistema incompatible*

$$\begin{pmatrix} 1 & x_1 & x_1^2 & \dots & x_1^m \\ 1 & x_2 & x_2^2 & \dots & x_2^m \\ 1 & x_3 & x_3^2 & \dots & x_3^m \\ \vdots & \vdots & \vdots & \dots & \vdots \\ 1 & x_n & x_n^2 & \dots & x_n^m \end{pmatrix} \begin{pmatrix} a_0 \\ a_1 \\ a_2 \\ \vdots \\ a_m \end{pmatrix} = \begin{pmatrix} y_1 \\ y_2 \\ y_3 \\ \vdots \\ y_n \end{pmatrix}$$

Si $m = n - 1$ *el sistema es compatible determinado (como ya vimos)*

Cuando $\begin{pmatrix} y_1 \\ y_2 \\ y_3 \\ \vdots \\ y_n \end{pmatrix}$ *no pertenece al espacio columna de* $V_m(x_1, x_2, \dots, x_n)$ *la mejor aproximación será la proyección ortogonal de dicho vector sobre el espacio columna de* V_m.

Podemos observar que como $x_i \neq x_j$ *cuando* $i \neq j$ *las columnas de la matriz* $V_m(x_1,\ x_2, \dots, x_n) \in R^{nx(m+1)}$ *forman un conjunto de vectores LI (recordar que n es mayor que m+1), Por lo tanto, el producto de* $V_m^T V_m$ *es una matriz* $\in R^{(m+1)x(m+1)}$ *que tiene inversa.*

Así como vimos en el ejercicio 21

$$V_m^T V_m \begin{pmatrix} a_0 \\ a_1 \\ a_2 \\ \vdots \\ a_m \end{pmatrix} = V_m^T \begin{pmatrix} y_1 \\ y_2 \\ y_3 \\ \vdots \\ y_n \end{pmatrix} \Longrightarrow \begin{pmatrix} a_0 \\ a_1 \\ a_2 \\ \vdots \\ a_m \end{pmatrix} = (V_m^T V_m)^{-1} V_m^T \begin{pmatrix} y_1 \\ y_2 \\ y_3 \\ \vdots \\ y_n \end{pmatrix}$$ *es la solución por cuadrados mínimos.*

$$\textit{Por lo tanto } V_m \begin{pmatrix} a_0 \\ a_1 \\ a_2 \\ \vdots \\ a_m \end{pmatrix} = V_m (V_m^T V_m)^{-1} V_m^T \begin{pmatrix} y_1 \\ y_2 \\ y_3 \\ \vdots \\ y_n \end{pmatrix} = P_{Col(V_m)} \begin{pmatrix} y_1 \\ y_2 \\ y_3 \\ \vdots \\ y_n \end{pmatrix}$$

No confundir la solución por cuadrados mínimos $\begin{pmatrix} a_0 \\ a_1 \\ a_2 \\ \vdots \\ a_m \end{pmatrix}$ *con la proyección del vector* $\begin{pmatrix} y_1 \\ y_2 \\ y_3 \\ \vdots \\ y_n \end{pmatrix}$ *sobre* $Col(V_m)$.

(b) Observar que la matriz $V_{m-1}(x_1, x_2, \ldots, x_m)$ *es la matriz de cambio de coordenadas* M_E^L *de la base canónica de* $\mathbb{R}_{m-1}[x]$ *en la base de los polinomios interpoladores de Lagrange correspondientes al conjunto de abscisas* $\{x_1, x_2, \ldots, x_m\}$ *:*

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

$L \coloneqq \{p_1(x), p_2(x), \ldots, p_m(x)\}$ *es un conjunto de m polinomios de grado m-1*

Recordemos como armábamos un polinomio a partir de los polinomios de Lagrange.

Dados $(x_1,\ y_1); (x_2,\ y_2); (x_3,\ y_3); \dots;\ (x_m,\ y_m)$

El polinomio que pasa por todos esos puntos se construye de la siguiente manera:

$$p(x) = p_1(x).\, y_1 + p_2(x).\, y_2 + \cdots + p_m(x).\, y_m$$

Entonces el polinomio constante que pasa por los puntos

$(x_1,\ 1); (x_2,\ 1); (x_3,\ 1); \dots;\ (x_m,\ 1)$

es

$$p(x) = 1 = p_1(x).\, 1 + p_2(x).\, 1 + \cdots + p_m(x).\, 1$$

Ya que para todo x , $p(x) = 1 \quad y_i = 1 \ \ \forall\, i$

$$\Rightarrow [1]^L = \begin{pmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{pmatrix}$$

El polinomio lineal $p(x) = x \quad y_i = x_i \ \ \forall \ i$ *que pasa por los puntos*

$$(x_1, x_1\,); (x_2,\ x_2); (x_3,\ x_3); \ldots;\ (x_m,\ x_m)$$

Es

$$p(x) = x = p_1(x).\,x_1 + p_2(x).\,x_2 + \cdots + p_m(x).\,x_m$$

$$\Rightarrow [x]^L = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_m \end{pmatrix}$$

Así podemos escribir cada uno de los polinomios de la base canónica como combinación lineal de los polinomios de Lagrange

Así el polinomio $p(x) = x^{m-1} \quad y_i = x_i^{m-1}$ *que pasa por los puntos*

$$(x_1, x_1^{m-1}\,); (x_2,\ x_2^{m-1}); (x_3,\ x_3^{m-1}); \ldots;\ (x_m,\ x_m^{m-1})$$

$$\Rightarrow [x^{m-1}]^L = \begin{pmatrix} x_1^{m-1} \\ x_2^{m-1} \\ \vdots \\ x_m^{m-1} \end{pmatrix}$$

Es $p(x) = x^{m-1} = p_1(x).\,x_1^{m-1} + p_2(x).\,x_2^{m-1} + \cdots + p_m(x).\,x_m^{m-1}$

$$M_E^L = \begin{pmatrix} 1 & x_1 & x_1^2 & \dots & x_1^{m-1} \\ 1 & x_2 & x_2^2 & \dots & x_2^{m-1} \\ 1 & x_3 & x_3^2 & \dots & x_3^{m-1} \\ \vdots & \vdots & \vdots & \dots & \vdots \\ 1 & x_m & x_m^2 & \dots & x_m^{m-1} \end{pmatrix}$$

Siendo $V_{m-1} = M_E^L$ *una matriz de cambio de base, tiene inversa por lo tanto su rango es máximo e igual a* m.

Además $V_{m-1}^{-1} = M_L^E$.

(**c**) Deducir que el rango de la matriz $V_m(x_1, x_2, \dots, x_n)$ es $m+1$.

$$V_m(x_1, x_2, \dots, x_n) = \begin{pmatrix} 1 & x_1 & x_1^2 & \dots & x_1^{m} \\ 1 & x_2 & x_2^2 & \dots & x_2^{m} \\ 1 & x_3 & x_3^2 & \dots & x_3^{m} \\ \vdots & \vdots & \vdots & \dots & \vdots \\ 1 & x_n & x_n^2 & \dots & x_n^{m} \end{pmatrix}$$

$$V_m(x_1, x_2, \ldots, x_n) = \begin{pmatrix} 1 & x_1 & x_1^2 & \ldots & x_1^m \\ 1 & x_2 & x_2^2 & \ldots & x_2^m \\ 1 & x_3 & x_3^2 & \ldots & x_3^m \\ \vdots & \vdots & \vdots & \ldots & \vdots \\ 1 & x_n & x_n^2 & \ldots & x_n^m \end{pmatrix}$$

Al ser $x_i \neq x_j \; \forall i \neq j$ las columnas de la matriz forman un conjunto LI (recordar que en los problemas de ajustes n es mucho menor que $m+1$) por lo tanto la matriz tiene rango $m+1$

(d) Concluir que $[p_m]^{\mathcal{E}} = \overset{\#}{V}_m(x_1, x_2, \ldots, x_n) \;\; y$.

Vimos que la solución por cuadrados mínimos es

$$\begin{pmatrix} a_0 \\ a_1 \\ a_2 \\ \vdots \\ a_m \end{pmatrix} = (V_m^T V_m)^{-1} V_m^T \begin{pmatrix} y_1 \\ y_2 \\ y_3 \\ \vdots \\ y_n \end{pmatrix} = V_m^{\#} \begin{pmatrix} y_1 \\ y_2 \\ y_3 \\ \vdots \\ y_n \end{pmatrix} \quad \text{y además} \quad [p_m]^{\varepsilon} = \begin{pmatrix} a_0 \\ a_1 \\ a_2 \\ \vdots \\ a_m \end{pmatrix}$$

Para el ejercicio 24

24. Usando la técnica de mínimos cuadrados, ajustar los siguientes datos

x	-5	-4	-3	-2	-1	0	1	2	3	4	5
y	2	7	9	12	13	14	14	13	10	8	4

mediante una recta $y = a_0 + a_1 x$, y mediante una cuadrática $y = a_0 + a_1 x + a_2 x^2$. ¿Cuál de estas dos curvas se ajusta mejor a los datos?

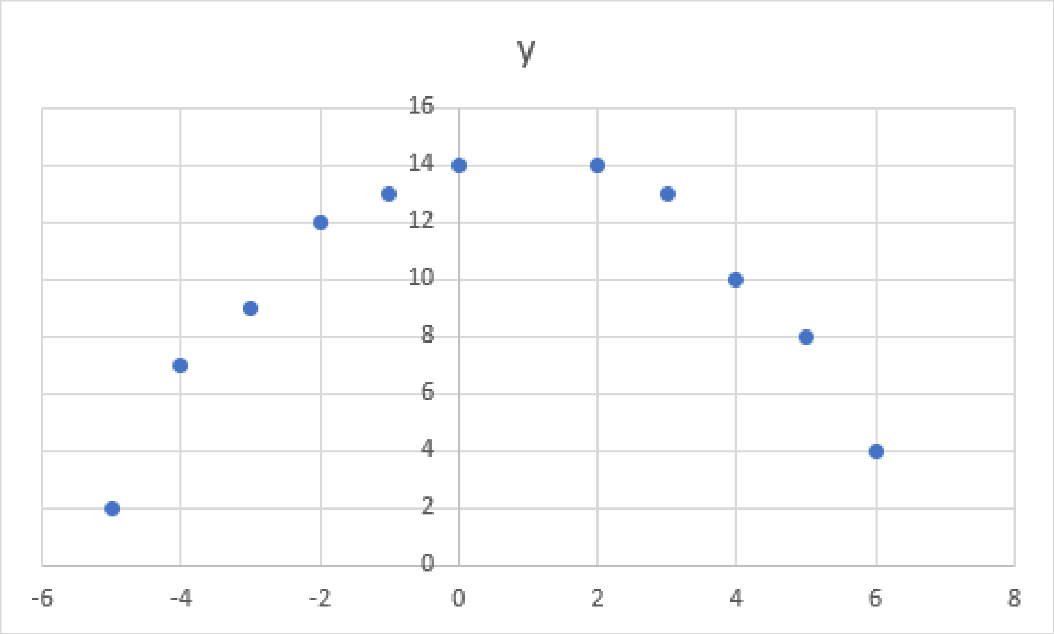
Pueden usar un archivo Excel

Primero escriban los datos en dos columnas

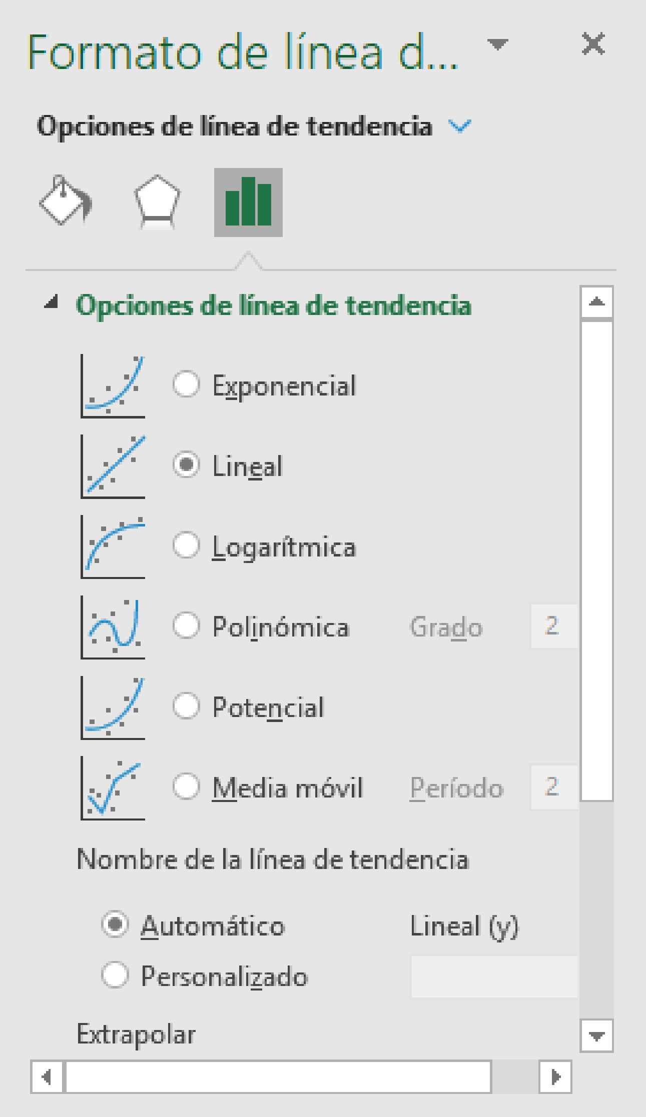
x	y
-5	2
-4	7
-3	9
-2	12
-1	13
0	14
2	14
3	13
4	10
5	8
6	4

Con los datos seleccionados vayan a → Insertar Gráfico de dispersión.

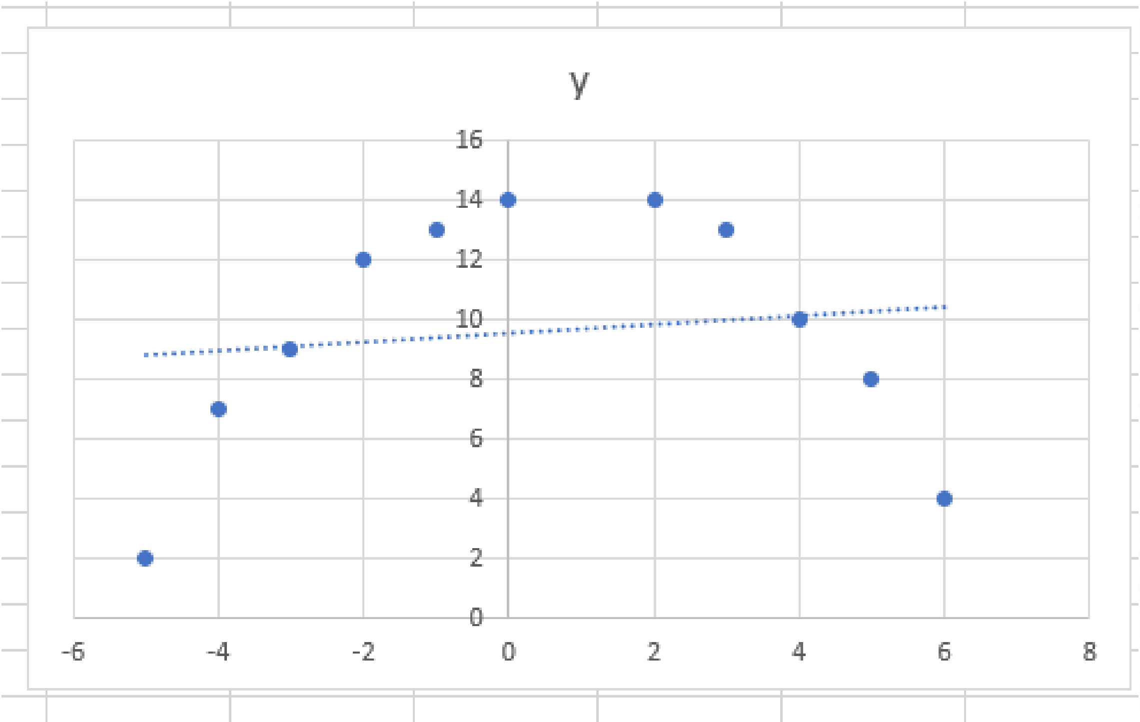
Obtienen entonces el siguiente gráfico



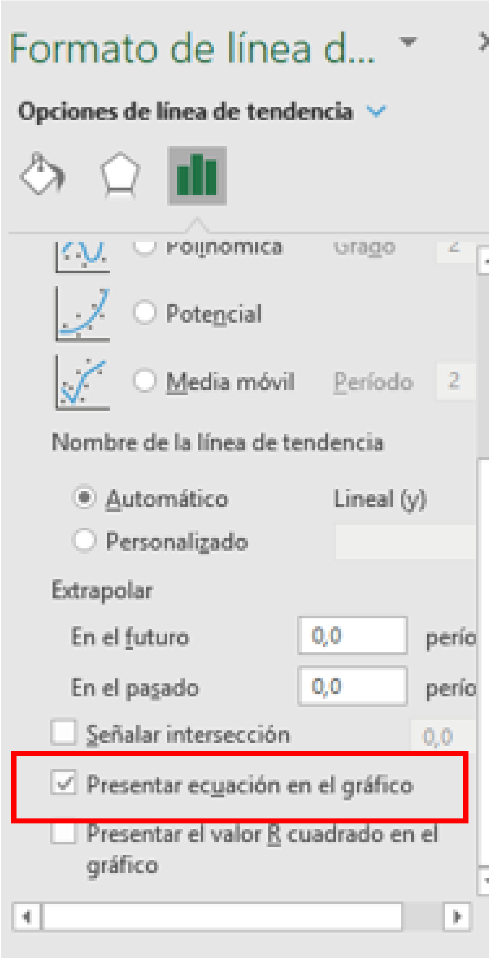
Luego seleccionan cualquiera de los puntos y hacen clic con el botón derecho del mouse y seleccionan agregar línea de tendencia, ahí podrán elegir entre varias opciones:

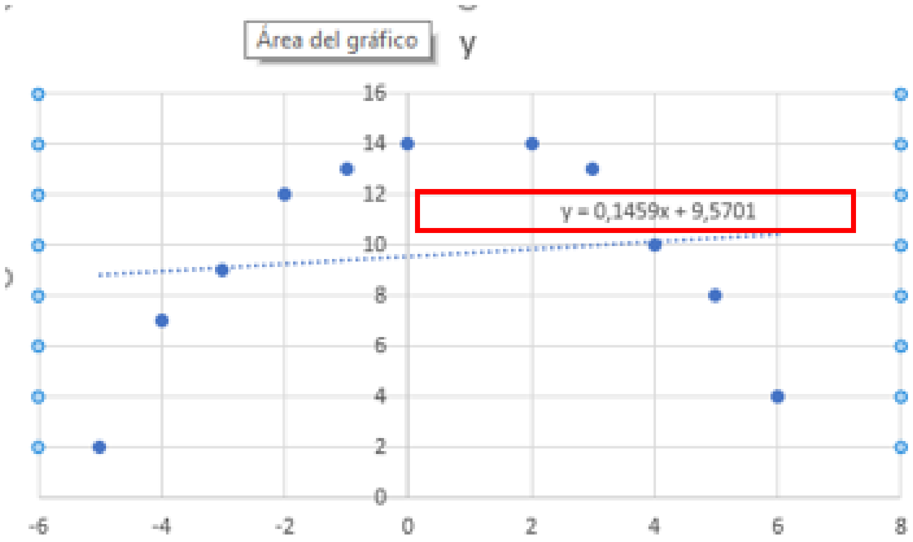


Esto se obtiene al seleccionar línea de tendencia lineal.

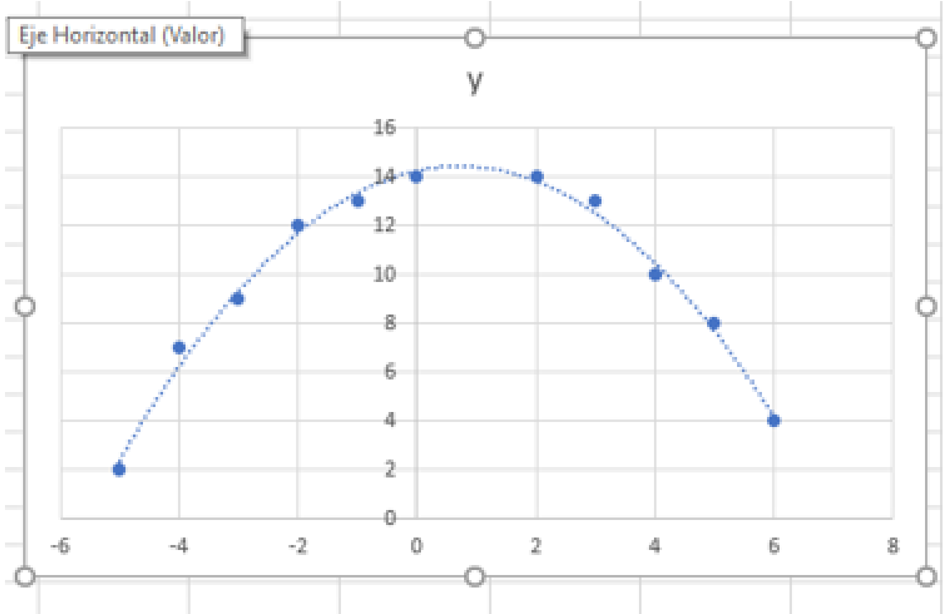


Si seleccionan también "Presentar ecuación en el gráfico"





Si luego seleccionan polinómica de grado 2 obtienen



Este ajuste es a simple vista mucho mejor….

Para el ejercicio 25 una pista…. Si $x = ae^{bt} \Rightarrow lnx = lna + bt$ *, así pueden hacer un ajuste lineal usando los puntos* (t, lnx)