

CAPÍTULO 6

Análisis de Datos

6.1. Ajuste por cuadrados mínimos. Es un procedimiento para ajustar a un conjunto de datos (pares ordenados y familia de funciones), la función que mejor se aproxime a los datos (línea de regresión o la línea de mejor ajuste), dentro de una familia de funciones dependiente de uno o varios parámetros. Concretamente, dada una tabla

x	x_1	x_2	\dots	x_N
y	y_1	y_2	\dots	y_N

buscamos una función $f(x; \alpha_1, \dots, \alpha_m)$ que realice el menor error de los residuos r_j definidos por $r_j = y_j - f(x_j; \alpha_1, \dots, \alpha_m)$ en el sentido de mínimos cuadrados. La idea es minimizar la suma S de cuadrados de los residuos:

$$S(\alpha_1, \dots, \alpha_m) = \frac{1}{2}(r_1^2 + \dots + r_N^2) = \frac{1}{2} \sum_{j=1}^N (y_j - f(x_j; \alpha_1, \dots, \alpha_m))^2$$

Este método se utiliza comúnmente para analizar una serie de datos que se obtengan de algún estudio, con el fin de expresar su comportamiento de manera lineal y así minimizar los errores de los datos obtenidos.

6.1.1. Modelo lineal. Vamos a estudiar el siguiente problema, supongamos que dos variables se relacionan de la forma $y = \alpha + \beta x$, pero los coeficientes α y β son desconocidos. Si mediante observaciones tenemos para distintos valores de x , los valores de la variable y correspondiente, podemos hallar los coeficientes desconocidos resolviendo el sistema

$$(6.1) \quad \begin{cases} y_1 = \alpha + \beta x_1, \\ y_2 = \alpha + \beta x_2, \\ \vdots \\ y_N = \alpha + \beta x_N. \end{cases}$$

Si $N = 1$ el sistema admite infinitas soluciones, para $N = 2$ hay una única solución supuesto que $x_1 \neq x_2$. Para $N > 2$, el sistema está sobredeterminado, por lo tanto tendremos solución sólo en el caso de tener los valores exactos de x_j e y_j (y asumiendo que la relación entre las variables es estrictamente lineal). Si dichos valores provienen de observaciones o de cálculos numéricos, esto no sucede y el sistema (6.1) no tiene solución. Vamos a replantear el problema de la siguiente forma: queremos hallar α, β que minimicen la distancia entre las observaciones y la predicción realizada con el modelo. Más concretamente, si definimos $r_j = y_j - \hat{y}_j$, donde $\hat{y}_j = \alpha + \beta x_j$, buscamos que $S = \frac{1}{2}(r_1^2 + \dots + r_N^2)$ sea mínimo. Podemos escribir S en función

de α y β :

$$S(\alpha, \beta) = \frac{1}{2} \sum_{j=1}^N (y_j - \alpha - \beta x_j)^2,$$

sus derivadas valen

$$\begin{aligned} \frac{\partial S}{\partial \alpha}(\alpha, \beta) &= - \sum_{j=1}^N (y_j - \alpha - \beta x_j) = - \sum_{j=1}^N y_j + \alpha N + \beta \sum_{j=1}^N x_j, \\ \frac{\partial S}{\partial \beta}(\alpha, \beta) &= - \sum_{j=1}^N (y_j - \alpha - \beta x_j) x_j = - \sum_{j=1}^N y_j x_j + \alpha \sum_{j=1}^N x_j + \beta \sum_{j=1}^N x_j^2. \end{aligned}$$

Para minimizar buscamos un punto (α, β) crítico, es decir donde ambas derivadas se anulen, esto nos lleva a las ecuaciones normales

$$\begin{cases} \sum_{j=1}^N y_j = \alpha N + \beta \sum_{j=1}^N x_j, \\ \sum_{j=1}^N y_j x_j = \alpha \sum_{j=1}^N x_j + \beta \sum_{j=1}^N x_j^2, \end{cases}$$

por ser la función S convexa, un punto crítico es necesariamente un mínimo (global).

Ejemplo 6.1. Consideremos la Tabla 6.1, las ecuaciones normales se escriben como

x	y	\hat{y}	$y - \hat{y}$
1	2.65707	2.75794	-0.10087
2	4.04218	3.86579	0.17639
3	4.76812	4.97365	-0.20553
4	6.36685	6.08150	0.28535
5	7.03401	7.18936	-0.15535

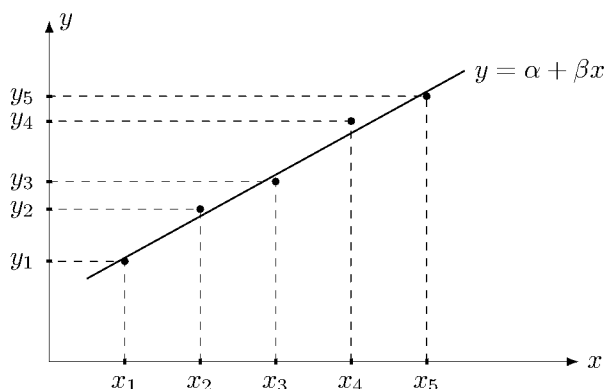
Tabla 6.1

$$\begin{cases} 24.8682 = 5\alpha + 15\beta \\ 85.6833 = 15\alpha + 55\beta \end{cases}$$

cuya solución es $\alpha = 1.65008$, $\beta = 1.10786$, $S \cong 0.095$. En el gráfico 6.1 mostramos los puntos de la tabla y la recta obtenida.

6.1.2. Interpretación geométrica. Vamos a darle un sentido geométrico al problema planteado, empezaremos por recordar la noción de producto interno. Dados dos vectores de $\mathbf{u}, \mathbf{v} \in \mathbb{R}^N$ se define el producto interno como $\mathbf{u} \cdot \mathbf{v} = u_1 v_1 + \cdots + u_N v_N \in \mathbb{R}$. De la definición se ve claramente que

- $\mathbf{u} \cdot (\mathbf{v} + \mathbf{w}) = \mathbf{u} \cdot \mathbf{v} + \mathbf{u} \cdot \mathbf{w}$,
- $\mathbf{u} \cdot \mathbf{v} = \mathbf{v} \cdot \mathbf{u}$,
- $\mathbf{u} \cdot \mathbf{u} = u_1^2 + \cdots + u_N^2 \geq 0$ y $\mathbf{u} \cdot \mathbf{u} = 0$ si y sólo si $\mathbf{u} = \mathbf{0}$.

Fig. 6.1: Recta de ajuste $y = \alpha + \beta x$.

Vemos que la norma de un vector se relaciona con el producto interno de la forma: $\|\mathbf{u}\| = \sqrt{\mathbf{u} \cdot \mathbf{u}}$. Se puede probar la desigualdad de Cauchy:

$$-\|\mathbf{u}\| \|\mathbf{v}\| \leq \mathbf{u} \cdot \mathbf{v} \leq \|\mathbf{u}\| \|\mathbf{v}\|,$$

lo que nos permite definir un ángulo $\theta \in [0, \pi]$ como el que verifica $\mathbf{u} \cdot \mathbf{v} = \|\mathbf{u}\| \|\mathbf{v}\| \cos(\theta)$.

Dada una recta en el plano generado por el vector \mathbf{u} y un punto \mathbf{v} fuera de la misma, sabemos que la mínima distancia entre la recta y el punto, se alcanza en la intersección $\hat{\mathbf{v}}$ con la perpendicular que pasa por el punto exterior (ver Figura 6.2(a)). De forma similar, si consideramos el plano generado por los vectores $\mathbf{u}_1, \mathbf{u}_2$ del espacio tridimensional y un punto \mathbf{v} exterior al mismo, la distancia mínima del punto al plano se obtiene en el punto $\hat{\mathbf{v}}$ proyección ortogonal de \mathbf{v} sobre el plano, como se muestra en la Figura 6.2(b). Si \mathbf{w} es otro punto del plano, por el Teorema de Pitágoras tenemos

$$\|\mathbf{v} - \mathbf{w}\|^2 = \|\mathbf{v} - \hat{\mathbf{v}}\|^2 + \|\hat{\mathbf{v}} - \mathbf{w}\|^2 \geq \|\hat{\mathbf{v}} - \mathbf{w}\|^2$$

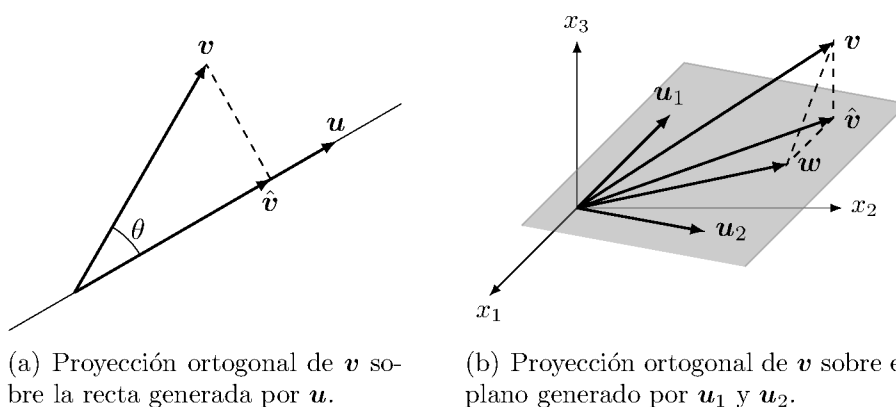


Fig. 6.2: Proyección ortogonal sobre subespacios de dimensión 1 y 2.

Podemos usar estas ideas para interpretar las ecuaciones normales, si $\mathbf{1}, \mathbf{x}, \mathbf{y} \in \mathbb{R}^N$ son los vectores definidos por $\mathbf{1} = (1 \dots 1)^T$, $\mathbf{x} = (x_1 \dots x_N)^T$ e $\mathbf{y} = (y_1 \dots y_N)^T$, buscamos el vector $\hat{\mathbf{y}}$ en el plano generado por los vectores $\mathbf{1}, \mathbf{x}$ que realice la mínima distancia al vector \mathbf{y} . Tenemos que $\hat{\mathbf{y}} = \alpha \mathbf{1} + \beta \mathbf{x}$ debe verificar $\mathbf{y} - \hat{\mathbf{y}} \perp \mathbf{1}$, $\mathbf{y} - \hat{\mathbf{y}} \perp \mathbf{x}$. Se deduce que α, β deben satisfacer el sistema (2×2) :

$$\begin{aligned} \mathbf{y} \cdot \mathbf{1} &= \alpha \mathbf{1} \cdot \mathbf{1} + \beta \mathbf{1} \cdot \mathbf{x}, \\ \mathbf{y} \cdot \mathbf{x} &= \alpha \mathbf{1} \cdot \mathbf{x} + \beta \mathbf{x} \cdot \mathbf{x}, \end{aligned}$$

que tiene única solución si y sólo si los vectores $\mathbf{1}, \mathbf{x}$ no son paralelos, es decir no son todos iguales los valores x_j . Se suele denominar residuo a $\mathbf{r} = \mathbf{y} - \hat{\mathbf{y}}$, el valor de S se puede calcular como $\frac{1}{2}\|\mathbf{r}\|^2$. Si \mathbf{z} es otro vector del plano generado por $\mathbf{1}$ y \mathbf{x} , $\mathbf{z} = \gamma\mathbf{1} + \delta\mathbf{x}$, tenemos

$$\|\mathbf{y} - \mathbf{z}\|^2 = \|\mathbf{y} - \hat{\mathbf{y}} + \hat{\mathbf{y}} - \mathbf{z}\|^2 = \|\mathbf{r}\|^2 + \|\hat{\mathbf{y}} - \mathbf{z}\|^2 + 2\mathbf{r} \cdot (\hat{\mathbf{y}} - \mathbf{z}).$$

Como \mathbf{r} es perpendicular a todo elemento generado por $\mathbf{1}$ y \mathbf{x} , el doble producto se anula por lo tanto $\|\mathbf{y} - \mathbf{z}\|^2 = \|\mathbf{r}\|^2 + \|\hat{\mathbf{y}} - \mathbf{z}\|^2$.

Si bien la función S no permite hallar los valores de α y β que hacen óptimo el ajuste del modelo en el sentido de mínimos cuadrados, su valor numérico no tiene significado. Observemos por ejemplo que S se modifica con cambios de escala (de unidades). Queremos un indicador de la bondad del ajuste invariante por transformaciones lineales de la variable y . Si consideramos $\bar{\mathbf{y}} = (\mathbf{y} \cdot \mathbf{1} / \mathbf{1} \cdot \mathbf{1})\mathbf{1} = (\bar{y} \dots \bar{y})^T$, donde $\bar{y} = (y_1 + \dots + y_N) / N$, podemos escribir

$$\|\mathbf{r}\|^2 = \|\mathbf{r} - \bar{\mathbf{y}}\|^2 - \|\hat{\mathbf{y}} - \bar{\mathbf{y}}\|^2 = \|\mathbf{y} - \bar{\mathbf{y}}\|^2 \left(1 - \frac{\|\hat{\mathbf{y}} - \bar{\mathbf{y}}\|^2}{\|\mathbf{y} - \bar{\mathbf{y}}\|^2}\right).$$

Como $\|\mathbf{y} - \bar{\mathbf{y}}\|^2$ es independiente de la elección los parámetros α y β , vemos que el ajuste depende del cociente $R^2 = \|\hat{\mathbf{y}} - \bar{\mathbf{y}}\|^2 / \|\mathbf{y} - \bar{\mathbf{y}}\|^2 \in [0, 1]$, mayores valores de R^2 reflejan un mejor ajuste. Al contrario de lo que sucede con S , el coeficiente R^2 no cambia por modificaciones lineales de la variable y , es decir que si consideramos la variable $Y = my + b$, tenemos $\alpha_Y = b + m\alpha_y$, $\beta_Y = m\beta_y$, $\chi_Y^2 = m^2\chi_y^2$ pero $R_Y^2 = R_y^2$.

Ejemplo 6.2. En la Tabla 6.2 (ver <https://www.engineeringtoolbox.com/>) se muestran la temperatura ($^{\circ}\text{C}$) de fusión del hielo para distinta presiones (MPa). Queremos establecer una relación lineal $T = \alpha + \beta P$.

Presión	Temp	Presión	Temp	Presión	Temp
6.1×10^{-4}	0.01	2.0×10^1	-1.54	9.0×10^1	-7.91
1.0×10^{-1}	0.003	3.0×10^1	-2.36	1.0×10^2	-8.94
1.0	-0.064	4.0×10^1	-3.21	1.2×10^2	-11.09
2.0	-0.14	5.0×10^1	-4.09	1.4×10^2	-13.35
5.0	-0.37	6.0×10^1	-5.00	1.6×10^2	-15.73
1.0×10^1	-0.75	7.0×10^1	-5.94	1.8×10^2	-18.22
1.5×10^1	-1.14	8.0×10^1	-6.91	2.0×10^2	-20.83

Tabla 6.2

Las ecuaciones normales son

$$\begin{cases} -1.2757 \times 10^2 = 21\alpha + 1.3731 \times 10^2 \beta, \\ -1.6498 \times 10^3 = 1.3731 \times 10^2 \alpha + 1.7075 \times 10^3 \beta, \end{cases}$$

de donde obtenemos los parámetros $\alpha = 0.5116$, $\beta = -1.007$ y los indicadores de bondad del ajuste $S = 2.71$, $R^2 = 0.993$. Si expresamos la temperatura en $^{\circ}\text{F}$ en lugar de hacerlo en $^{\circ}\text{C}$, $T (^{\circ}\text{F}) = 32 + 1.8T (^{\circ}\text{C})$, obtenemos $\alpha = 32.920$, $\beta = -1.813$ y $S = 8.77$, $R^2 = 0.993$. En la Figura 6.3, los puntos presión-temperatura y la recta de ajuste.

6.2. Linealización. En general la ley que relaciona las variables de interés x e y , $y = f(x)$ no es lineal, pero en muchos casos existen transformaciones simples de cada una de ellas que convierten al problema en lineal. Es decir, si consideramos nuevas variables $X = g(x)$, $Y = h(y)$, la relación anterior se escribe como $Y = \alpha + \beta X$. En la Tabla 6.3 mostramos distintos modelos no lineales que aparecen con frecuencia en las aplicaciones y las transformaciones para linealizarlos.

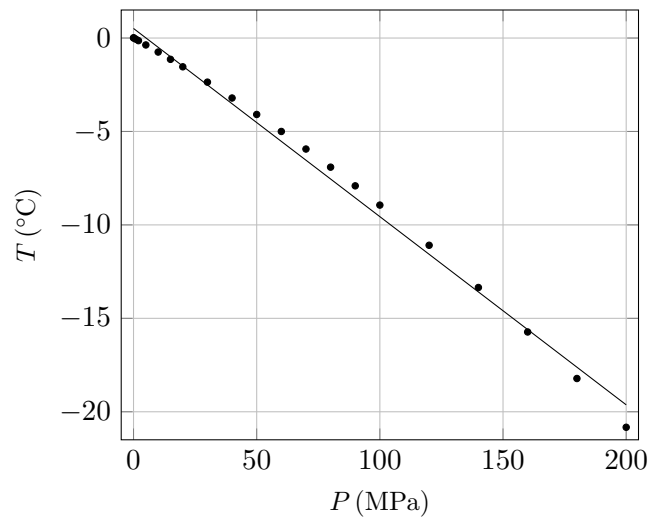


Fig. 6.3

Modelo	Transformación		
	$Y = h(y)$	$X = g(x)$	$Y = \alpha + \beta X$
Exponencial: $y = k e^{bx}$	$Y = \ln(y)$	$X = x$	$Y = \ln(k) + b X$
Potencia: $y = k x^b$	$Y = \ln(y)$	$X = \ln(x)$	$Y = \ln(k) + b X$
Logarítmica: $y = k + b \ln(x)$	$Y = y$	$X = \ln(x)$	$Y = k + b X$
Hiperbólico: $y = k x / (b + x)$	$Y = 1/y$	$X = 1/x$	$Y = 1/k + b/k X$

Tabla 6.3: Modelos no lineales y las linealizaciones.

Ejemplo 6.3 (Modelo potencial). La tercera ley de Kepler establece una relación potencial entre el período orbital (año) de cada planeta y el semieje mayor de su órbita. En la Figura 6.4 mostramos la órbita elíptica del planeta alrededor del sol, el cual ocupa un foco de la elipse (primera ley de Kepler). En la Tabla 6.4 mostramos los valores de las semiejes a medida en metros y el período orbital medido en días. La tercera ley de Kepler establece la relación

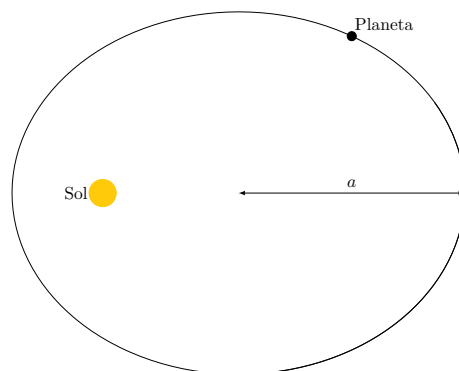


Fig. 6.4: Órbita elíptica y semieje mayor.

$\tau = k a^b$, que se transforma en $T = \alpha + \beta A$, donde $T = \ln(\tau)$, $\alpha = \ln(k)$ y $b = \beta$. Las ecuaciones normales son

$$\begin{cases} 61.3237 = 8.0 \alpha + 215.260 b, \\ 1677.42 = 215.260 \alpha + 5810.35 b, \end{cases}$$

Planeta	a (m)	τ (d)
Mercurio (☿)	$5.790\,922\,70 \times 10^{10}$	$8.796\,934 \times 10^1$
Venus (♀)	$1.082\,094\,75 \times 10^{11}$	$2.247\,010 \times 10^2$
Tierra (♁)	$1.495\,982\,62 \times 10^{11}$	$3.652\,570 \times 10^2$
Marte (♂)	$2.279\,388\,24 \times 10^{11}$	$6.869\,601 \times 10^2$
Júpiter (♃)	$7.783\,408\,21 \times 10^{11}$	$4.335\,355 \times 10^3$
Saturno (♄)	$1.426\,666\,42 \times 10^{12}$	$1.075\,774 \times 10^4$
Urano (♅)	$2.870\,658\,19 \times 10^{12}$	$3.079\,910 \times 10^4$
Neptuno (♆)	$4.498\,396\,44 \times 10^{12}$	$6.022\,490 \times 10^4$

Tabla 6.4: Valores del semieje mayor a y el período orbital τ .

que tienen como solución $\alpha = -32.7041$ y $b = 1.50031$, es decir

$$\tau = 6.26317 \times 10^{-15} a^{1.50031}.$$

Obsevemos que Johannes Kepler no disponía de datos sobre Urano y Neptuno dado que no se conocían estos planetas al momento de enunciar esta ley (1618).¹ De la ley de gravitación de Newton (1685) se deduce que $\tau = 2\pi (GM)^{-1/2} a^{3/2}$, donde G es la constante de gravitación universal y M es la masa del Sol. Usando que $G = 6.674 \times 10^{-11} \text{ N m}^2 \text{ kg}^{-2}$ y $M = 1.989 \times 10^{30} \text{ kg}$, obtenemos el valor $k = 6.31183 \times 10^{-15} \text{ d/m}^{3/2}$ ($b = 1.5$).

Notemos que de las observaciones astronómicas, sólo podemos obtener el valor del producto GM , dado que en la fórmula solo interviene el producto de ambas cantidades. No fue hasta 1798 que Henry Cavendish obtuvo implícitamente el valor de G como resultado de los experimentos realizados con balanzas de torsión que permitieron determinar la masa de la Tierra y, como consecuencia, el valor de G .

<http://statpages.info/nonlin.html>

6.2.1. Mínimos cuadrados generalizados y no lineales*. Debemos observar que al linealizar el problema tomando $X = g(x)$ y $Y = h(y)$ y luego aplicar el método de regresión lineal, estamos minimizando $(Y_1 - \hat{Y}_1)^2 + \dots + (Y_N - \hat{Y}_N)^2$ que no equivalente a minimizar $(y_1 - \hat{y}_1)^2 + \dots + (y_N - \hat{y}_N)^2$. Supongamos que \hat{Y} son los valores del modelo de mínimos cuadrados linealizado $Y = \alpha + \beta X$, si definimos $\tilde{y}_j = h^{-1}(\hat{Y}_j)$, suponiendo $\tilde{y}_j \cong y_j$ tenemos $Y_j - \hat{Y}_j = h(y_j) - h(\tilde{y}_j) \cong h'(y_j)(y_j - \tilde{y}_j)$, de donde se deduce

$$(Y_1 - \hat{Y}_1)^2 + \dots + (Y_N - \hat{Y}_N)^2 \cong (h'(y_1))^2 (y_1 - \tilde{y}_1)^2 + \dots + (h'(y_N))^2 (y_N - \tilde{y}_N)^2.$$

Podemos considerar una función S modificada

$$S_G = \frac{1}{2} (w_1 (Y_1 - \hat{Y}_1)^2 + \dots + w_N (Y_N - \hat{Y}_N)^2),$$

donde $w_j = (h'(y_j))^{-2}$. Al minimizar S_G obtenemos las nuevas ecuaciones normales

$$\begin{cases} \sum_{j=1}^N w_j y_j = \alpha \sum_{j=1}^N w_j + \beta \sum_{j=1}^N w_j x_j, \\ \sum_{j=1}^N w_j y_j x_j = \alpha \sum_{j=1}^N w_j x_j + \beta \sum_{j=1}^N w_j x_j^2, \end{cases}$$

asumiendo $(h'(y_j))^{-2} (Y_j - \hat{Y}_j)^2 \cong (y_j - h^{-1}(\hat{Y}_j))^2$

Ejemplo 6.4. Queremos estudiar el modelo exponencial $y = k e^{bx}$ que mejor ajuste a los datos de la Tabla 6.5.

Si consideramos el mejor ajuste de S , donde

$$S = \frac{1}{2} \sum_{j=1}^{16} (\ln(y_j) - \alpha - \beta x_j)^2$$

obtenemos las ecuaciones normales

$$\begin{cases} 34.6158 = 16.0 \alpha + 17.1337 \beta, \\ 47.9919 = 17.1337 \alpha + 24.0476 \beta, \end{cases}$$

que tienen como solución $\alpha = 0.111283$ y $\beta = 1.91642$, lo que corresponde a $y = 1.11771 e^{1.91642x}$. Por otro lado, considerando que $(h'(y))^{-2} = y^2$, la función a minimizar es

$$S_G = \frac{1}{2} \sum_{j=1}^{16} y_j^2 (\ln(y_j) - \alpha_G - \beta_G x_j)^2$$

y las ecuaciones normales generalizadas

$$\begin{cases} 31\,512.0 = 8553.6 \alpha_G + 15\,761.2 \beta_G, \\ 59\,012.9 = 15\,761.2 \alpha_G + 29\,515.6 \beta_G, \end{cases}$$

que da como resultado $\alpha_G = -4.01117 \times 10^{-3}$ y $\beta_G = 2.00152$, es decir $y = 0.995997 e^{2.00152x}$. Para comparar ambos modelos consideramos la suma de cuadrados de la función no lineal:

$$S_N(\alpha, \beta) = \frac{1}{2} ((y_1 - e^{\alpha+\beta x_1})^2 + \dots + (y_{16} - e^{\alpha+\beta x_{16}})^2),$$

para cada uno de los modelos anteriores obtenemos $S_N(\alpha, \beta) = 11.5404$ y $S_N(\alpha_G, \beta_G) = 2.73312$, donde vemos la mejora del segundo método.

Podemos plantear también directamente la minimización de la función $S_N(\alpha, \beta)$, para esto buscamos un punto crítico de S_N

$$\begin{aligned} 0 &= \frac{\partial S_N}{\partial \alpha}(\alpha, \beta) = \sum_{j=1}^{16} (y_j - e^{\alpha+\beta x_j}) e^{\alpha+\beta x_j}, \\ 0 &= \frac{\partial S_N}{\partial \beta}(\alpha, \beta) = \sum_{j=1}^{16} (y_j - e^{\alpha+\beta x_j}) e^{\alpha+\beta x_j} x_j, \end{aligned}$$

¹ Los planetas Mercurio, Venus, Marte, Júpiter y Saturno, son los más cercanos a la Tierra y pueden ser vistos sin instrumentos ópticos. Urano fue descubierto en 1781 y Neptuno en 1846.

x	y	$\ln(y)$	x	y	$\ln(y)$
0.068 005	1.816 13	0.596 71	1.110 41	9.0719	2.205 18
0.251 188	1.850 42	0.615 41	1.296 46	12.6341	2.536 40
0.293 132	1.458 38	0.377 33	1.326 62	13.8117	2.625 52
0.559 454	3.399 98	1.223 77	1.524 57	20.1724	3.004 32
0.667 651	3.160 39	1.150 70	1.651 31	27.8109	3.325 43
0.710 296	5.287 71	1.665 39	1.801 58	37.2007	3.616 33
0.955 559	6.373 66	1.852 17	1.974 99	51.1806	3.935 36
0.962 632	6.811 48	1.918 61	1.979 86	52.8378	3.967 23

Tabla 6.5

resolviendo estas ecuaciones no lineales obtenemos $\alpha_N = -2.89922 \times 10^{-2}$ y $\beta_N = 2.01455$, $S_N = 2.68944$, lo que nos permite concluir que el método de mínimos cuadrados generalizados nos da una buena aproximación.

Resolución de las ecuaciones no lineales Para hallar un punto crítico de S planteamos la iteración del método de Newton

$$\begin{pmatrix} \alpha^{(n)} \\ \beta^{(n)} \end{pmatrix} = \begin{pmatrix} \alpha^{(n-1)} \\ \beta^{(n-1)} \end{pmatrix} - \begin{pmatrix} \frac{\partial^2 S_N}{\partial \alpha^2} & \frac{\partial^2 S_N}{\partial \alpha \partial \beta} \\ \frac{\partial^2 S_N}{\partial \alpha \partial \beta} & \frac{\partial^2 S_N}{\partial \beta^2} \end{pmatrix}^{-1} \cdot \begin{pmatrix} \frac{\partial S_N}{\partial \alpha} \\ \frac{\partial S_N}{\partial \beta} \end{pmatrix} \Bigg|_{(\alpha^{(n-1)}, \beta^{(n-1)})}$$

Partiendo del punto inicial $(\alpha, \beta) = (0.111283, 1.91642)$ obtenido mediante la linealización obtenemos las iteraciones que mostramos en siguiente tabla:

n	α	β	S_N
0	0.111 283 384	1.916 415 381	11.540 371 940
1	-0.033 322 616	2.018 800 281	2.746 016 039
2	-0.029 027 752	2.014 584 857	2.689 441 908
3	-0.028 992 228	2.014 554 909	2.689 440 046
4	-0.028 992 226	2.014 554 908	2.689 440 046
5	-0.028 992 226	2.014 554 908	2.689 440 046

Ejemplo 6.5 (Modelo hiperbólico). Cinética de Michaelis-Menten

6.3. Mínimos cuadrados para suma de funciones. Consideremos el caso general, supongamos que queremos aproximar la relación $y = f(x)$ como combinación de funciones conocidas de la forma $\hat{f}(x) = \alpha_1 \phi_1(x) + \dots + \alpha_n \phi_n(x)$, donde $\phi_1(x), \dots, \phi_n(x)$ son funciones dadas. Planteamos la elección de los coeficientes basados en el criterio de mínimos cuadrados. Esto implica hallar $\alpha_1, \dots, \alpha_n$ de forma que sea mínimo $S = (y_1 - \hat{y}_1)^2 + \dots + (y_N - \hat{y}_N)^2$, donde $\hat{y}_j = \hat{f}(x_j)$. Si derivamos respecto a cada coeficiente, en el mínimo se debe verificar

$$0 = \frac{\partial S}{\partial \alpha_i} = 2(y_1 - \hat{y}_1) \frac{\partial \hat{y}_1}{\partial \alpha_i} + \dots + 2(y_N - \hat{y}_N) \frac{\partial \hat{y}_N}{\partial \alpha_i},$$

para $i = 1, \dots, n$, que se puede reescribir de la forma

$$(6.2) \quad y_1 \frac{\partial \hat{y}_1}{\partial \alpha_i} + \dots + y_N \frac{\partial \hat{y}_N}{\partial \alpha_i} = \hat{y}_1 \frac{\partial \hat{y}_1}{\partial \alpha_i} + \dots + \hat{y}_N \frac{\partial \hat{y}_N}{\partial \alpha_i}.$$

Dado que $\frac{\partial \hat{y}_j}{\partial \alpha_i} = \phi_i(x_j)$, el lado izquierdo vale

$$y_1 \frac{\partial \hat{y}_1}{\partial \alpha_i} + \dots + y_N \frac{\partial \hat{y}_N}{\partial \alpha_i} = y_1 \phi_i(x_1) + \dots + y_N \phi_i(x_N)$$

De la misma forma, el lado derecho se puede plantear

$$\begin{aligned} \hat{y}_1 \frac{\partial \hat{y}_1}{\partial \alpha_i} + \dots + \hat{y}_N \frac{\partial \hat{y}_N}{\partial \alpha_i} &= (\alpha_1 \phi_1(x_1) + \dots + \alpha_n \phi_n(x_1)) \phi_i(x_1) + \dots \\ &+ (\alpha_1 \phi_1(x_N) + \dots + \alpha_n \phi_n(x_N)) \phi_i(x_N) \\ &= (\phi_i(x_1) \phi_1(x_1) + \dots + \phi_i(x_N) \phi_1(x_N)) \alpha_1 + \dots \\ &+ (\phi_i(x_1) \phi_n(x_1) + \dots + \phi_i(x_N) \phi_n(x_N)) \alpha_n. \end{aligned}$$

Si definimos $\phi_{ij} = \phi_i(x_j)$, las ecuaciones (6.2) se escriben como

$$\sum_{j=1}^N \phi_{ij} y_j = \sum_{k=1}^n \left(\sum_{j=1}^N \phi_{ij} \phi_{kj} \right) \alpha_k$$

las que se conocen como las ecuaciones normales. Si $\Phi \in \mathbb{R}^{N \times n}$ es la matriz cuyos elementos son $(\Phi)_{jk} = \phi_k(x_j)$ y los vectores $\mathbf{y} \in \mathbb{R}^N$, $\boldsymbol{\alpha} \in \mathbb{R}^n$ dados por $\mathbf{y} = (y_1 \dots y_N)^T$, $\boldsymbol{\alpha} = (\alpha_1 \dots \alpha_n)^T$, las ecuaciones normales se escriben matricialmente como $\Phi^T \cdot \mathbf{y} = \Phi^T \cdot \Phi \cdot \boldsymbol{\alpha}$. Si la matriz $\Phi^T \cdot \Phi$ es inversible podemos despejar los coeficientes $\boldsymbol{\alpha} = (\Phi^T \cdot \Phi)^{-1} \Phi^T \mathbf{y}$.

6.3.1. Modelo polinomial. En muchas situaciones, la relación entre dos variables x e y no se puede expresar a través de una función lineal sino que es necesario utilizar otro tipo de funciones, como por ejemplo funciones polinomiales de mayor orden:

$$y = f(x) = \alpha_0 + \alpha_1 x + \dots + \alpha_n x^n.$$

Igual que en el caso lineal, buscamos minimizar el error $S = \frac{1}{2}((y_1 - \hat{y}_1)^2 + \dots + (y_N - \hat{y}_N)^2)$, donde $\hat{y}_k = f(x_k)$. Si definimos $\boldsymbol{\phi}_0 = (1 \dots 1)^T$, $\boldsymbol{\phi}_1 = (x_1 \dots x_N)^T$, \dots , $\boldsymbol{\phi}_n = (x_1^n \dots x_N^n)^T$, entonces $\hat{\mathbf{y}} = \alpha_0 \boldsymbol{\phi}_0 + \dots + \alpha_n \boldsymbol{\phi}_n$ y $S = \frac{1}{2} \|\mathbf{y} - \hat{\mathbf{y}}\|^2$. Como $\mathbf{y} - \hat{\mathbf{y}} \perp \boldsymbol{\phi}_k$, $k = 0, \dots, n$, obtenemos las ecuaciones normales

$$\begin{aligned} \mathbf{y} \cdot \boldsymbol{\phi}_0 &= \alpha_0 \boldsymbol{\phi}_0 \cdot \boldsymbol{\phi}_0 + \dots + \alpha_n \boldsymbol{\phi}_n \cdot \boldsymbol{\phi}_0, \\ \mathbf{y} \cdot \boldsymbol{\phi}_1 &= \alpha_0 \boldsymbol{\phi}_0 \cdot \boldsymbol{\phi}_1 + \dots + \alpha_n \boldsymbol{\phi}_n \cdot \boldsymbol{\phi}_1, \\ &\vdots \\ \mathbf{y} \cdot \boldsymbol{\phi}_n &= \alpha_0 \boldsymbol{\phi}_0 \cdot \boldsymbol{\phi}_n + \dots + \alpha_n \boldsymbol{\phi}_n \cdot \boldsymbol{\phi}_n. \end{aligned} \quad (6.3)$$

Para que el sistema tenga solución única, los vectores $\boldsymbol{\phi}_0, \dots, \boldsymbol{\phi}_n$ tienen que ser independientes. Si $\alpha_0 \boldsymbol{\phi}_0 + \dots + \alpha_n \boldsymbol{\phi}_n = \mathbf{0}$, tenemos $f(x_1) = 0, \dots, f(x_N) = 0$. Como $f(x)$ es un polinomio de grado n , $f(x)$ tiene que ser idénticamente nula, lo que equivale a $\alpha_0 = 0, \dots, \alpha_n = 0$.

Ejemplo 6.6. La ecuación de estado para un sistema compuesto de un mol de un gas ideal es $pV = RT$, donde p es la presión del gas, V el volumen y T la temperatura absoluta. R es una constante universal cuyo valor es $R = 8.314472 \text{ J mol}^{-1} \text{ K}^{-1}$. Dicho de otra forma $Z = pV/(RT)$ es una constante. Sin embargo, no existe ningún gas real que cumpla exactamente la ecuación anterior. Las diferencias son más notables a medida que la densidad del gas es mayor, lo que ocurre a temperaturas bajas y altas presiones. En [18] se estudia en forma experimental la ecuación de estado del neón para temperaturas en el rango 80 – 130 K y presiones de hasta $2 \times 10^8 \text{ Pa}$. En la Tabla I de [18] se muestran valores experimentales de p y Z para distintas isothermas (T constante), reproducimos en la Tabla 6.6 los valores correspondientes a $T = 100 \text{ K}$ ^{2 3}. En [18] se ajusta por cuadrados mínimos un polinomio de grado 10. Por simplicidad y estabilidad numérica, vamos a considerar un ajuste quíntico. Proponemos una aproximación de la forma $Z = \alpha_0 + \alpha_1 x + \dots + \alpha_5 x^5$, donde x es la variable adimensional $x = p/1.0 \times 10^8 \text{ Pa}$. Resolviendo las ecuaciones normales, obtenemos

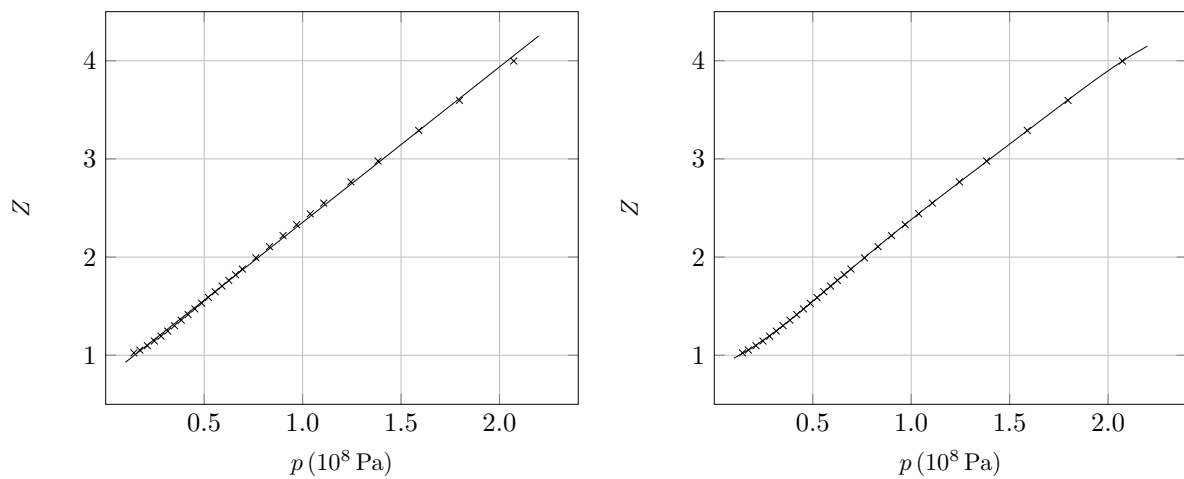
$$Z = 0.888 + 0.598882x + 2.37695x^2 - 2.33479x^3 + 1.01929x^4 - 0.166425x^5.$$

El valor del función error S es 2.748×10^{-4} . En la Figura 6.5 se muestran los resultados, comparamos la aproximación quíntica con la aproximación lineal $Z = 0.768940 + 1.58423x$, cuyo error es $S = 1.127 \times 10^{-2}$.

²La presión está expresada en atmósferas, recordemos $1 \text{ atm} = 101325 \text{ Pa}$.

³Observar que en la Tabla I de [18], los datos correspondientes a $P = 209.013 \text{ atm}$ ($2.11782 \times 10^7 \text{ Pa}$) y $T = 100 \text{ K}$ aparecen repetidos.

p (Pa)	Z	p (Pa)	Z
$1.431\,78 \times 10^7$	1.023 97	$6.251\,05 \times 10^7$	1.762 59
$1.732\,78 \times 10^7$	1.052 36	$6.595\,53 \times 10^7$	1.820 34
$2.117\,82 \times 10^7$	1.097 05	$6.940\,02 \times 10^7$	1.878 00
$2.471\,87 \times 10^7$	1.144 61	$7.628\,99 \times 10^7$	1.992 51
$2.812\,12 \times 10^7$	1.193 88	$8.317\,94 \times 10^7$	2.106 23
$3.151\,05 \times 10^7$	1.245 90	$9.006\,95 \times 10^7$	2.218 83
$3.495\,48 \times 10^7$	1.301 04	$9.696\,03 \times 10^7$	2.330 30
$3.839\,90 \times 10^7$	1.357 25	$1.038\,51 \times 10^8$	2.440 45
$4.184\,32 \times 10^7$	1.414 25	$1.107\,41 \times 10^8$	2.549 83
$4.528\,76 \times 10^7$	1.471 68	$1.245\,23 \times 10^8$	2.765 54
$4.873\,22 \times 10^7$	1.529 74	$1.383\,07 \times 10^8$	2.978 11
$5.217\,66 \times 10^7$	1.588 02	$1.589\,83 \times 10^8$	3.290 10
$5.562\,12 \times 10^7$	1.646 41	$1.796\,61 \times 10^8$	3.596 27
$5.906\,58 \times 10^7$	1.704 55	$2.072\,35 \times 10^8$	3.997 17

Tabla 6.6: Datos de p y Z correspondientes a $T = 100$ K.(a) Aproximación lineal: $S = 1.127 \times 10^{-2}$.(b) Aproximación quintica: $S = 2.748 \times 10^{-4}$.Fig. 6.5: Gráfico de Z en función de p .