Estadística [continuación]

Santander, 2017-2018

Regresión lineal con un “feature”: el caso más simple (I)

- Consideremos un problema de regresión lineal donde tan sólo existe una **“feature”**
- El modelo de regresión es por lo tanto de la forma siguiente:

$$y = f(x) = \alpha_0 + \alpha_1 x$$

- Si agrupamos las diferentes medidas de **x** e **y** en la matriz y vectores correspondientes

$$\begin{bmatrix} y^{(1)} \\ y^{(2)} \\ \dots \\ y^{(N)} \end{bmatrix} = \begin{bmatrix} 1 & x^{(1)} \\ 1 & x^{(2)} \\ \dots & \dots \\ 1 & x^{(N)} \end{bmatrix} \begin{bmatrix} \alpha_0 \\ \alpha_1 \end{bmatrix}$$

- La forma de la función de coste, utilizando el cuadrado de la distancia euclídea es por lo tanto

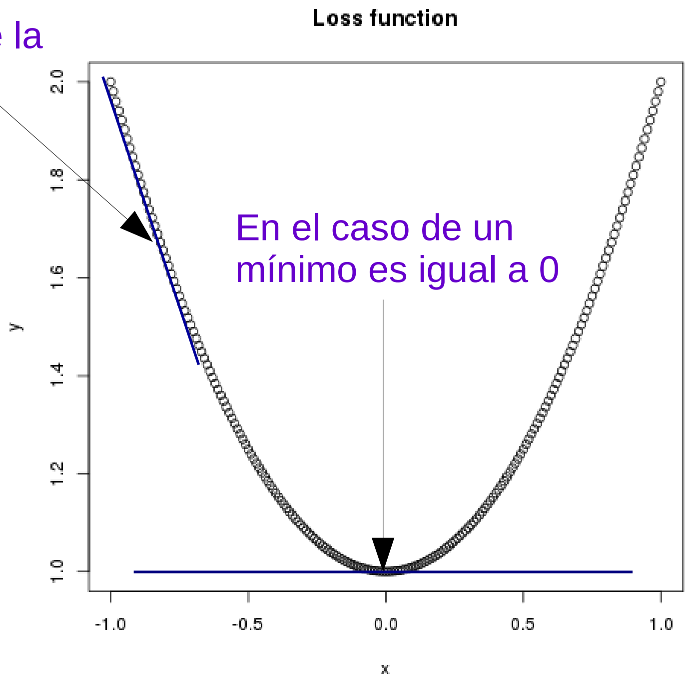
$$Loss = (y - X\alpha)^T (y - X\alpha) = \sum_{j=1}^{N} (y_j - \alpha_0 - \alpha_1 x_j)^2$$

Regresión lineal con un “feature”: el caso más simple (II)

- Para el caso lineal la minimización de la función de coste puede hacerse analíticamente
- Basta recordar que los mínimos de una función cumplen que la **derivada en ese punto vale 0**

La derivada es la inclinación de la recta tangente en cada punto

$$\nabla_{\alpha} Loss = 0 \Rightarrow \begin{array}{l} \dfrac{\partial\, Loss}{\partial\, \alpha_0} = 0 \\ \dfrac{\partial\, Loss}{\partial\, \alpha_1} = 0 \end{array}$$



(*) Los máximos y los puntos de inflexión también cumplen esta condición. En el caso de regresión lineal sabemos que siempre será un mínimo (¿por qué?).

Regresión lineal con un "feature": el caso más simple (III)

- Recuperando la expresión de la función de coste:

$$Loss=(y-X\,\alpha)^T(y-X\,\alpha)=\sum_{j=1}^{N}(y_j-\alpha_0-\alpha_1 x_j)^2$$

- Resulta sencillo ver cuáles son las condiciones que han de cumplir los parámetros

$$\nabla_\alpha Loss=0\Rightarrow \begin{array}{l} \dfrac{\partial\, Loss}{\partial\,\alpha_0}=-2\sum_{j=0}^{N}(y_j-\alpha_0-\alpha_1 x_j)=0 \\ \dfrac{\partial\, Loss}{\partial\,\alpha_1}=-2\sum_{j=0}^{N}(y_j-\alpha_0-\alpha_1 x_j)\,x_j=0 \end{array}$$

- Es posible despejar alpha_0 en la primera ecuación para obtener:

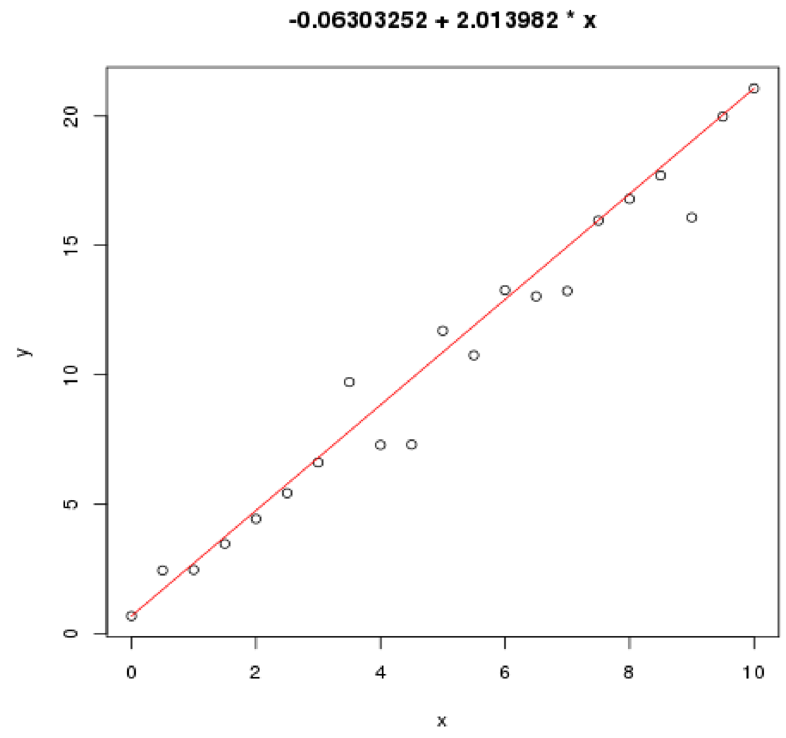
$$\alpha_0=\bar{y}-\alpha_1\bar{x}$$

- Sustituyendo en la segunda ecuación:

$$\alpha_1=\frac{\overline{xy}-\bar{x}\,\bar{y}}{\overline{x^2}-\bar{x}^2}$$

Regresión lineal con un "feature": Ejemplo

- Generamos un modelo en el que **y = f(x) = 2 x**
- Y en el que **p(y | x)** tiene una forma como la siguiente: $p(y|x)=\frac{1}{\sqrt{2\pi}\sigma}e^{\frac{-1}{2}\frac{(y-2x)^2}{\sigma^2}}$



Regresión lineal en notación matricial: caso general

- El caso general en el que hay M features puede resolverse fácilmente usando matrices

$$\begin{bmatrix} y^{(1)} \\ y^{(2)} \\ \dots \\ y^{(N)} \end{bmatrix} = \begin{bmatrix} 1 & x_1^{(1)} & x_2^{(1)} & \dots & x_M^{(1)} \\ 1 & x_1^{(2)} & x_2^{(2)} & \dots & x_M^{(2)} \\ \dots & \dots & \dots & \dots & \dots \\ 1 & x_1^{(N)} & x_2^{(N)} & \dots & x_M^{(N)} \end{bmatrix} \begin{bmatrix} \alpha_0 \\ \alpha_1 \\ \dots \\ \alpha_M \end{bmatrix}$$

- Como ya hemos visto puede expresarse la función de coste como:

$$Loss = (y - X\alpha)^T (y - X\alpha)$$

- Que puede ser derivada con respecto a $\boldsymbol{\alpha}$ de la siguiente forma:

$$\nabla_\alpha Loss = -X^T (y - X\alpha) = -X^T y + X^T X \alpha = 0$$

- Multiplicando por la matrix inversa correspondiente podemos despejar el vector α

$$\alpha = (X^T X)^{-1} X^T y$$

Covarianza de un vector (I)

- El siguiente paso en el que vamos a estar interesados es en calcular **la matriz de covarianza de α**
- Para ello supongamos un vector genérico de dimensión M al que llamaremos **b**

$$b=\begin{bmatrix} b_1 \\ b_2 \\ \dots \\ b_M \end{bmatrix}$$

- La matriz de covarianza es aquella que contiene en la diagonal las varianzas de b_j …
- …y las covarianzas correspondientes cruzadas cov(b_i, b_j) para los elementos no diagonales

$$Cov(b)=E[b\,b^T]-E[b]E[b]^T$$

- En donde se entiende que:

$$b\,b^T=\begin{bmatrix} b_1 \\ b_2 \\ \dots \\ b_M \end{bmatrix}[b_1 b_2 \dots b_M]=\begin{bmatrix} b_1^2 & b_1 b_2 & \dots & b_1 b_M \\ b_2 b_1 & b_2^2 & \dots & b_2 b_M \\ \dots & \dots & \dots & \dots \\ b_M b_1 & b_M b_2 & \dots & b_M^2 \end{bmatrix}$$

Covarianza de un vector (II)

- Volviendo a la definición de covarianza de un vector podemos escribir por lo tanto

$$Cov(b)=E[bb^T]-E[b]E[b]^T=\begin{bmatrix} E[b_1^2]-E[b_1]^2 & E[b_1 b_2]-E[b_1]E[b_2] & \ldots & E[b_1 b_M]-E[b_1]E[b_M] \\ E[b_2 b_1]-E[b_2]E[b_1] & E[b_2^2]-E[b_2]^2 & \ldots & E[b_2 b_M]-E[b_2]E[b_M] \\ \ldots & \ldots & \ldots & \ldots \\ E[b_M b_1]-E[b_M]E[b_1] & E[b_M b_2]-E[b_M]E[b_2] & \ldots & E[b_M^2]-E[b_M]^2 \end{bmatrix}$$

- Que no es más que la definición de covarianza para un vector:

$$Cov(b)=\begin{bmatrix} Var(b_1) & Cov(b_{1,}b_2) & \ldots & Cov(b_{1,}b_M) \\ Cov(b_{2,}b_1) & Var(b_2) & \ldots & Cov(b_{2,}b_M) \\ \ldots & \ldots & \ldots & \ldots \\ Cov(b_M,b_1) & Cov(b_M,b_2) & \ldots & Var(b_M) \end{bmatrix}$$

- Supongamos ahora que el vector **b** es el producto de una matriz constante por un vector **a**, **b=Ma**

$$Cov(b)=E[bb^T]-E[b]E[b]^T=Cov(Ma)=E[Ma(Ma)^T]-E[Ma]E[Ma]^T=E[Maa^T M^T]-E[Ma]E[a^T M^T]$$

$$Cov(b)=M\,E[aa^T]M^T-M\,E[a]E[a^T]M^T=M(E[aa^T]-E[a]E[a^T])M^T=M\,Cov(a)M^T$$

Cálculo de la covarianza

- Con estas herramientas en la mano vamos a calcular la matriz de covarianza del vector α.

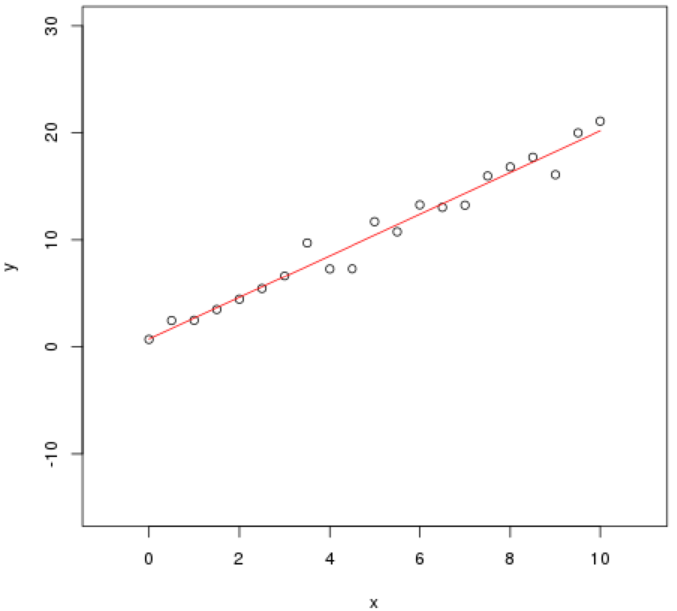
$$\alpha=(X^T X)^{-1} X^T y$$

- La primera pregunta que debemos responder es ¿cuál es el origen de la covarianza de α?.
- Hemos asumido que la coordenada independiente x es fija (su varianza es exactamente 0).
- Sin embargo la coordenada dependiente está distribuida como p(y | x) y su varianza no es 0.
- De hecho hay otras dos asunciones que hemos hecho en el modelo lineal
  - Las medidas de y son independientes para cada valor de x
  - La varianza de y es la misma para todas las p(y | x) con diferentes valores de x.
- En estas condiciones se tiene que $Var(y_i) = \sigma^2$ y también que $Cov(y_i,y_j)=0$. Por lo tanto:

$$Cov(y)=\begin{bmatrix} \sigma^2 & 0 & \dots & 0 \\ 0 & \sigma^2 & \dots & 0 \\ \dots & \dots & \dots & \dots \\ 0 & 0 & \dots & \sigma^2 \end{bmatrix} \qquad Cov(\alpha)=\left[(X^T X)^{-1} X^T\right] Cov(y) \left[(X^T X)^{-1} X^T\right]^T$$

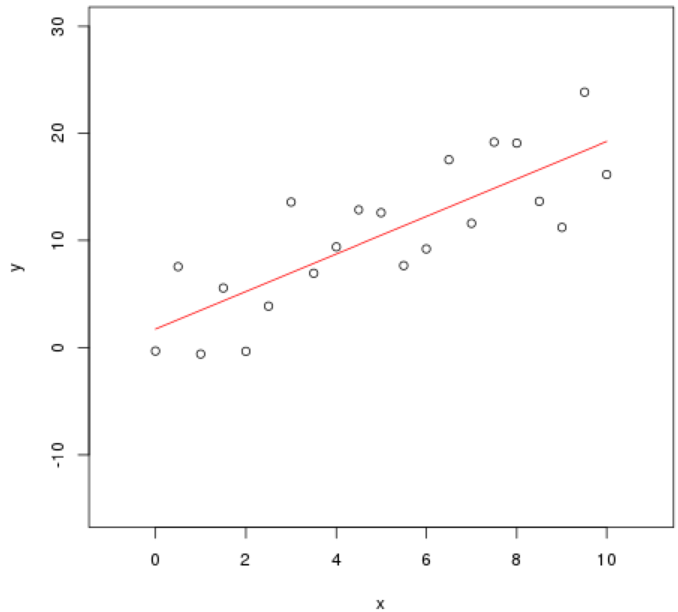
Ejemplo: efecto de la covarianza en y en alpha

Modelo real: y = 2x + N(0, 1)



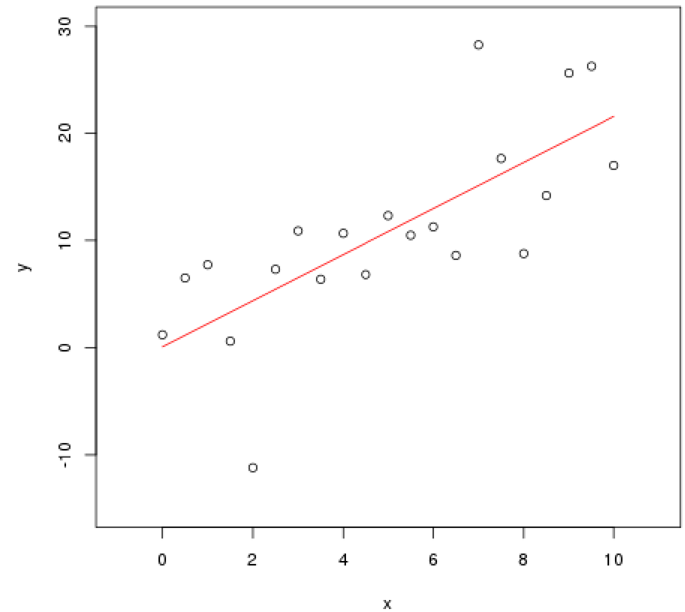
Ajuste: y = (0.72±0.42) + (1.95±0.07) x

Modelo real: y = 2x + N(0, 3)



Ajuste: y = (1.7±1.3) + (1.75±0.22) x

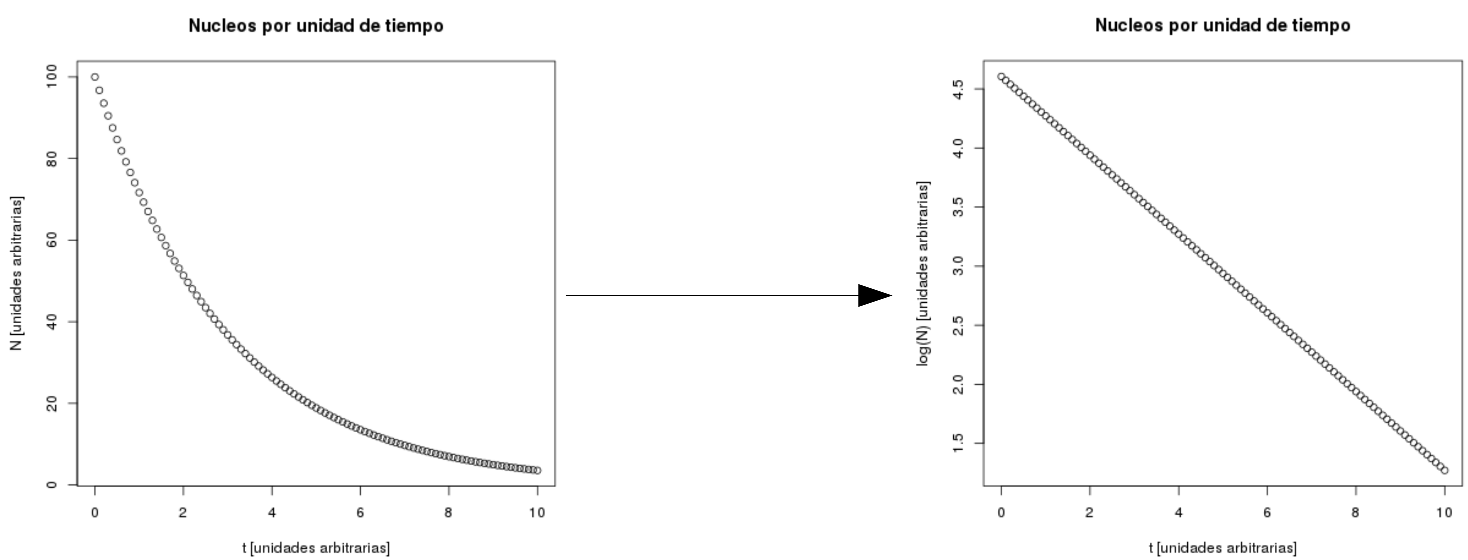
Modelo real: y = 2x + N(0, 6)



Ajuste: y = (0.1±2.5) + (2.15±0.43) x

Regresión lineal para funciones no lineales

- Muchos de los problemas de la vida cotidiana no guardan una relación lineal entre sus variables.
- En ocasiones es posible "traducir" nuestro problema no lineal haciendo un cambio de variable
- Ejemplo: supongamos que medimos el número de núcleos de uranio en una muestra en función de t
- Este tipo de proceso viene dado por una ley exponencial $N = N_0 \exp(-t/\tau)$
- Obviamente no podemos hacer un ajuste lineal entre N y "t" pero si redefinimos:
  - $Y = \log(N)$ entonces tenemos que: $y = \log(N_0) - t / \tau$ que es lineal en la variable t



Regresión lineal para polinomios

- Otra aplicación interesante que podemos considerar a la hora de modelar es usar **polinomios**
- Supongamos el esquema habitual $\{(y^{(1)}, x^{(1)}), (y^{(2)}, x^{(2)}),\ldots,(y^{(N)}, x^{(N)})\}$ con **y** y **x** de dimensión 1
- Podemos artificialmente expandir la dimensión de **x** simplemente añadiendo potencias de las **x**

$$\{(y^{(1)},x^{(1)}),(y^{(2)},x^{(2)}),\ldots,(y^{(N)},x^{(N)})\}\Rightarrow\{(y^{(1)},x^{(1)},x^{(1)2},\ldots,x^{(1)M}),(y^{(2)},x^{(2)},x^{(2)2},\ldots,x^{(2)M}),\ldots,(y^{(N)},x^{(N)},x^{(N)2},\ldots,x^{(N)M})\}$$

- Ahora podemos aplicar la regresión lineal con estas nuevas coordenadas de manera que el modelo:

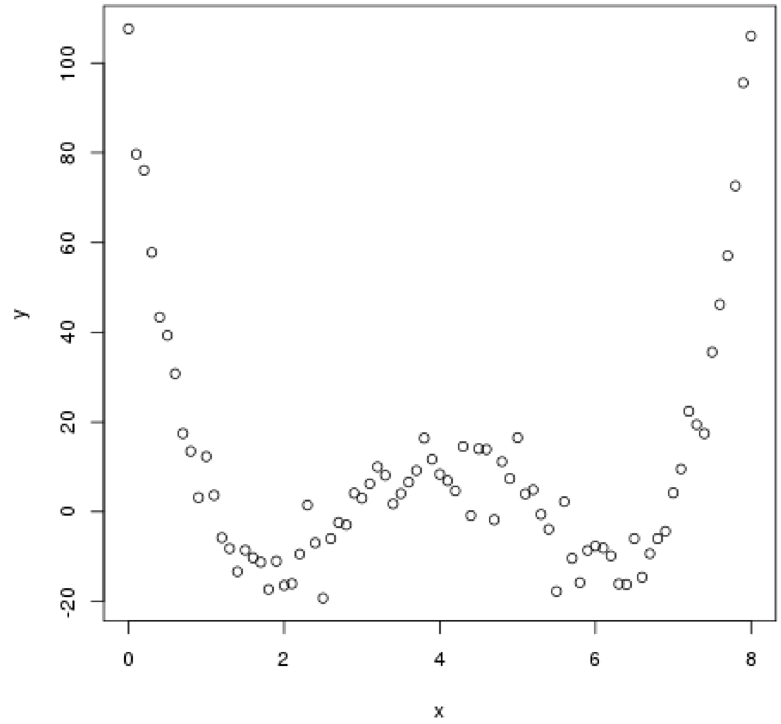
$$y=f(x)=\alpha_0+\alpha_1 x_1+\ldots+\alpha_M x_M\Rightarrow y=f(x)=\alpha_0+\alpha_1 x+\ldots+\alpha_M x^M$$

- Con lo cual podemos ajustar un polinomio usando exactamente las mismas herramientas vistas.

$$\begin{bmatrix} y^{(1)} \\ y^{(2)} \\ \ldots \\ y^{(N)} \end{bmatrix}=\begin{bmatrix} 1 & x_1^{(1)} & x_2^{(1)} & \ldots & x_M^{(1)} \\ 1 & x_1^{(2)} & x_2^{(2)} & \ldots & x_M^{(2)} \\ \ldots & \ldots & \ldots & \ldots & \ldots \\ 1 & x_1^{(N)} & x_2^{(N)} & \ldots & x_M^{(N)} \end{bmatrix}\begin{bmatrix} \alpha_0 \\ \alpha_1 \\ \ldots \\ \alpha_M \end{bmatrix} \longrightarrow \begin{bmatrix} y^{(1)} \\ y^{(2)} \\ \ldots \\ y^{(N)} \end{bmatrix}=\begin{bmatrix} 1 & x^{(1)} & x^{(1)2} & \ldots & x^{(1)M} \\ 1 & x^{(2)} & x^{(2)2} & \ldots & x^{(2)M} \\ \ldots & \ldots & \ldots & \ldots & \ldots \\ 1 & x^{(N)} & x^{(N)2} & \ldots & x^{(N)M} \end{bmatrix}\begin{bmatrix} \alpha_0 \\ \alpha_1 \\ \ldots \\ \alpha_M \end{bmatrix}$$

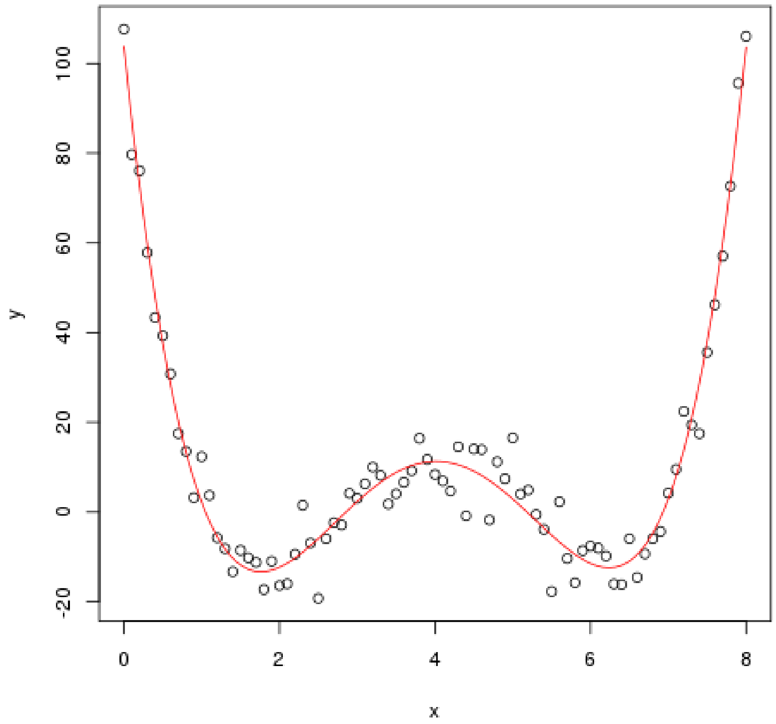
Regresión lineal con polinomios: ejemplo

- Generamos un modelo en el que **y = f(x) = 2 + (x-1)(x-3)(x-5)(x-7)**
- Y en el que **p(y | x)** tiene una forma como la siguiente: $p(y|x)=\frac{1}{\sqrt{2\pi}\sigma}e^{\frac{-1}{2}\frac{(y-(2+(x-1)(x-3)(x-5)(x-7)))^2}{\sigma^2}}$



Regresión lineal con polinomios: ejemplo

- Generamos un modelo en el que **y = f(x) = 2 + (x-1)(x-3)(x-5)(x-7)**
- Y en el que **p(y | x)** tiene una forma como la siguiente: $p(y|x)=\frac{1}{\sqrt{2\pi}\sigma}e^{\frac{-1}{2}\frac{(y-(2+(x-1)(x-3)(x-5)(x-7)))^2}{\sigma^2}}$



Ejercicio 4

1) Crea una función como la del ejercicio 3 en la que se pase como input: un vector x con la variable independiente, un parámetro “a”, un parámetro “b” y un valor “sigma”; y que devuelva un vector “y” que esté distribuido como una función normal con media = a*x + b y sigma = “sigma”.
2) Construye una función que reciba dos vectores “x” e “y” supuestamente relacionados linealmente y calcule los valores de “a” y “b” que minimizan la función de coste.
3) Genera un vector x aleatorio con valores entre 0 y 8 y N = 100 puntos. Usa la función creada en 1 con valores a=1, b=2 y sigma=2, y la función creada en 2 para encontrar el mínimo de la función de coste. Pinta en un mismo plot “x” e “y” representados con puntos, y la recta “a*x + b”.
4) Construye una función que calcule la matriz de covarianza asociada al ajuste lineal anterior. Utilízala con los vectores “x” e “y” anteriores y calcula dicha matriz para ese caso particular.
5) Crea una función como la del apartado 1 donde se añadan 3 parámetros más (c, d, e) y en dónde todo sea igual salvo que media = a + b*x + *x^2 + d*x^3 + e*x^4
6) Repite 3 y 4 con la función generada en 5 y usando: **a = 107, b=-176, c=86, d=-16, e=1**