

Estadística [continuación]

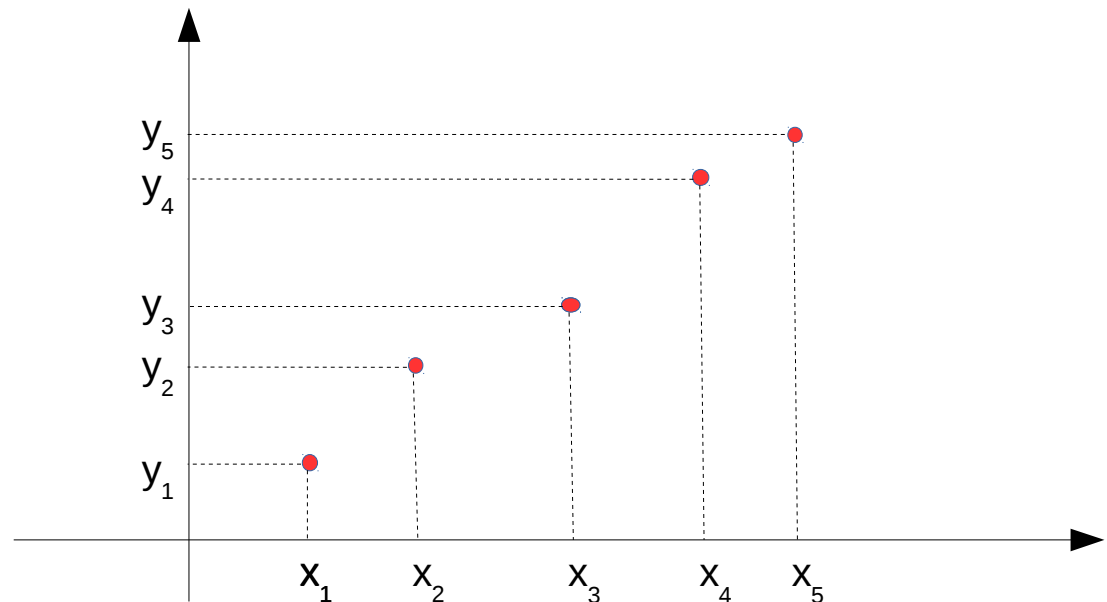
Santander, 2018-2019

Modelado estadístico

- Uno de los objetivos de la estadística consiste en modelizar procesos que involucren variables estadísticas.
- Supongamos un estadístico “**y**” que depende de un parámetro “**x**” tal y como se muestra en la gráfica.
- Puesto que son variables estadísticas, para una x dada, su comportamiento viene dado por $\text{pdf}(y | x)$.
- Si además conocemos cómo es la distribución de “**x**” tenemos que $\text{pdf}(y, x) = \text{pdf}(y | x) \text{pdf}(x)$.
- Y el último ingrediente que tenemos es que los valores de x e y están correlacionados por alguna función.

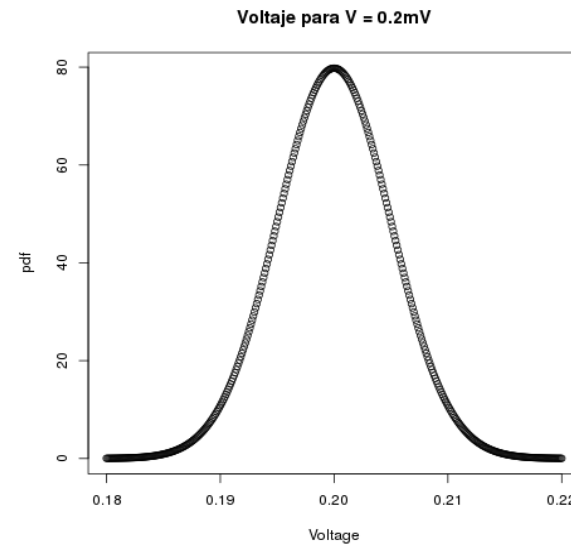
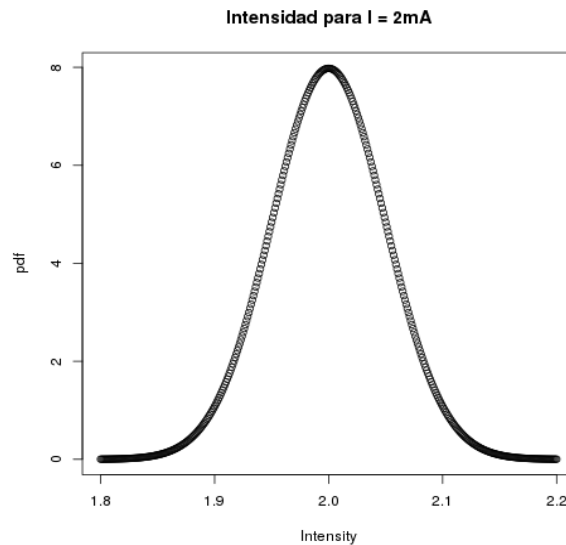
$$E[y|x] = f(x)$$

$$M = \{(x_1, y_1), (x_2, y_2), (x_3, y_3), (x_4, y_4), (x_5, y_5)\}$$



Modelado estadístico: ejemplo (I)

- Supongamos dos variables estadísticas de origen físico: **Voltaje** e **Intensidad** medidas en un circuito.
- Sabemos que ambas magnitudes están relacionadas a través de la resistencia: **$V = I R$ (ley de Ohm)**.
- Supongamos un circuito en el que fijamos la intensidad al valor I_0 y medimos ambos V e I .
- Debido a la acumulación de errores/resolución lo más probable es que las medidas sigan distribuciones:

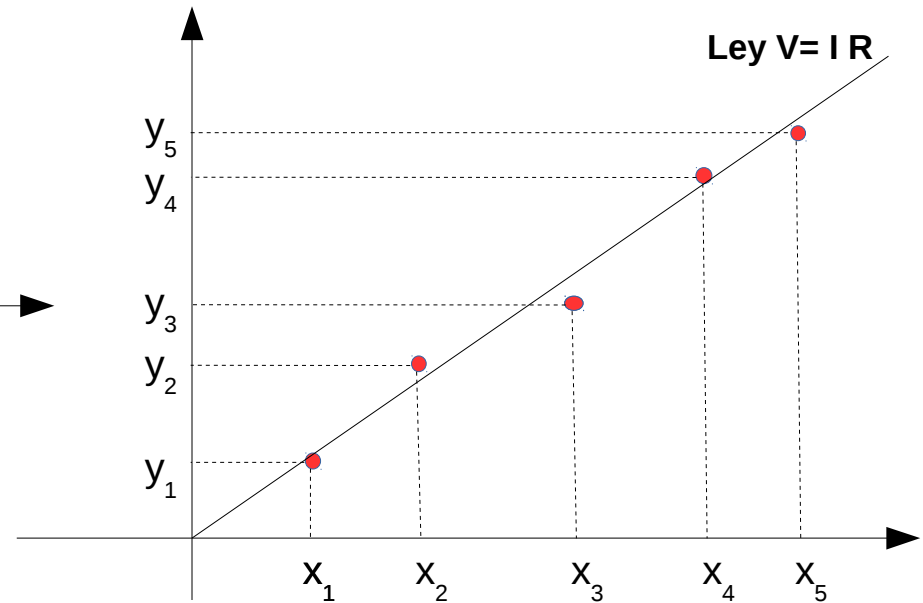
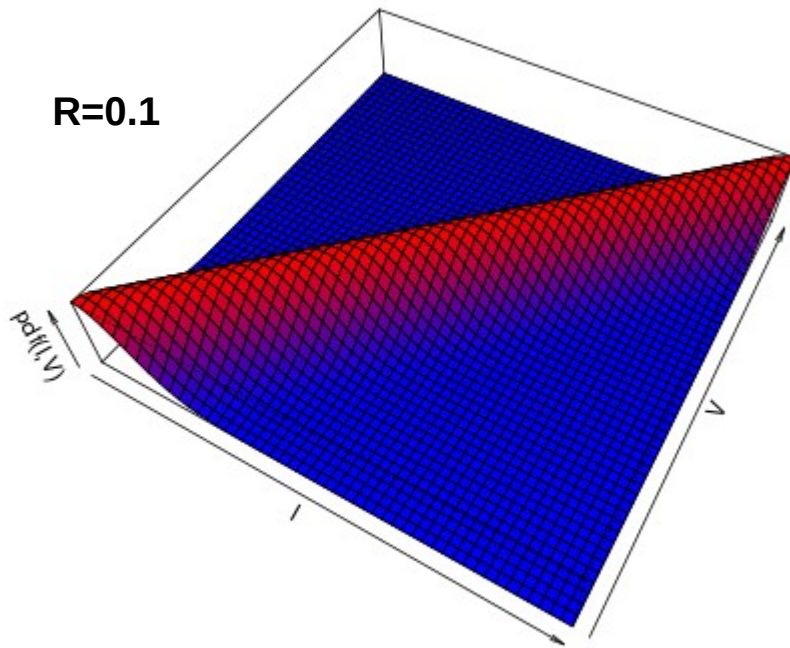


$R=0.1$

Modelado estadístico: ejemplo (II)

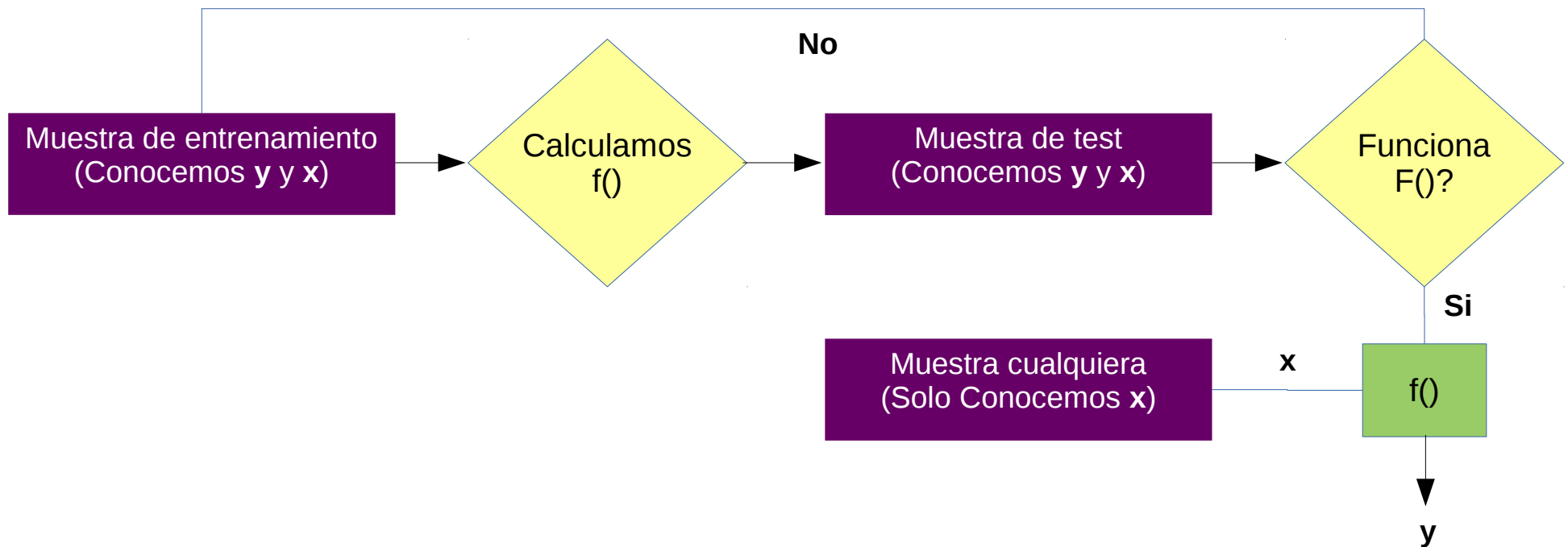
→ La pdf total puede escribirse de la siguiente forma:

$$pdf(V, I) = pdf(V|I) pdf(I) \propto N(\mu_I = I, \sigma_I = 0.05 \text{ mA}) N(\mu_V = RI, \sigma_V = 0.005 \text{ mV})$$



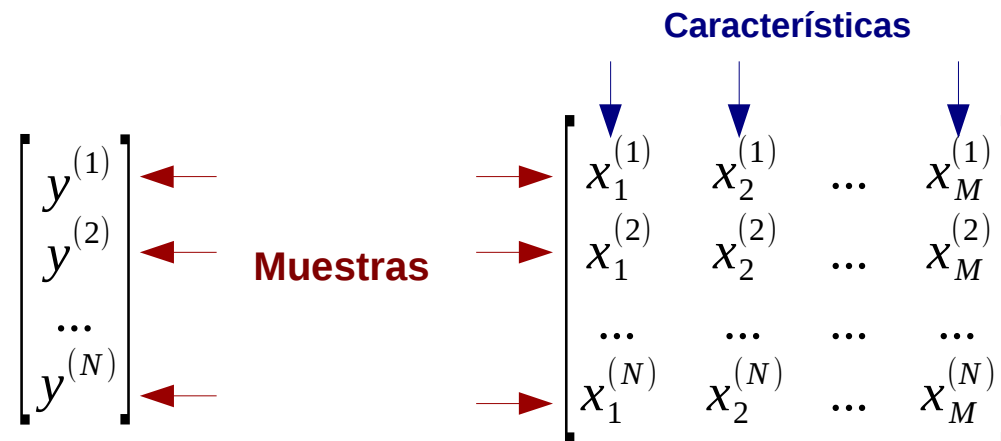
Modelado estadístico: objetivo

- El modelado estadístico persigue encontrar la relación $y = f(x)$ a partir de una muestra concreta de datos.
- Esto con frecuencia consiste en encontrar los parámetros que caracterizan a $f \rightarrow y = f(x, \Theta)$
- El objetivo final del modelado tiene que ver con la capacidad posterior de predecir nuevas situaciones.



Algunas definiciones

- Una vez visto el contexto estadístico del problema a resolver vamos con algunas definiciones
- A la variable bajo estudio la llamaremos la **“variable dependiente”** y
- A las variables de las que depende y las llamaremos características o **“features”** $x_1, x_2, x_3, x_4, x_5, x_6, \dots, x_M$
- Cuando tenemos una muestra concreta de N datos diferentes con frecuencia agrupamos en matrices



- Un modelo es cualquier función $f()$ que asigna un valor y a un vector de características $y = f(x_1, \dots, x_M)$.
- Un **parámetro** del modelo es una cantidad fija de la que depende el modelo. **Ejemplo anterior: R**

Métricas y función de coste

- El objetivo del modelado es encontrar la función que mejor “describa” nuestros datos.
- Resulta evidente que para ello necesitamos cuantificar la adecuación de un modelo a los datos.
- Para ello comenzamos definiendo una **métrica** que nos permita cuantificar la similitud entre 2 datos.
- Supongamos por ejemplo que tenemos dos realizaciones de la variable dependiente $y^{(1)}$ y $y^{(2)}$
- Podemos cuantificar su similitud utilizando la **distancia entre ambos** (euclídea por ejemplo):

$$d = \|y^{(1)} - y^{(2)}\|$$

- Si conocemos para un experimento x y y podemos ver como de bien predice nuestro modelo

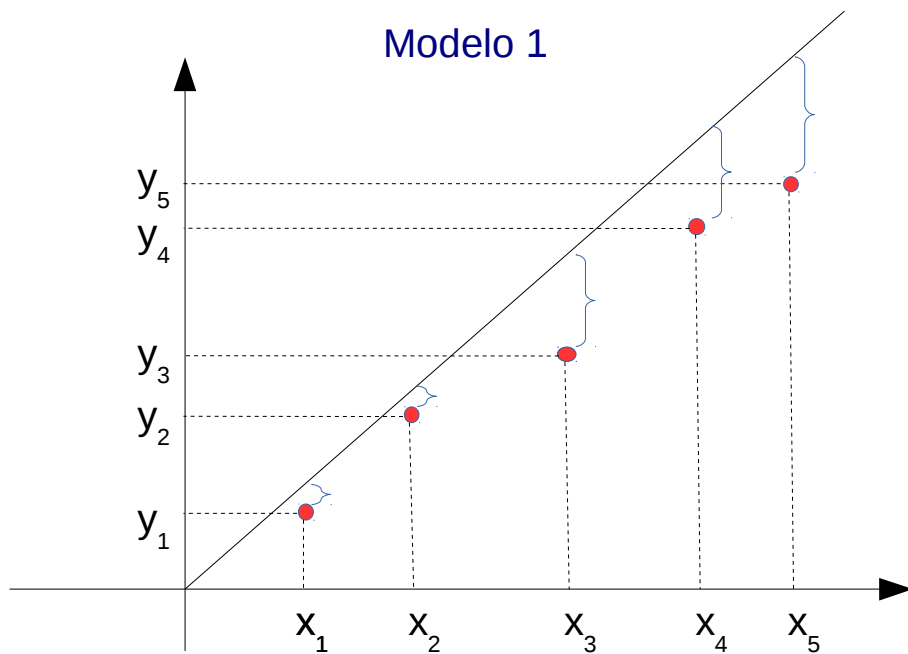
$$d = \|y^{(1)} - f(x_1^{(1)}, \dots, x_M^{(1)}, \Theta)\|$$

- Pero si además disponemos de una muestra de N elementos podemos calcular la distancia total

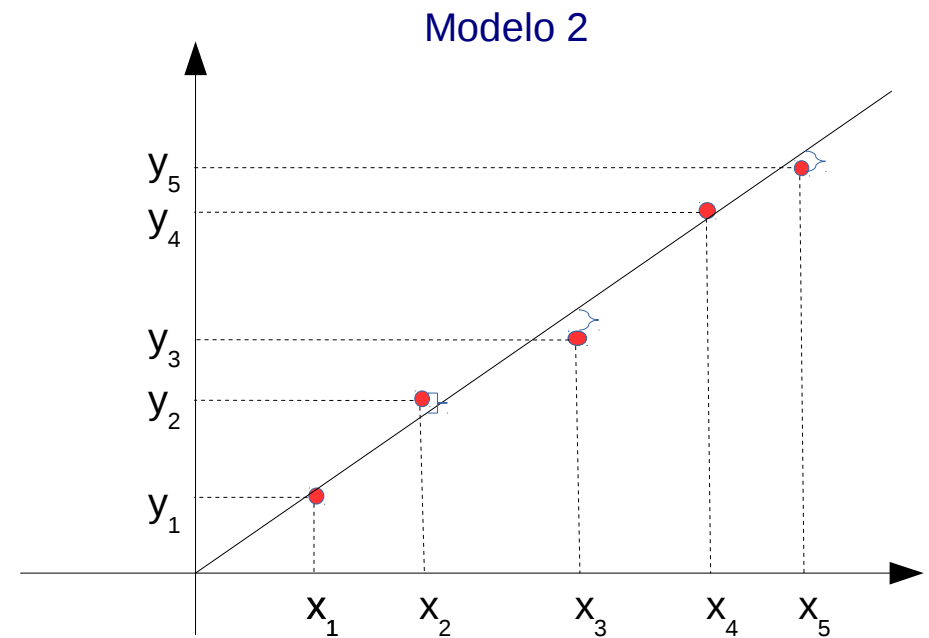
$$\text{cost} = \sum_{i=1}^N \|y^{(i)} - f(x_1^{(i)}, \dots, x_M^{(i)}, \Theta)\| \quad \longleftarrow \text{Funcion de coste}$$

Métricas y función de coste: Ejemplos

- Supongamos que tenemos dos modelos lineales: $y = a_1 x$ y $y = a_2 x$
- La función de coste mide la adecuación de un modelo concreto a los datos observados.



Función de coste alta para el modelo con $y = a_1 * x$



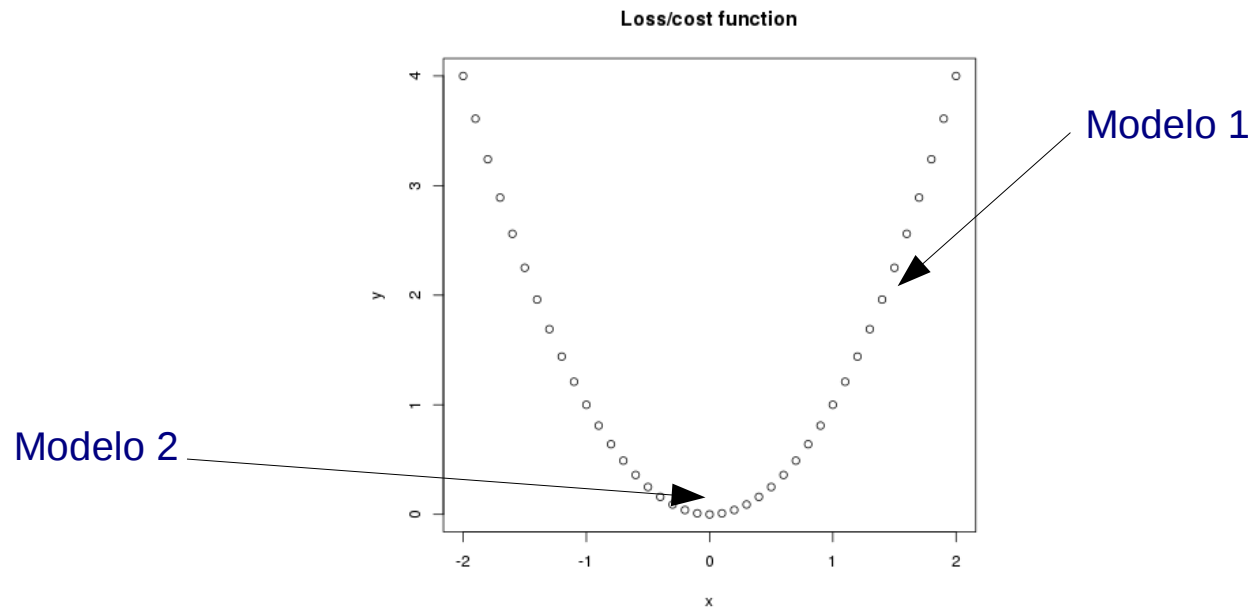
Función de coste baja para el modelo con $y = a_2 * x$

Minimización de la función de coste

- Una vez obtenida la función de coste, lo que necesitamos **es encontrar aquel modelo que la minimice**
- Recordemos que la función de coste depende de la muestra de entrenamiento y del parámetro/s

$$\text{loss}(y^{(1)}, y^{(2)}, \dots, y^{(N)}, x_1^{(1)}, x_2^{(1)}, \dots, x_M^{(N)}, \Theta_1, \Theta_2, \Theta_3)$$

- Cuando hablamos de minimizar nos referimos a minimizar en relación a los parámetros.

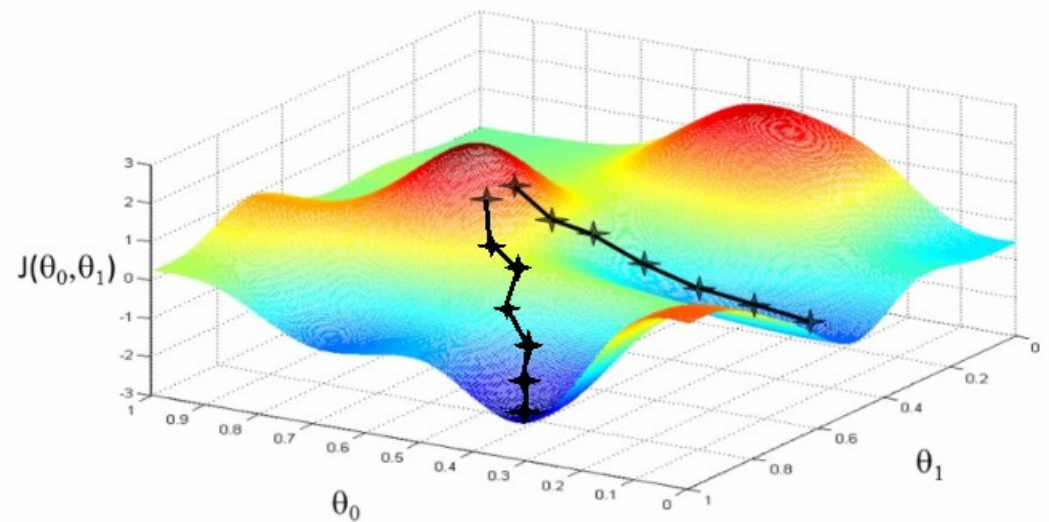
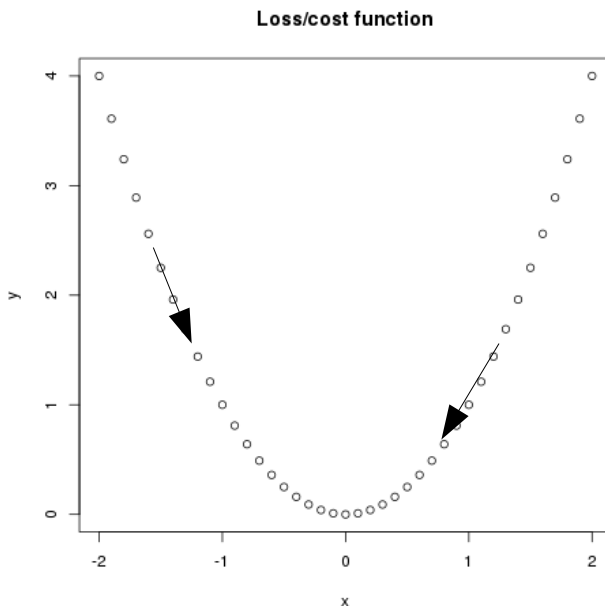


¿Cómo se minimiza una función?

- La técnica más utilizada para minimizar una función se conoce como **“gradient descent”**
- Utiliza una propiedad matemática: el gradiente de una función apunta en la dirección de máxima variación.
- Si calculamos por lo tanto el gradiente de la función de loss nos dirá en qué dirección la función disminuye.

$$\nabla_{\Theta} \text{loss} = \frac{\partial \text{loss}}{\partial \Theta_i}$$

$$\Theta_{n+1} = \Theta_n + \Lambda \nabla_{\Theta} \text{loss}$$



- Para problemas con muchas dimensiones se utiliza aún más el **“stochastic gradient descent”**

Regresión lineal

- Un caso muy particular de regresión es la conocida como regresión lineal en la que el modelo es lineal.

$$y = f(x_1, \dots, x_M) = \alpha_0 \cdot 1 + \alpha_1 \cdot x_1 + \alpha_2 \cdot x_2 + \dots + \alpha_M \cdot x_M$$

- Si tuviésemos una muestra de varios pares (y, x) podemos escribir la predicción para cada uno de ellos:

$$\begin{bmatrix} 1 & x_1^{(1)} & x_2^{(1)} & \dots & x_M^{(1)} \\ 1 & x_1^{(2)} & x_2^{(2)} & \dots & x_M^{(2)} \\ 1 & \dots & \dots & \dots & \dots \\ 1 & x_1^{(N)} & x_2^{(N)} & \dots & x_M^{(N)} \end{bmatrix} \cdot \begin{bmatrix} \alpha_1 \\ \alpha_2 \\ \dots \\ \alpha_M \end{bmatrix}$$

- Llamando genericamente a esa matriz X y al vector α podemos escribir el vector predicho como $y_p = X\alpha$
- Y por lo tanto, si usamos como **función de coste** el cuadrado de la distancia euclídea tenemos que:

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

Ejercicio 3

- Consideremos un sistema en el que existe una característica x y una variable dependiente y que se relacionan como $y = a * x$.
Supongamos también que para cada valor fijo de x la medida y viene dada por una pdf gaussiana centrada en $a * x$ y con σ .
Escribe una función de R que tome como input un vector aleatorio de N elementos de x que tome valores entre $[0, 10]$, el factor a , y σ ; y genere el correspondiente vector y .
- Una vez terminada la función genera el vector x con $N = 100$, y un vector y con $a = 2$ y $\sigma = 0.2$. Píntalos en una gráfica.
- Escribe una función en R a la que le pases un vector y dependiente, un vector x con la característica, y un parámetro a y te devuelva el valor de la loss function para ese valor del parámetro a .
- Llama a la función anterior con diferentes valores del parámetro a , y pinta la curva de la loss function en función de a . Comprueba visualmente que el mínimo está en la posición que nos esperamos.