

# Regresión univariable

- Muestreo
  - Hold-out
  - Cros-validación
- Hipótesis
  - Regresión lineal (sin *offset*)
  - Regresión lineal
  - Regresión polinómica
- Función de coste
  - Cuadrática
  - Regularización (regresión de arista)
- Optimización
  - Ecuación normal
  - Gradiente descendente
- Evaluación
  - Bootstrap

# Ejemplo en el sector eléctrico

- Una compañía eléctrica tiene una extensa base de datos con información muy diversa sobre sus clientes. Se desea explotar dicha información para dar respuesta a 3 cuestiones:
  - ¿Que energía demandará un nuevo cliente?  
(predicción)
  - ¿Es potencial comprador de un vehículo eléctrico?  
(clasificación)

# Ejemplo en el sector eléctrico

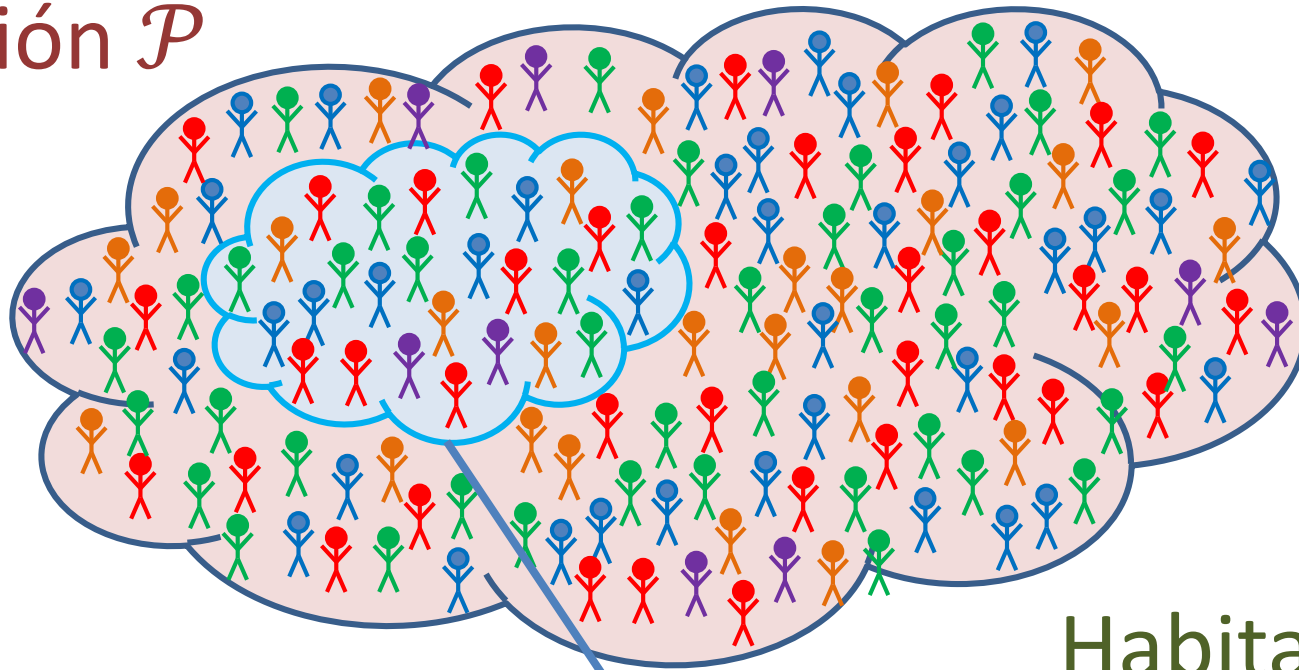
## Predicción

- En el momento de hacer un contrato con un cliente nuevo, la empresa realiza un perfil del usuario para lo que le solicita diversos datos.
  - Con esa información y en base a la experiencia acumulado sobre otros clientes, desea conocer el consumo previsto de energía a lo largo de un año.
  - Ello le permitirá realizar una mejor planificación y gestión de la red eléctrica

# Población y muestra

## 1. Extracción de una muestra

Población  $\mathcal{P}$



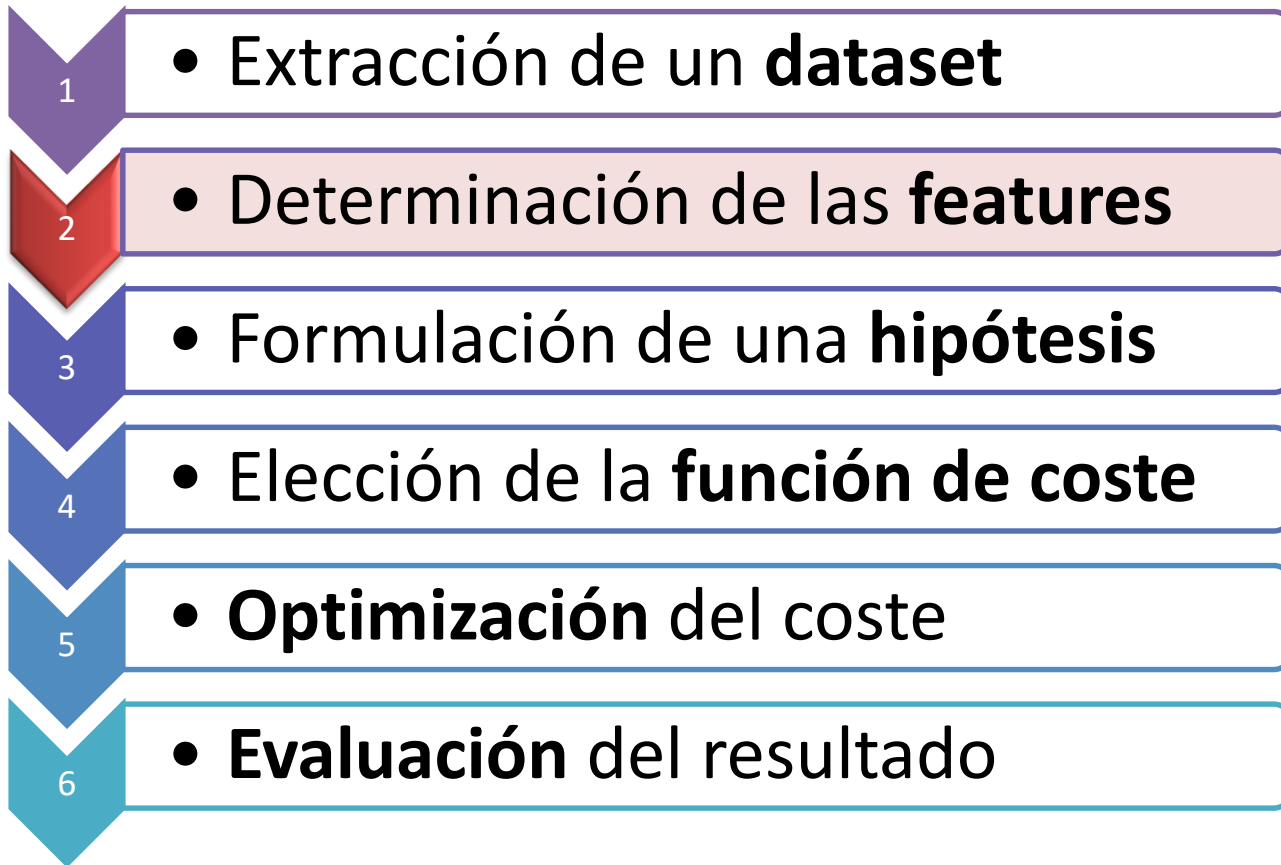
Muestra  $\mathcal{M}$   
(dataset)



Clientes

Habitantes

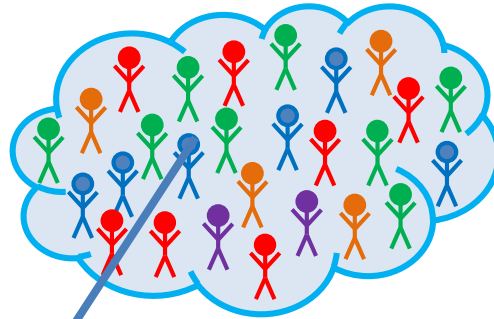
# Pasos del *machine learning*



# Elementos de la muestra y objetivos

## 2. Determinación de *features*

Muestra  $\mathcal{M}$   
(dataset)



$n$  elementos

$i$ -ésimo elemento

Caracterizado por 1 feature (rasgo):

- $x^{(i)}$ : ingresos anuales

Valor del objetivo (target) :

- $y^{(i)}$ : consumo anual de electricidad

$$x = \begin{bmatrix} x^{(1)} \\ x^{(2)} \\ \vdots \\ x^{(n)} \end{bmatrix}$$

$$y^{(i)} = \begin{bmatrix} y^{(1)} \\ y^{(2)} \\ \vdots \\ y^{(n)} \end{bmatrix}$$

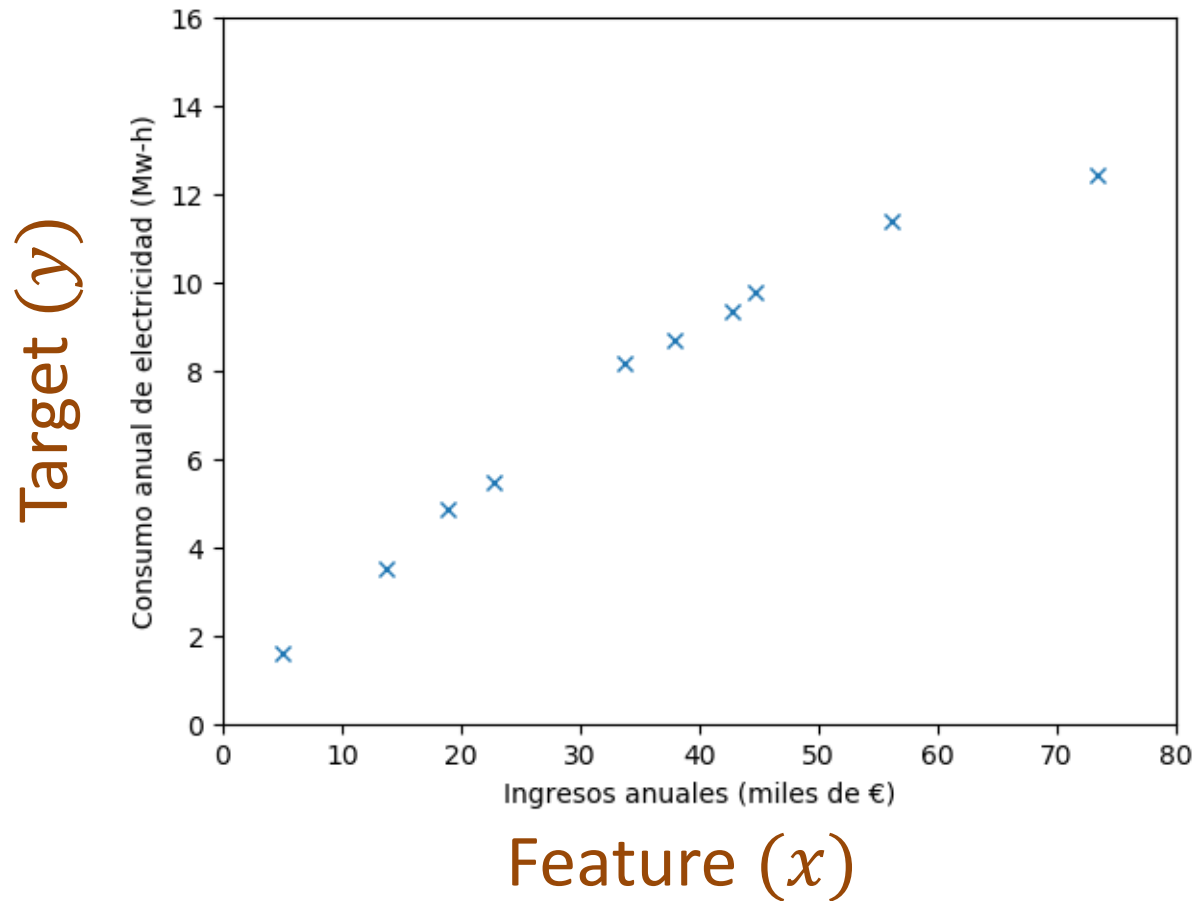
# Elementos de la muestra y objetivos

## 2. Determinación de *features*

Cliente	Ingresos anuales (miles de €) $x^{(i)}$	Consumo anual de electricidad (MW-h) $y^{(i)}$
1	44.6	9.81
2	73.4	12.4
3	5.0	1.63
4	33.7	8.17
5	18.9	4.87
⋮	⋮	⋮

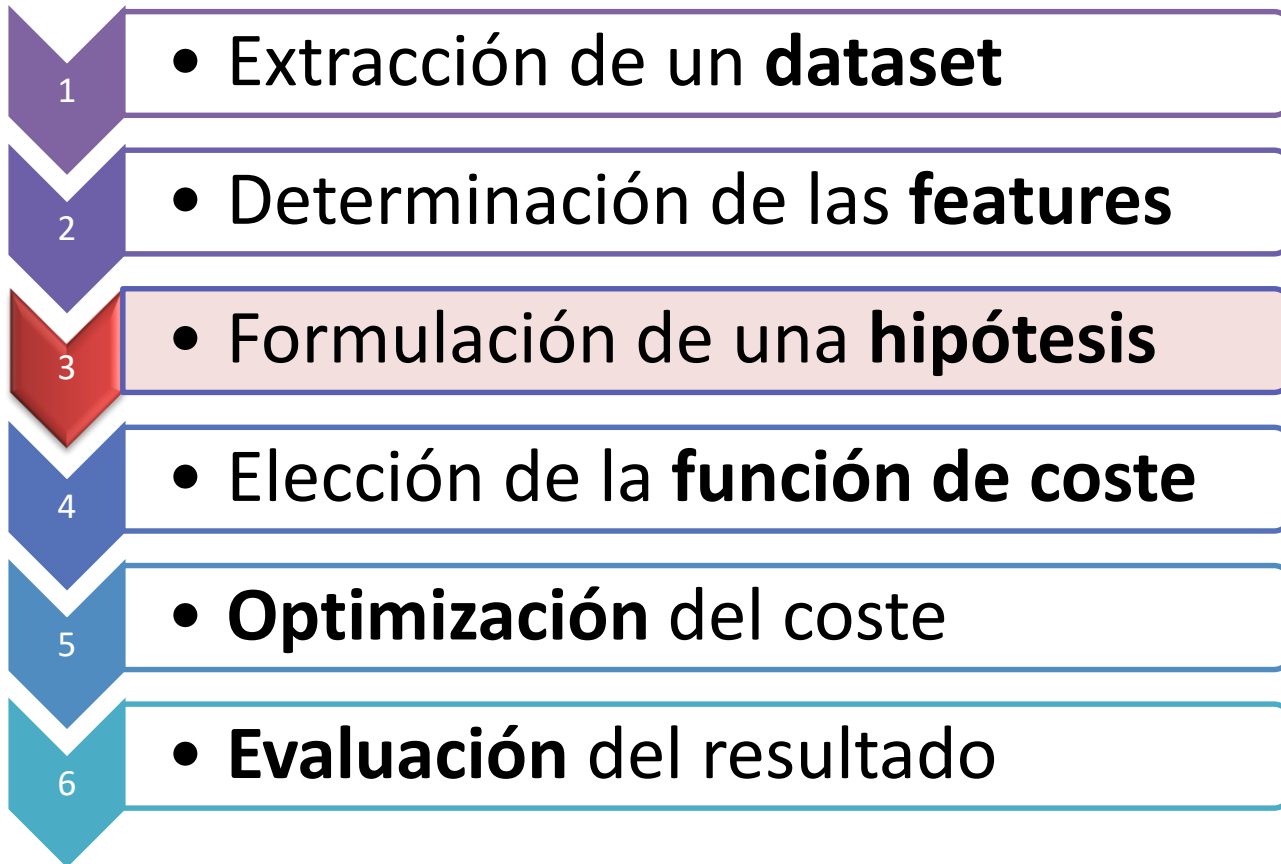
# Elementos de la muestra y objetivos

## 2. Determinación de *features*



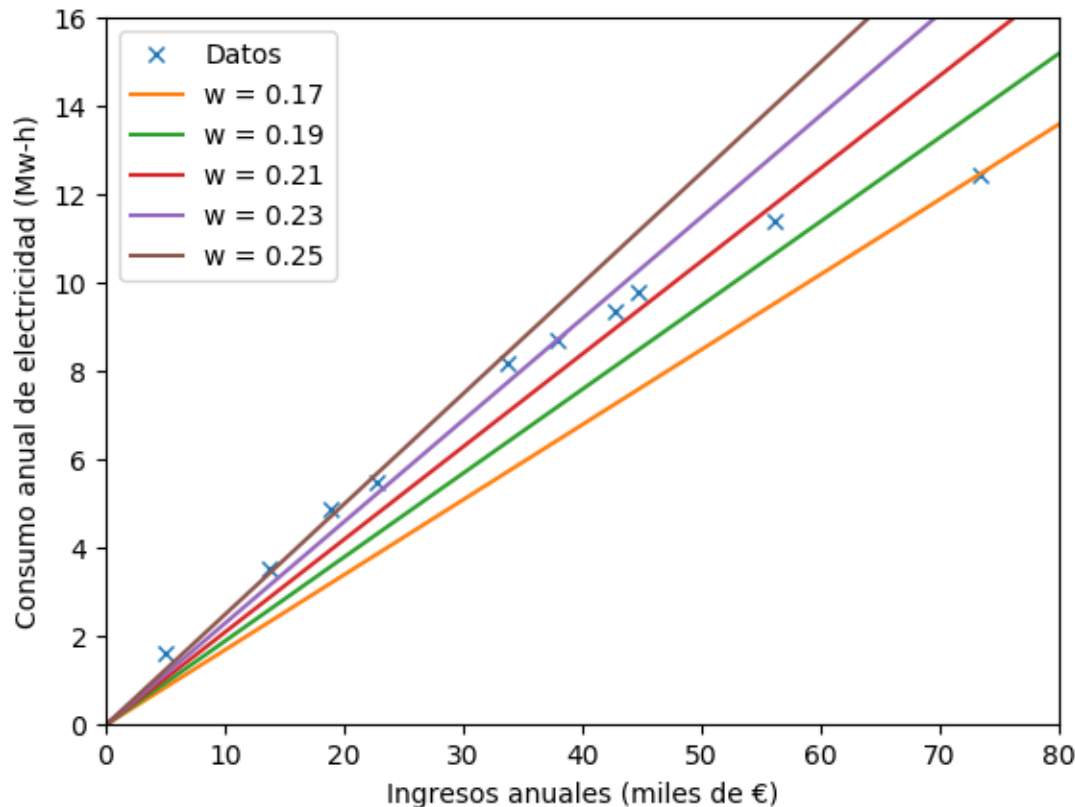


# Pasos del *machine learning*



# Regresión lineal (sin *offset*)

## 3. Formulación de hipótesis

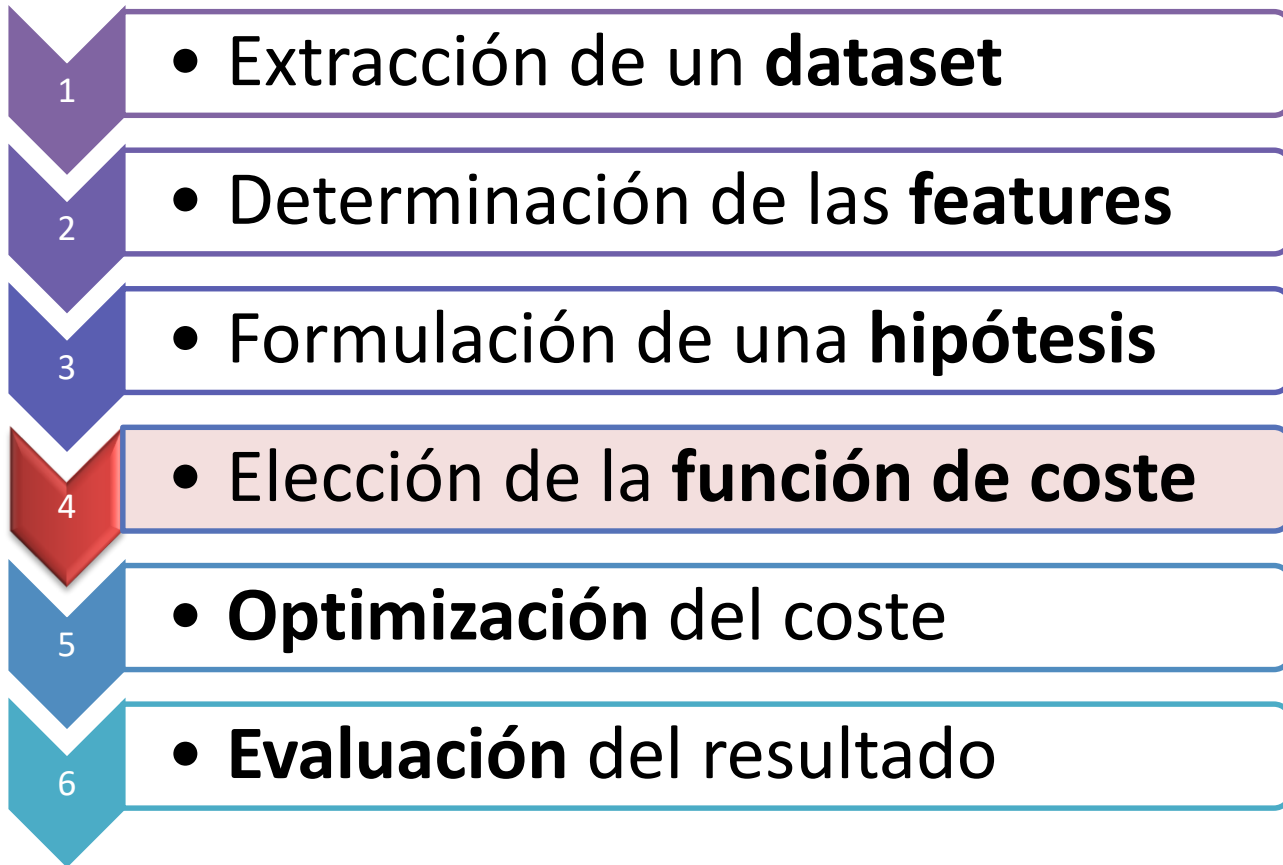


La posición de la recta depende del parámetro  $w$

$$\text{Hipótesis } h_w(x) = wx$$

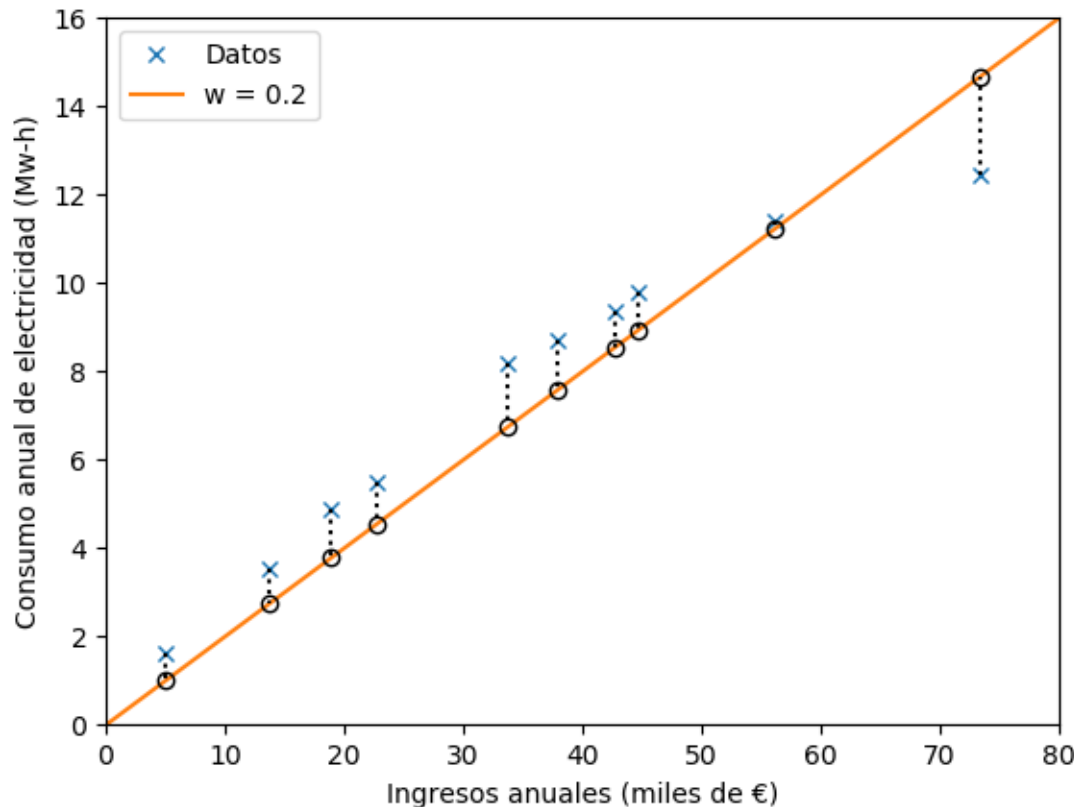
Recta que pasa por el origen

# Pasos del *machine learning*



# Error cuadrático medio

## 4. Elección de la función de coste



$$J(h_w(x), y) = \frac{1}{n} \sum_{i=1}^n (h_w(x^{(i)}) - y^{(i)})^2$$

# Error cuadrático medio

## 4. Elección de la función de coste

$$J(h_w(x), y) = \frac{1}{n} \sum_{i=1}^n (h_w(x^{(i)}) - y^{(i)})^2$$

The diagram shows the cost function  $J(h_w(x), y) = \frac{1}{n} \sum_{i=1}^n (w x^{(i)} - y^{(i)})^2$  with annotations. A blue label "Datos (constantes)" has arrows pointing to the  $x^{(i)}$  and  $y^{(i)}$  terms, which are enclosed in blue boxes. A red label "Parámetro (variable)" has an arrow pointing to the  $w$  term, which is enclosed in a red box.

$$J(h_w(x), y) = \frac{1}{n} \sum_{i=1}^n (\boxed{w} \boxed{x^{(i)}} - \boxed{y^{(i)}})^2$$

Datos (constantes)

Parámetro (variable)

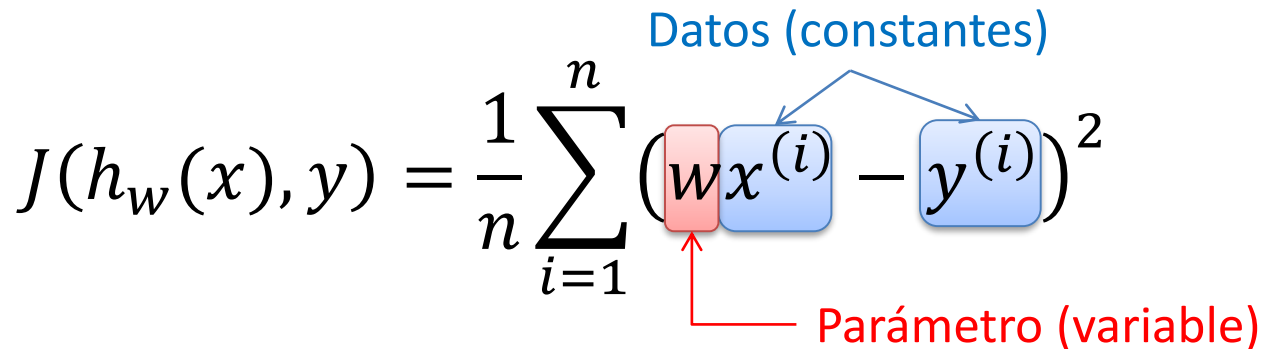
# Error cuadrático medio

## 4. Elección de la función de coste

$$J(h_w(x), y) = \frac{1}{n} \sum_{i=1}^n (w x^{(i)} - y^{(i)})^2$$

Datos (constantes)

Parámetro (variable)

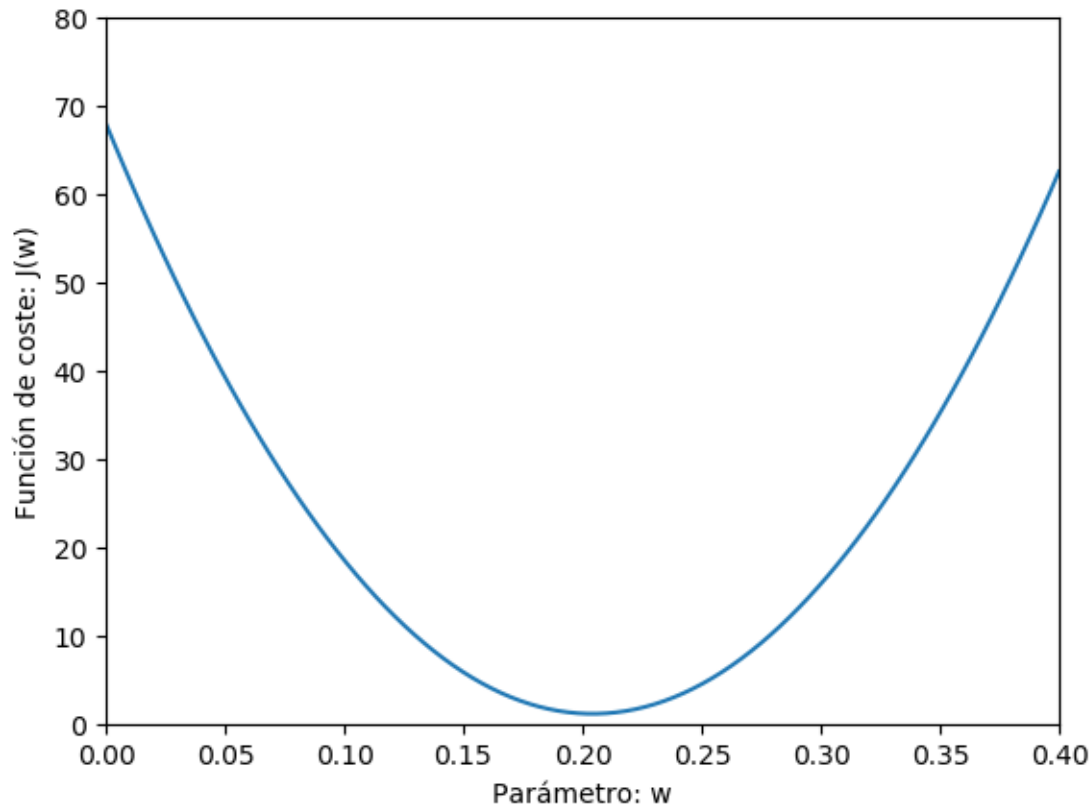


$x^{(i)}$	$y^{(i)}$
44.6	9.81
73.4	12.4
5.0	1.63
33.7	8.17
18.9	4.87
$\vdots$	$\vdots$

$$J(h_w(x), y) = \frac{1}{n} \left[ \begin{aligned} &(w \cdot 44.6 - 9.81)^2 + \\ &(w \cdot 73.4 - 12.4)^2 + \\ &(w \cdot 5.0 - 1.63)^2 + \\ &(w \cdot 33.7 - 8.17)^2 + \\ &(w \cdot 18.9 - 4.87)^2 + \\ &\vdots \end{aligned} \right]$$

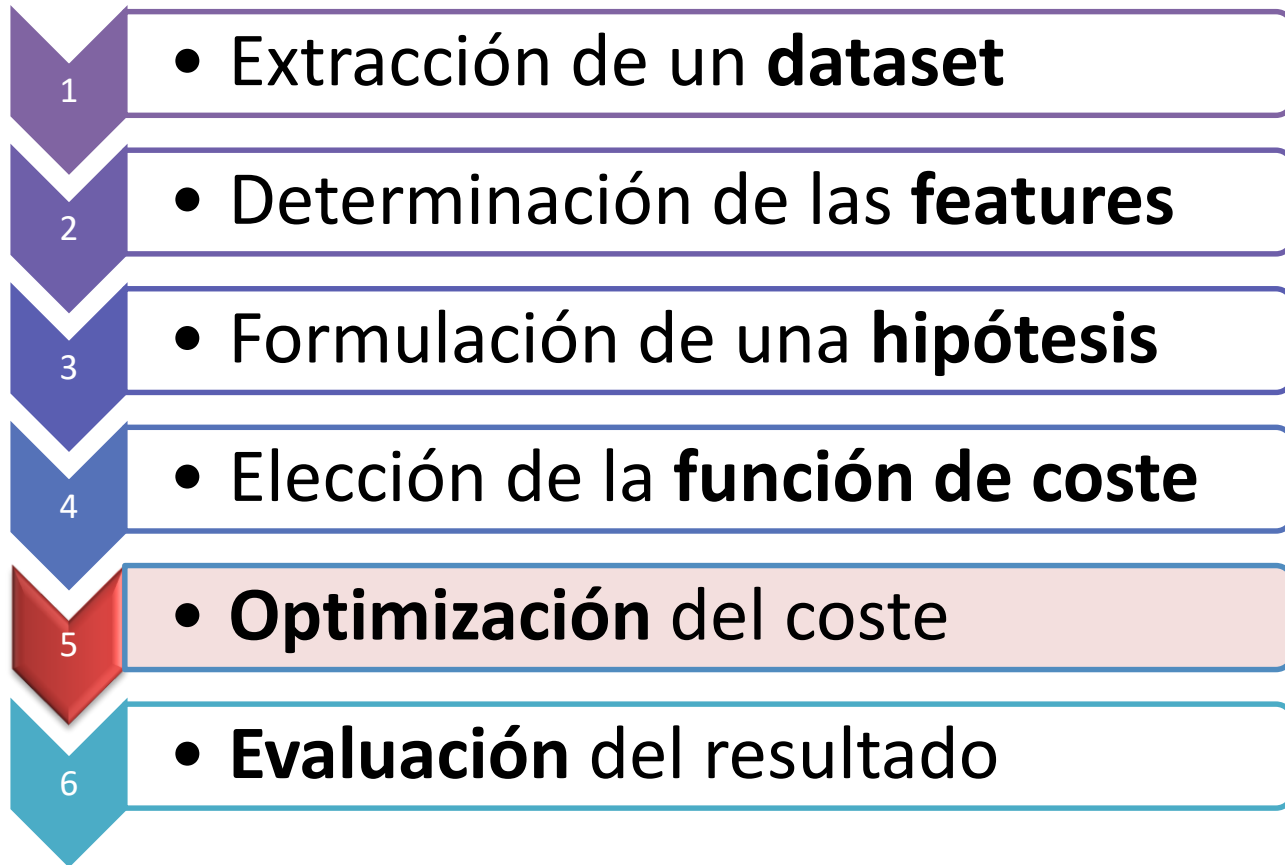
# Error cuadrático medio

## 4. Elección de la función de coste



$$J(h_w(x), y) = J(w) = \frac{1}{n} \sum_{i=1}^n (wx^{(i)} - y^{(i)})^2$$

# Pasos del *machine learning*





# Normal Equation

## 5. Optimización del coste

$$w^* = \arg \min_w J(h_w(x), y) = \arg \min_w \frac{1}{n} \sum_{i=1}^n (wx^{(i)} - y^{(i)})^2$$

$$\frac{d}{dw} J(h_w(x), y) = 0$$

$$\frac{d}{dw} \frac{1}{n} \sum_{i=1}^n (wx^{(i)} - y^{(i)})^2 = 0$$

# Normal Equation

## 5. Optimización del coste

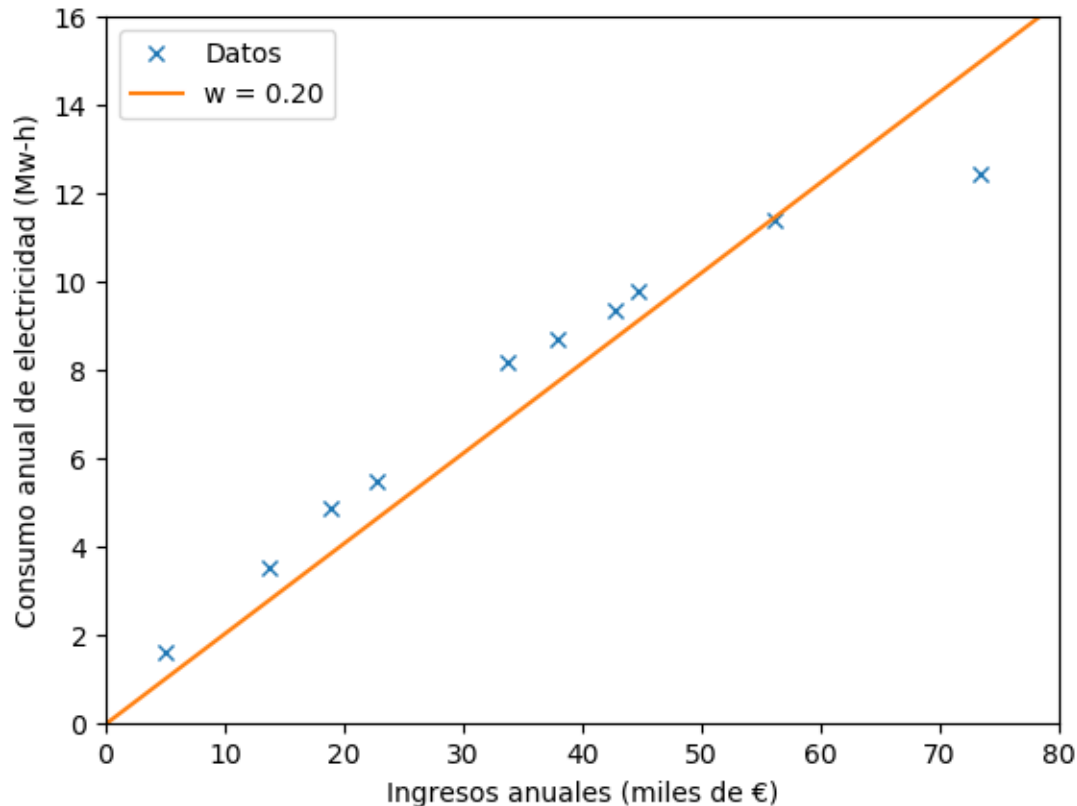
$$\frac{1}{n} \sum_{i=1}^n \frac{d}{dw} (wx^{(i)} - y^{(i)})^2 = 0$$

$$\sum_{i=1}^n 2(w^*x^{(i)} - y^{(i)})x^{(i)} = 0$$

$$w^* \sum_{i=1}^n x^{(i)2} - \sum_{i=1}^n x^{(i)}y^{(i)} = 0$$

# Normal Equation

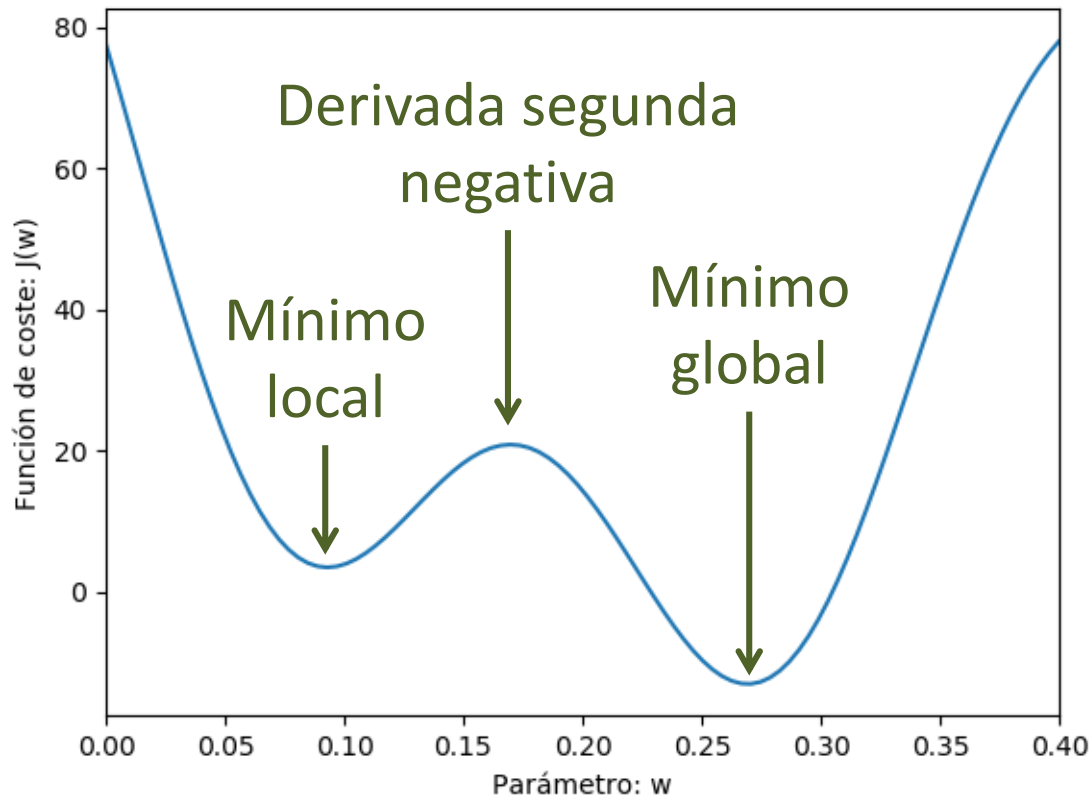
## 5. Optimización del coste



$$w^* = \frac{\sum_{i=1}^n x^{(i)} y^{(i)}}{\sum_{i=1}^n x^{(i)2}} = 0.20 \rightarrow J = 1.30$$

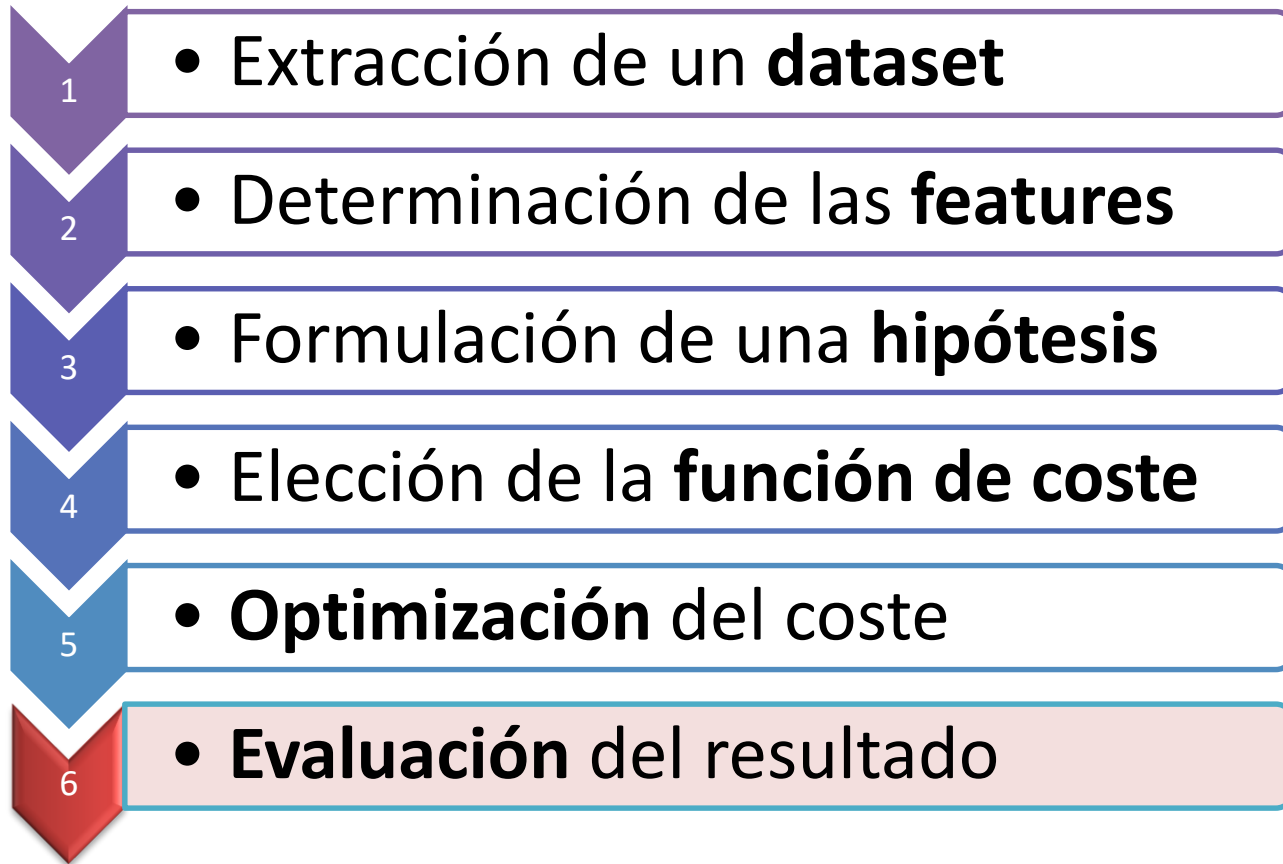
# Normal Equation

## 5. Optimización del coste



Algunas funciones de coste pueden ser NO CONVEXAS  
El error cuadrático medio es una función de coste CONVEXA

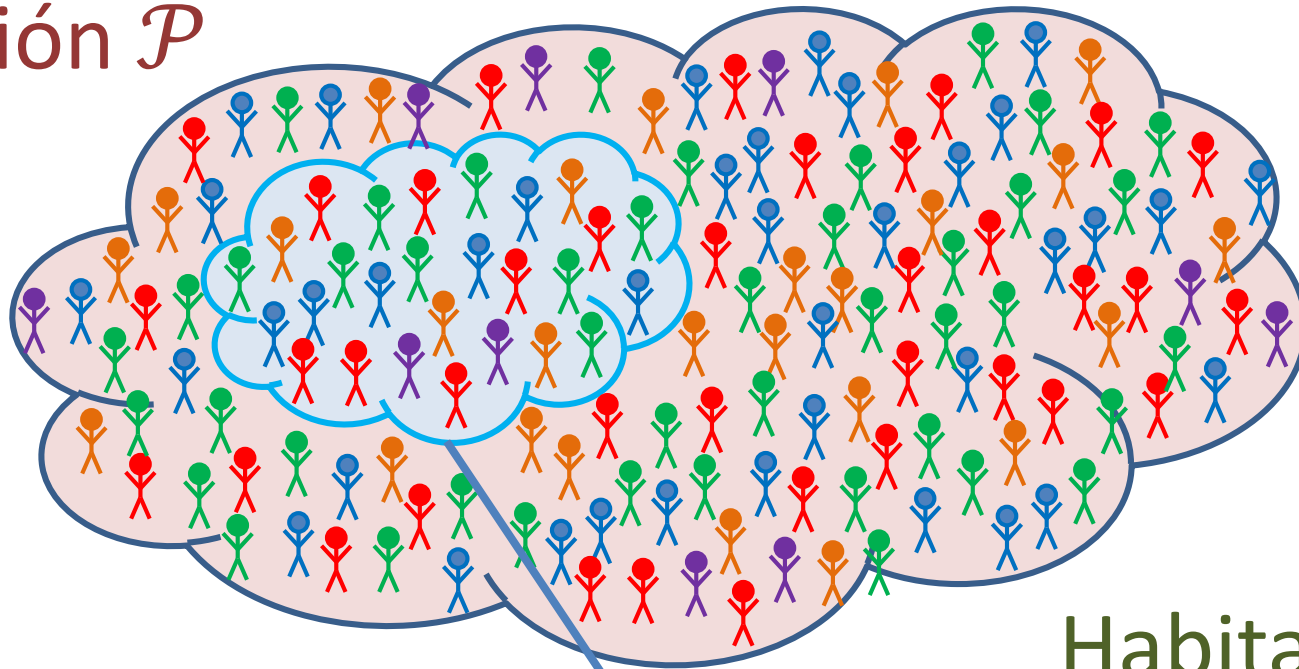
# Pasos del *machine learning*



# Generalización

## 6. Evaluación del resultado

Población  $\mathcal{P}$



Habitantes  
 $N$

Muestra  $\mathcal{M}$   
(dataset)



Cientes  
 $n$

# Generalización

## 6. Evaluación del resultado

Riesgo:  $R_{\mathcal{P}}[h] = E_{\mathcal{P}}[J(h(x), y)] = \frac{1}{N} \sum_{i=1}^N J(h(x_i), y_i)$

Riesgo experimental

$$R_{\mathcal{M}}[h] = E_{\mathcal{M}}[J(h(x), y)] = \frac{1}{n} \sum_{i=1}^n J(h(x_i), y_i)$$

$$R_{\mathcal{P}}[h] = (R_{\mathcal{P}}[h] - R_{\mathcal{M}}[h]) + R_{\mathcal{M}}[h]$$

Riesgo

Error de  
generalización

Riesgo  
experimental

# Generalización

## 6. Evaluación del resultado

$$R_{\mathcal{P}}[h] = (R_{\mathcal{P}}[h] - R_{\mathcal{M}}[h]) + R_{\mathcal{M}}[h]$$

$$R_{\mathcal{P}}[h] = R_{\mathcal{M}}[h] + \varepsilon_{gen}[h]$$

Riesgo  
experimental

Error de  
generalización

$\sim$ Test  
error

Training  
error

Bias

Variance



# Análisis bootstrap

## 6. Evaluación del resultado



Cliente	Ingresos $x$	Consumo $y$
1	44.6	9.81
2	73.4	12.4
3	5.0	1.63
4	33.7	8.17
5	18.9	4.87
6	13.8	3.54
7	22.7	5.94
8	37.8	8.59
9	42.7	9.47
10	56.2	11.1

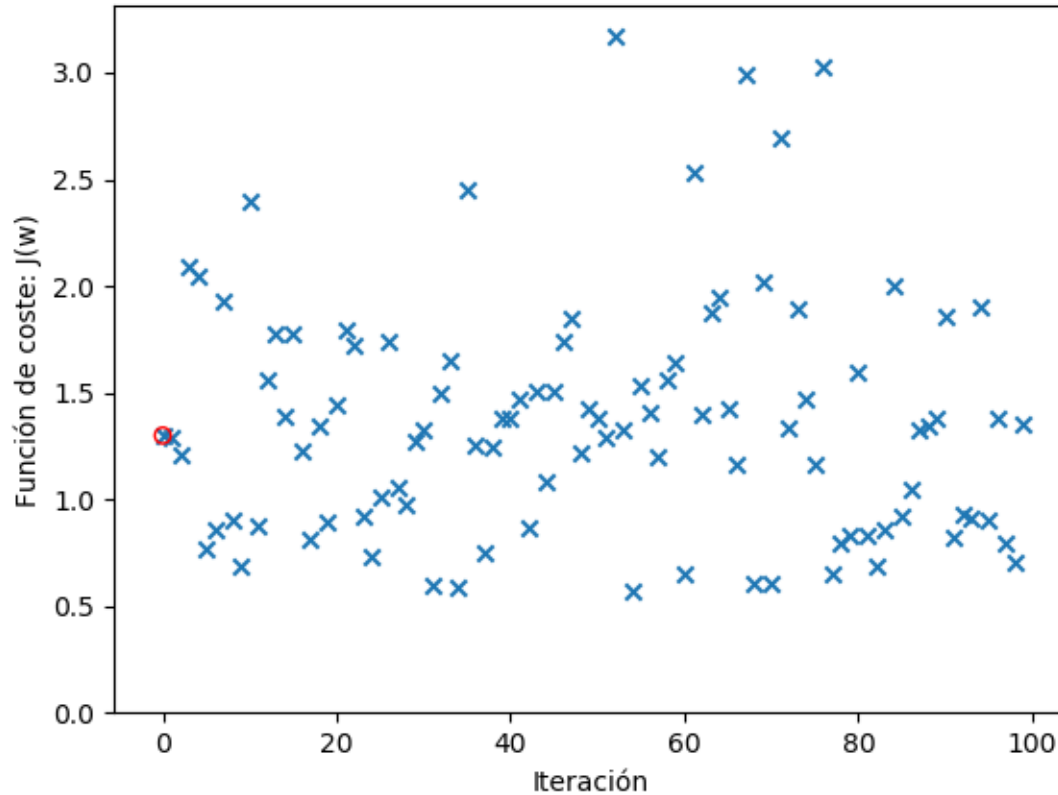
$$J = 1.30$$

Cliente	Ingresos $x$	Consumo $y$
6	13.8	3.54
9	42.7	9.47
6	13.8	3.54
1	44.6	9.81
1	44.6	9.81
2	73.4	12.4
8	37.8	8.59
7	22.7	5.94
3	5.0	1.63
4	33.7	8.17

$$J = 1.29$$

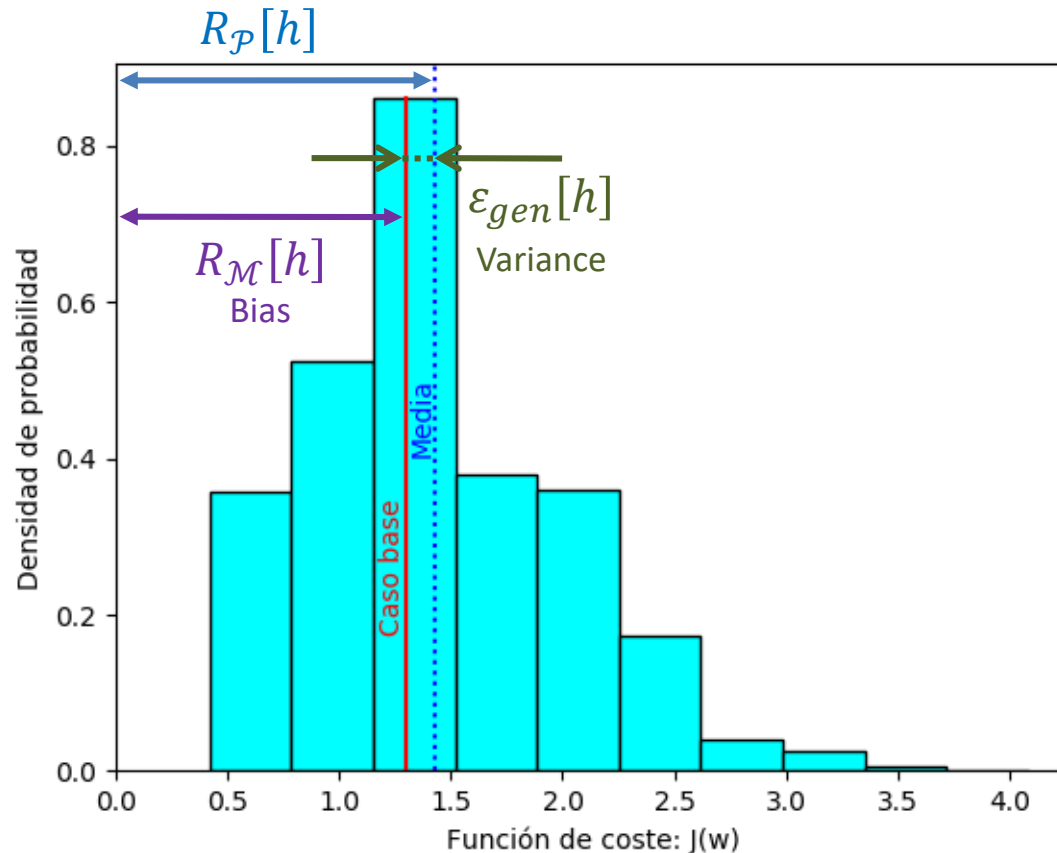
# Análisis bootstrap

## 6. Evaluación del resultado



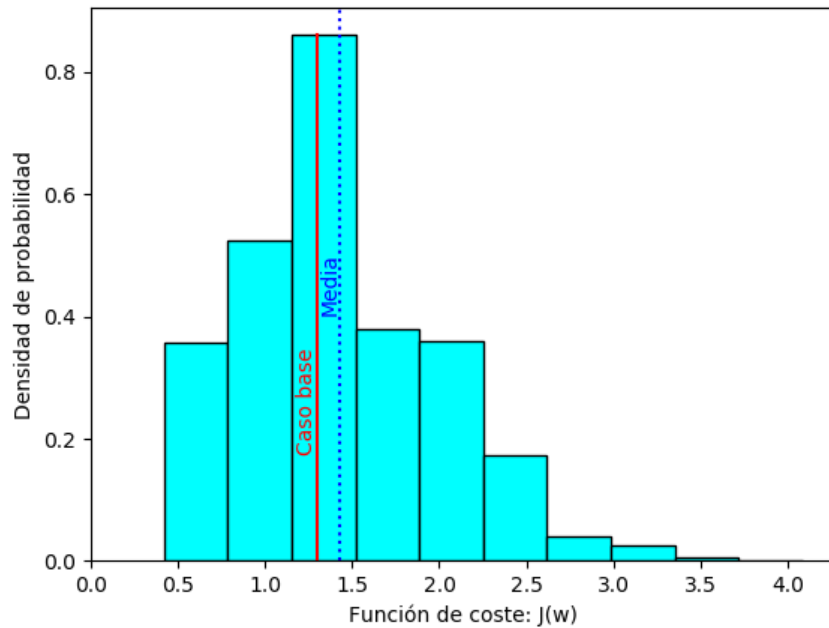
# Análisis bootstrap

## 6. Evaluación del resultado

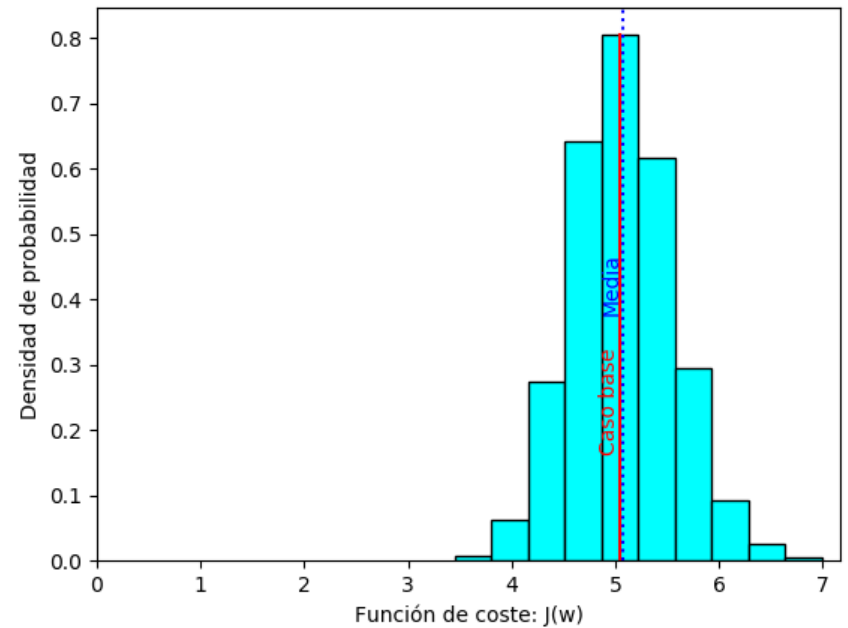


# Análisis bootstrap

## 6. Evaluación del resultado



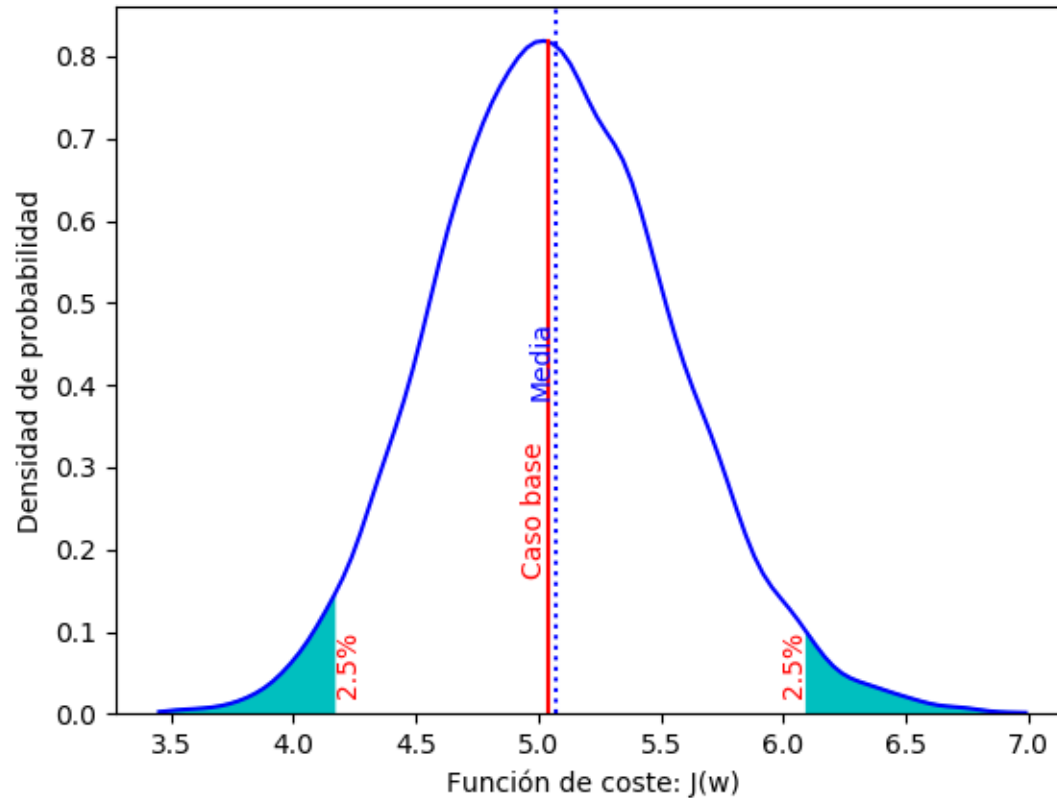
$n = 10$



$n = 100$

# Análisis bootstrap

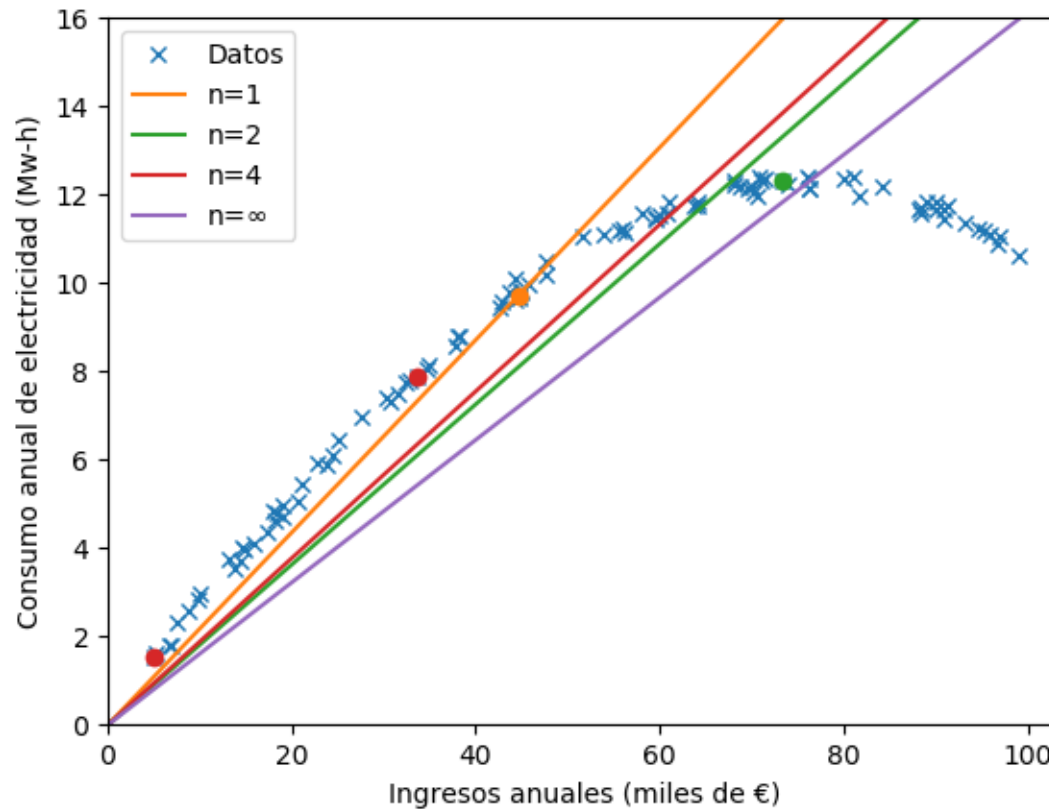
## 6. Evaluación del resultado



$$n = 100$$

# Análisis bootstrap

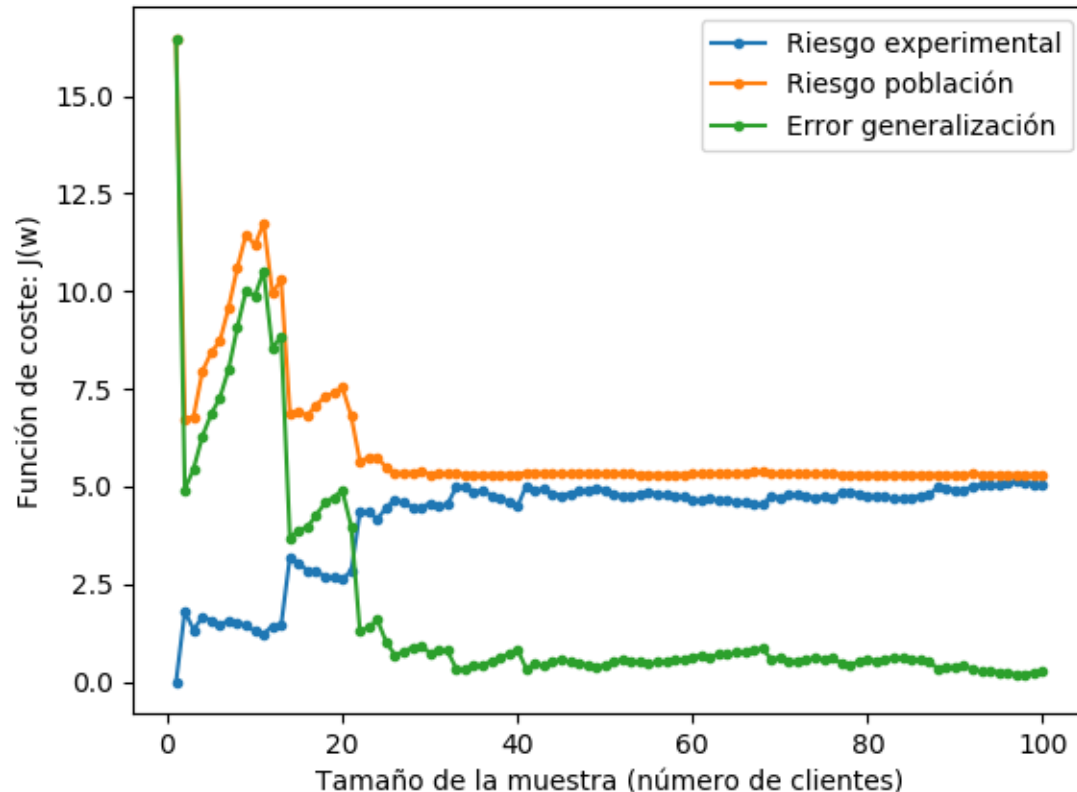
## 6. Evaluación del resultado



# Análisis bootstrap

## 6. Evaluación del resultado

Learning curve



$$R_{\mathcal{P}}[h] = R_{\mathcal{M}}[h] + \varepsilon_{gen}[h]$$