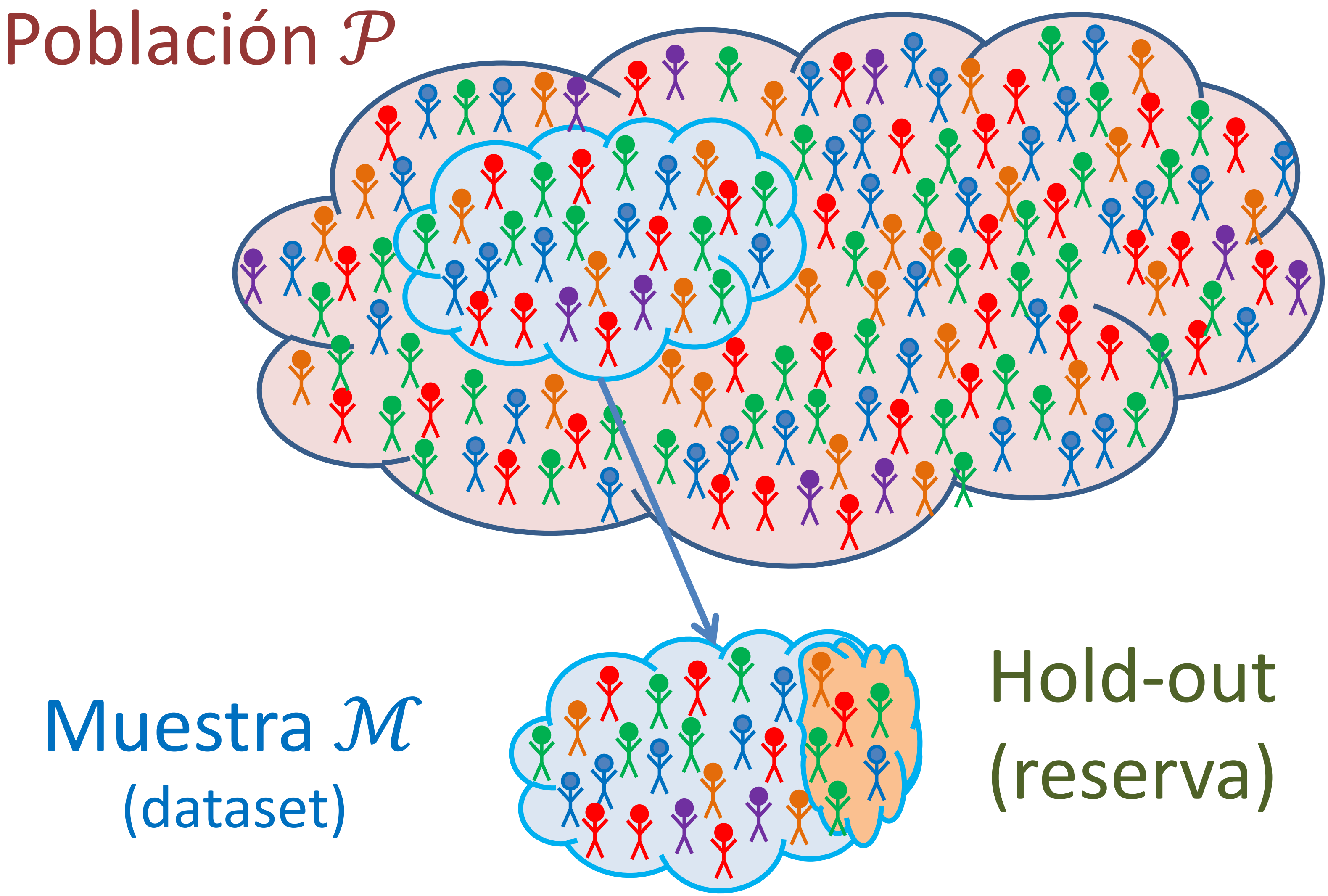
Pasos del *machine learning*

1. Extracción de un **dataset**
2. Determinación de las **features**
3. Formulación de una **hipótesis**
4. Elección de la **función de coste**
5. **Optimización** del coste
6. **Evaluación** del resultado

Hold-out

Hold-out

1. Extracción de una muestra

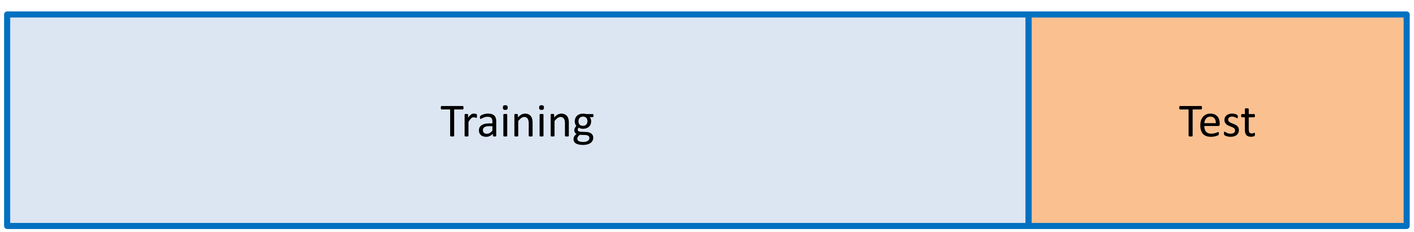


Hold-out

1. Extracción de una muestra

Muestra $\mathcal{M}$
(dataset)

Hold-out
(reserva)

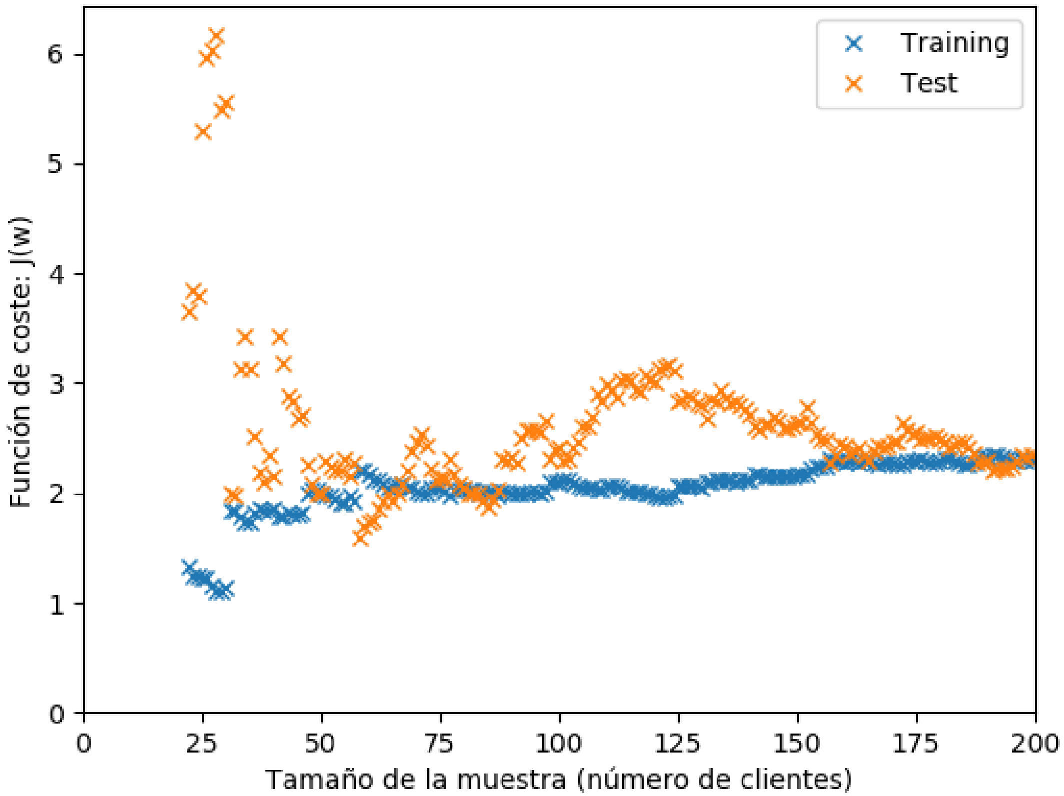


$\sim 70\%$ $\sim 30\%$

Evaluación de resultados

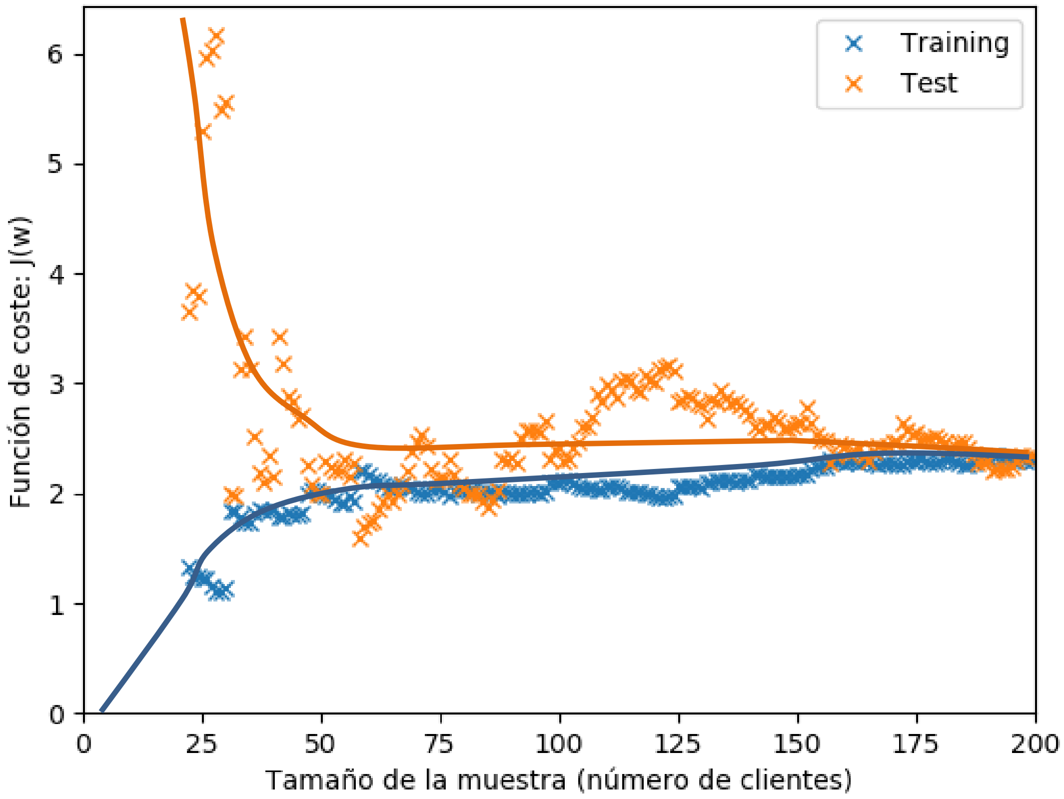
Hold-out

6. Evaluación del resultado



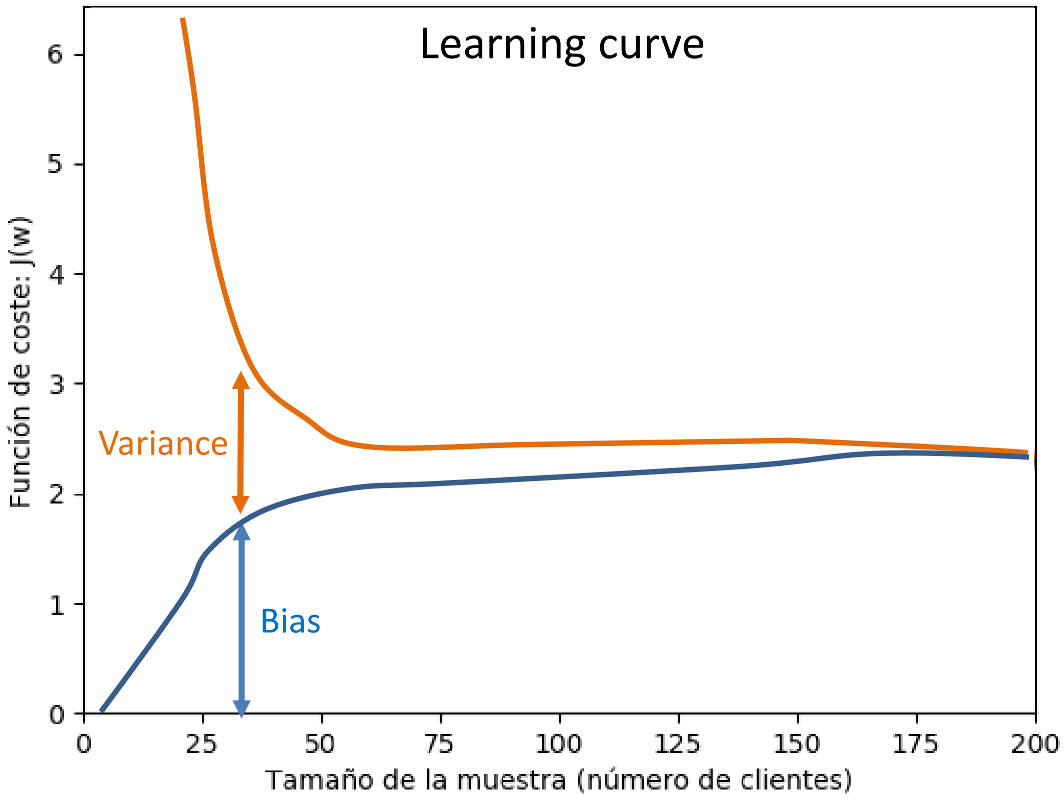
Hold-out

6. Evaluación del resultado



Hold-out

6. Evaluación del resultado



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

Pasos del *machine learning*

1. Extracción de un **dataset**
2. Determinación de las **features**
3. Formulación de una **hipótesis**
4. Elección de la **función de coste**
5. **Optimización** del coste
6. **Evaluación** del resultado

Gradient Descent

Gradient Descent (GD)

5. Optimización del coste

- Problemas de la optimización por Normal Equation
  - El tiempo de cálculo crece con el número de features
  - No es aplicable a todas las funciones de coste
    - A veces no hay solución analítica a la ecuación:

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

Gradient Descent (GD)

5. Optimización del coste

- Alternativa: Gradient Descent (GD)

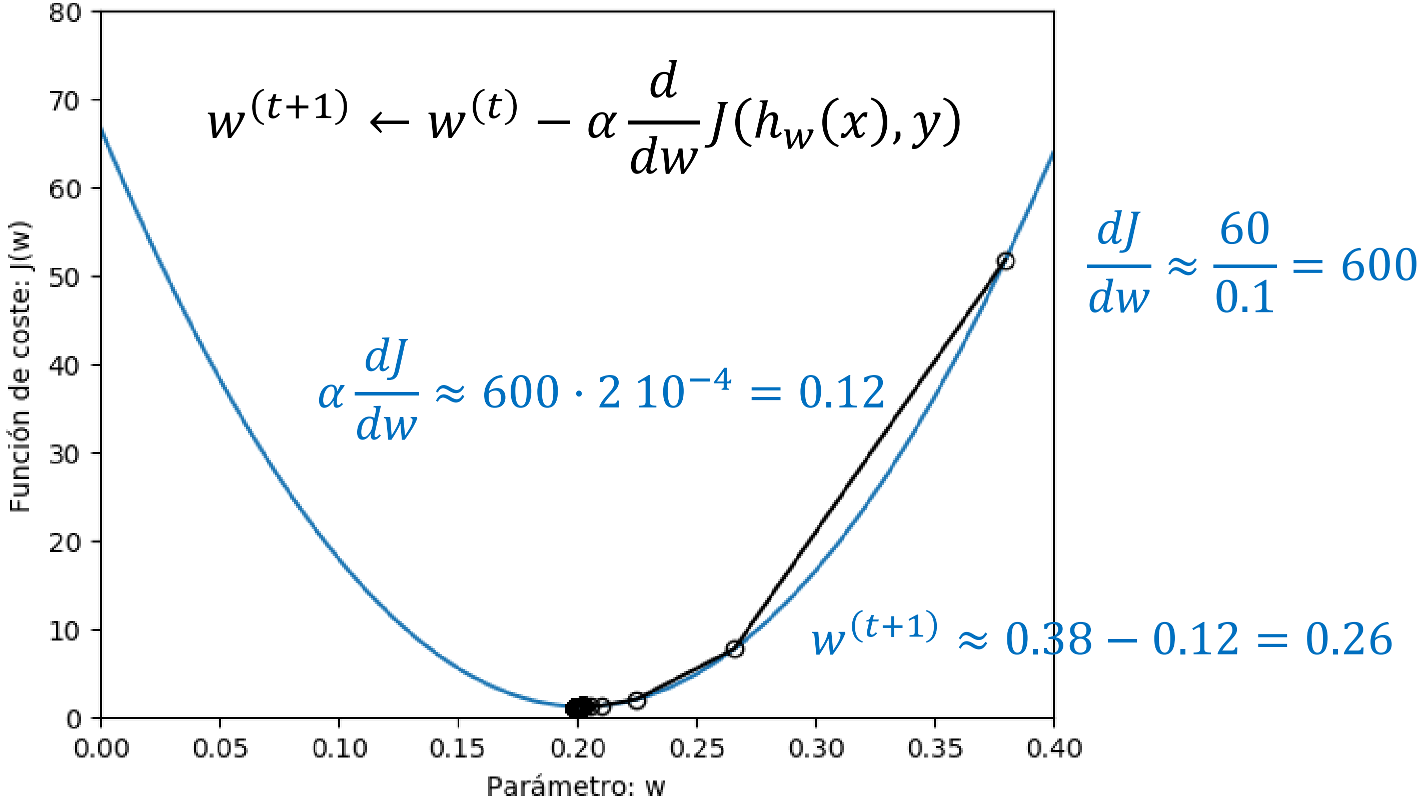
$$w^{(t+1)} \leftarrow w^{(t)} - \alpha \frac{d}{dw} J(h_w(x), y)$$

$$w^{(t+1)} \leftarrow w^{(t)} - \alpha \frac{d}{dw} J(w)$$

α: learning rate

Gradient Descent (GD)

5. Optimización del coste



$\alpha = 2 \cdot 10^{-4}$: (learning rate) pequeño

Gradient Descent (GD)

5. Optimización del coste

$$\frac{d}{dw}J(w) = \frac{d}{dw}\frac{1}{n}\sum_{i=1}^{n}\left(wx^{(i)} - y^{(i)}\right)^2$$

$$\frac{d}{dw}J(w) = \frac{1}{n}\sum_{i=1}^{n}\frac{d}{dw}\left(wx^{(i)} - y^{(i)}\right)^2$$

$$\boxed{\frac{d}{dw}J(w) = \frac{1}{n}\sum_{i=1}^{n}2\left(wx^{(i)} - y^{(i)}\right)x^{(i)}}$$

$$\frac{d^2}{dw^2}J(w) = \frac{2}{n}\sum_{i=1}^{n}x^{(i)^2}$$

Positiva

(y constante)

Gradient Descent (GD)

5. Optimización del coste

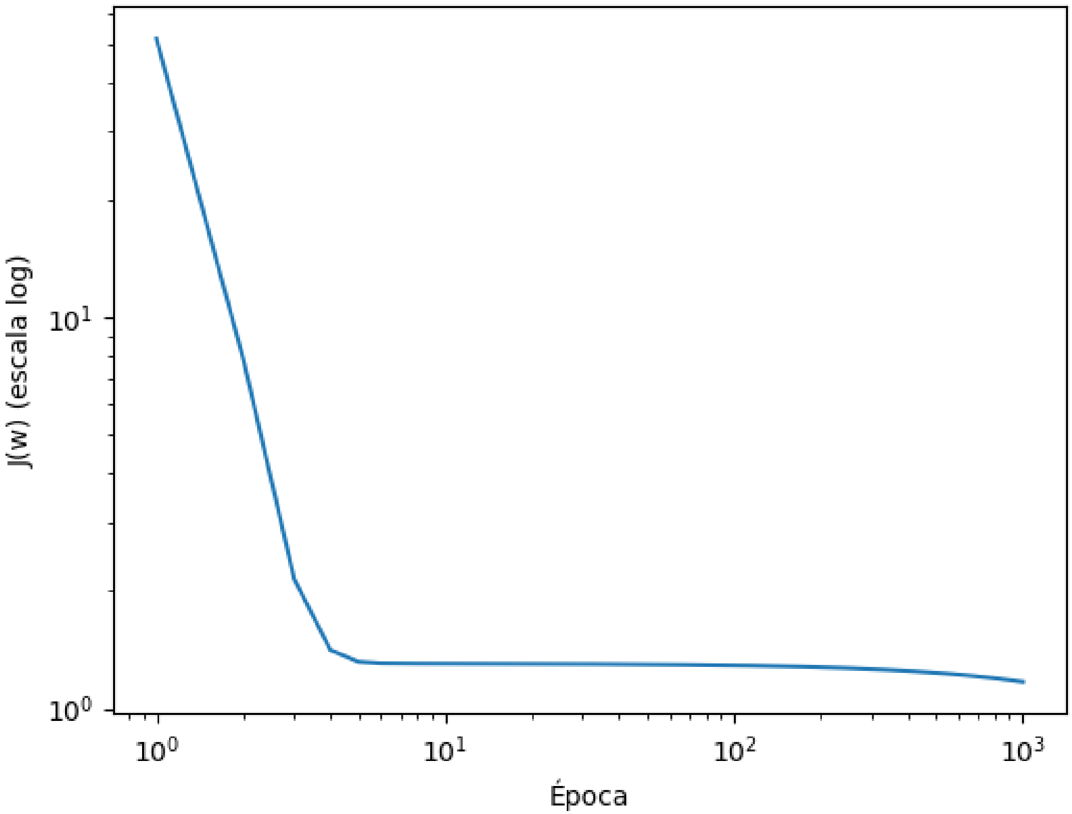
- Repetir hasta convergencia

Es decir, hasta alcanzar un mínimo de $J(w)$

$$w^{(t+1)} \leftarrow w^{(t)} - \alpha \underbrace{\frac{1}{n}\sum_{i=1}^{n} \overbrace{2\left(wx^{(i)} - y^{(i)}\right)x^{(i)}}^{\text{Derivada de la función de coste correspondiente al } i\text{-ésimo punto}}}_{\text{Media de la derivada de la función de coste para todos los puntos}}$$

Gradient Descent (GD)

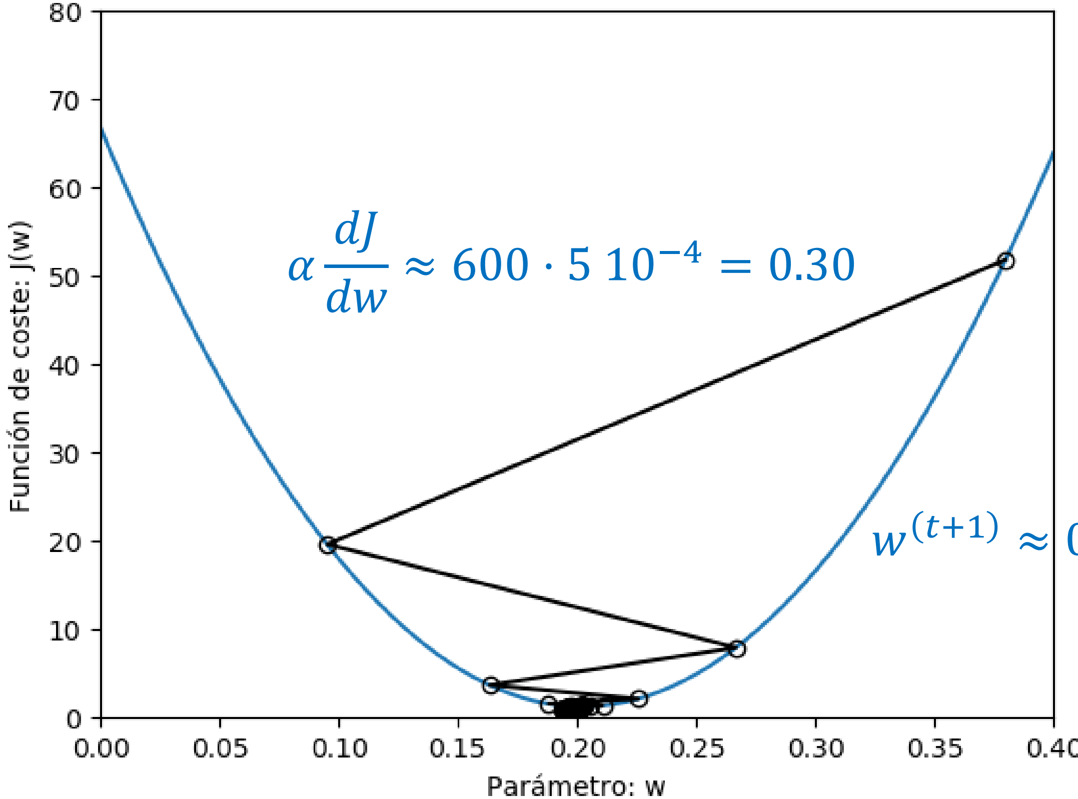
5. Optimización del coste



$\alpha = 2 \cdot 10^{-4}$:(learning rate) pequeño

Gradient Descent (GD)

5. Optimización del coste



$$\alpha \frac{dJ}{dw} \approx 600 \cdot 5\ 10^{-4} = 0.30$$

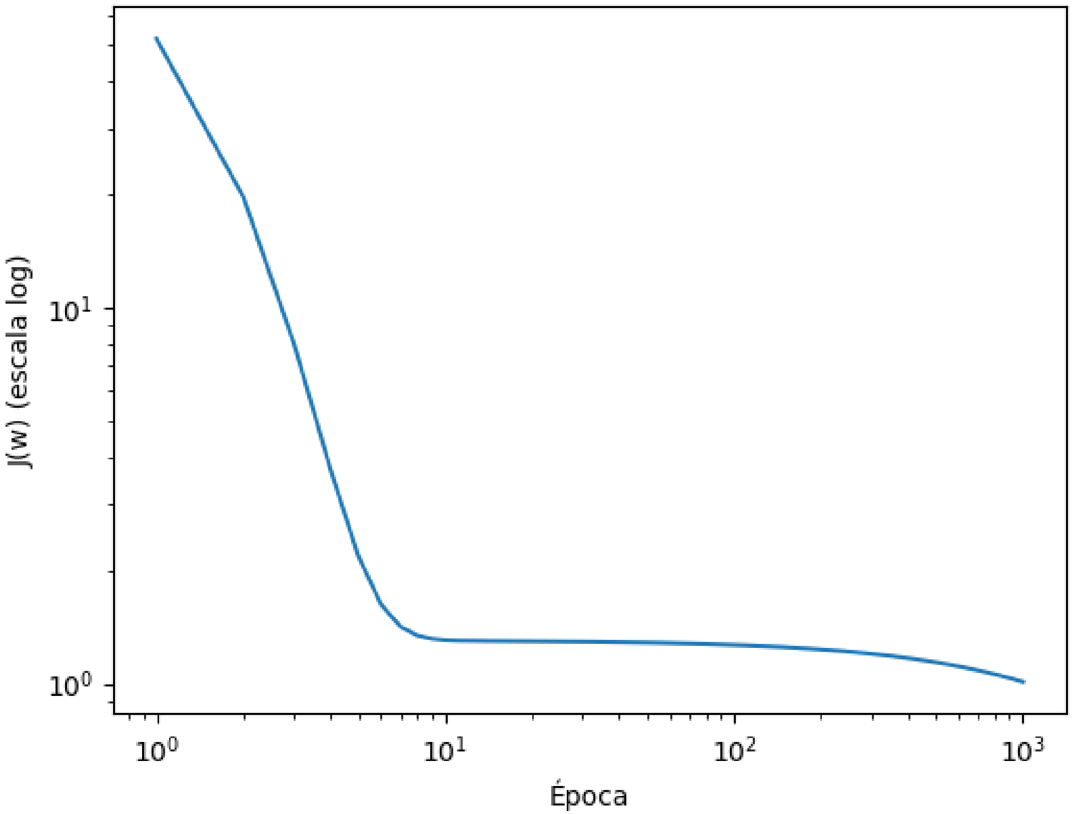
$$\frac{dJ}{dw} \approx \frac{60}{0.1} = 600$$

$$w^{(t+1)} \approx 0.38 - 0.30 = 0.08$$

$\alpha = 5 \cdot 10^{-4}$:(learning rate) grande

Gradient Descent (GD)

5. Optimización del coste



$\alpha = 5 \cdot 10^{-4}$:(learning rate) grande

Stochastic Gradient Descent (SGD)

5. Optimización del coste

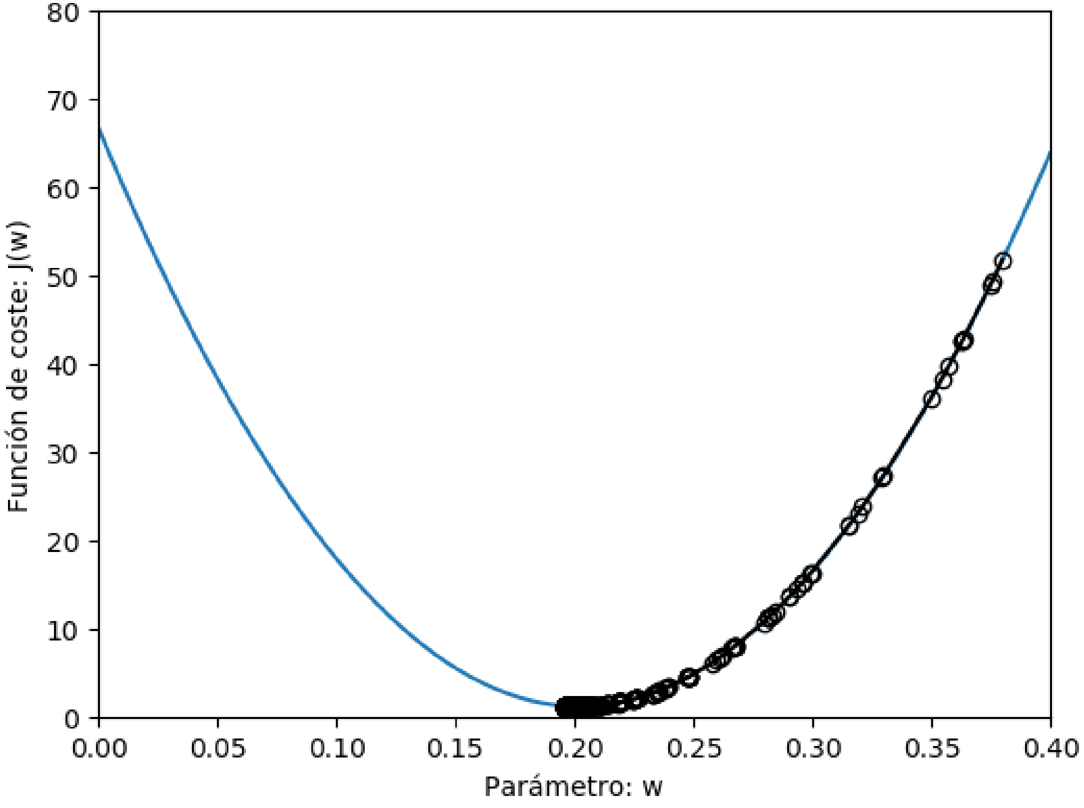
- Repetir hasta convergencia
  - Ordenar aleatoriamente los pares $\left(x^{(i)}, y^{(i)}\right)$
  - Para $i = 1,2,\cdots,n$ hacer

$$w^{(t+1)} \leftarrow w^{(t)} - \alpha \underbrace{2\left(wx^{(i)} - y^{(i)}\right)x^{(i)}}_{\substack{\text{Aproximación a la derivada} \\ \text{de la función de coste} \\ \text{Correspondiente a} \\ \text{un único punto}}}$$

Una época

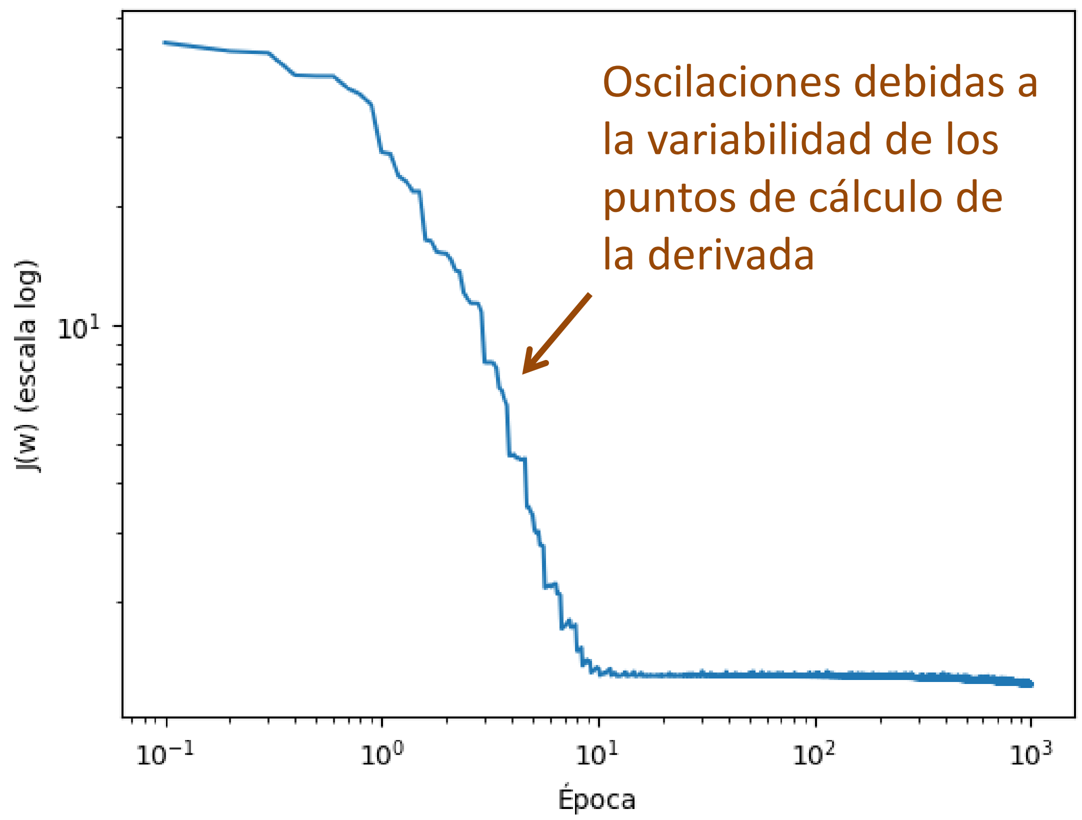
Stochastic Gradient Descent (SGD)

5. Optimización del coste



Stochastic Gradient Descent (SGD)

5. Optimización del coste



Batch Gradient Descent (BGD)

5. Optimización del coste

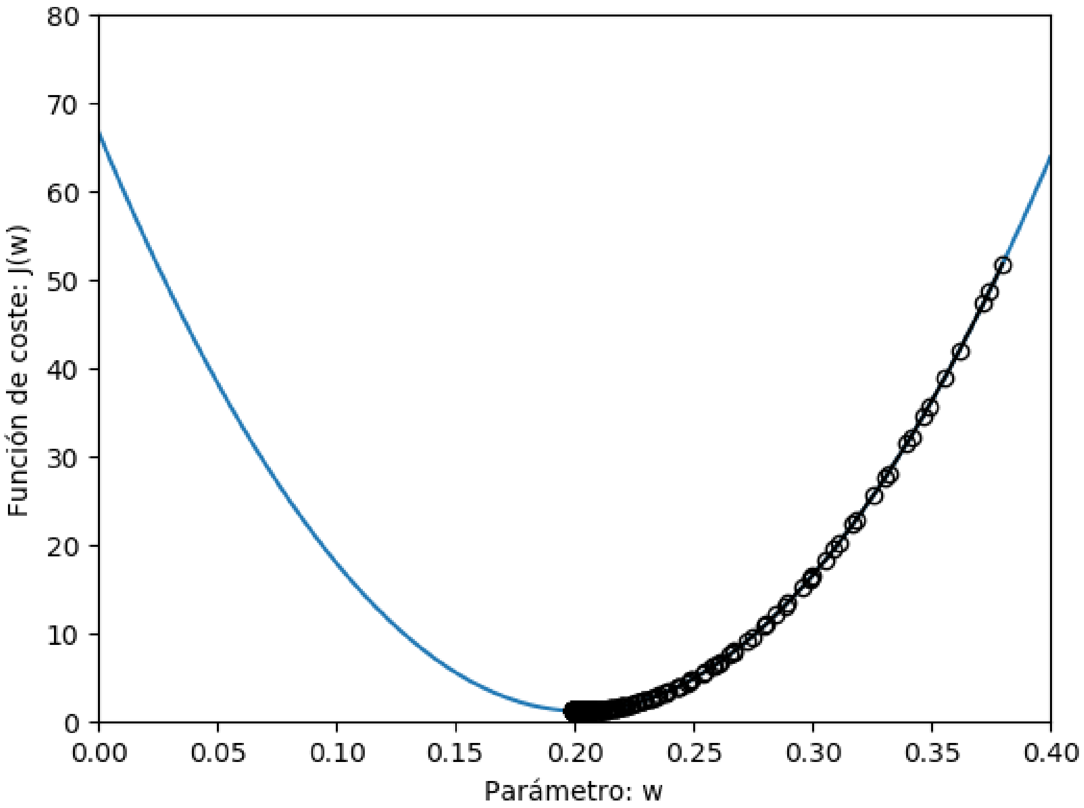
- Repetir hasta convergencia
  - Ordenar aleatoriamente los pares $\left(x^{(i)}, y^{(i)}\right)$
  - Establecer m tandas de datos de dimensión b
  - Para tanda $= 1,2,\cdots,m$ hacer

$$w^{(t+1)} \leftarrow w^{(t)} - \alpha \underbrace{\frac{1}{b}\sum_{i=1}^{b} 2\left(wx^{(i)} - y^{(i)}\right)x^{(i)}}_{\substack{\text{Aproximación a la derivada} \\ \text{de la función de coste} \\ \text{Media correspondiente a} \\ \text{una tanda de puntos}}}$$

Una época

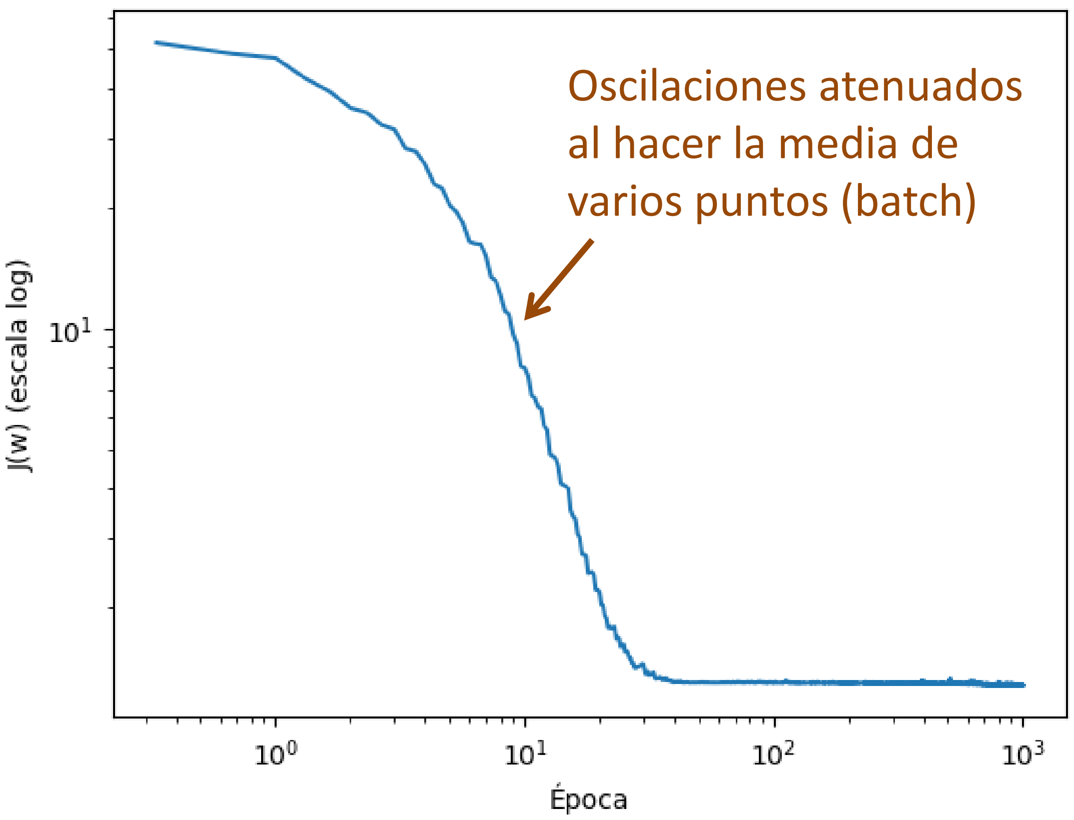
Batch Gradient Descent (SGD)

5. Optimización del coste



Batch Gradient Descent (SGD)

5. Optimización del coste



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

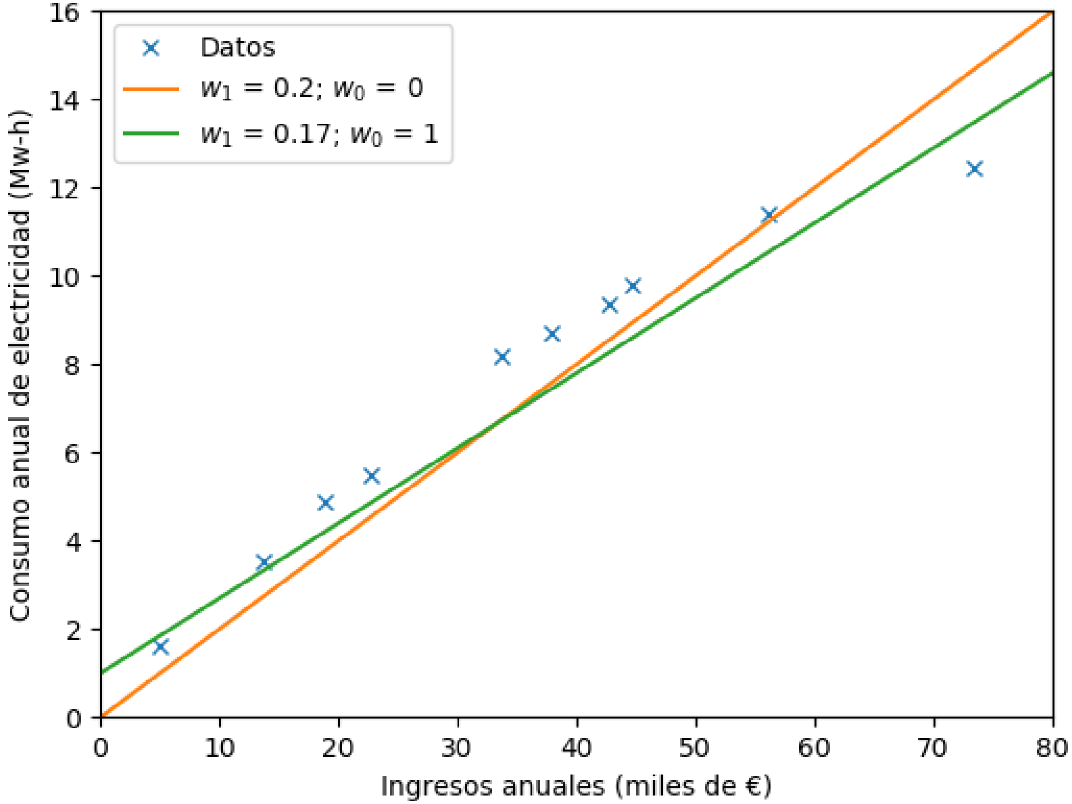
Pasos del *machine learning*

1. Extracción de un **dataset**
2. Determinación de las **features**
3. Formulación de una **hipótesis**
4. Elección de la **función de coste**
5. **Optimización** del coste
6. **Evaluación** del resultado

Regresión lineal

Regresión lineal

3. Formulación de hipótesis



La posición de la recta depende de los parámetros w_0 y w_1

Hipótesis $h_w(x) = w_0 + w_1 x$

Regresión lineal

3. Formulación de hipótesis

$$h_w(x) = w_0 + w_1 x$$

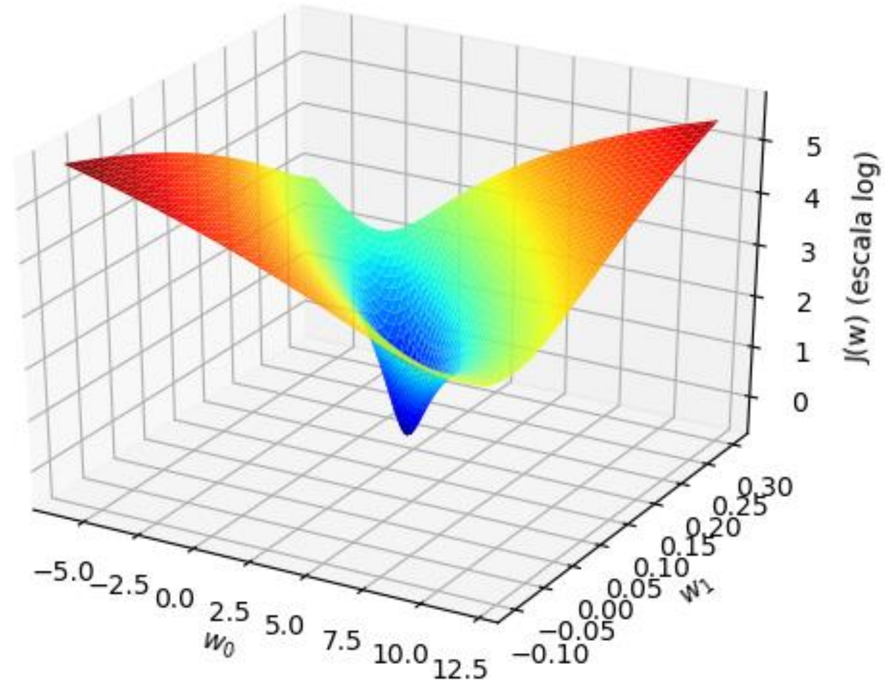
$$x = \begin{bmatrix} x^{(1)} \\ x^{(2)} \\ \vdots \\ x^{(n)} \end{bmatrix} \Longrightarrow \mathbb{X} = \begin{bmatrix} 1 & x^{(1)} \\ 1 & x^{(2)} \\ \vdots & \vdots \\ 1 & x^{(n)} \end{bmatrix}$$

$$w = [w_0 \quad w_1]; \quad w^T = \begin{bmatrix} w_0 \\ w_1 \end{bmatrix}$$

$$h_w(x) = \mathbb{X} w^T$$

Regresión lineal

4. Elección de la función de coste

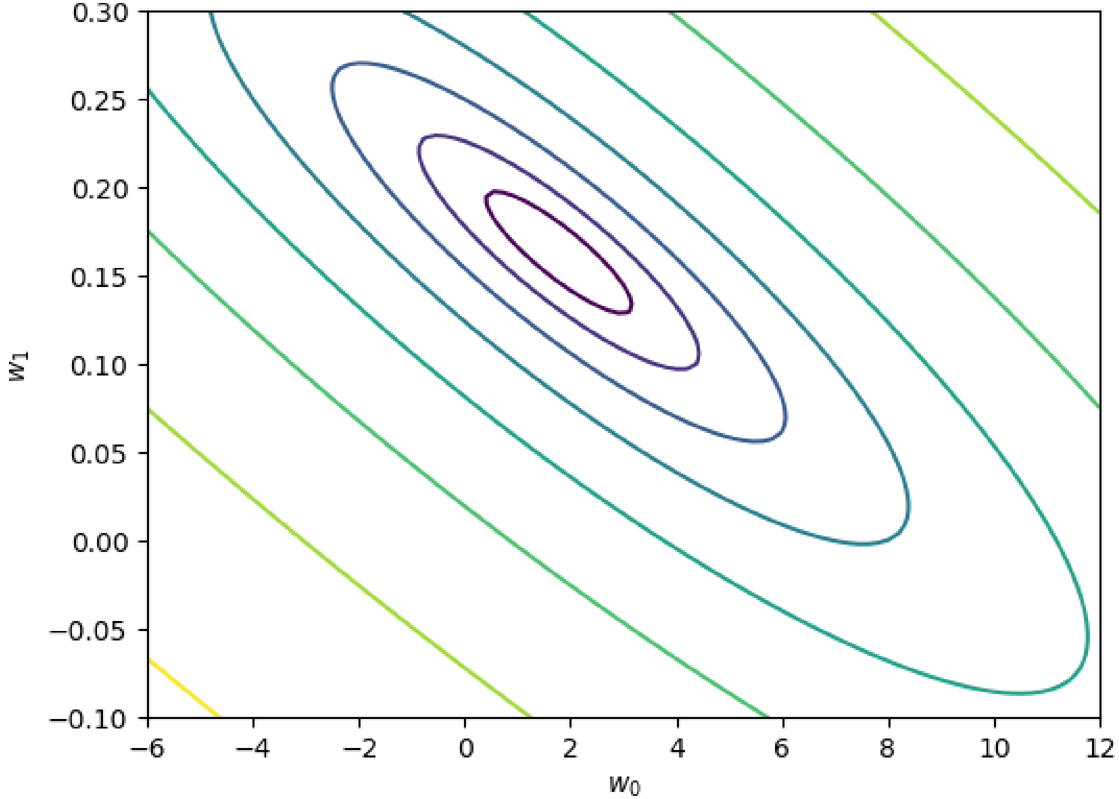


Error cuadrático medio

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

Regresión lineal

4. Elección de la función de coste



Error cuadrático medio

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

Regresión lineal

4. Elección de la función de coste

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

$$h_w(x) = w_0 + w_1 x$$

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

Regresión lineal

4. Elección de la función de coste

$$\mathbb{X}w^T - y = \begin{bmatrix} w_0 + w_1 x^{(1)} \\ w_0 + w_1 x^{(2)} \\ \vdots \\ w_0 + w_1 x^{(n)} \end{bmatrix} - \begin{bmatrix} y^{(1)} \\ y^{(2)} \\ \vdots \\ y^{(n)} \end{bmatrix} = \begin{bmatrix} w_0 + w_1 x^{(1)} - y^{(1)} \\ w_0 + w_1 x^{(2)} - y^{(2)} \\ \vdots \\ w_0 + w_1 x^{(n)} - y^{(n)} \end{bmatrix}$$

$$(\mathbb{X}w^T - y)^T(\mathbb{X}w^T - y) = \sum_{i=1}^{n} \left(w_0 + w_1 x^{(i)} - y^{(i)}\right)^2$$

$$J(h_w(x), y) = \frac{1}{n}(\mathbb{X}w^T - y)^T(\mathbb{X}w^T - y)$$

Regresión lineal

5. Optimización mediante Normal Equation

$$w^* = \arg\min_{w} J(h_w(x), y)$$

$$w^* = \arg\min_{w} \left[\frac{1}{n}(\mathbb{X}w^T - y)^T(\mathbb{X}w^T - y)\right]$$

$$\frac{\partial}{\partial w_j} J(h_w(x), y) = 0, \forall j$$

∇: operador nabla

∇J: gradiente

$$\nabla J(h_w(x), y) \equiv \begin{bmatrix} \frac{\partial J}{\partial w_0} \\ \frac{\partial J}{\partial w_1} \end{bmatrix} = 0$$

Regresión lineal

5. Optimización mediante Normal Equation

$$\nabla J \equiv \begin{bmatrix} \frac{\partial J}{\partial w_0} \\ \frac{\partial J}{\partial w_1} \end{bmatrix}$$

Jacobiano

$$H \equiv \nabla^2 J \equiv \begin{bmatrix} \frac{\partial^2 J}{\partial w_0^2} & \frac{\partial^2 J}{\partial w_0 \partial w_1} \\ \frac{\partial^2 J}{\partial w_1 \partial w_0} & \frac{\partial^2 J}{\partial w_1^2} \end{bmatrix}$$

Hessiano

Regresión lineal

5. Optimización mediante Normal Equation

$$J = \frac{1}{n}(\mathbb{X}w^T - y)^T(\mathbb{X}w^T - y)$$

$$J = \frac{1}{n}[(\mathbb{X}w^T)^T - y^T](\mathbb{X}w^T - y)$$

$$J = \frac{1}{n}(w\mathbb{X}^T - y^T)(\mathbb{X}w^T - y)$$

$$J = \frac{1}{n}(w\mathbb{X}^T\mathbb{X}w^T - w\mathbb{X}^Ty - y^T\mathbb{X}w^T + y^Ty)$$

https://eli.thegreenplace.net/2014/derivation-of-the-normal-equation-for-linear-regression

Regresión lineal

5. Optimización mediante Normal Equation

$$(1\times 1)\quad \overset{(1\times 2)}{w}\,\overset{(2\times n)}{\mathbb{X}^T}\,\overset{(n\times 1)}{y} = (w\mathbb{X}^T y)^T = y^T(w\mathbb{X}^T)^T = \overset{(1\times n)}{y^T}\,\overset{(n\times 2)}{\mathbb{X}}\,\overset{(2\times 1)}{w^T}\quad (1\times 1)$$

$$J = \frac{1}{n}\left(w\mathbb{X}^T\mathbb{X}w^T - 2w\mathbb{X}^T y + y^T y\right)$$

$$\mathbb{X} = \begin{bmatrix} 1 & x^{(1)} \\ 1 & x^{(2)} \\ \vdots & \vdots \\ 1 & x^{(n)} \end{bmatrix}; y = \begin{bmatrix} y^{(1)} \\ y^{(2)} \\ \vdots \\ y^{(n)} \end{bmatrix}$$

$$w = [w_0 \quad w_1]; \quad w^T = \begin{bmatrix} w_0 \\ w_1 \end{bmatrix}$$

Regresión lineal

5. Optimización mediante Normal Equation

$$\mathbb{X}^T\mathbb{X} = \begin{bmatrix} 1 & 1 & \cdots & 1 \\ x^{(1)} & x^{(2)} & \cdots & x^{(n)} \end{bmatrix} \begin{bmatrix} 1 & x^{(1)} \\ 1 & x^{(2)} \\ \vdots & \vdots \\ 1 & x^{(n)} \end{bmatrix}$$

$$\mathbb{X}^T\mathbb{X} = \begin{bmatrix} 1^2 + 1^2 + \cdots + 1^2 & x^{(1)} + x^{(2)} + \cdots + x^{(n)} \\ x^{(1)} + x^{(2)} + \cdots + x^{(n)} & {x^{(1)}}^2 + {x^{(2)}}^2 + \cdots + {x^{(n)}}^2 \end{bmatrix}$$

Regresión lineal

5. Optimización mediante Normal Equation

$$\mathbb{X}^T\mathbb{X} = \begin{bmatrix} \sum_{i=1}^{n} 1^2 & \sum_{i=1}^{n} x^{(i)} \\ \sum_{i=1}^{n} x^{(i)} & \sum_{i=1}^{n} {x^{(i)}}^2 \end{bmatrix}$$

$$\mathbb{X}^T\mathbb{X}w^T = \begin{bmatrix} \sum_{i=1}^{n} 1^2 & \sum_{i=1}^{n} x^{(i)} \\ \sum_{i=1}^{n} x^{(i)} & \sum_{i=1}^{n} {x^{(i)}}^2 \end{bmatrix} \begin{bmatrix} w_0 \\ w_1 \end{bmatrix}$$

Regresión lineal

5. Optimización mediante Normal Equation

$$\mathbb{X}^T\mathbb{X}w^T = \begin{bmatrix} w_0 \sum_{i=1}^{n} 1^2 + w_1 \sum_{i=1}^{n} x^{(i)} \\ w_0 \sum_{i=1}^{n} x^{(i)} + w_1 \sum_{i=1}^{n} {x^{(i)}}^2 \end{bmatrix}$$

$$w\mathbb{X}^T\mathbb{X}w^T = [w_0 \quad w_1]\mathbb{X}^T\mathbb{X}w^T = w_0 w_0 \sum_{i=1}^{n} 1^2 +$$

$$w_0 w_1 \sum_{i=1}^{n} x^{(i)} + w_1 w_0 \sum_{i=1}^{n} x^{(i)} + w_1 w_1 \sum_{i=1}^{n} {x^{(i)}}^2$$

Regresión lineal

5. Optimización mediante Normal Equation

$$w\mathbb{X}^T\mathbb{X}w^T = w_0^2\sum_{i=1}^{n}1^2 + 2w_0w_1\sum_{i=1}^{n}x^{(i)} + w_1^2\sum_{i=1}^{n}{x^{(i)}}^2$$

$$w\mathbb{X}^T y = [w_0 \quad w_1]\begin{bmatrix} 1 & 1 & \cdots & 1 \\ x^{(1)} & x^{(2)} & \cdots & x^{(n)} \end{bmatrix} y$$

$$w\mathbb{X}^T y = [w_0 + w_1x^{(1)} \quad w_0 + w_1x^{(2)} \quad \cdots \quad w_0 + w_1x^{(n)}]y$$

$$w\mathbb{X}^T y = \left(w_0 + w_1x^{(1)}\right)y^{(1)} + \left(w_0 + w_1x^{(1)}\right)y^{(2)} + \cdots + \left(w_0 + w_1x^{(1)}\right)y^{(n)}$$

Regresión lineal

5. Optimización mediante Normal Equation

$$w\mathbb{X}^T y = \sum_{i=1}^{n} \left(w_0 + w_1 x^{(i)}\right) y^{(i)}$$

$$y^T y = \sum_{i=1}^{n} {y^{(i)}}^2$$

$$nJ = w_0^2 \sum_{i=1}^{n} 1^2 + 2w_0 w_1 \sum_{i=1}^{n} x^{(i)} + w_1^2 \sum_{i=1}^{n} {x^{(i)}}^2 - 2\sum_{i=1}^{n} \left(w_0 + w_1 x^{(i)}\right) y^{(i)} + \sum_{i=1}^{n} {y^{(i)}}^2$$

Regresión lineal

5. Optimización mediante Normal Equation

$$\frac{\partial J}{\partial w} = \begin{bmatrix} \frac{\partial J}{\partial w_0} \\ \frac{\partial J}{\partial w_1} \end{bmatrix}$$

$$n\left(\frac{\partial J}{\partial w}\right) = \begin{bmatrix} 2w_0 \sum_{i=1}^{n} 1^2 + 2w_1 \sum_{i=1}^{n} x^{(i)} - 2 \sum_{i=1}^{n} y^{(i)} \\ 2\mathrm{w}_1 \sum_{i=1}^{n} {x^{(i)}}^2 + 2w_0 \sum_{i=1}^{n} x^{(i)} - 2 \sum_{i=1}^{n} x^{(i)} y^{(i)} \end{bmatrix}$$

Regresión lineal

5. Optimización mediante Normal Equation

$$\boxed{\frac{\partial J}{\partial w_0} = \frac{2}{n}\sum_{i=1}^{n}\left(w_0 + w_1 x^{(i)} - y^{(i)}\right)}$$

$$\frac{\partial J}{\partial w_1} = \frac{2\mathrm{w}_1}{n}\sum_{i=1}^{n} {x^{(i)}}^2 + \frac{2w_0}{n}\sum_{i=1}^{n} x^{(i)} - \frac{2}{n}\sum_{i=1}^{n} x^{(i)} y^{(i)}$$

$$\boxed{\frac{\partial J}{\partial w_1} = \frac{2}{n}\sum_{i=1}^{n}\left(w_0 + w_1 x^{(i)} - y^{(i)}\right)x^{(i)}}$$

Regresión lineal

5. Optimización mediante Normal Equation

$$\left(\frac{\partial J}{\partial w}\right) = \begin{bmatrix} \dfrac{\partial J}{\partial w_0} \\ \dfrac{\partial J}{\partial w_1} \end{bmatrix}$$

$$n\left(\frac{\partial J}{\partial w}\right) = 2\begin{bmatrix} w_0 \sum_{i=1}^{n} 1^2 + w_1 \sum_{i=1}^{n} x^{(i)} \\ w_1 \sum_{i=1}^{n} {x^{(i)}}^2 + w_0 \sum_{i=1}^{n} x^{(i)} \end{bmatrix} - 2\begin{bmatrix} \sum_{i=1}^{n} y^{(i)} \\ \sum_{i=1}^{n} x^{(i)} y^{(i)} \end{bmatrix}$$

Regresión lineal

5. Optimización mediante Normal Equation

$$\mathbb{X}^T y = \begin{bmatrix} 1 & 1 & \cdots & 1 \\ x^{(1)} & x^{(2)} & \cdots & x^{(n)} \end{bmatrix} \begin{bmatrix} y^{(1)} \\ y^{(2)} \\ \vdots \\ y^{(n)} \end{bmatrix} = \begin{bmatrix} \sum_{i=1}^{n} y^{(i)} \\ \sum_{i=1}^{n} x^{(i)} y^{(i)} \end{bmatrix}$$

$$n\left(\frac{\partial J}{\partial w}\right) = 2\mathbb{X}^T \mathbb{X} w^T - 2\mathbb{X}^T y$$

$$\nabla J = \frac{\partial J}{\partial w} = \frac{2}{n}\left(\mathbb{X}^T \mathbb{X} w^T - \mathbb{X}^T y\right) = 0$$

Regresión lineal

5. Optimización mediante Normal Equation

$$\nabla J = \frac{2}{n}(\mathbb{X}^T\mathbb{X}w^T - \mathbb{X}^T y)^T$$

$$\nabla J = \frac{2}{n}[(\mathbb{X}^T\mathbb{X}w^T)^T - (\mathbb{X}^T y)^T]$$

$$\nabla J = \frac{2}{n}[w(\mathbb{X}^T\mathbb{X})^T - y^T\mathbb{X}]$$

$$\nabla J = \frac{2}{n}(w\mathbb{X}^T\mathbb{X} - y^T\mathbb{X}) = 0$$

$$w^*\mathbb{X}^T\mathbb{X} = y^T\mathbb{X}$$

$$\boxed{w^* = y^T\mathbb{X}(\mathbb{X}^T\mathbb{X})^{-1}}$$

Regresión lineal

5. Optimización mediante Normal Equation

$$H \equiv \nabla^2 J = \begin{bmatrix} \dfrac{\partial^2 J}{\partial w_0^2} & \dfrac{\partial^2 J}{\partial w_0 \partial w_1} \\ \dfrac{\partial^2 J}{\partial w_1 \partial w_0} & \dfrac{\partial^2 J}{\partial w_1^2} \end{bmatrix}$$

Condición de convexidad:

El Hessiano es una matriz semidefinida positiva

$$z = \begin{bmatrix} z_0 \\ z_1 \end{bmatrix} \in \mathbb{R}^2 \qquad z^T H z \geq 0$$

https://wiki.math.ntnu.no/_media/tma4180/2016v/note2.pdf

Regresión lineal

5. Optimización mediante Normal Equation

$$\frac{\partial J}{\partial w_0} = \frac{2}{n}\sum_{i=1}^{n}\left(w_0 + w_1 x^{(i)} - y^{(i)}\right)$$

$$\frac{\partial^2 J}{\partial w_0^2} = \frac{2}{n}\sum_{i=1}^{n} 1 = 2$$

$$\frac{\partial^2 J}{\partial w_1 \partial w_0} = \frac{2}{n}\sum_{i=1}^{n} x^{(i)}$$

Regresión lineal

5. Optimización mediante Normal Equation

$$\frac{\partial J}{\partial w_1} = \frac{2}{n}\sum_{i=1}^{n}\left(w_0 + w_1 x^{(i)} - y^{(i)}\right)x^{(i)}$$

$$\frac{\partial^2 J}{\partial w_0 \partial w_1} = \frac{2}{n}\sum_{i=1}^{n} x^{(i)}$$

$$\frac{\partial^2 J}{\partial w_1^2} = \frac{2}{n}\sum_{i=1}^{n} {x^{(i)}}^2$$

Regresión lineal

5. Optimización mediante Normal Equation

$$H \equiv \nabla^2 J = \frac{2}{n} \begin{bmatrix} n & \sum_{i=1}^{n} x^{(i)} \\ \sum_{i=1}^{n} x^{(i)} & \sum_{i=1}^{n} {x^{(i)}}^2 \end{bmatrix} = \frac{2}{n} \mathbb{X}^T \mathbb{X}$$

Condición de convexidad:

El Hessiano es una matriz semidefinida positiva

$$z = \begin{bmatrix} z_0 \\ z_1 \end{bmatrix} \in \mathbb{R}^2 \qquad z^T H z \geq 0$$

Regresión lineal

5. Optimización mediante Normal Equation

$$z^T H z = [z_0 \quad z_1] \frac{2}{n} \begin{bmatrix} n & \sum_{i=1}^{n} x^{(i)} \\ \sum_{i=1}^{n} x^{(i)} & \sum_{i=1}^{n} {x^{(i)}}^2 \end{bmatrix} \begin{bmatrix} z_0 \\ z_1 \end{bmatrix}$$

$$z^T H z = \frac{2}{n} \left[z_0 n + z_1 \sum_{i=1}^{n} x^{(i)} \quad z_0 \sum_{i=1}^{n} x^{(i)} + z_1 \sum_{i=1}^{n} {x^{(i)}}^2 \right] \begin{bmatrix} z_0 \\ z_1 \end{bmatrix}$$

$$z^T H z = \frac{2}{n} \left(z_0^2 n + z_0 z_1 \sum_{i=1}^{n} x^{(i)} + z_1 z_0 \sum_{i=1}^{n} x^{(i)} + z_1^2 \sum_{i=1}^{n} {x^{(i)}}^2 \right)$$

Regresión lineal

5. Optimización mediante Normal Equation

$$z^T H z = \frac{2}{n}\left(z_0^2 \sum_{i=1}^{n} 1 + 2 z_0 z_1 \sum_{i=1}^{n} x^{(i)} + z_1^2 \sum_{i=1}^{n} {x^{(i)}}^2 \right)$$

$$z^T H z = \frac{2}{n}\left(\sum_{i=1}^{n} z_0^2 + \sum_{i=1}^{n} 2 z_0 z_1 x^{(i)} + \sum_{i=1}^{n} z_1^2 {x^{(i)}}^2 \right)$$

$$z^T H z = \frac{2}{n} \sum_{i=1}^{n} \left(z_0^2 + 2 z_0 z_1 x^{(i)} + z_1^2 {x^{(i)}}^2 \right)$$

Regresión lineal

5. Optimización mediante Normal Equation

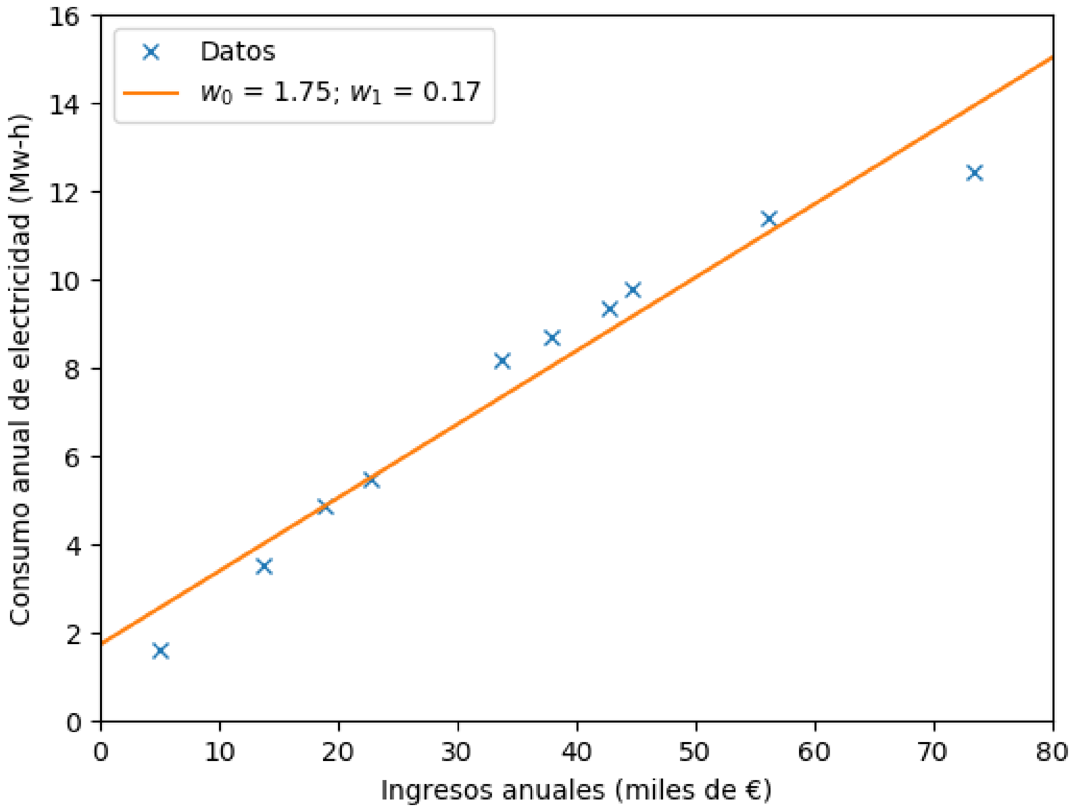
$$z^T H z = \frac{2}{n}\sum_{i=1}^{n}\left(z_0 + z_1 x^{(i)}\right)^2$$

$$\left(z_0 + z_1 x^{(i)}\right)^2 \geq 0, \forall z, x^{(i)} \in \mathbb{R}^2$$

$$\boxed{z^T H z \geq 0}$$

Regresión lineal

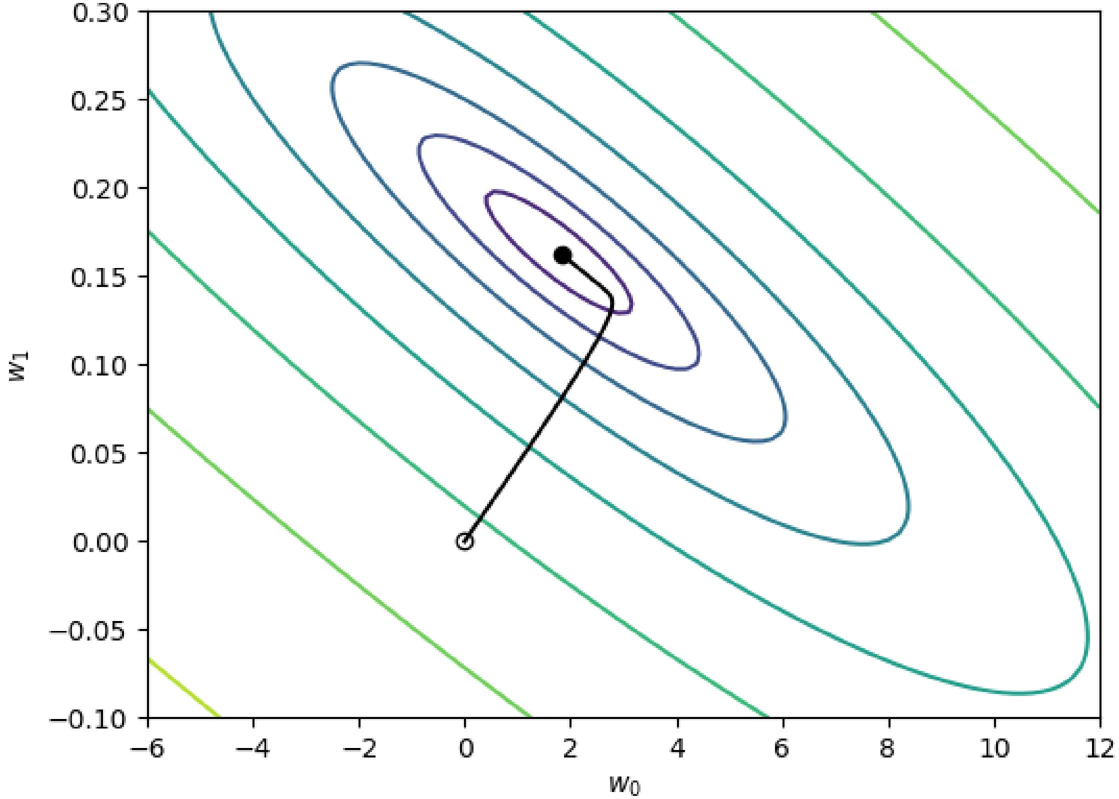
5. Optimización mediante Normal Equation



$$w^* = y^T \mathbb{X}(\mathbb{X}^T \mathbb{X})^{-1} = [1.75 \quad 0.17]$$

Regresión lineal

5. Optimización mediante Gradient Descent



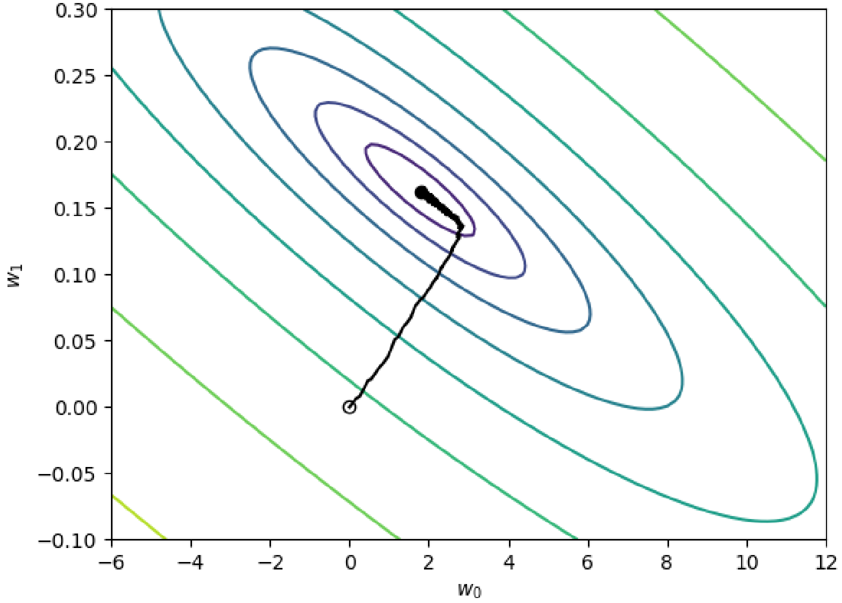
$$w_0^{(t+1)} \leftarrow w_0^{(t)} - \alpha \frac{\partial J(w)}{dw_0}$$

$$w_1^{(t+1)} \leftarrow w_1^{(t)} - \alpha \frac{\partial J(w)}{dw_1}$$

Regresión lineal

5. Optimización mediante Gradient Descent

SGD



BGD