Contenido

Regresión multivariable

- Representación gráfica multivariable
- Codificación 1-hot
- Normalización de variables
- Descomposición en Valores Singulares
  - Matrices de covarianza y Gramm
  - Autovalores y autovectores
  - Definición de la SVD
  - Obtención de la SVD
  - Aplicación a la solución de la Normal Equation

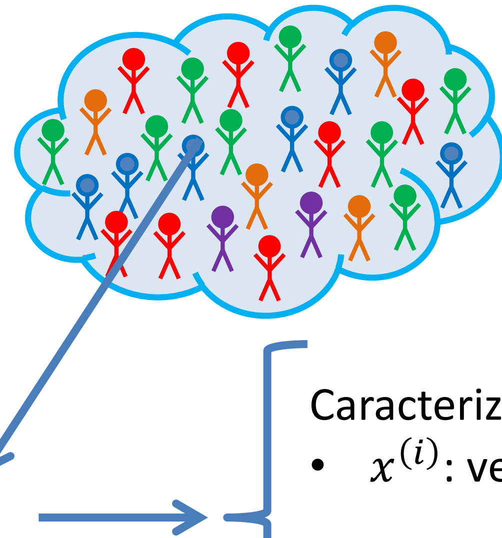
Regresión multivariable
Ejemplo en el sector eléctrico

- En otra zona en la que opera la empresa los datos disponibles de los clientes para hacer la predicción de su consumo de energía son los siguientes:
  1. Ingresos anuales (€)
  2. Número de habitantes en la vivienda (promedio anual de personas)
  3. Potencia contratada (kW)
  4. Tipo de vivienda (piso, adosado, casa)
  5. Superficie de la vivienda (m^2)
  6. Horas de ocupación en invierno (horas/día)
  7. Horas de ocupación en verano (horas/día)
  8. Edad (años)
  9. Estado civil (soltero, casado, separado, viudo)
  10. Nivel de estudios alcanzado (sin estudios, primaria, secundaria, bachiller, grado, máster, doctor)

Regresión multivariable

2. Determinación de *features*

Muestra $\mathcal{M}$
(dataset)



n elementos

i-ésimo elemento

Caracterizado por d features (rasgos):

- $x^{(i)}$: vector fila con d features

Valor del objetivo (target) :

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

$$(n \times d) \quad X = \begin{bmatrix} x_1^{(1)} & x_2^{(1)} & \cdots & x_d^{(1)} \\ x_1^{(2)} & x_2^{(2)} & \cdots & x_d^{(2)} \\ \vdots & \vdots & \ddots & \vdots \\ x_1^{(n)} & x_2^{(n)} & \cdots & x_d^{(n)} \end{bmatrix} \qquad \begin{matrix} (n \times 1) \\ y^{(i)} \end{matrix} = \begin{bmatrix} y^{(1)} \\ y^{(2)} \\ \vdots \\ y^{(n)} \end{bmatrix}$$

Regresión multivariable

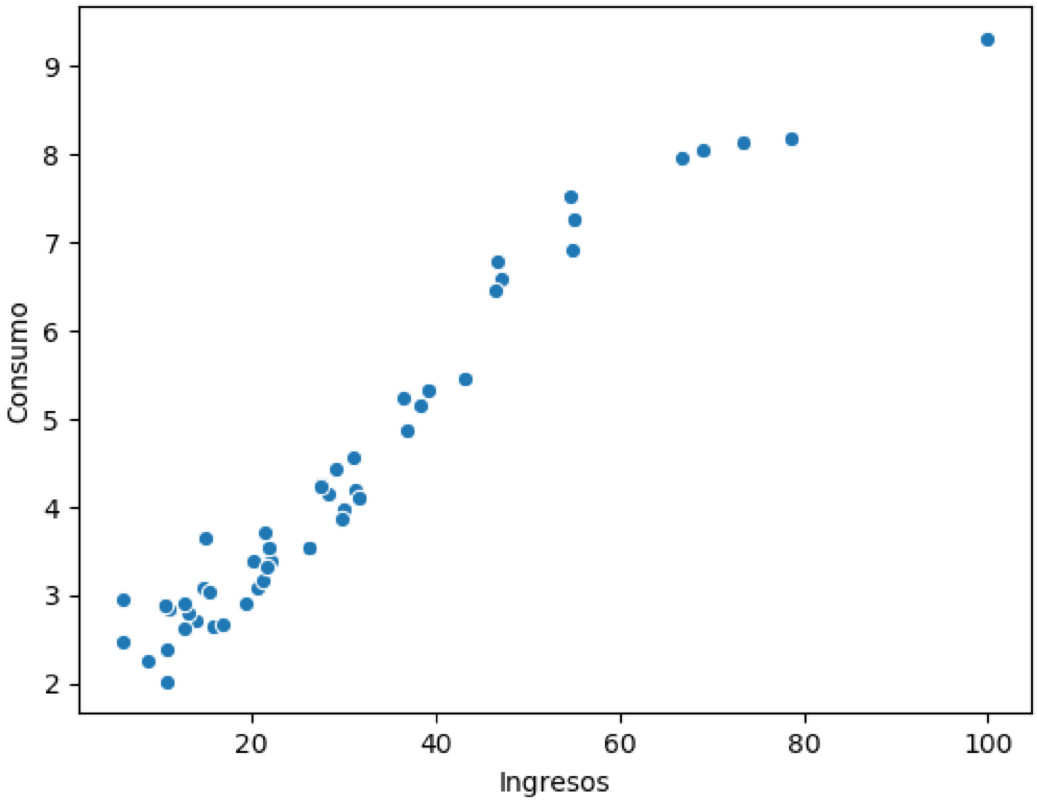
2. Determinación de *features*

	Ingresos (k€)	Habitantes	Potencia (kW)	Tipo Vivienda	Superficie (m^2)	Horas invierno	Horas verano	Edad	Estado civil	Estudios	Consumo (MW-h)
i	$x_1^{(i)}$	$x_2^{(i)}$	$x_3^{(i)}$	$x_4^{(i)}$	$x_5^{(i)}$	$x_6^{(i)}$	$x_7^{(i)}$	$x_8^{(i)}$	$x_9^{(i)}$	$x_{10}^{(i)}$	$y^{(i)}$
1	68.4	2	7.5	Ad.	136	6.5	4.1	26.3	Sep.	Mas.	9.81
2	38.1	3	4.8	Piso	101	17.1	6.5	23.1	Cas.	Bac.	12.4
3	39.2	3	4.6	Casa	82	15.6	7.1	75.3	Viud.	Bac.	1.63
4	31.8	4	3.9	Piso	72	16.7	7.2	21.7	Sol.	Sec.	8.17
5	58.1	2	5.5	Piso	132	13.2	5.2	32.4	Cas.	Sec.	4.87
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮

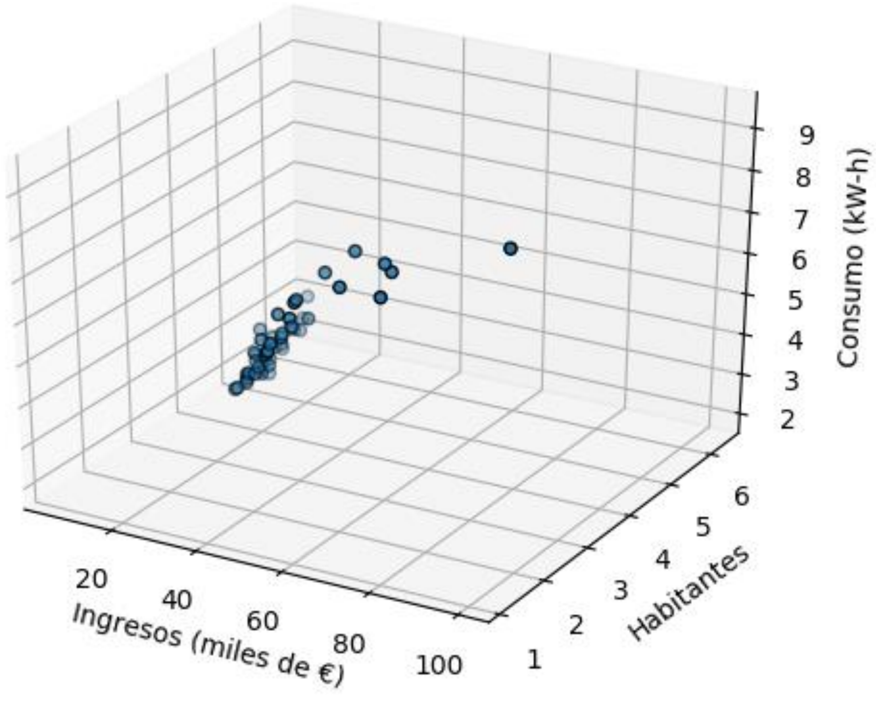
Regresión multivariable

- Representación gráfica multivariable
- Codificación 1-hot
- Normalización de variables
- Descomposición en Valores Singulares
  - Matrices de covarianza y Gramm
  - Autovalores y autovectores
  - Definición de la SVD
  - Obtención de la SVD
  - Aplicación a la solución de la Normal Equation

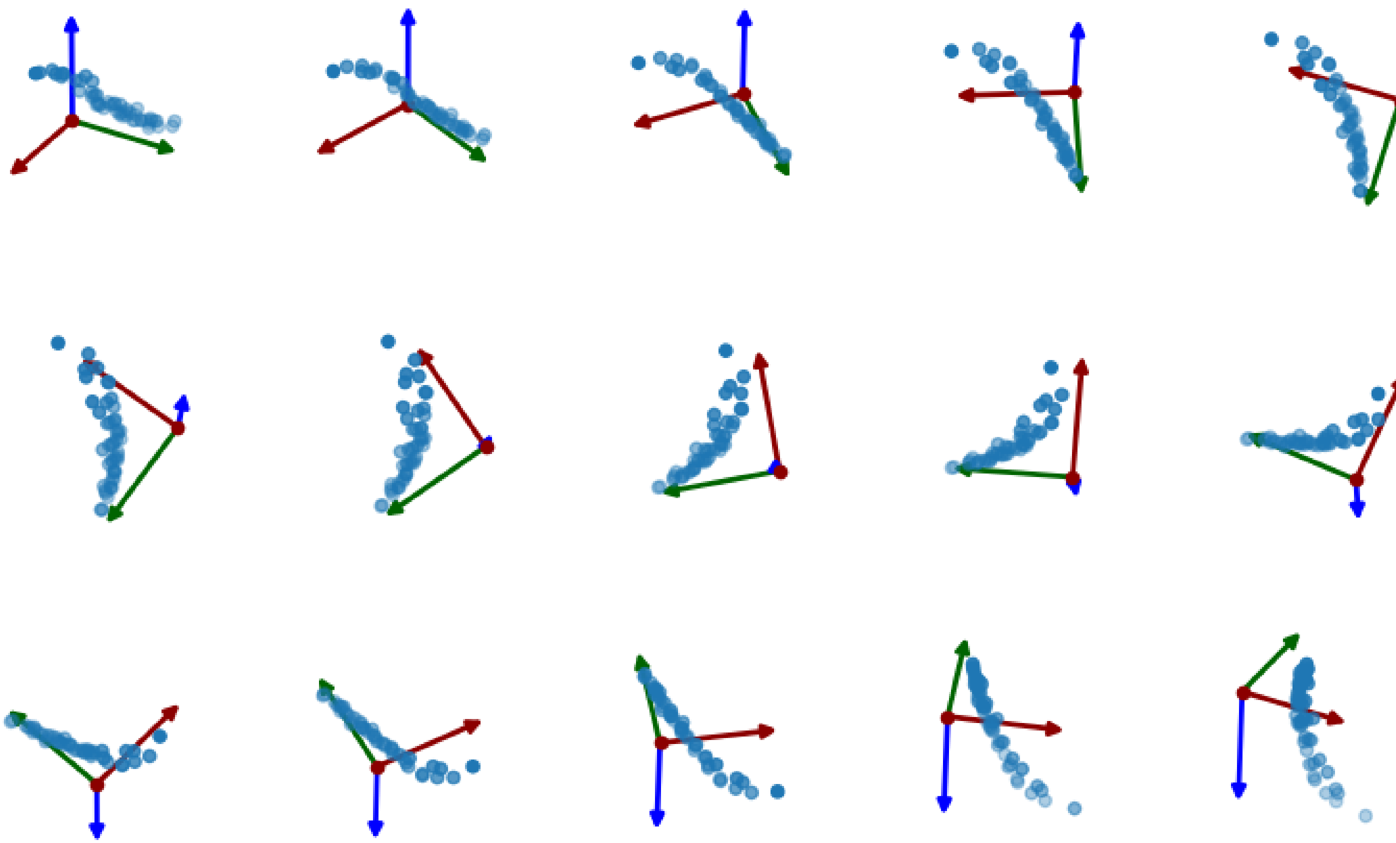
Representación multivariable



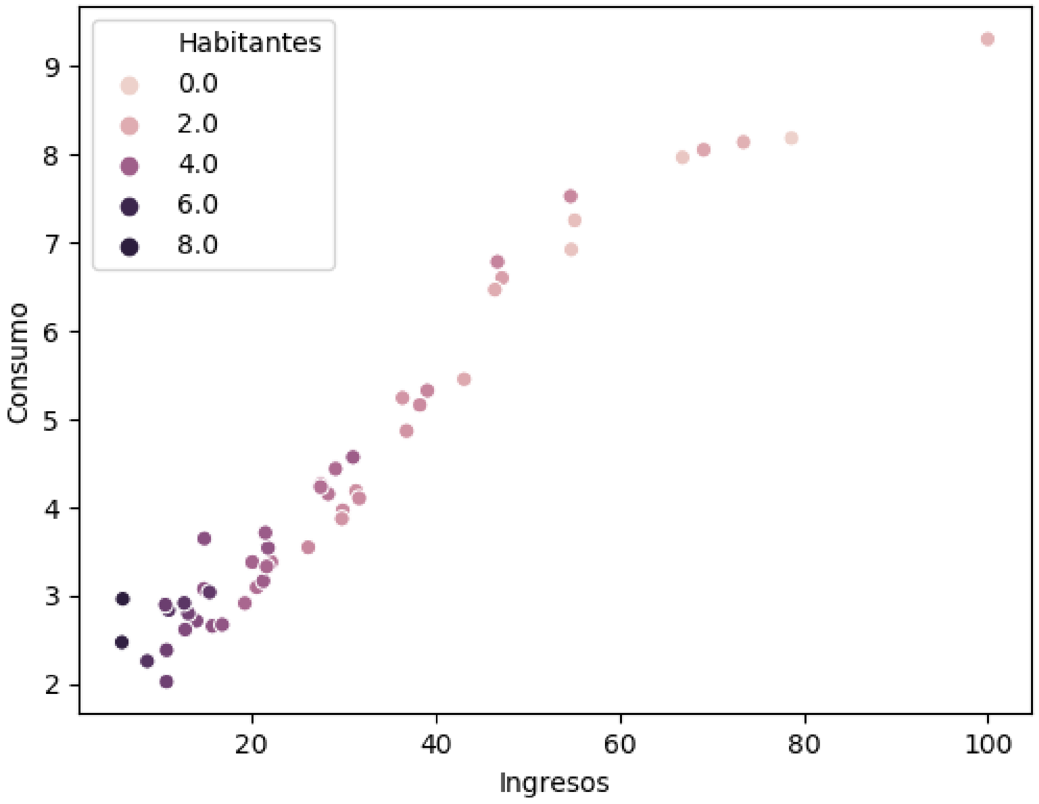
Representación multivariable



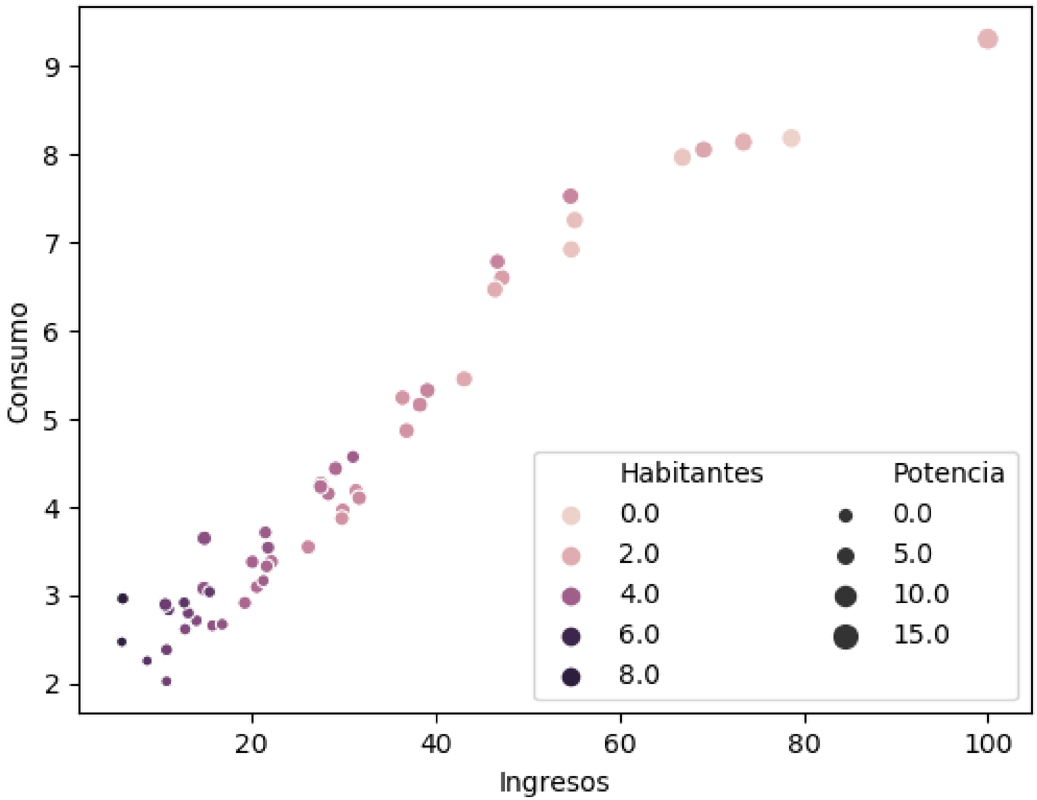
Representación multivariable



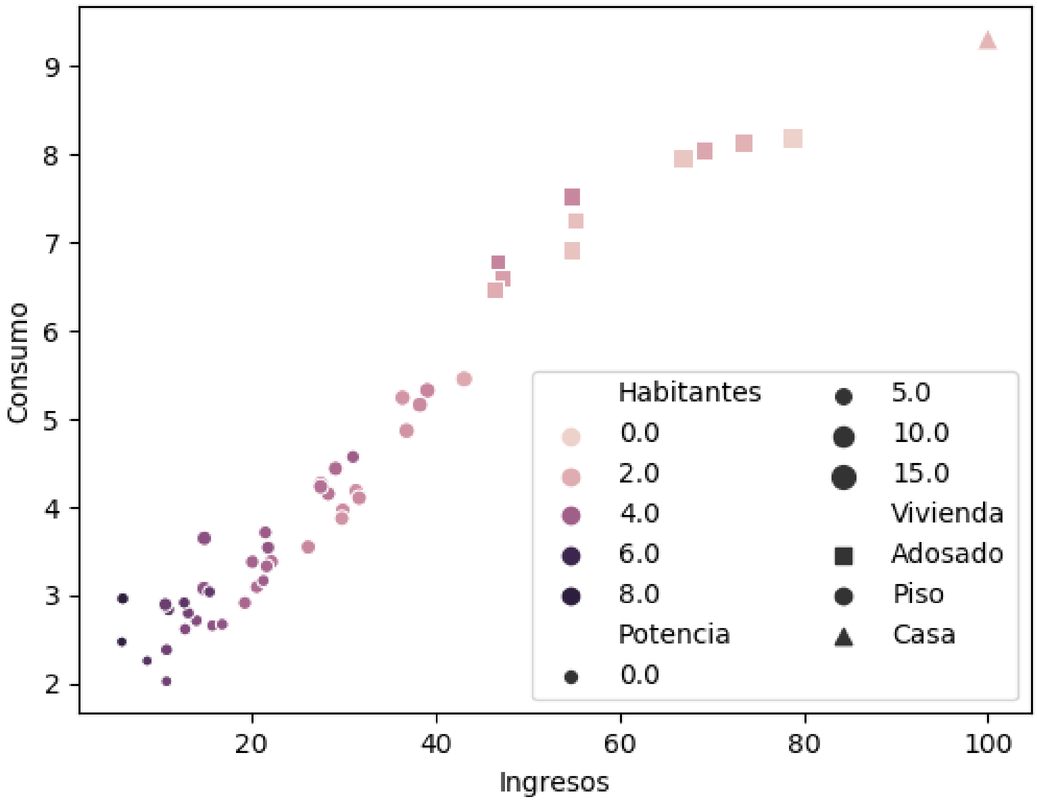
Representación multivariable



Representación multivariable



Representación multivariable



Regresión multivariable

- Representación gráfica multivariable
- Codificación 1-hot
- Normalización de variables
- Descomposición en Valores Singulares
  - Matrices de covarianza y Gramm
  - Autovalores y autovectores
  - Definición de la SVD
  - Obtención de la SVD
  - Aplicación a la solución de la Normal Equation

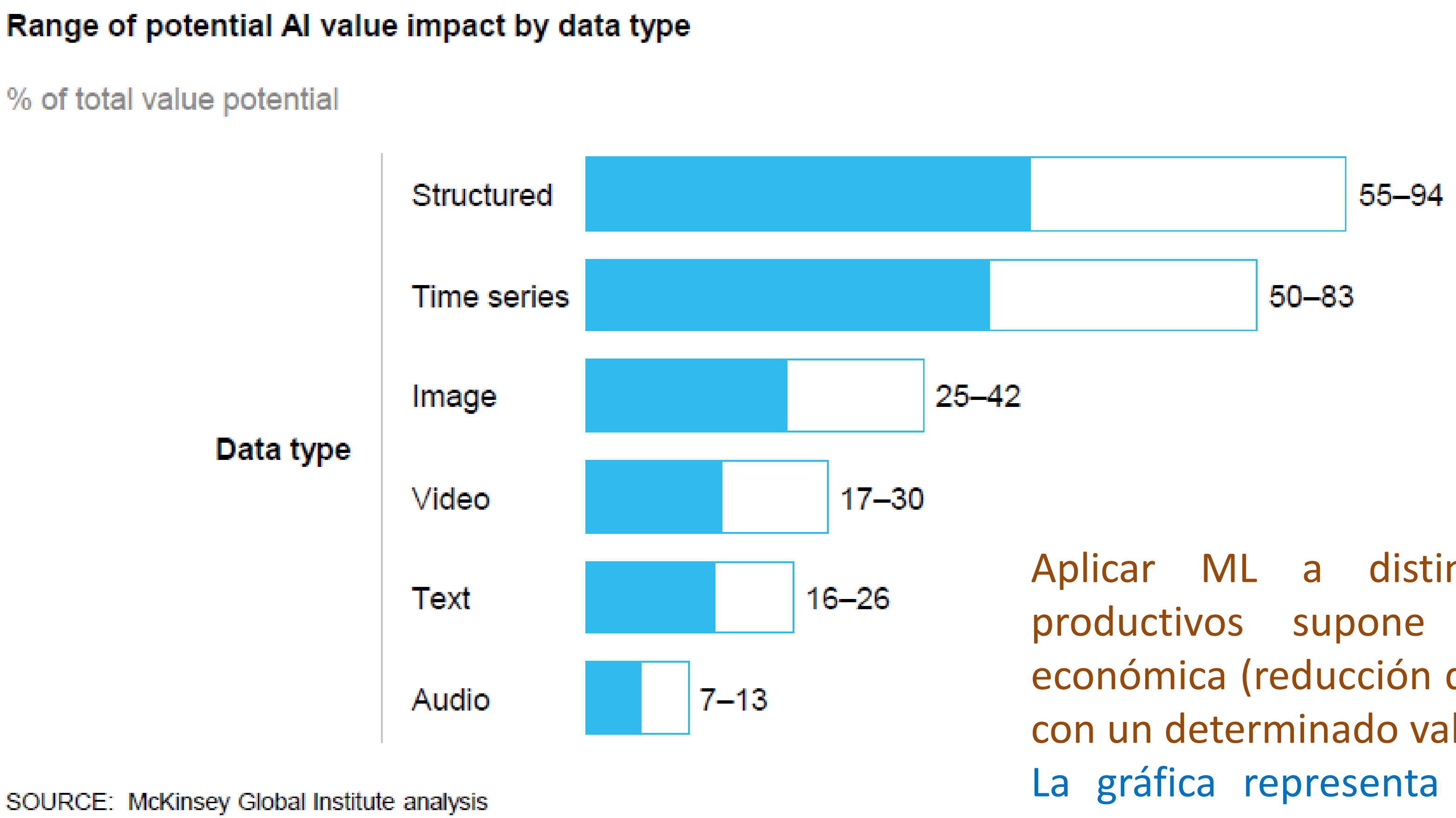
Regresión multivariable

2. Determinación de *features*

- Estructurados
  - Numéricos
    - Continuos
    - Discretos
  - Categóricos
    - Ordinales
    - Nominales
- Series temporales
- Imagen
- Vídeo
- Audio
- Texto

Regresión multivariable

2. Determinación de *features*



Aplicar ML a distintos sectores productivos supone una mejora económica (reducción de costes, etc.) con un determinado valor total (€).

La gráfica representa el porcentaje que en dicha mejora global juega cada tipo de datos.

https://www.mckinsey.com/~/media/mckinsey/featured%20insights/artificial%20intelligence/notes%20from%20the%20ai%20frontier%20applications%20and%20value%20of%20deep%20learning/notes-from-the-ai-frontier-insights-from-hundreds-of-use-cases-discussion-paper.ashx

Regresión multivariable
2. Determinación de *features*

- Numérica
  - Continua
    - Ingresos; promedio de habitantes; superficie; horas invierno; horas verano; edad
  - Discreta
    - Potencia
- Categórica
  - Ordinal (con orden; tiene sentido la media)
    - Estudios
  - Nominal (sin orden; no tiene sentido la media)
    - Tipo de vivienda; estado civil
    - 1-hot encoding

1-hot encoding

Tipo de vivienda	Piso	Adosado	Casa
Piso	1	0	0
Adosado	0	1	0
Casa	0	0	1

Estado civil	Soltero	Casado	Separado	Viudo
Soltero	1	0	0	0
Casado	0	1	0	0
Separado	0	0	1	0
Viudo	0	0	0	1

1-hot encoding

Cliente	Ingresos (k€)	Habitantes	Potencia (kW)	Piso	Adosado	Casa	Superficie (m²)	Horas invierno	Horas verano	Edad	Soltero	Casado	Separado	Viudo	Estudios	Consumo (MW-h)
i	$x_1^{(i)}$	$x_2^{(i)}$	$x_3^{(i)}$	$x_4^{(i)}$	$x_5^{(i)}$	$x_6^{(i)}$	$x_7^{(i)}$	$x_8^{(i)}$	$x_9^{(i)}$	$x_{10}^{(i)}$	$x_{11}^{(i)}$	$x_{12}^{(i)}$	$x_{13}^{(i)}$	$x_{14}^{(i)}$	$x_{15}^{(i)}$	$y^{(i)}$
1	68.4	2	7.5	0	1	0	136	6.5	4.1	26.3	0	0	1	0	6	9.81
2	38.1	3	4.8	1	0	0	101	17.1	6.5	23.1	0	1	0	0	4	12.4
3	39.2	3	4.6	0	0	1	82	15.6	7.1	75.3	0	0	0	1	4	1.63
4	31.8	4	3.9	1	0	0	72	16.7	7.2	21.7	1	0	0	0	3	8.17
5	58.1	2	5.5	1	0	0	132	13.2	5.2	32.4	0	1	0	0	3	4.87
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮

Regresión multivariable

3. Formulación de hipótesis

$$h_w(x) = w_1 x_1 + w_2 x_2 + \cdots + w_d x_d$$

$$X = \begin{bmatrix} x_1^{(1)} & x_2^{(1)} & \cdots & x_d^{(1)} \\ x_1^{(2)} & x_2^{(2)} & \cdots & x_d^{(2)} \\ \vdots & \vdots & \ddots & \vdots \\ x_2^{(n)} & x_2^{(n)} & \cdots & x_d^{(n)} \end{bmatrix} \qquad w = \begin{bmatrix} w_1 & w_2 & \cdots & w_d \end{bmatrix}$$

Regresión multivariable

4. Elección de la función de coste

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

$$Xw^T - y = \begin{bmatrix} w_1 x_1^{(1)} + w_2 x_2^{(1)} + \cdots + w_d x_d^{(1)} - y^{(1)} \\ w_1 x_1^{(2)} + w_2 x_2^{(2)} + \cdots + w_d x_d^{(2)} - y^{(2)} \\ \vdots \\ w_1 x_1^{(n)} + w_2 x_2^{(n)} + \cdots + w_d x_d^{(n)} - y^{(n)} \end{bmatrix}$$

Regresión multivariable

4. Elección de la función de coste

$$(X - y)^T (Xw^T - y) =$$
$$\sum_{i=1}^{n} \left(w_1 x_1^{(i)} + w_2 x_2^{(i)} + \cdots + w_d x_d^{(i)} - y^{(1)} \right)^2$$

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

$$J(h_w(x), y) = \frac{1}{n} (X - y)^T (Xw^T - y) + \lambda \|w\|_2^2$$

Regresión multivariable

5. Optimización mediante Normal Equation

$$\nabla J(h_w(x), y) = 0$$

$$\nabla \left[\frac{1}{n} (X - y)^T (Xw^T - y) + \lambda \|w\|_2^2 \right] = 0$$

$$\nabla J(h_w(x), y) = \frac{2}{n} (wX^T X - y^T X) + 2\lambda w = 0$$

$$w^* = y^T X (X^T X + n\lambda I_d)^{-1}$$

Regresión multivariable

5. Optimización mediante Gradient Descent

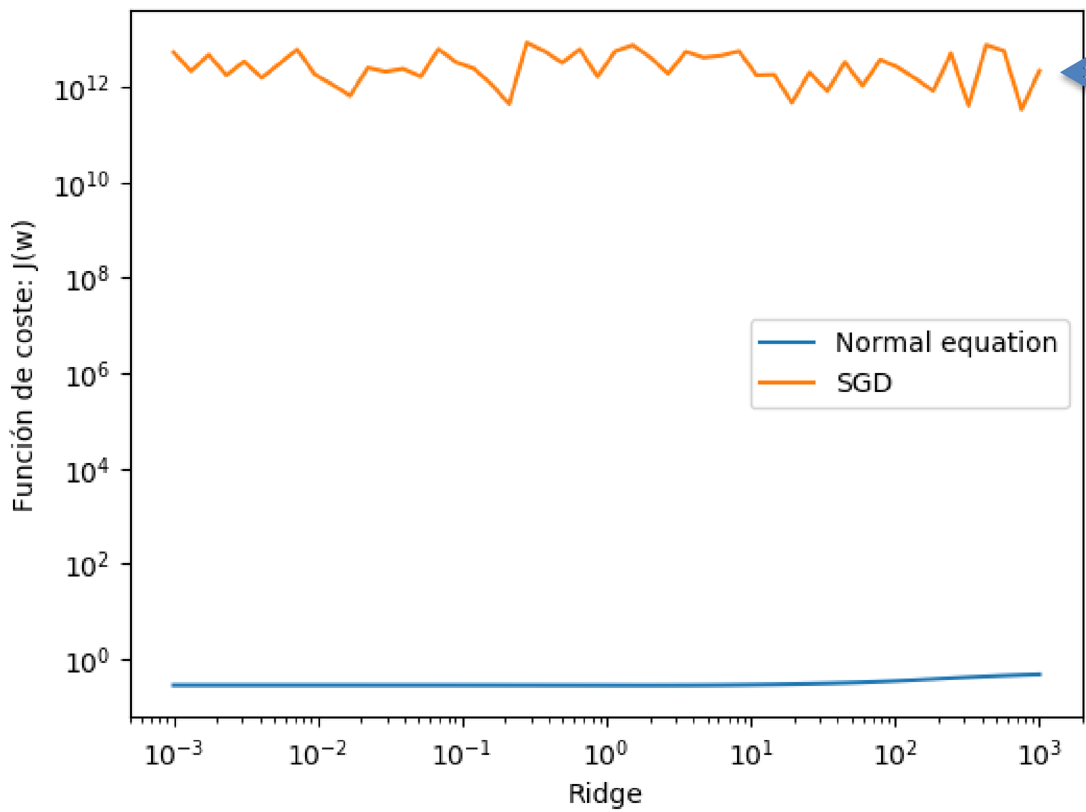
$$\begin{bmatrix} w_1^{(t+1)} \\ \vdots \\ w_d^{(t+1)} \end{bmatrix} \leftarrow \begin{bmatrix} w_1^{(t)} \\ \vdots \\ w_d^{(t)} \end{bmatrix} - \alpha \begin{bmatrix} \dfrac{\partial J(h_w(x), y)}{\partial w_1} \\ \vdots \\ \dfrac{\partial J(h_w(x), y)}{\partial w_d} \end{bmatrix}$$

$$w^{(t+1)^T} \leftarrow w^{(t)^T} - \alpha \nabla J$$

$$w^{(t+1)^T} \leftarrow w^{(t)^T} - \alpha \left(\frac{2}{n} (wX^TX - y^TX) + 2\lambda w \right)$$

Regresión multivariable

6. Evaluación del resultado

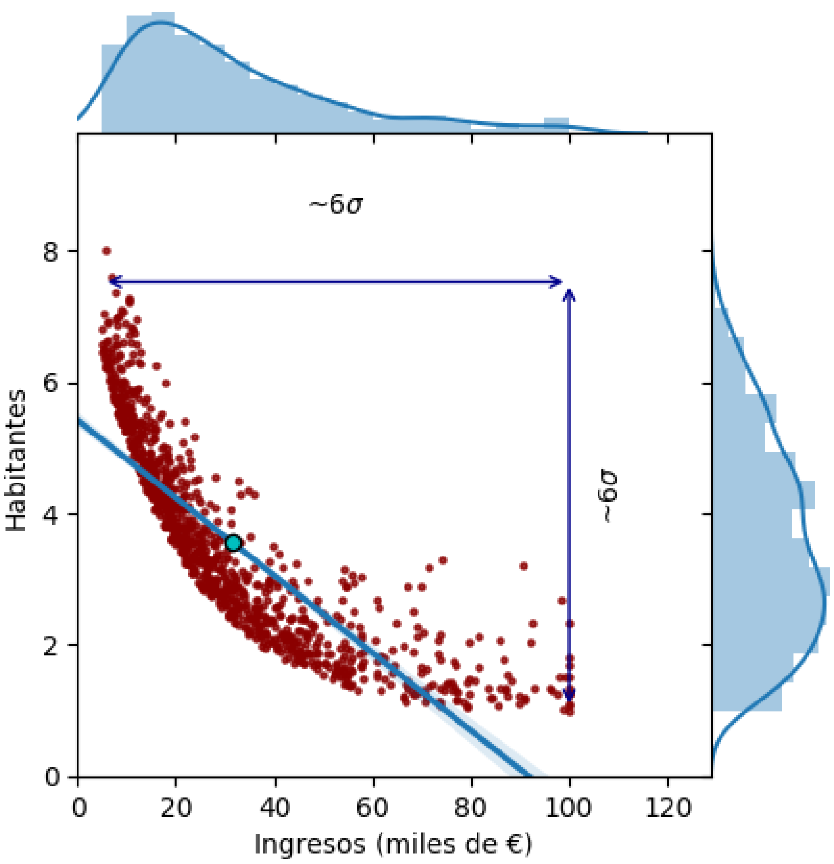


No converge

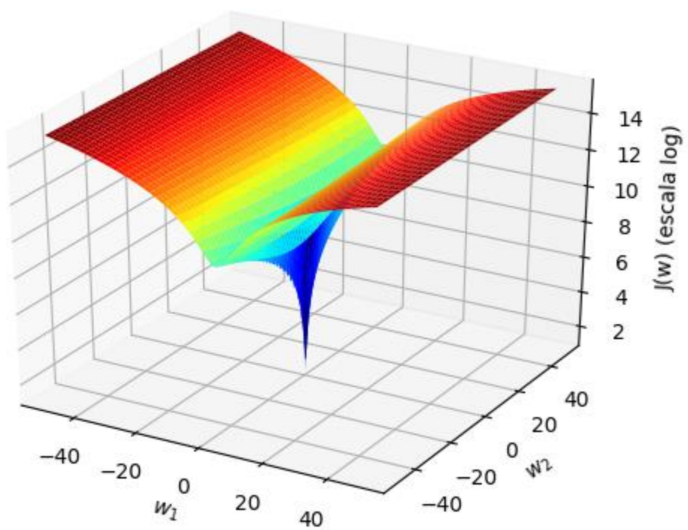
Regresión multivariable

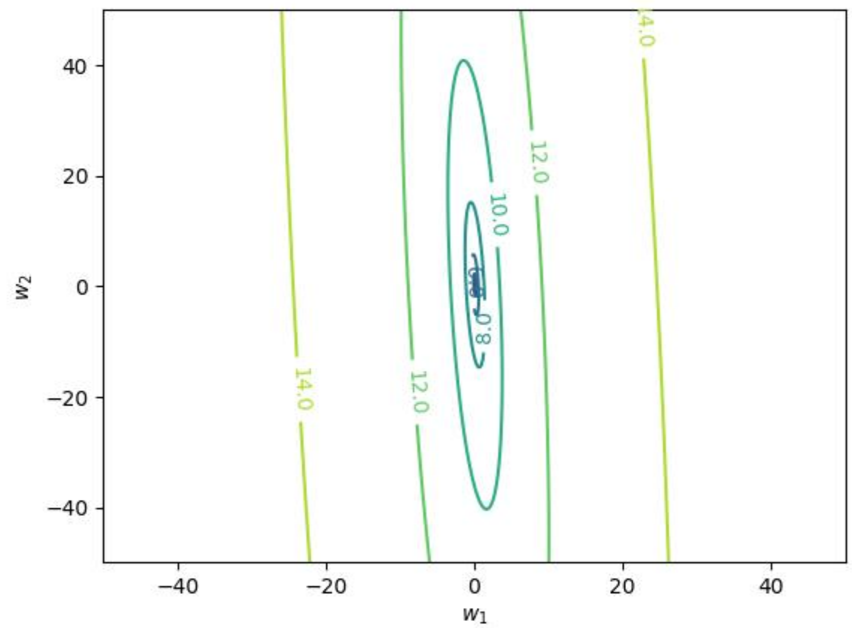
- Representación gráfica multivariable
- Codificación 1-hot
- Normalización de variables
- Descomposición en Valores Singulares
  - Matrices de covarianza y Gramm
  - Autovalores y autovectores
  - Definición de la SVD
  - Obtención de la SVD
  - Aplicación a la solución de la Normal Equation

Normalización de variables



Normalización de variables





Coste escala logarítmica

$\lambda = 0$

Normalización de variables

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

$$J(h_w(x), y) = \frac{1}{n}\sum_{i=1}^{n}\left(w_1 x_1^{(i)} + w_2 x_2^{(i)} - y^{(i)}\right)^2$$

Orden de magnitud: $O_m(\cdot)$

$$\underset{10^3}{O_m(w_1)} = \frac{\overset{10^3}{O_m(y)}}{\underset{10^0}{O_m(x_1)}}; \underset{10^1}{O_m(w_2)} = \frac{\overset{10^3}{O_m(y)}}{\underset{10^2}{O_m(x_2)}}; \underset{10^6}{O_m(J)} = \left(\underset{10^3}{O_m(y)}\right)^2$$

Normalización de variables

$$\frac{\partial J}{\partial w_1} = \frac{1}{n}\sum_{i=1}^{n} \frac{\partial}{\partial w_1}\left(w_1 x_1^{(i)} + w_2 x_2^{(i)} - y^{(i)}\right)^2$$

$$\frac{\partial J}{\partial w_1} = \frac{2}{n}\sum_{i=1}^{n} x_1^{(i)}\left(w_1 x_1^{(i)} + w_2 x_2^{(i)} - y^{(i)}\right)$$

$$\frac{\partial J}{\partial w_2} = \frac{2}{n}\sum_{i=1}^{n} x_2^{(i)}\left(w_1 x_1^{(i)} + w_2 x_2^{(i)} - y^{(i)}\right)$$

Normalización de variables

$$\begin{bmatrix} w_1^{(t+1)} \\ w_2^{(t+1)} \end{bmatrix} \leftarrow \begin{bmatrix} w_1^{(t)} \\ w_2^{(t)} \end{bmatrix} - \alpha \begin{bmatrix} \frac{\partial J(h_w(x), y)}{\partial w_1} \\ \frac{\partial J(h_w(x), y)}{\partial w_2} \end{bmatrix}$$

$$\begin{bmatrix} \Delta w_1^{(t)} \\ \Delta w_2^{(t)} \end{bmatrix} = -\alpha \begin{bmatrix} \frac{\partial J(h_w(x), y)}{\partial w_1} \\ \frac{\partial J(h_w(x), y)}{\partial w_2} \end{bmatrix}$$

Normalización de variables

$$O_m\left(\Delta w_1^{(t)}\right) = O_m(\alpha)O_m\left(\frac{\partial J}{\partial w_1}\right) = O_m(\alpha)O_m(x_1)O_m(y)$$

10^2 10^{-1} 10^3 10^{-1} 10^0 10^3

10^2

$$O_m\left(\frac{\Delta w_1^{(t)}}{w_1}\right) = \frac{O_m(\alpha)O_m(x_1)O_m(y)}{\dfrac{O_m(y)}{O_m(x_1)}} = O_m(\alpha)O_m(x_1)^2$$

10^3 10^{-1} $(10^0)^2$

Normalización de variables

$$\underset{10^4}{O_m\left(\Delta w_2^{(t)}\right)} = \underset{10^{-1}}{O_m(\alpha)}\,\underset{10^5}{O_m\left(\frac{\partial J}{\partial w_2}\right)} = \underset{10^{-1}}{O_m(\alpha)}\,\underset{10^2}{O_m(x_2)}\,\underset{10^3}{O_m(y)}$$

$$O_m\left(\frac{\overset{10^4}{\Delta w_2^{(t)}}}{\underset{10^1}{w_2}}\right) = \frac{O_m(\alpha)O_m(x_2)O_m(y)}{\frac{O_m(y)}{O_m(x_2)}} = \underset{10^{-1}}{O_m(\alpha)}\,\underset{(10^2)^2}{O_m(x_2)^2}$$

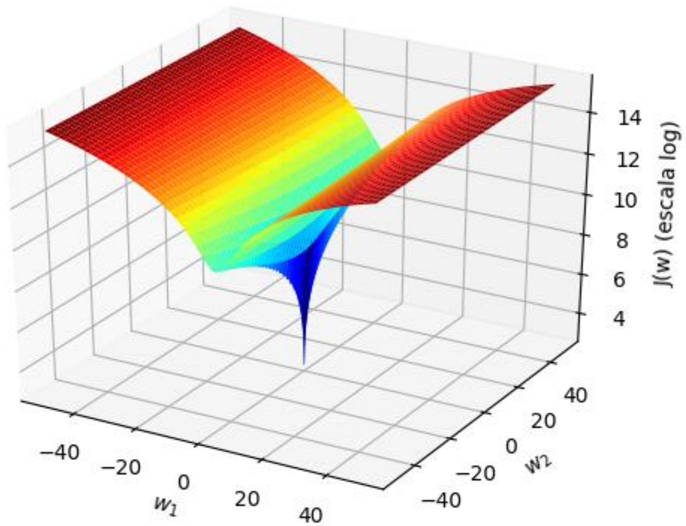
Normalización de variables

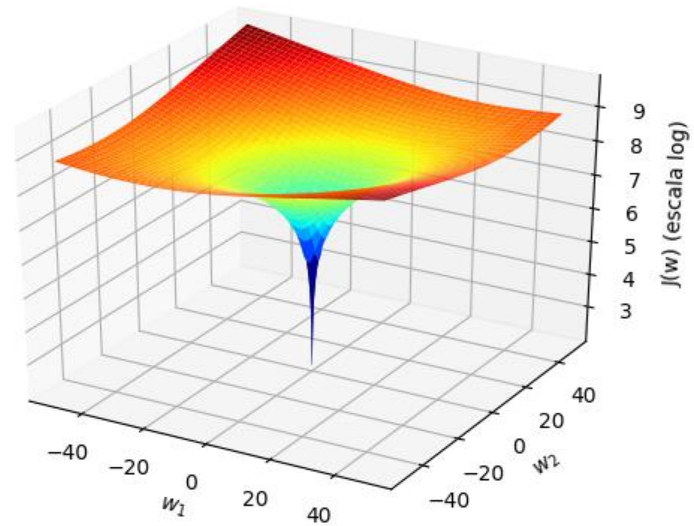
$$\frac{\overset{10^{-1}}{O_m}\left(\frac{\Delta w_1^{(t)}}{w_1}\right)}{\underset{10^{3}}{O_m}\left(\frac{\Delta w_2^{(t)}}{w_2}\right)} = \frac{O_m(\alpha)O_m(x_1)^2}{O_m(\alpha)O_m(x_2)^2}$$

$$\boxed{\frac{O_m\left(\frac{\Delta w_1^{(t)}}{w_1}\right)}{O_m\left(\frac{\Delta w_2^{(t)}}{w_2}\right)} = \left(\frac{\overset{(10^0)^2}{O_m(x_1)}}{\underset{(10^2)^2}{O_m(x_2)}}\right)^2 = 10^{-4}}$$

Normalización de variables

Coste escala logarítmica
$\lambda = 100$



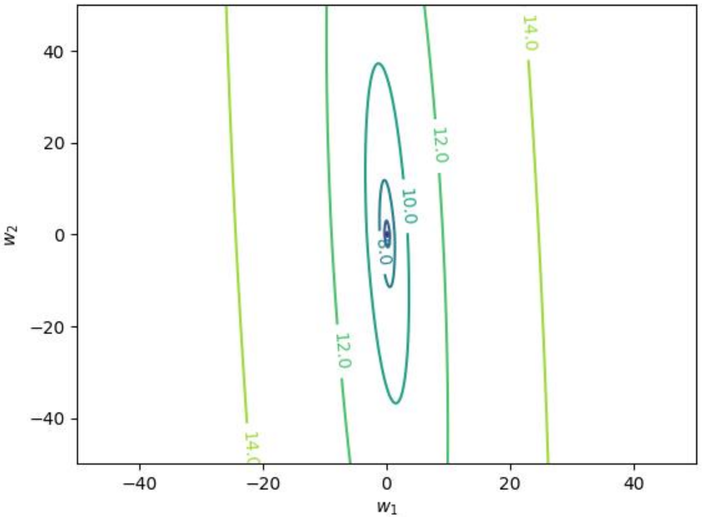


x

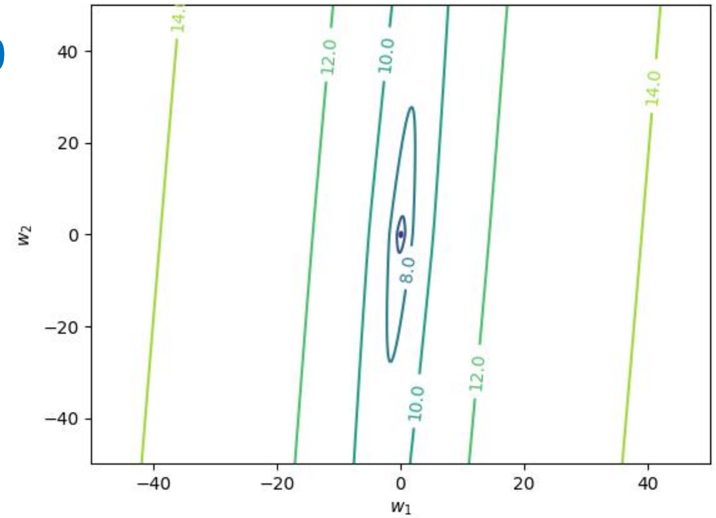
$$\frac{x-\mu}{\sigma}$$

Normalización de variables

$\lambda = 100$

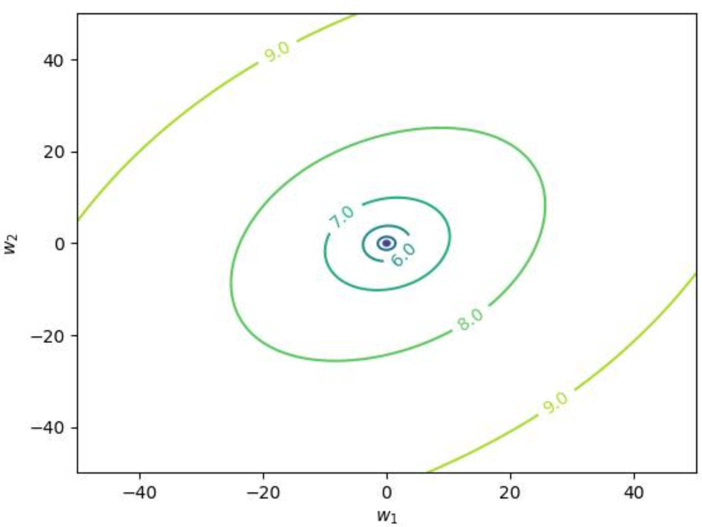


x

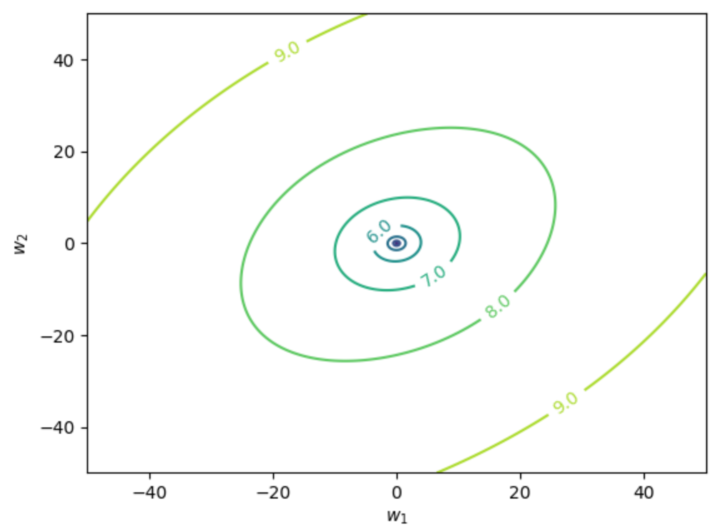


$x - \mu$

Coste escala logarítmica

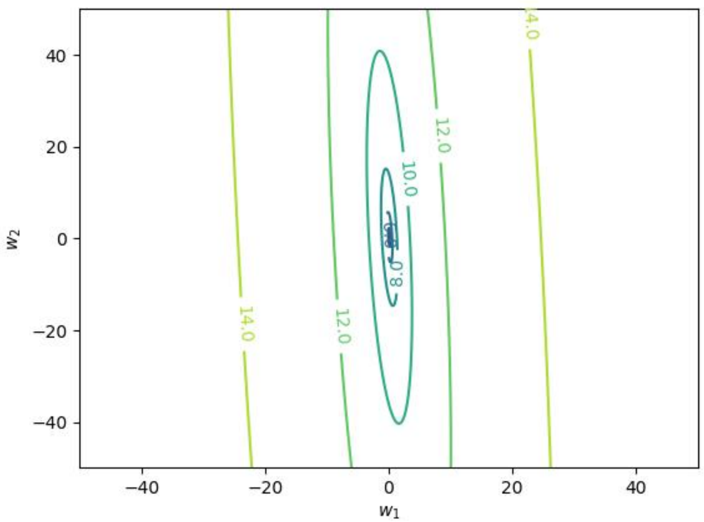


$\frac{x}{\sigma}$



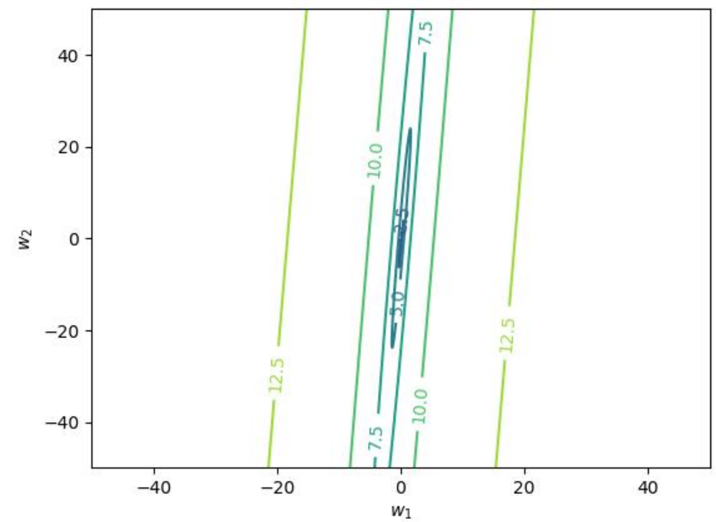
$\frac{x - \mu}{\sigma}$

Normalización de variables



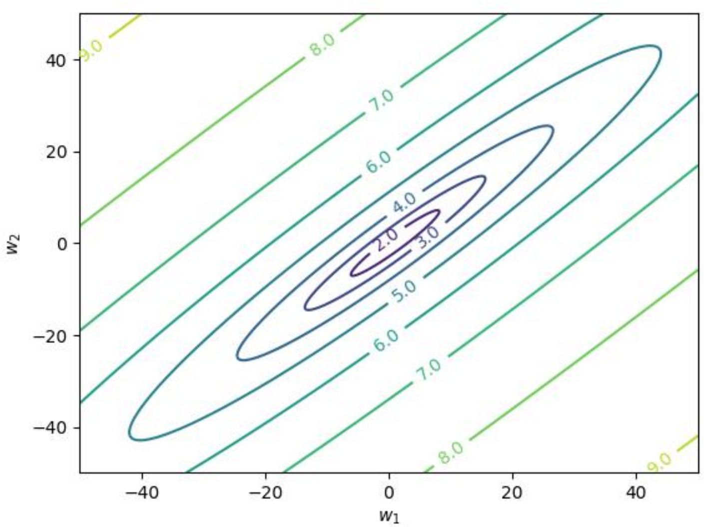
$\lambda = 0$

x

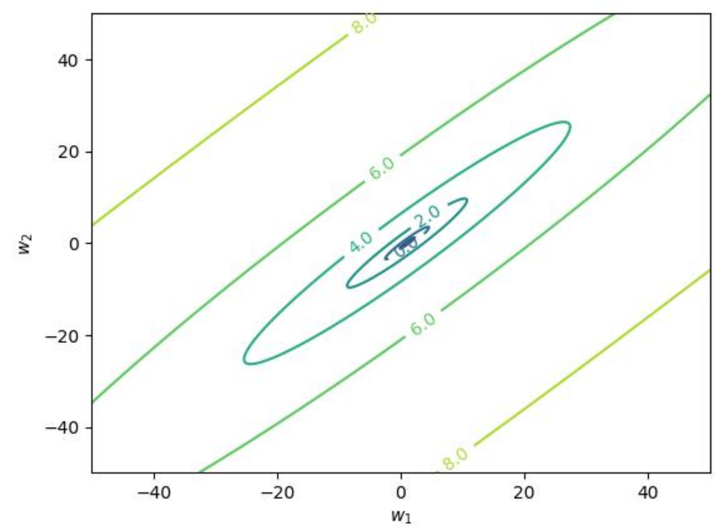


$x - \mu$

Coste escala logarítmica

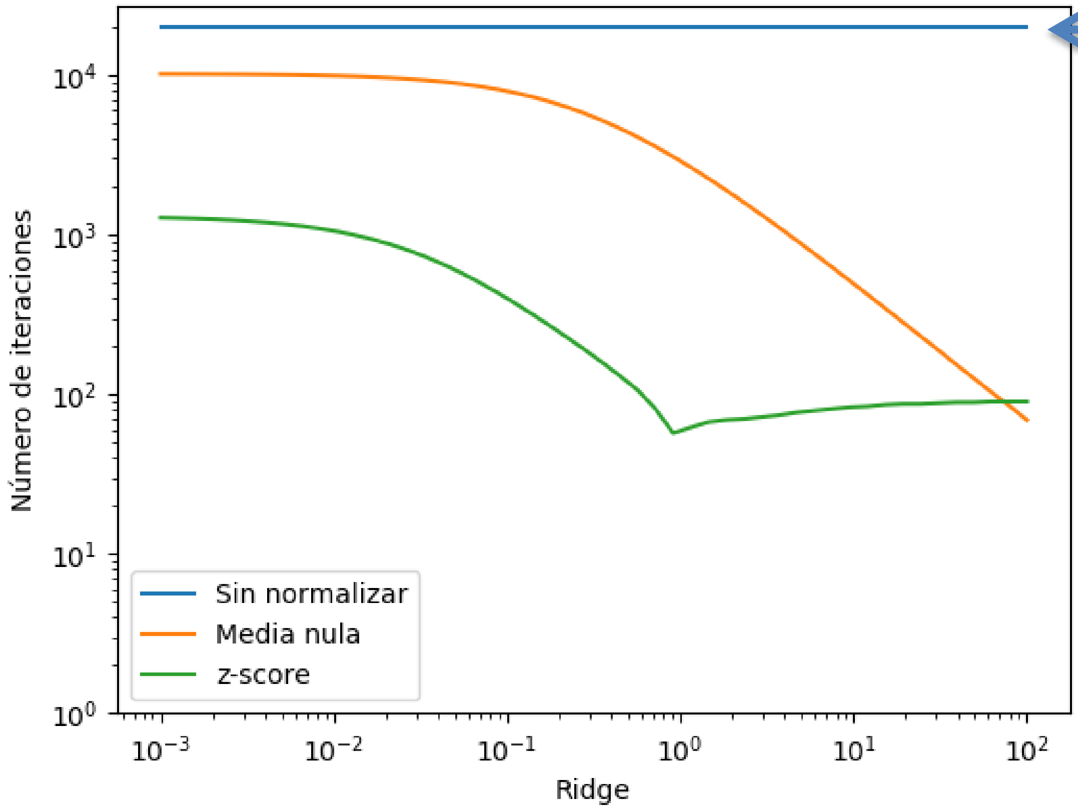


$\frac{x}{\sigma}$



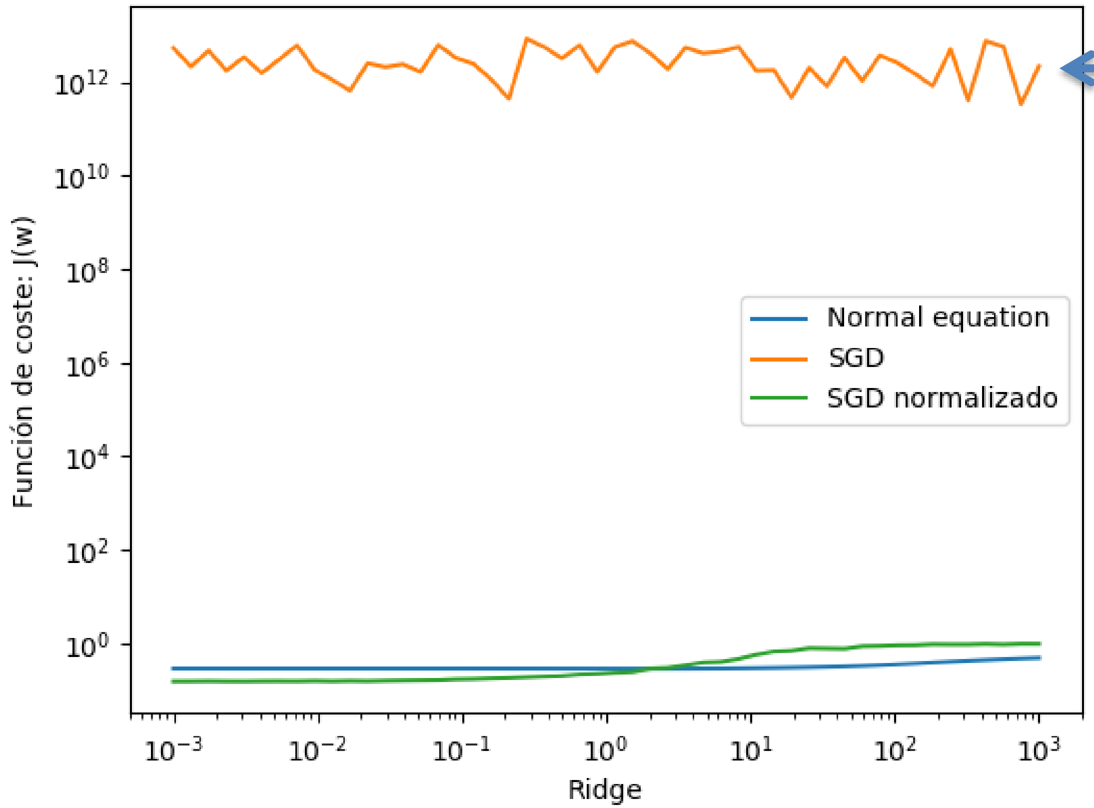
$\frac{x - \mu}{\sigma}$

Normalización de variables



No converge

Normalización de variables



No converge

Normalización de variables

z-score
Media nula
Varianza unitaria

$$x' = \frac{x - \mu_x}{\sigma_x}$$

Media nula

$$x' = \frac{x - \mu_x}{\max(x) - \min(x)}$$

Reescalado
min-max

$$x' = \frac{x - \min(x)}{\max(x) - \min(x)}$$

Reescalado
por rango

$$x' = \frac{x - \mu_x}{rango(x)}$$