UNIVERSIDAD ICESI | A OTRO NIVEL | VIRTUAL

REGRESIÓN LINEAL SIMPLE, MULTIVARIADA, POLINOMIAL Y MÉTRICAS

UNIDAD 1: Regresión Lineal

Agenda

01 **Acercamiento intuitivo**

02 **Descripción matemática**

03 **Aplicación**

04 **Estimación de parámetros**

05 **Lecturas complementarias**

06 **Práctica en Google Colab**

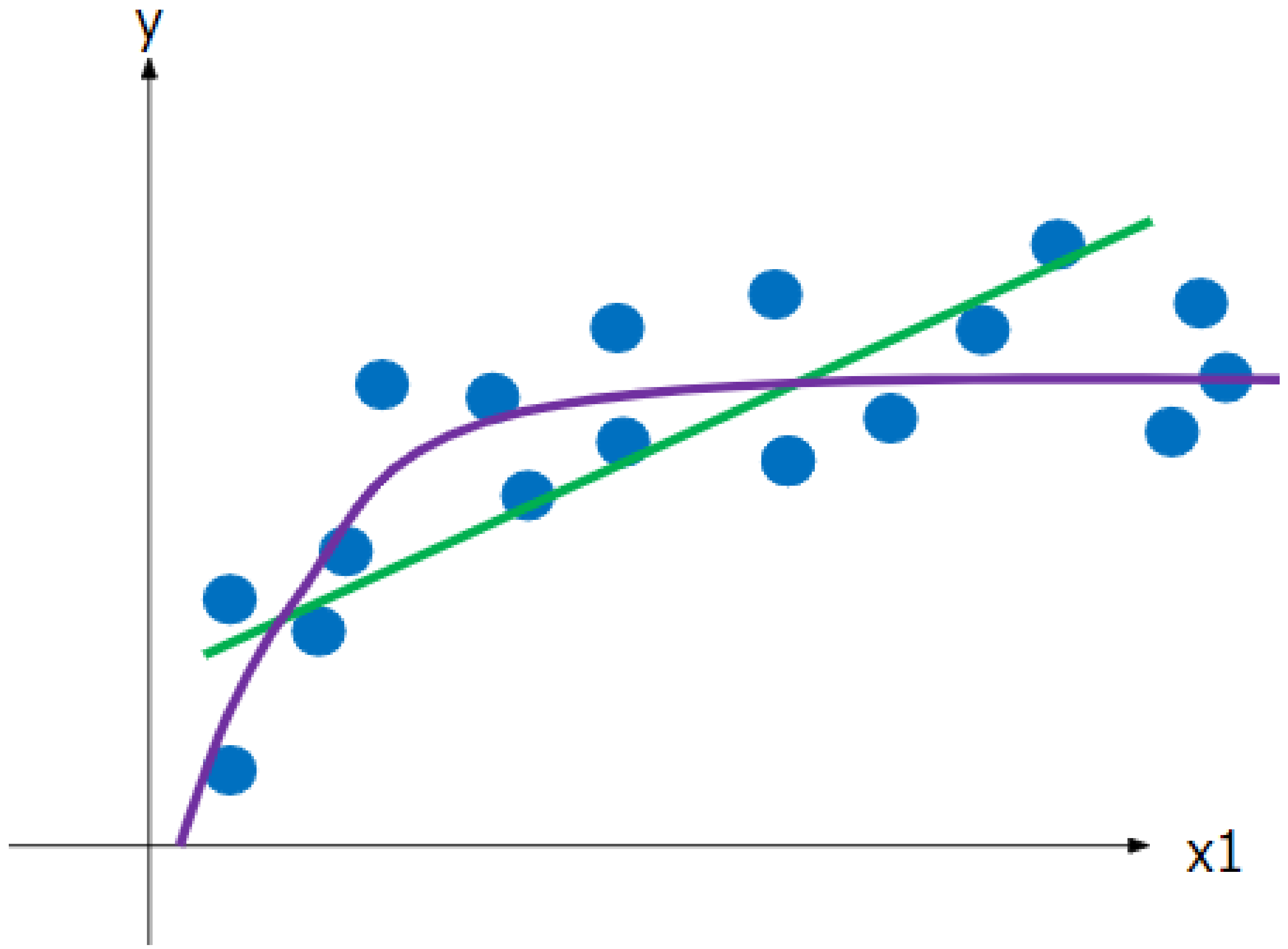
07 **¡Comprueba tu conocimiento!**

01

Acercamiento intuitivo

Regresión lineal simple, multivariada, polinomial y métricas

¿Qué es la Regresión?

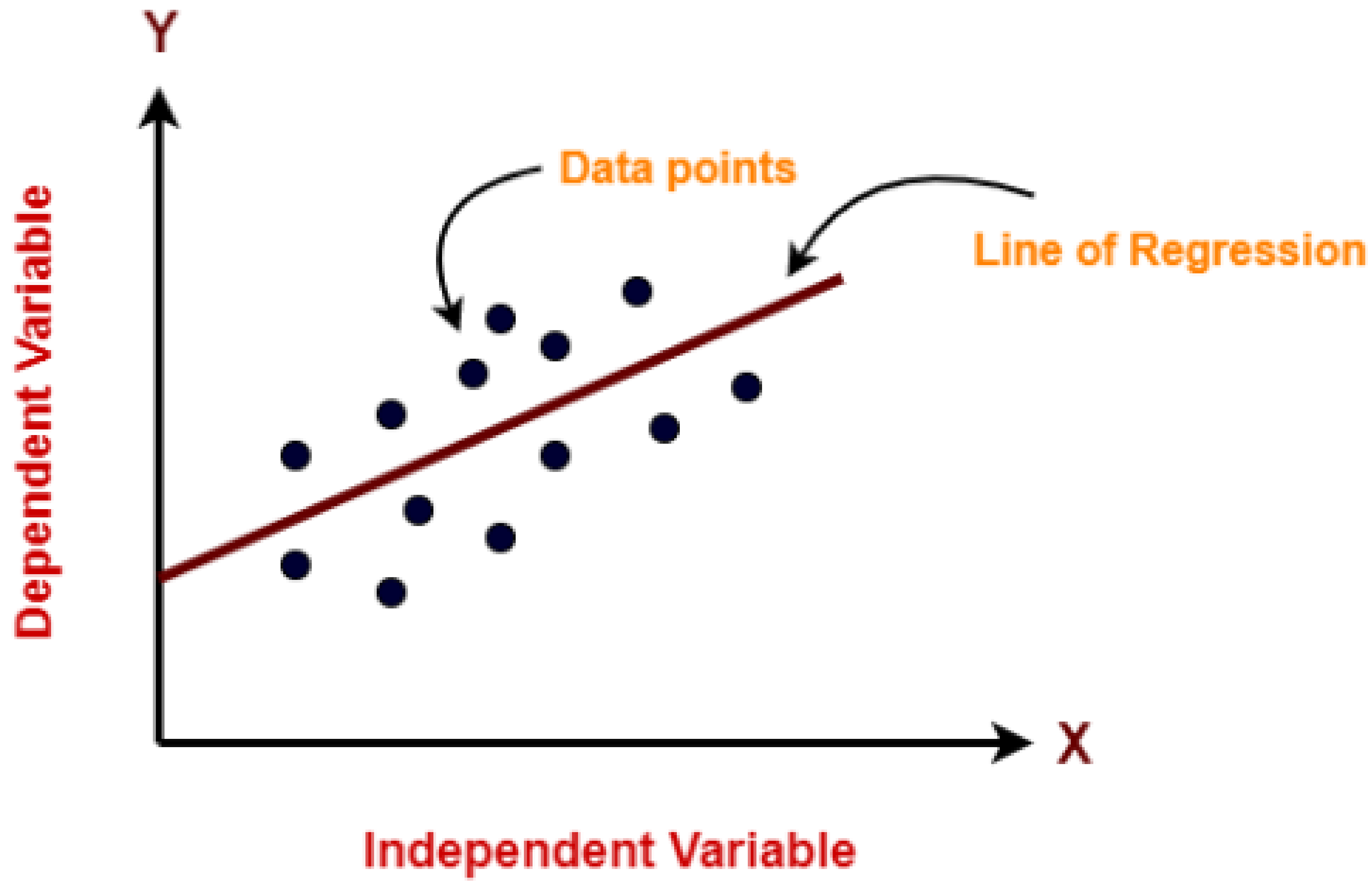


Objetivo: Ajustar modelos para predecir valores continuos de la variable objetivo (**target**) con respecto a una o varias variables independientes (**predictores**).

Métodos:
Regresión lineal (simple y múltiple).
Regresión polinomial.
KNN.
Árboles de regresión.

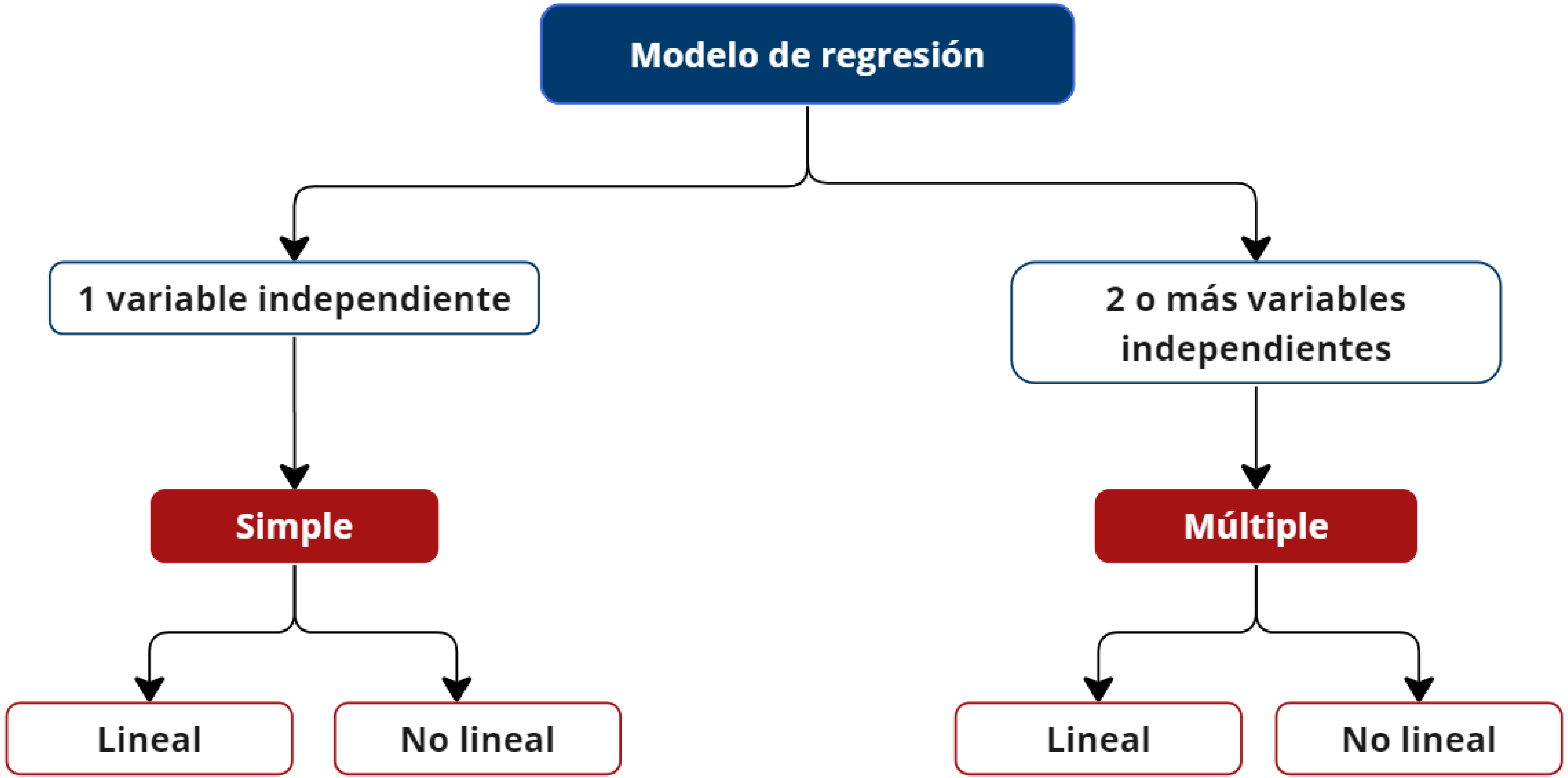
Línea base (baseline): evaluación dada por un modelo que predice una medida de tendencia central (e.g.: el promedio).

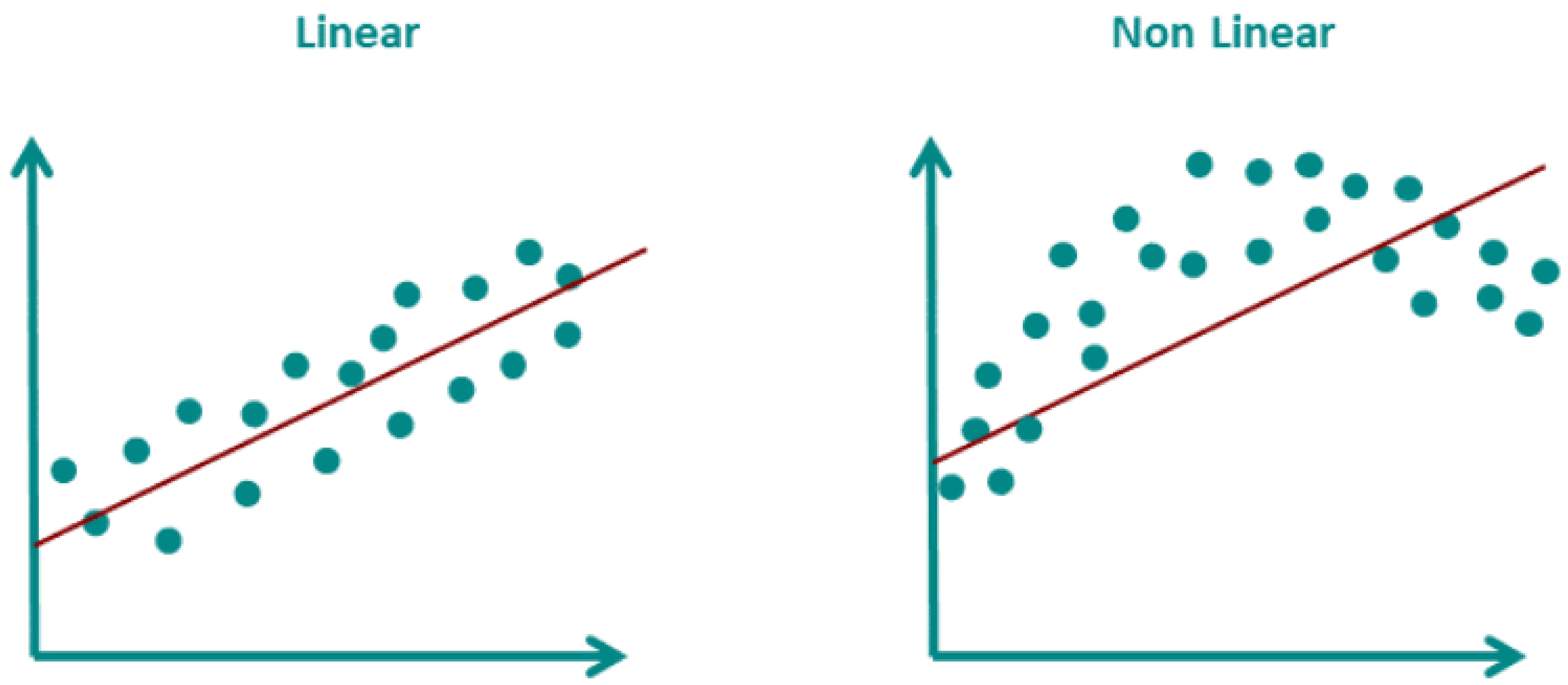
¿Qué es la Regresión?



Construye una **línea o curva que pasa a través de todos los puntos de datos** en el gráfico de predicción objetivo de tal manera que la distancia vertical entre los puntos de datos y la curva de regresión es **mínima**.

Tipos de modelos de Regresión





En el gráfico de la derecha se presenta una relación no lineal entre la variable dependiente y la independiente. Por lo tanto, la **línea de regresión no pasa a través de los puntos** de manera significativa.
Dado que esto no es posible, el modelo de regresión no puede interpretar significativamente los coeficientes o podría haber errores en la predicción.

02

Descripción matemática

Regresión lineal y Residuo
Métricas

Regresión lineal Simple

Intercepto

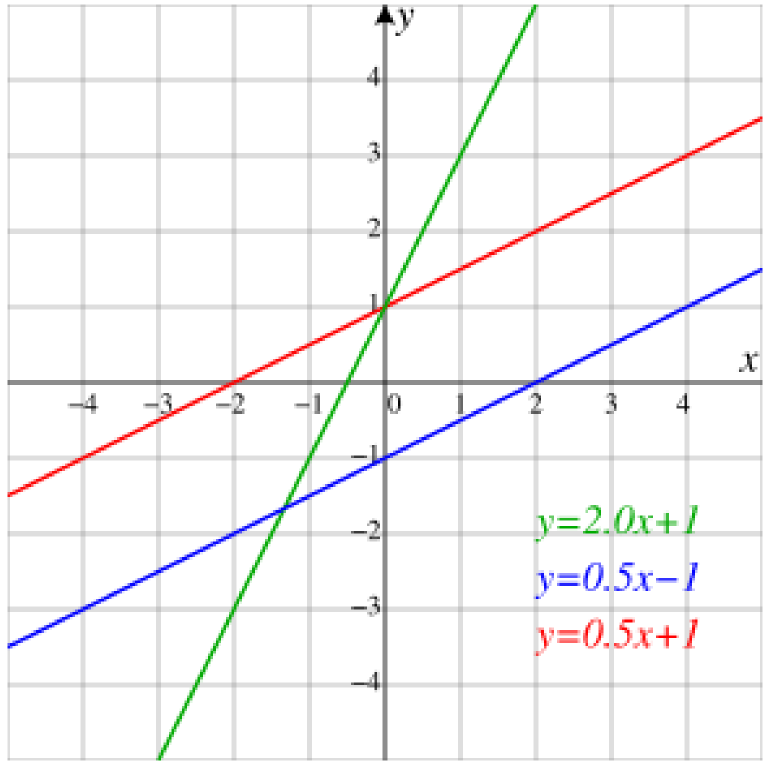
Pendiente

Error aleatorio

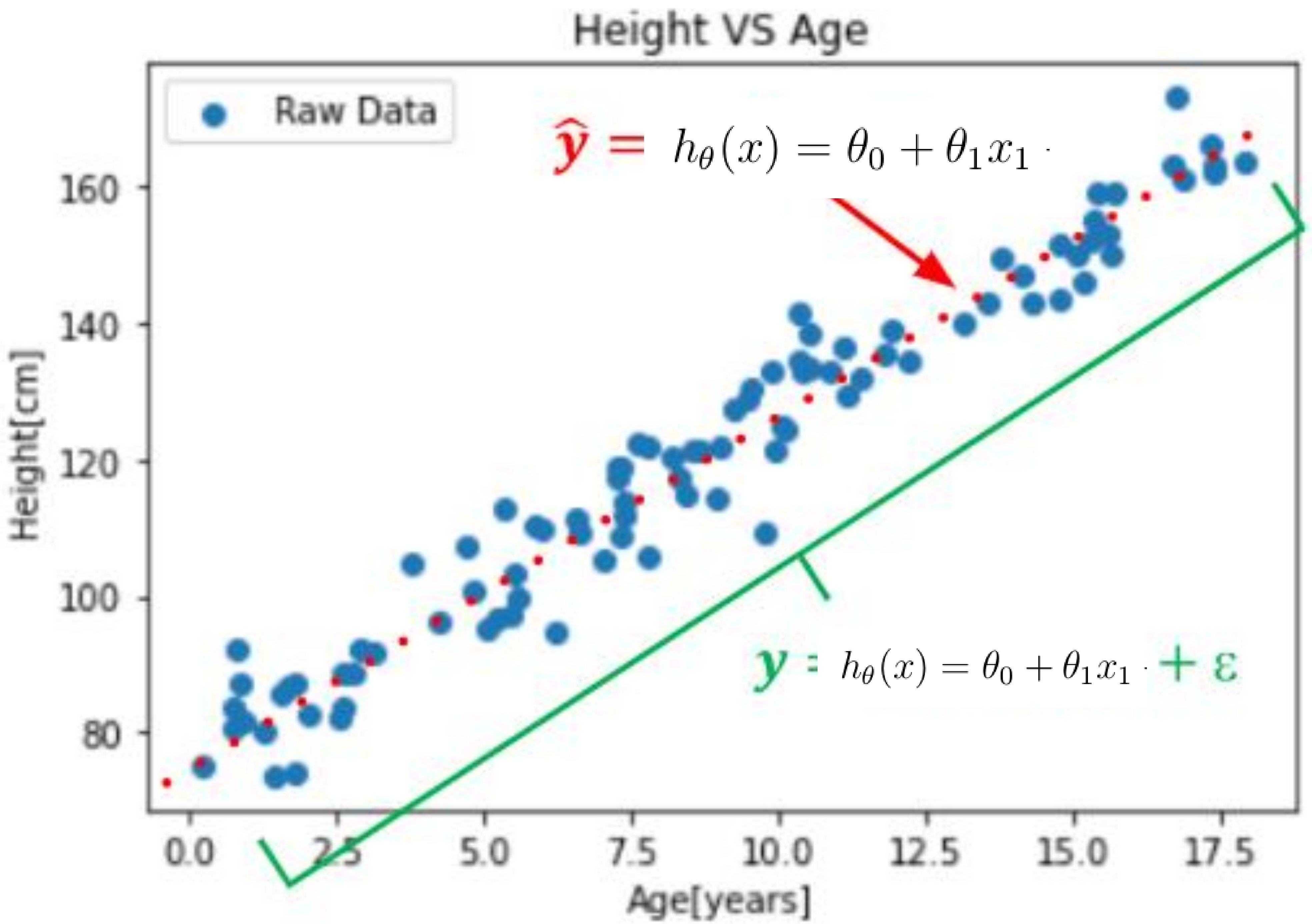
$$h(x) = \theta_0 + \theta_1 \cdot x + \epsilon$$

Variable Dependiente (respuesta - y)

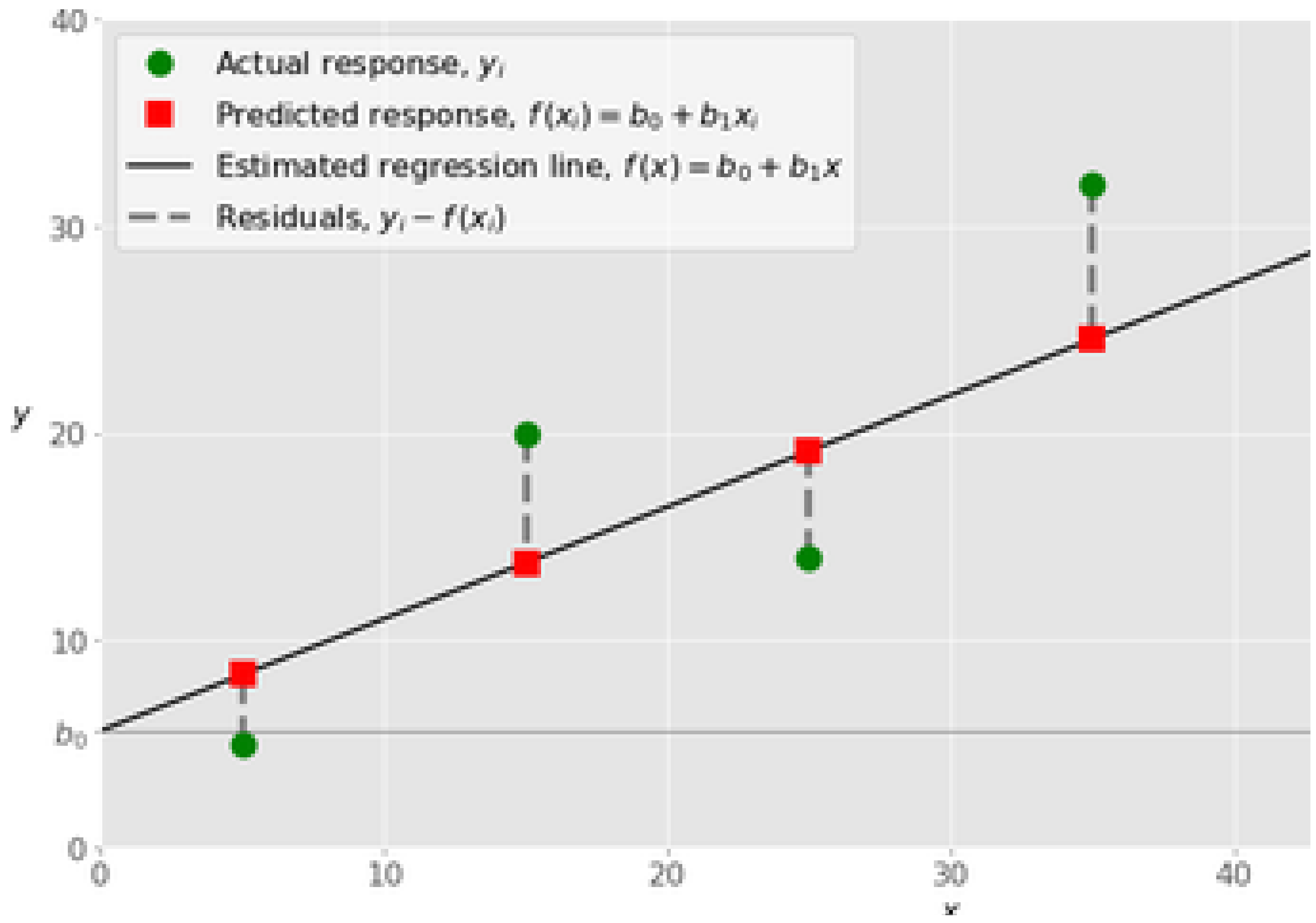
Variable Independiente (Explicatoria)



Regresión lineal Simple



Residuo



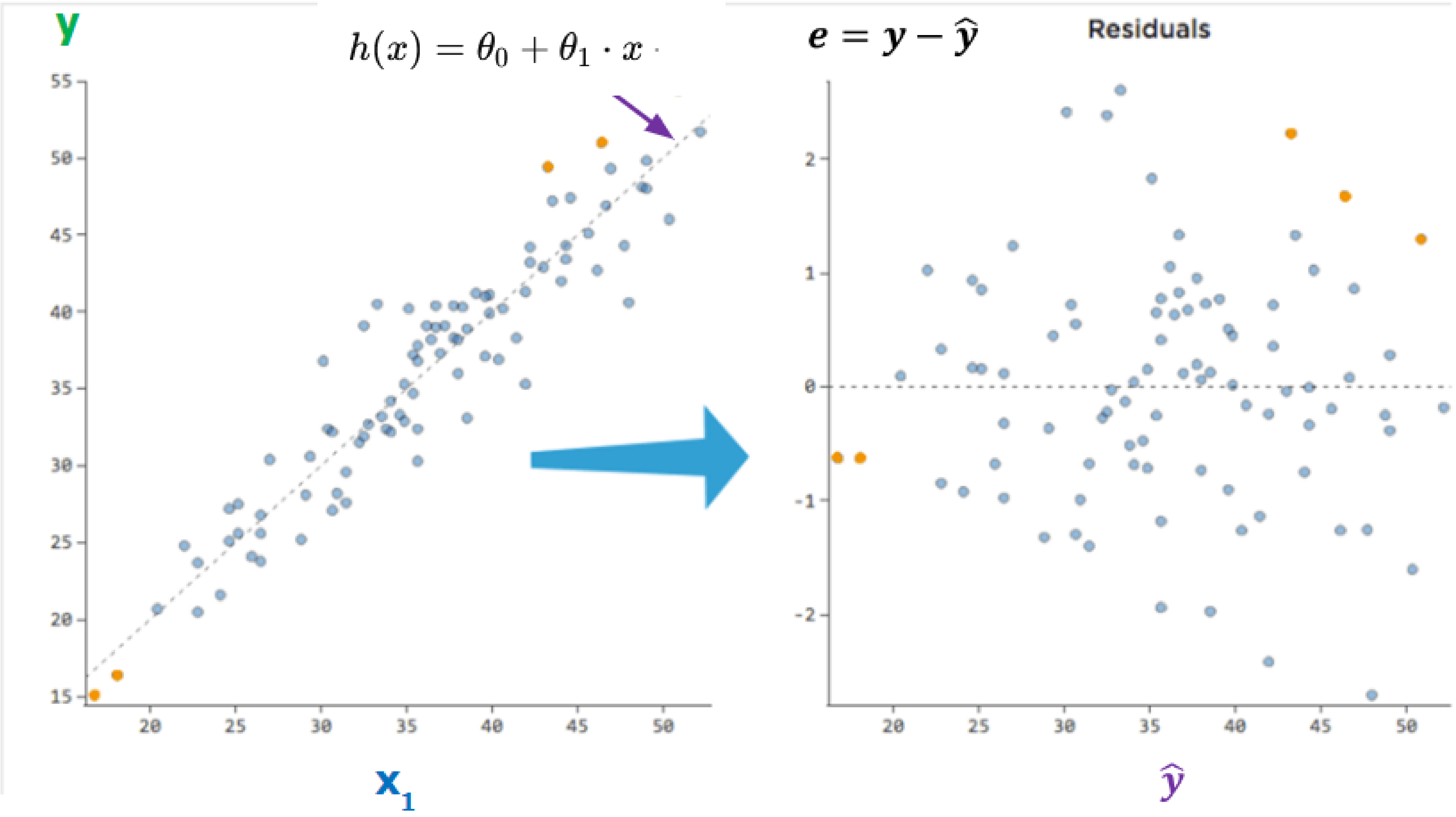
- La distancia entre los datos y la curva construida.
- Indica si el modelo ha capturado la relación entre los predictores y la variable objetivo.

Residuo (e) =
valor observado de salida - valor predicho

$$e = y - \hat{y}$$

Los modelos de regresión buscan minimizar el valor de **e** para el conjunto de predictores de entrenamiento.

Residuo



La gráfica de los residuos puede ayudar a identificar si el modelo de regresión ha capturado la relación entre la variable objetivo y los predictores.

Función de costo

Dado un **conjunto de datos** con m ejemplos:

$$\{(x^{(1)}, y^{(1)}), (x^{(2)}, y^{(2)}), \ldots, (x^{(m)}, y^{(m)})\}$$

donde:

- $x^{(i)}$ representa la variable de entrada del **i-ésimo** ejemplo,
- $y^{(i)}$ es la salida real (valor esperado),
- $h_\theta(x^{(i)})$ es la **predicción del modelo** para el ejemplo i.

$$h_\theta(x^{(i)}) = \theta_0 + \theta_1 x^{(i)}$$

Función de costo

$$h_\theta(x^{(i)}) = \theta_0 + \theta_1 x^{(i)}$$

Para cada punto $(x^{(i)}, y^{(i)})$, la diferencia entre la predicción $h_\theta(x^{(i)})$ y el valor real $y^{(i)}$ es el **error residual**:

$$\boxed{e^{(i)} = h_\theta(x^{(i)}) - y^{(i)}}$$

Si queremos medir el **error total** en todos los datos, sumamos estos errores para todos los ejemplos:

$$\boxed{\sum_{i=1}^{m} e^{(i)} = \sum_{i=1}^{m} (h_\theta(x^{(i)}) - y^{(i)})}$$

Función de costo

¿Podemos usar la suma directa de los errores?

$$\sum_{i=1}^{m} e^{(i)} = \sum_{i=1}^{m} (h_\theta(x^{(i)}) - y^{(i)})$$

Función de costo

Elevar al cuadrado los errores:

$$\sum_{i=1}^{m} (e^{(i)})^2$$

Esto garantiza que todos los errores sean positivos y penaliza mas los errores grandes.

UNIVERSIDAD ICESI | A OTRO NIVEL | VIRTUAL

Función de costo OLS

La función de costo $J(\theta_0, \theta_1)$ mide el error cuadrático medio entre las predicciones del modelo $h(x)$ y los valores observados y.

$$J(\theta_0, \theta_1) = \frac{1}{2m} \sum_{i=1}^{m} \left(h(x^{(i)}) - y^{(i)} \right)^2$$

- **¿Por qué dividir por $2m$?**

- El término $\frac{1}{m}$ obtiene el **promedio** del error cuadrático.
- El $\frac{1}{2}$ se usa para simplificar la derivada en **gradiente descendente.**

Función de costo OLS

Conversión de la Función de Costo OLS a Notación Matricial

En regresión lineal simple, el modelo se define como:

$$h_\theta(x) = \theta_0 + \theta_1 x$$

Para representar esta ecuación en forma **matricial**, escribimos:

$$h_\theta(X) = X\theta$$

Función de costo OLS

Extendemos la matriz de características a la siguiente forma

Matriz de características X de dimensión $(m \times 2)$ incluye el término de sesgo (1) y la única variable x:

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

Expresamos el vector de parámetros de la siguiente forma

θ de dimensión (2×1)

$$\theta = \begin{bmatrix} \theta_0 \\ \theta_1 \end{bmatrix}$$

Función de costo OLS

Expresamos la salida o variable objetivo de la siguiente forma:

Vector de valores reales y de dimensión $(m \times 1)$:

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

Entonces el vector de predicciones para todos los ejemplos es:

$$h_\theta(X) = X\theta = \begin{bmatrix} 1 & x^{(1)} \\ 1 & x^{(2)} \\ \vdots & \vdots \\ 1 & x^{(m)} \end{bmatrix} \begin{bmatrix} \theta_0 \\ \theta_1 \end{bmatrix} = \begin{bmatrix} \theta_0 + \theta_1 x^{(1)} \\ \theta_0 + \theta_1 x^{(2)} \\ \vdots \\ \theta_0 + \theta_1 x^{(m)} \end{bmatrix}$$

Función de costo OLS

La función de costo en su forma escalar es:

$$J(\theta) = \frac{1}{2m} \sum_{i=1}^{m} (h_\theta(x^{(i)}) - y^{(i)})^2$$

Podemos escribir esta suma como un producto de matrices:

$$J(\theta) = \frac{1}{2m} (X\theta - y)^T (X\theta - y)$$

Métricas de evaluación

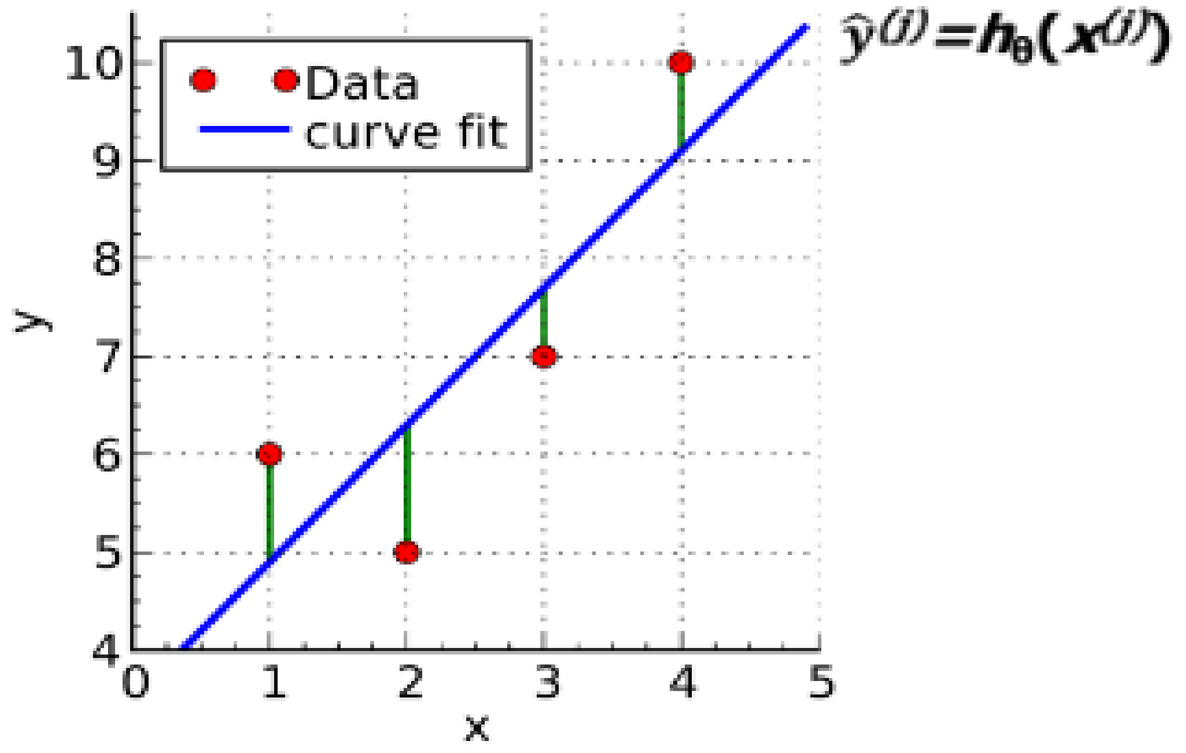
En regresión lineal, los residuales son las diferencias entre los valores observados de la variable dependiente y y los valores predichos por el modelo $h(x)$. Se calculan de la siguiente manera:

$$\text{Residual} = y - h(x)$$

- y: Valor observado de la variable dependiente.
- $h(x)$: Valor predicho por el modelo, calculado como:

$$h(x) = \theta_0 + \theta_1 \cdot x_1 + \theta_2 \cdot x_2 + \cdots + \theta_n \cdot x_n$$

- $y - h(x)$: El residual, que mide el error de predicción para cada observación.



Métricas más utilizadas

MSE

$$\frac{1}{m}\sum_{j=1}^{m}\left[\left(y^{(j)}-h_\theta\left(x^{(j)}\right)\right)^2\right]$$

RMSE

$$\sqrt{\frac{1}{m}\sum_{j=1}^{m}\left[\left(y^{(j)}-h_\theta\left(x^{(j)}\right)\right)^2\right]}$$

R^2 (coeficiente de determinación)

$$1-\frac{\sum_{j=1}^{m}\left[\left(y^{(j)}-h_\theta\left(x^{(j)}\right)\right)^2\right]}{\sum_{j=1}^{m}\left[\left(y^{(j)}-\bar{y}\right)^2\right]}$$

03

Aplicación

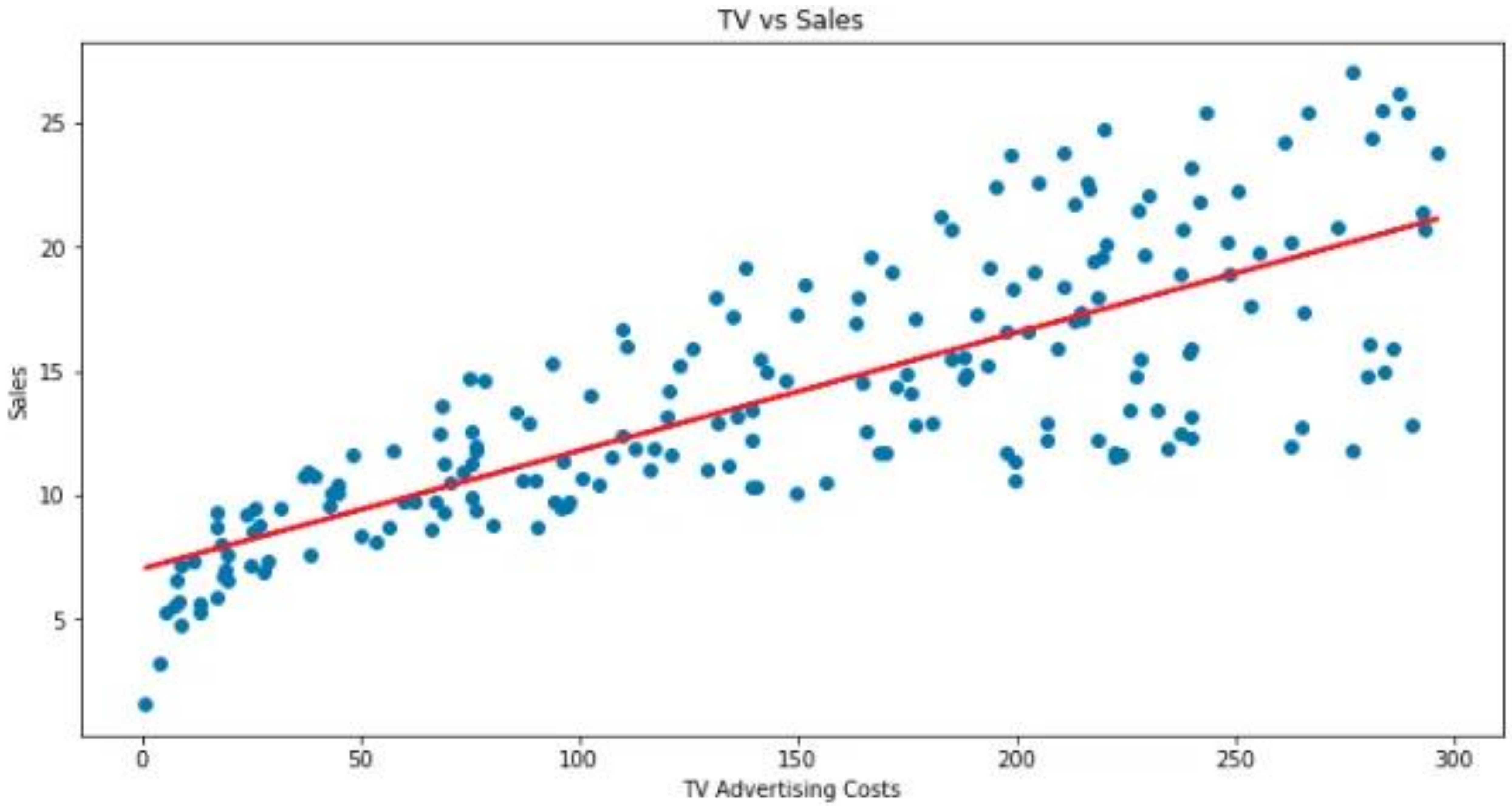
Dataset para predecir las ventas del siguiente año de 200 tiendas

Predicción de ventas

Una reconocida app lo ha contratado para que los ayude a predecir las ventas del año siguiente de 200 tiendas en toda España.

	TV	Radio	Newspaper	Sales
0	230.1	37.8	69.2	2210.0
1	44.5	39.3	45.1	1040.0
2	17.2	45.9	69.3	930.0
3	151.5	41.3	58.5	1850.0
4	180.8	10.8	58.4	1290.0
...	...	...	...	...
195	38.2	3.7	13.8	760.0
196	94.2	4.9	8.1	970.0
197	177.0	9.3	6.4	1280.0

Predicción de ventas

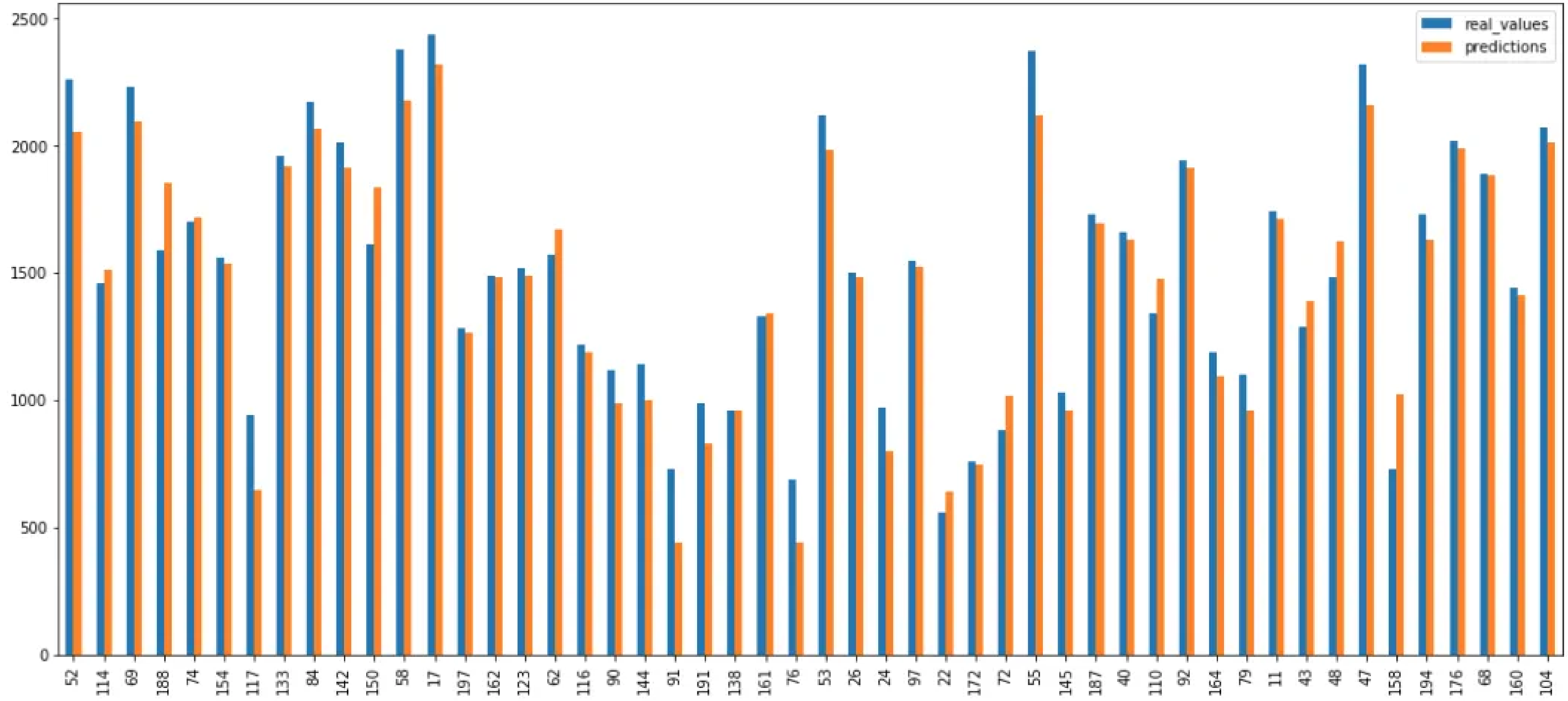


Predicción de ventas

	Coefficient
TV	4.604620
Radio	18.615614
Newspaper	0.087046

- Por cada euro que invertimos en **Televisión** suben 4.6€ las ventas.
- Por cada euro que invertimos en **radio**, suben 18,6€ las ventas.
- Por cada euro que invertimos en **prensa**, suben 0.08€ las ventas.

Ventas reales y las predicciones realizadas



01

Estimación de parámetros

Algoritmo OLS y Gradient descent

Actualización de parámetros

El algoritmo de **gradiente descendente** ajusta los parámetros θ de una función de costo $J(\theta)$ utilizando la regla de actualización:

$$\theta = \theta - \alpha \frac{\partial J}{\partial \theta}$$

Donde:

- α es la **tasa de aprendizaje**,
- $\frac{\partial J}{\partial \theta}$ es el **gradiente de la función de costo** respecto a los parámetros.

UNIVERSIDAD ICESI | A OTRO NIVEL | VIRTUAL

Actualización de parámetros

1. Derivada en funciones compuestas

Si tenemos una función compuesta $f(g(x))$, su derivada se expresa con la **regla de la cadena**:

$$\frac{d}{dx} f(g(x)) = f'(g(x)) \cdot g'(x)$$

$$J(\theta) = L(h_\theta(x), y)$$

$$\frac{\partial J}{\partial \theta} = \frac{\partial J}{\partial h} \cdot \frac{\partial h}{\partial \theta}$$

Actualización de parámetros

Los parámetros se ajustan iterativamente utilizando las derivadas parciales de la función de costo con respecto a θ_0 y θ_1.

$$\theta_0 := \theta_0 - \alpha \frac{\partial}{\partial \theta_0} J(\theta_0, \theta_1)$$

$$\theta_1 := \theta_1 - \alpha \frac{\partial}{\partial \theta_1} J(\theta_0, \theta_1)$$

α: Tasa de aprendizaje, que controla el tamaño de los pasos en la actualización de los parámetros.

$\frac{\partial}{\partial \theta_0}$ y $\frac{\partial}{\partial \theta_1}$: Gradientes de la función de costo respecto a θ_0 y θ_1.

Derivación del Gradiente

Derivada respecto a θ_0:

$$\frac{\partial J(\theta_0,\theta_1)}{\partial \theta_0} = \frac{\partial}{\partial \theta_0}\left(\frac{1}{2m}\sum_{i=1}^{m}\left(h(x^{(i)}) - y^{(i)}\right)^2\right)$$

Expandiendo $h(x^{(i)}) = \theta_0 + \theta_1 \cdot x_1^{(i)}$:

$$\frac{\partial J(\theta_0,\theta_1)}{\partial \theta_0} = \frac{1}{m}\sum_{i=1}^{m}\left(h(x^{(i)}) - y^{(i)}\right)$$

Derivada respecto a θ_1:

$$\frac{\partial J(\theta_0,\theta_1)}{\partial \theta_1} = \frac{\partial}{\partial \theta_1}\left(\frac{1}{2m}\sum_{i=1}^{m}\left(h(x^{(i)}) - y^{(i)}\right)^2\right)$$

Expandiendo $h(x^{(i)}) = \theta_0 + \theta_1 \cdot x_1^{(i)}$:

$$\frac{\partial J(\theta_0,\theta_1)}{\partial \theta_1} = \frac{1}{m}\sum_{i=1}^{m}\left(h(x^{(i)}) - y^{(i)}\right) \cdot x_1^{(i)}$$

$$\theta_j := \theta_j - \alpha \cdot \frac{\partial J(\theta_0,\theta_1)}{\partial \theta_j}$$

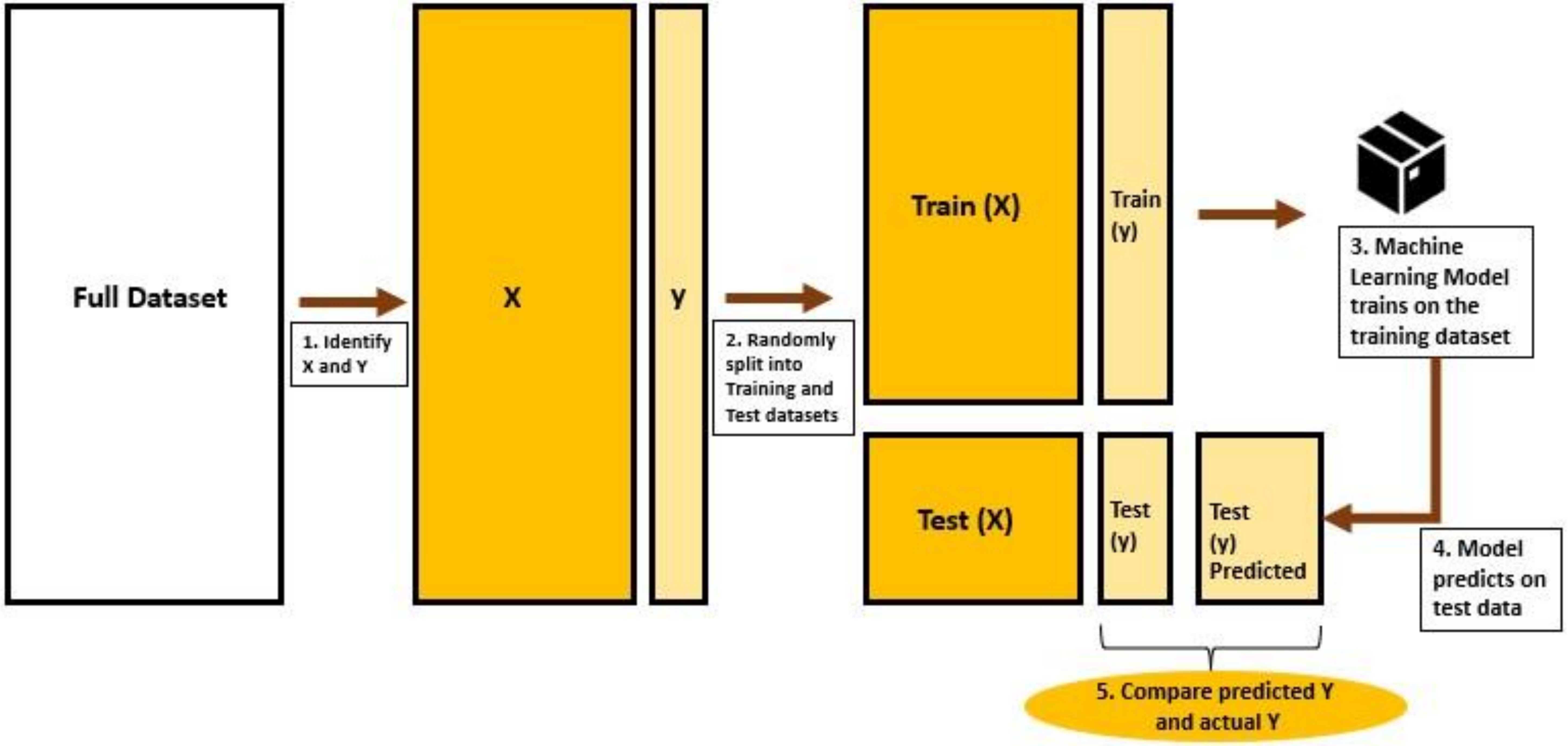
- Actualización de θ_0:

$$\theta_0 := \theta_0 - \alpha \cdot \frac{1}{m}\sum_{i=1}^{m}\left(h(x^{(i)}) - y^{(i)}\right)$$

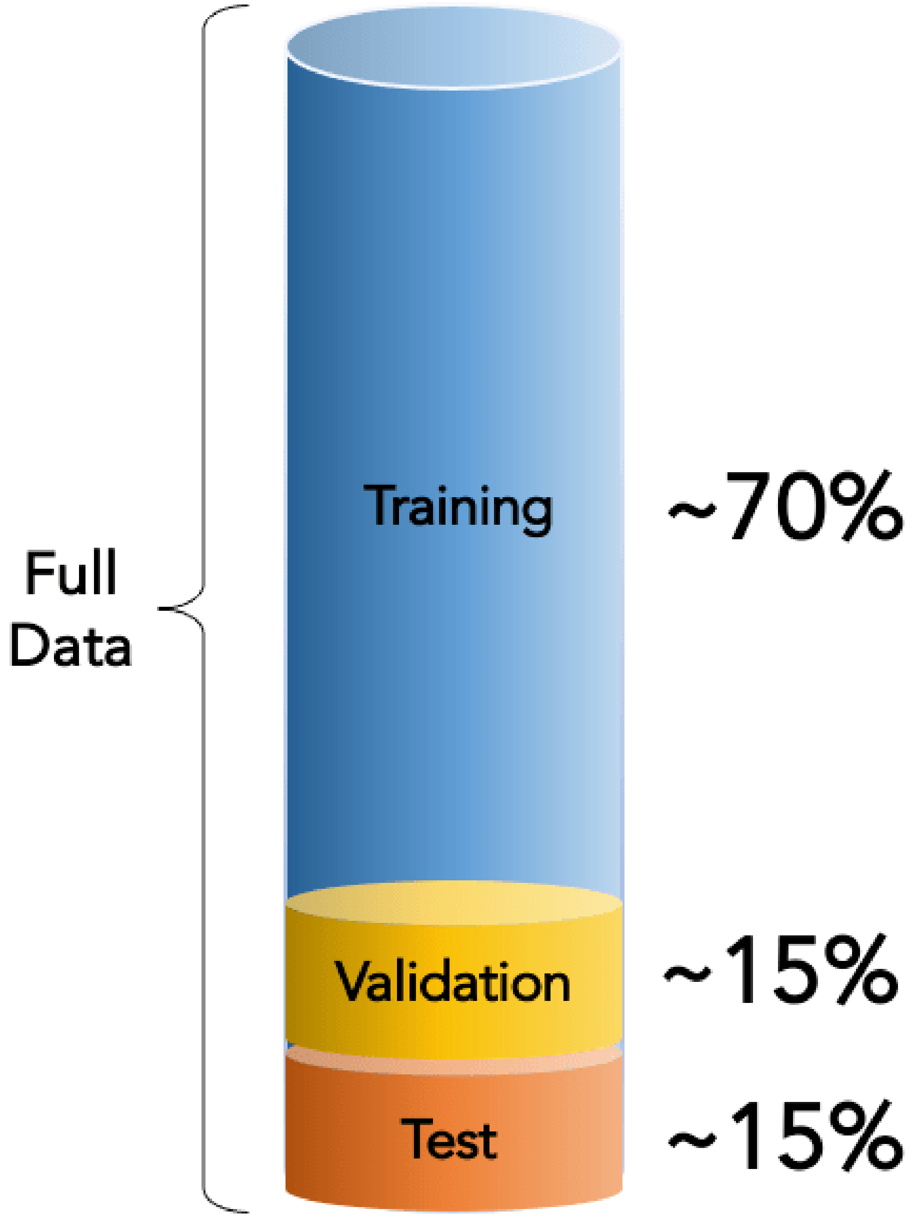
- Actualización de θ_1:

$$\theta_1 := \theta_1 - \alpha \cdot \frac{1}{m}\sum_{i=1}^{m}\left(h(x^{(i)}) - y^{(i)}\right) \cdot x_1^{(i)}$$

Metodología para entrenar un modelo



Metodología para entrenar un modelo



Regresión Lineal Múltiple

La regresión lineal múltiple es un modelo estadístico que describe la relación entre una variable dependiente y y múltiples variables independientes $x_1, x_2, \ldots, x_n$. El modelo se expresa como:

$$h(x) = \theta_0 + \theta_1 \cdot x_1 + \theta_2 \cdot x_2 + \cdots + \theta_n \cdot x_n + \epsilon$$

- $h(x)$: Predicción del modelo.
- θ_0: Término de intersección.
- $\theta_1, \theta_2, \ldots, \theta_n$: Coeficientes que indican la influencia de cada variable independiente x_i sobre y .
- ϵ: Término de ruido, que representa la variabilidad no explicada por el modelo.

Regresión Lineal Múltiple

El modelo se puede expresar de la forma

$$h(x) = \sum_{i=0}^{n} \theta_i x_i = \theta^T x,$$

Y la función objetivo se expresa de la siguiente forma

$$J(\theta) = \frac{1}{2} \sum_{i=1}^{m} (h_\theta(x^{(i)}) - y^{(i)})^2.$$

Algoritmo LMS

Queremos elegir θ para minimizar J(θ).

Específicamente, consideremos el algoritmo de descenso de gradiente, que comienza con un θ inicial y realiza la actualización repetidamente:

$$\theta_j := \theta_j - \alpha \frac{\partial}{\partial \theta_j} J(\theta).$$

Algoritmo LMS

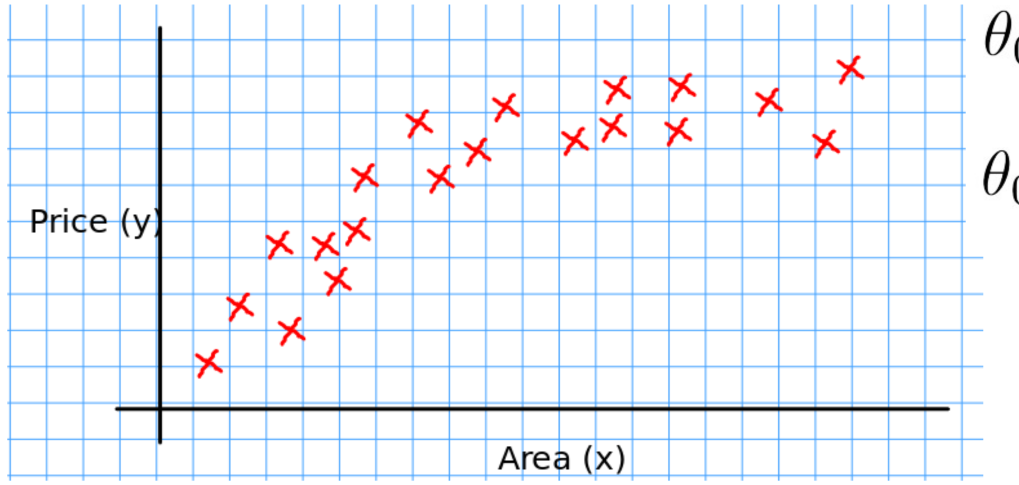
$$
\begin{aligned}
\frac{\partial}{\partial \theta_j} J(\theta) &= \frac{\partial}{\partial \theta_j} \frac{1}{2} \left(h_\theta(x) - y\right)^2 \\
&= 2 \cdot \frac{1}{2} \left(h_\theta(x) - y\right) \cdot \frac{\partial}{\partial \theta_j} \left(h_\theta(x) - y\right) \\
&= \left(h_\theta(x) - y\right) \cdot \frac{\partial}{\partial \theta_j} \left(\sum_{i=0}^{n} \theta_i x_i - y \right) \\
&= \left(h_\theta(x) - y\right) x_j
\end{aligned}
$$

$$
\boxed{\theta_j := \theta_j + \alpha \left(y^{(i)} - h_\theta(x^{(i)})\right) x_j^{(i)}.}
$$

Regresión Polinomial

- Regresión polinomial en una variable (**predicción**):

$$h_\theta(x) = \theta_0 + \theta_1 \times frontage + \theta_2 \times depth$$



$\theta_0 + \theta_1 x + \theta_2 x^2$

$\theta_0 + \theta_1 x + \theta_2 x^2 + \theta_3 x^3$

$$\begin{aligned} h_\theta(x) &= \theta_0 + \theta_1 x_1 + \theta_2 x_2 + \theta_3 x_3 \\ &= \theta_0 + \theta_1 (size) + \theta_2 (size)^2 + \theta_3 (size)^3 \end{aligned}$$

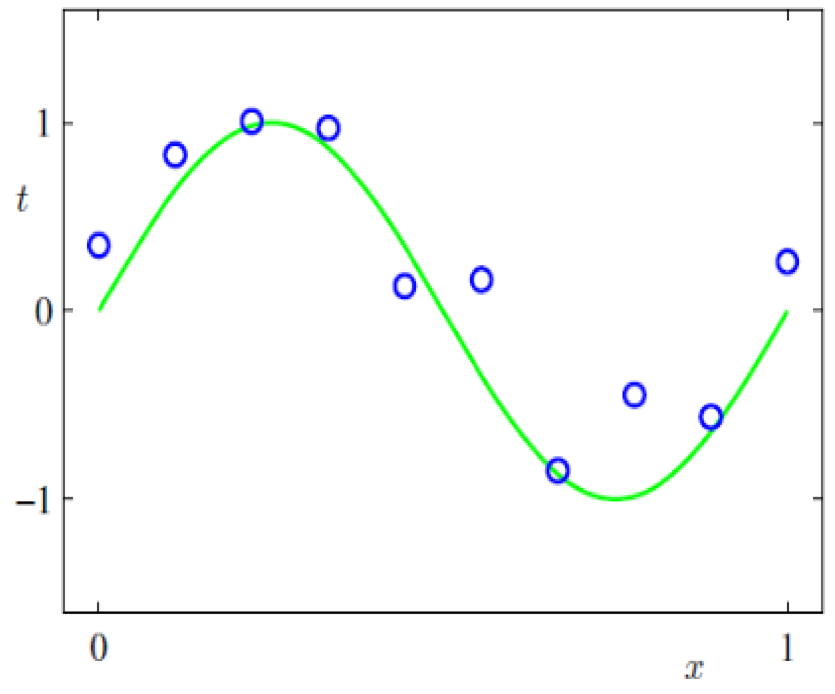
$$x_1 = (size)$$
$$x_2 = (size)^2$$
$$x_3 = (size)^3$$

Regresión Polinomial

- Regresión polinomial en una variable (**predicción**):

La regresión polinomial es una extensión de la regresión lineal que permite modelar la relación entre una variable dependiente y y una variable independiente x mediante un polinomio de grado n. El modelo se expresa como:

$$h(x) = \theta_0 + \theta_1 \cdot x + \theta_2 \cdot x^2 + \cdots + \theta_n \cdot x^n + \epsilon$$



- $h(x)$: Predicción del modelo.
- θ_0: Término de intersección.
- $\theta_1, \theta_2, \ldots, \theta_n$: Coeficientes del polinomio que determinan la forma de la curva.
- x^n: Potencias de la variable independiente x, que permiten capturar relaciones no lineales.
- ϵ: Término de ruido, que representa la variabilidad no explicada por el modelo.