Regresión

Regresión

- Es una forma estadística de establecer una relación entre una variable dependiente y un conjunto de variables independientes.

- El objetivo es predecir una variable dependiente (y) en función de los valores independientes (x)

Regresión Lineal

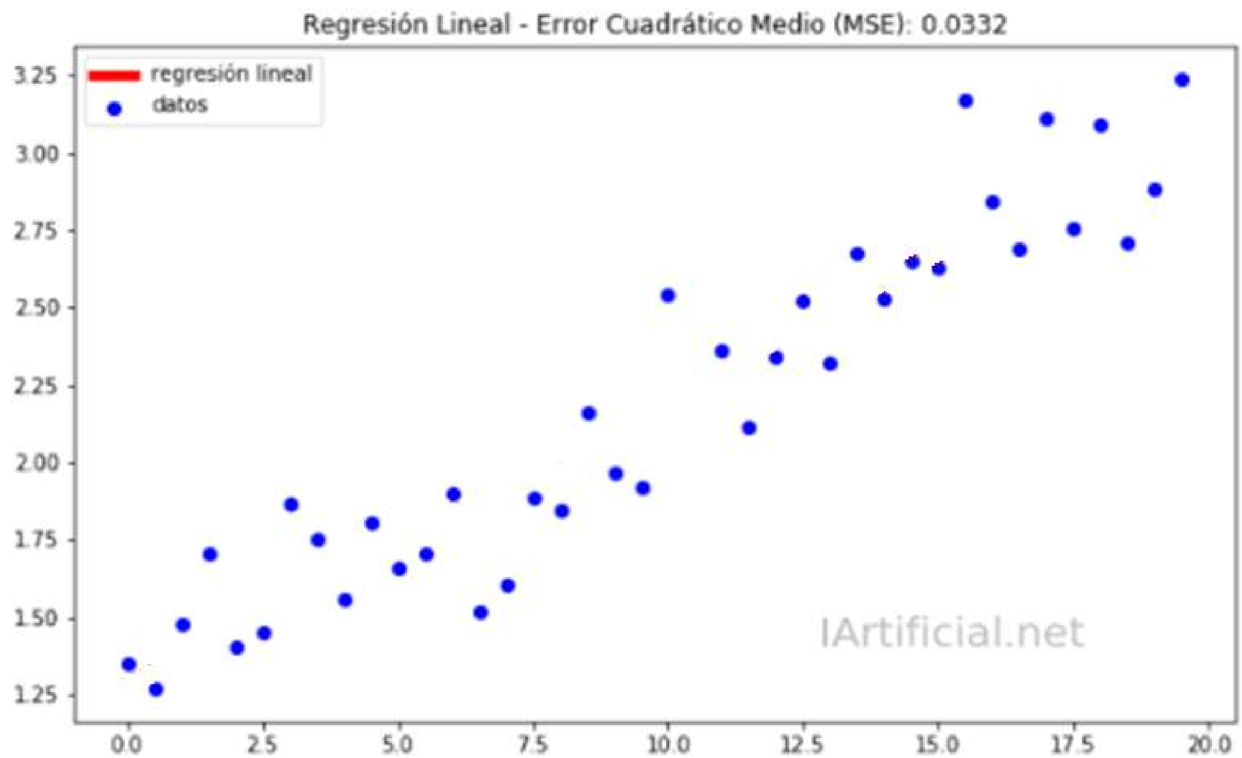
Es una técnica paramétrica de machine learning.

Paramétrica se refiere que de entrada conocemos el número de coeficientes (# parámetros) que debemos calcular.

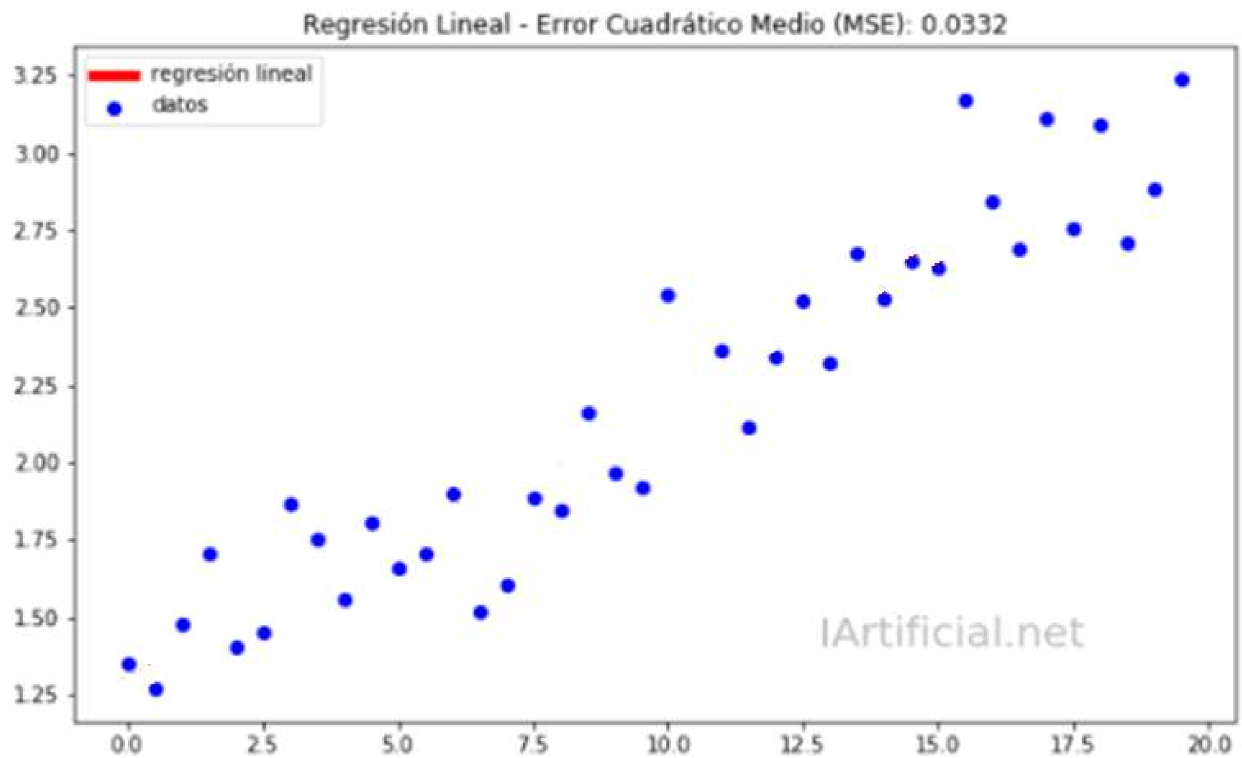
$$y = wx + b$$

El aprendizaje se basa en encontrar los coeficientes que presentan el mejor comportamiento dentro de los datos que se están tratando.

Regresión Lineal



Regresión Lineal

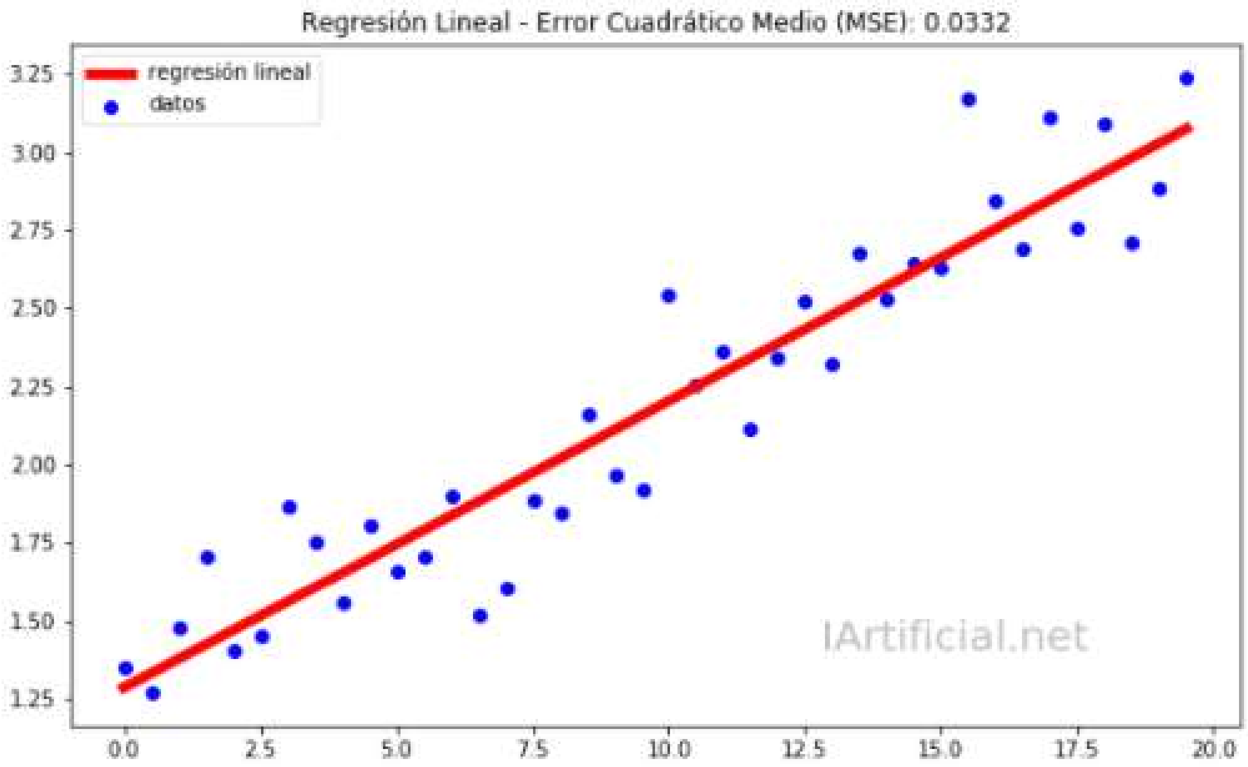


$$y = wx + b$$

$$w = 0.0918$$
$$b = 1.2859$$

$$y = 0.0918x + 1.2859$$

Regresión Lineal



$$y = wx + b$$

$$w = 0.0918$$
$$b = 1.2859$$

$$y = 0.0918x + 1.2859$$

https://phet.colorado.edu/sims/html/least-squares-regression/latest/least-squares-regression_es.html

Consideraciones

- Debe existir una relación lineal entre las variables dependientes e independientes.
- El modelo se ajustará a los datos cuando tenga variables de entrada altamente correlacionadas
- El modelo hará predicciones mas confiables si sus variables tienen distribución normal

Regresión Lineal

Aclaremos

$$y = wx + b$$

$$y = b + wx$$

$$y = b + w_1 x_1$$

Regresión Lineal

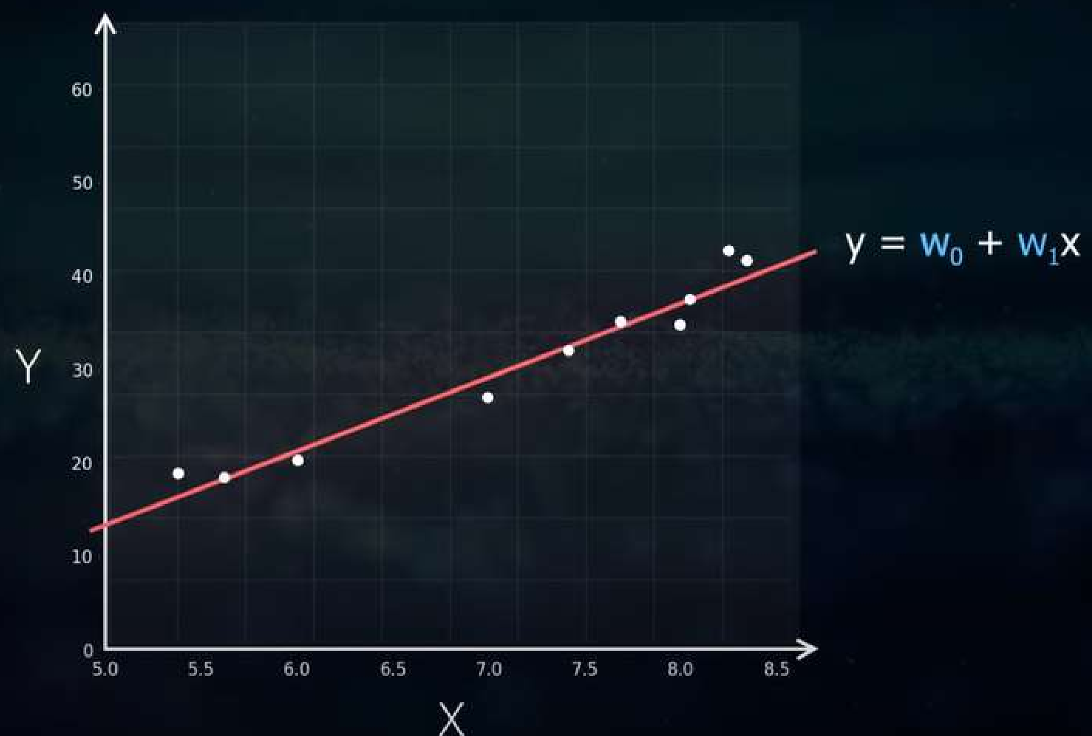
- Sigamos aclarando

$$y = b + w_1 x_1$$

$$w_0 = b$$
$$x_0 = 1$$

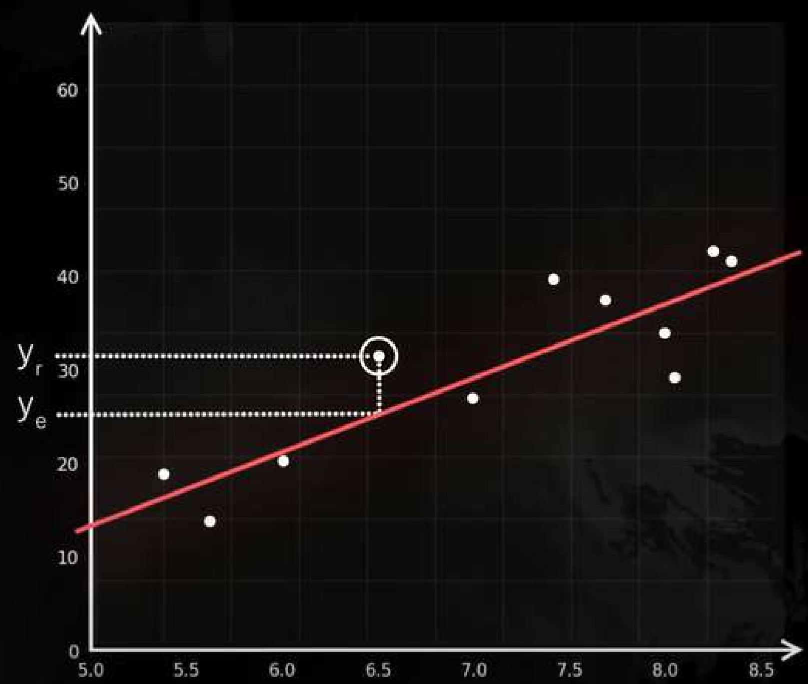
$$y = w_0 x_0 + w_1 x_1$$

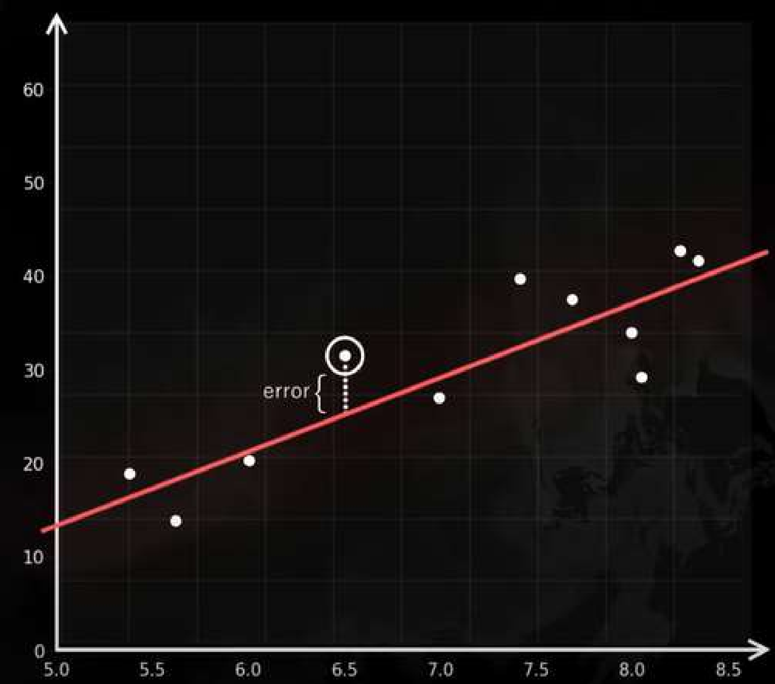
$$y = w_0 + w_1 x$$

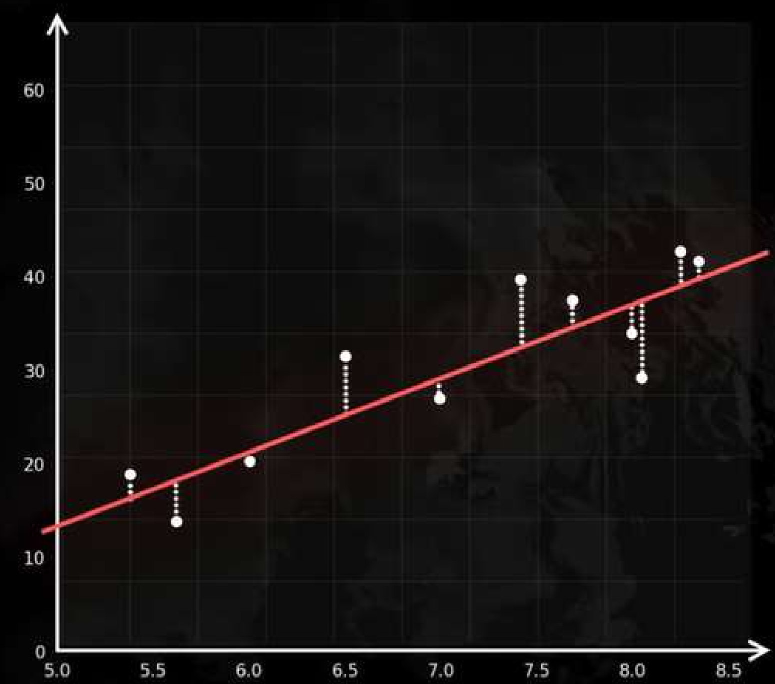


60
50
40
30
20
10
0
5.0
5.5
6.0
6.5
7.0
7.5
8.0
8.5
Y
X
$y = w_0 + w_1x$

El Error


60
50
40
y_r
30
y_e
20
10
0
5.0
5.5
6.0
6.5
7.0
7.5
8.0
8.5

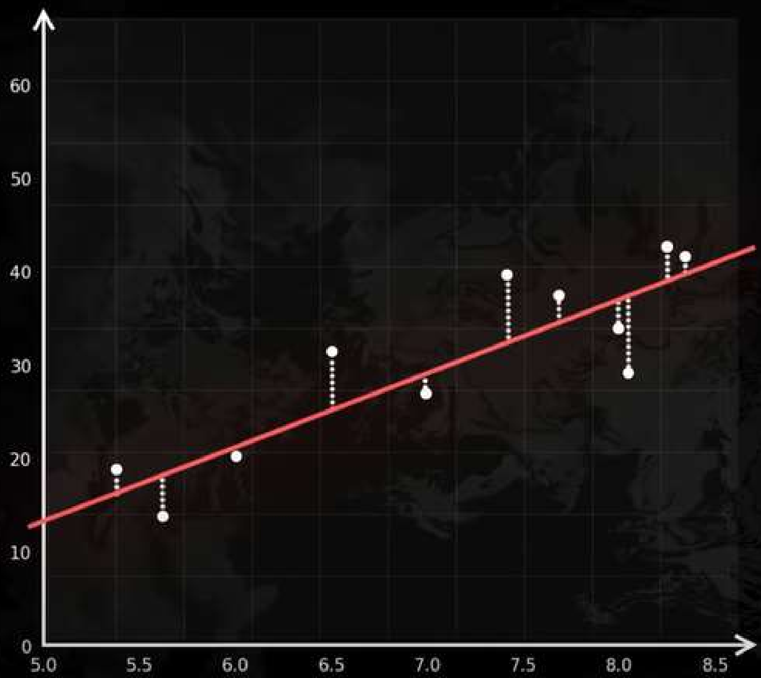

60
50
40
30
20
10
0
error
5.0
5.5
6.0
6.5
7.0
7.5
8.0
8.5



60
50
40
30
20
10
0
5.0
5.5
6.0
6.5
7.0
7.5
8.0
8.5

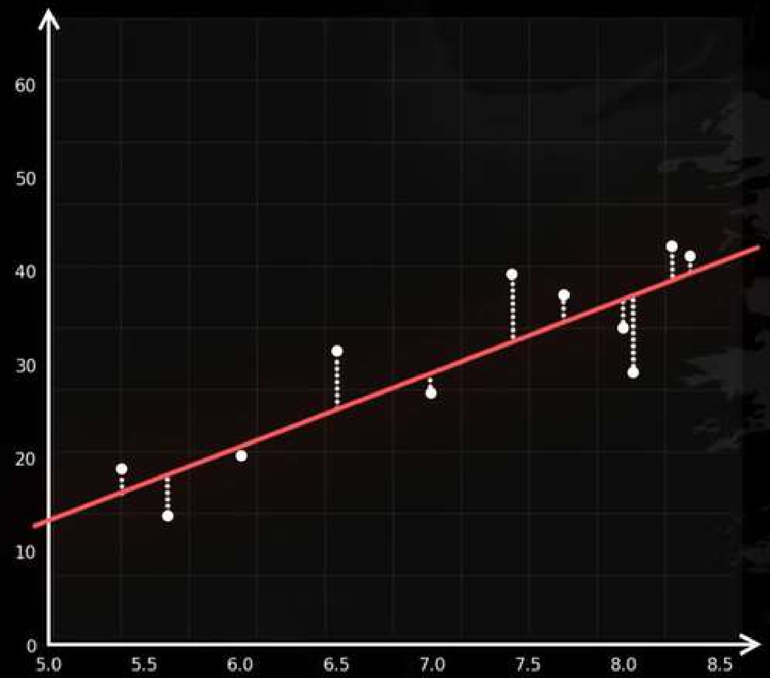
FUNCION DE COSTE:

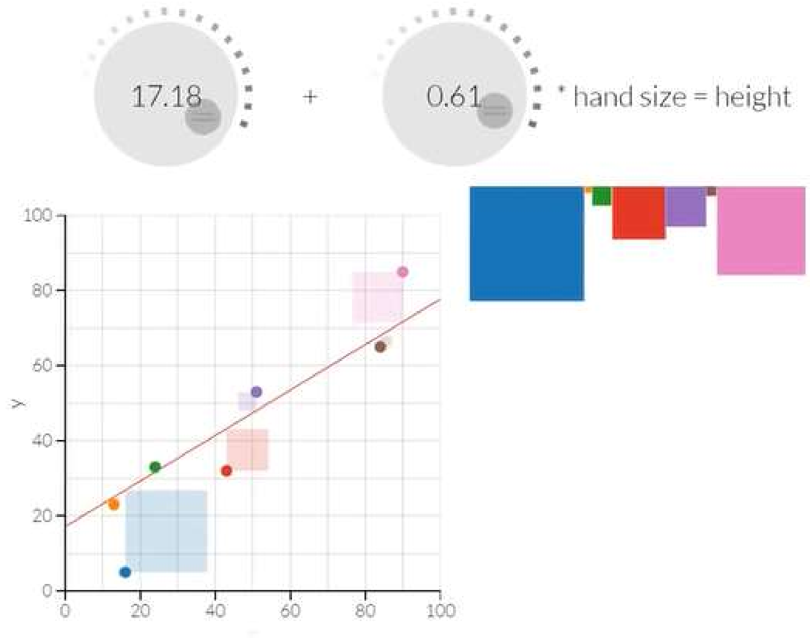
media(y_r - y_e)



ERROR CUADRATICO MEDIO:

$$\text{media}((y_r - y_e)^2)$$

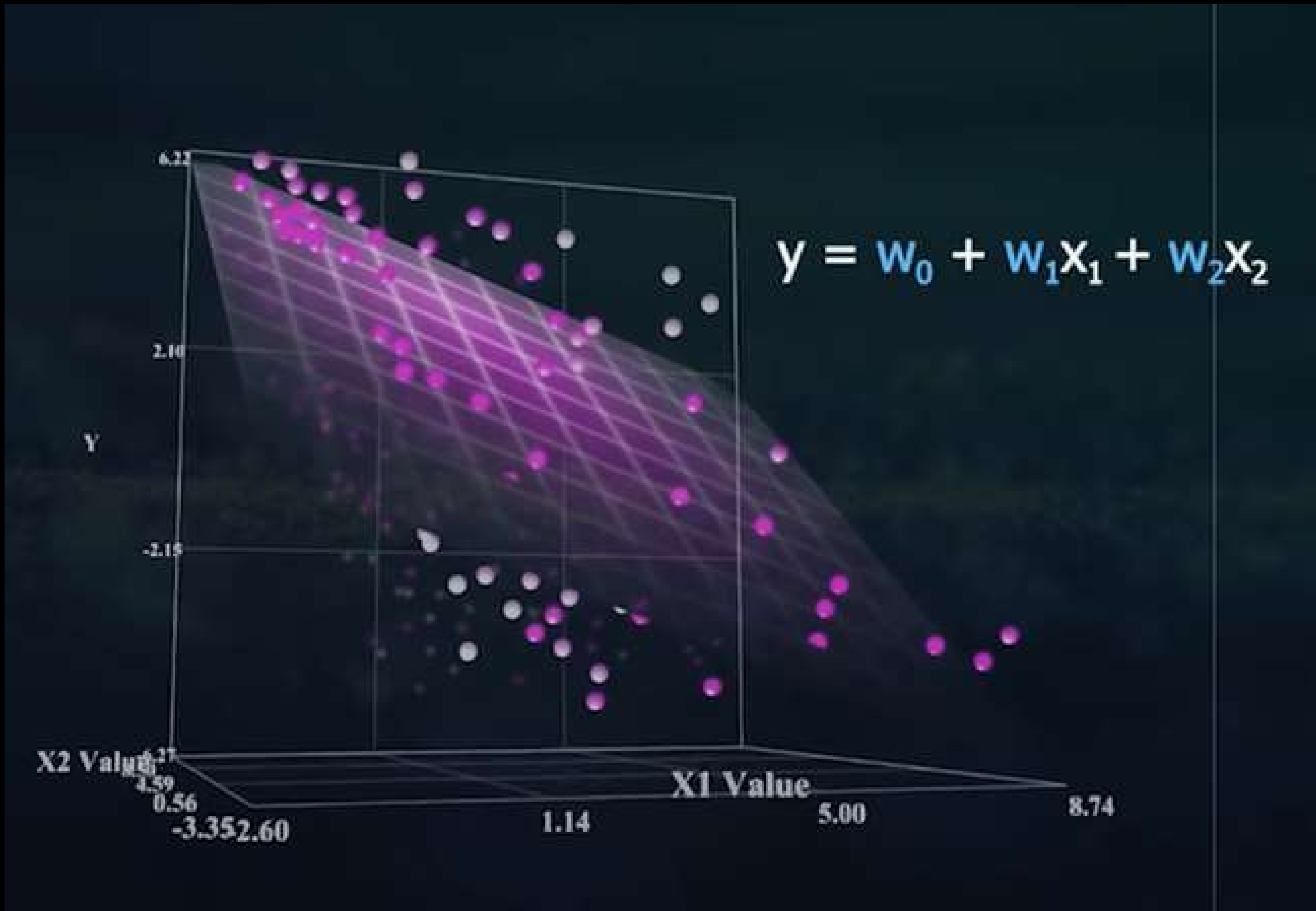




https://phet.colorado.edu/sims/html/least-squares-regression/latest/least-squares-regression_es.html

Tridimensional

$$y = w_0 + w_1x_1 + w_2x_2$$


y = w0 + w1x1 + w2x2
6.22
2.10
-2.15
Y
X2 Value
4.59
0.56
-3.35
X1 Value
2.60
1.14
5.00
8.74

Múltiples dimensiones

Regresión Lineal

- X es cada uno de los parámetros del data set que vamos a tener en cuenta

- Si tenemos múltiples parámetros de entrada tendremos múltiples X por tanto múltiples W (pesos)

$$X = [x_0, x_1, x_2, ..., x_N]$$
$$W = [w_0, w_1, w_2, ..., w_N]$$

$$y = w_0 + w_1x_1 + w_2x_2 + w_3x_3 + \ldots$$

HIPERPLANOS EN ESPACIOS MULTIDIMENSIONALES

Entradas

Estructura de los datos de entrenamiento

Como estamos tratando de una técnica de aprendizaje supervisado, el dataset de entrenamiento deberá otorgar los parámetros de entrada y los resultados esperados.

- Los datos de entrada (X) estarán almacenados en una matriz
- Los datos objetivos –resultados (Y) estarán dados en un vector
- Los pesos (W) serán los datos que serán calculados y que se espera que obtengan la mejor recta

$$y_1 = w_0 + w_1 x_{11} + w_2 x_{12} + w_3 x_{13} + \ldots$$

$$y_2 = w_0 + w_1 x_{21} + w_2 x_{22} + w_3 x_{23} + \ldots$$

$$y_3 = w_0 + w_1 x_{31} + w_2 x_{32} + w_3 x_{33} + \ldots$$

$$y_4 = w_0 + w_1 x_{41} + w_2 x_{42} + w_3 x_{43} + \ldots$$

$$y_5 = w_0 + w_1 x_{51} + w_2 x_{52} + w_3 x_{53} + \ldots$$

Estructura de los datos de entrenamiento

- Los datos de entrada (X) estarán almacenados en una matriz

$$
\begin{aligned}
y_1 &= w_0 + w_1x_{11} + w_2x_{12} + w_3x_{13} + \dots \\
y_2 &= w_0 + w_1x_{21} + w_2x_{22} + w_3x_{23} + \dots \\
y_3 &= w_0 + w_1x_{31} + w_2x_{32} + w_3x_{33} + \dots \\
y_4 &= w_0 + w_1x_{41} + w_2x_{42} + w_3x_{43} + \dots \\
y_5 &= w_0 + w_1x_{51} + w_2x_{52} + w_3x_{53} + \dots \\
&\vdots
\end{aligned}
\qquad
\begin{bmatrix}
x_{11} & x_{12} & x_{13} & \dots \\
x_{21} & x_{22} & x_{23} & \dots \\
x_{31} & x_{32} & x_{33} & \dots \\
x_{41} & x_{42} & x_{43} & \dots \\
x_{51} & x_{52} & x_{53} & \dots
\end{bmatrix}
$$

Estructura de los datos de entrenamiento

- Los datos objetivos- resultados (Y) estarán dados en un vector

$$X = \begin{bmatrix} x_{11} & x_{12} & x_{13} & \dots \\ x_{21} & x_{22} & x_{23} & \dots \\ x_{31} & x_{32} & x_{33} & \dots \\ x_{41} & x_{42} & x_{43} & \dots \\ x_{51} & x_{52} & x_{53} & \dots \end{bmatrix}$$

$$\begin{aligned} y_1 &= w_0 + w_1x_{11} + w_2x_{12} + w_3x_{13} + \dots \\ y_2 &= w_0 + w_1x_{21} + w_2x_{22} + w_3x_{23} + \dots \\ y_3 &= w_0 + w_1x_{31} + w_2x_{32} + w_3x_{33} + \dots \\ y_4 &= w_0 + w_1x_{41} + w_2x_{42} + w_3x_{43} + \dots \\ y_5 &= w_0 + w_1x_{51} + w_2x_{52} + w_3x_{53} + \dots \\ &\vdots \end{aligned}$$

$$Y = \begin{bmatrix} y_1 \\ y_2 \\ y_3 \\ y_4 \\ y_5 \end{bmatrix}$$

Estructura de los datos de entrenamiento

- Los Los pesos (W) estarán dados en un vector

$$
\begin{aligned}
y_1 &= w_0 + w_1x_{11} + w_2x_{12} + w_3x_{13} + \ldots \\
y_2 &= w_0 + w_1x_{21} + w_2x_{22} + w_3x_{23} + \ldots \\
y_3 &= w_0 + w_1x_{31} + w_2x_{32} + w_3x_{33} + \ldots \\
y_4 &= w_0 + w_1x_{41} + w_2x_{42} + w_3x_{43} + \ldots \\
y_5 &= w_0 + w_1x_{51} + w_2x_{52} + w_3x_{53} + \ldots \\
&\vdots
\end{aligned}
$$

$$
X = \begin{bmatrix} x_{11} & x_{12} & x_{13} & \ldots \\ x_{21} & x_{22} & x_{23} & \ldots \\ x_{31} & x_{32} & x_{33} & \ldots \\ x_{41} & x_{42} & x_{43} & \ldots \\ x_{51} & x_{52} & x_{53} & \ldots \end{bmatrix}
$$

$$
Y = \begin{bmatrix} y_1 \\ y_2 \\ y_3 \\ y_4 \\ y_5 \end{bmatrix}
$$

$$
W = \begin{bmatrix} w_0 & w_1 & w_2 & w_3 \end{bmatrix}
$$

$$\begin{aligned}
y_1 &= w_0 + w_1x_{11} + w_2x_{12} + w_3x_{13} + \dots \\
y_2 &= w_0 + w_1x_{21} + w_2x_{22} + w_3x_{23} + \dots \\
y_3 &= w_0 + w_1x_{31} + w_2x_{32} + w_3x_{33} + \dots \\
y_4 &= w_0 + w_1x_{41} + w_2x_{42} + w_3x_{43} + \dots \\
y_5 &= w_0 + w_1x_{51} + w_2x_{52} + w_3x_{53} + \dots \\
&\vdots
\end{aligned}$$

$$\boxed{Y = XW}$$

$$X = \begin{bmatrix} x_{11} & x_{12} & x_{13} & \dots \\ x_{21} & x_{22} & x_{23} & \dots \\ x_{31} & x_{32} & x_{33} & \dots \\ x_{41} & x_{42} & x_{43} & \dots \\ x_{51} & x_{52} & x_{53} & \dots \end{bmatrix}$$

$$Y = \begin{bmatrix} y_1 \\ y_2 \\ y_3 \\ y_4 \\ y_5 \end{bmatrix}$$

$$W = \begin{bmatrix} w_0 & w_1 & w_2 & w_3 \end{bmatrix}$$

Regresión Lineal

- La ecuación lineal quedaría como

$$y = w_0x_0 + \ w_1x_1 \ldots + w_nx_n$$

$$y = WX$$

Entrenamiento

Regresión Lineal Simple	ERROR CUADRATICO MEDIO $\text{media}((y_r - y_e)^2)$
Regresión Lineal Múltiple	MINIMO ERROR CUADRATICO MEDIO $W = (X^T X)^{-1} X^T Y$

Entrenamiento

- El proceso de aprendizaje consiste en averiguar qué parámetros W minimizan el error cuadrático medio entre los resultados reales y los estimados.

$$\hat{W} = (X^T X)^{-1} X^T y$$

ERROR CUADRATICO MEDIO

$$\text{media}((y_r - y_e)^2)$$

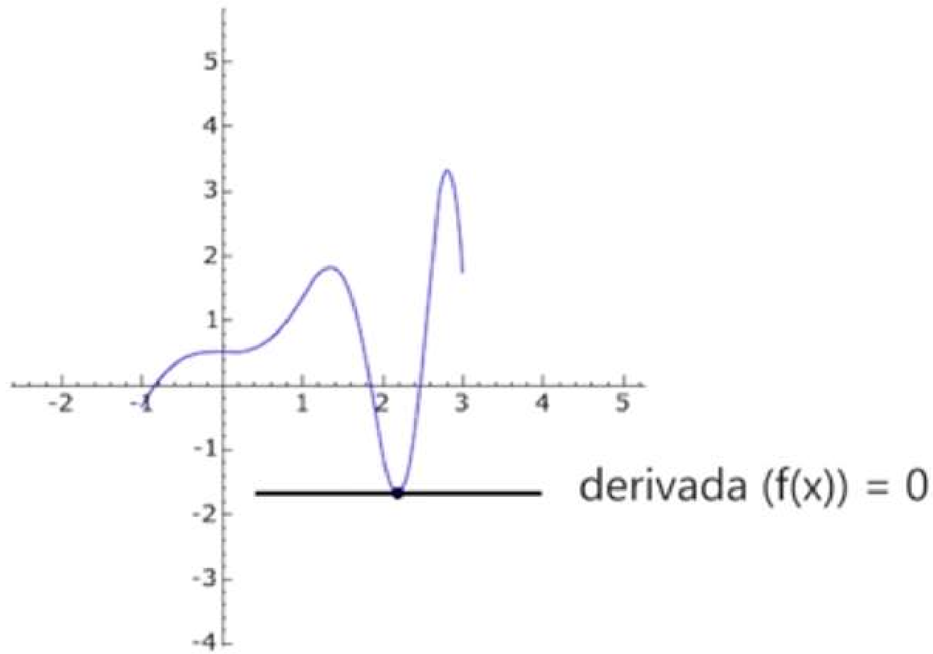
ERROR CUADRATICO MEDIO VECTORIAL

$$(Y - XW)^T(Y - XW)$$

ERROR CUADRATICO MEDIO VECTORIAL

$$(Y - XW)^T(Y - XW)$$

$$Y^TY - W^TX^TY - Y^TXW + W^TX^TXW$$


derivada (f(x)) = 0

ERROR CUADRATICO MEDIO VECTORIAL

$$(Y - XW)^T(Y - XW)$$

$$Y^TY - W^TX^TY - Y^TXW + W^TX^TXW$$

MINIMO ERROR CUADRATICO MEDIO

$$-2X^TY + 2X^TXW = 0$$

MINIMO ERROR CUADRATICO MEDIO

$$W = (X^T X)^{-1} X^T Y$$

$$\hat{W} = (X^T X)^{-1} X^T y$$

$$X = \begin{bmatrix} 1 & 3 & 5 \\ 7 & 9 & 11 \\ 13 & 15 & 17 \end{bmatrix}$$

$$Y = \begin{bmatrix} 2 & 4 & 6 \end{bmatrix}$$

$$\left(\begin{bmatrix} 1 & 7 & 13 \\ 3 & 9 & 15 \\ 5 & 11 & 17 \end{bmatrix} \begin{bmatrix} 1 & 3 & 5 \\ 7 & 9 & 11 \\ 13 & 15 & 17 \end{bmatrix} \right)^{-1} \begin{bmatrix} 1 & 7 & 13 \\ 3 & 9 & 15 \\ 5 & 11 & 17 \end{bmatrix} \begin{bmatrix} 2 & 4 & 6 \end{bmatrix}$$

$$\hat{W} = (X^T X)^{-1} X^T y$$

$$X = \begin{bmatrix} 1 & 3 & 5 \\ 7 & 9 & 11 \\ 13 & 15 & 17 \end{bmatrix}$$

$$Y = \begin{bmatrix} 2 & 4 & 6 \end{bmatrix}$$

$$\left(\begin{pmatrix} 219 & 261 & 303 \\ 261 & 315 & 369 \\ 303 & 369 & 435 \end{pmatrix} \right)^{-1} \begin{pmatrix} 1 & 7 & 13 \\ 3 & 9 & 15 \\ 5 & 11 & 17 \end{pmatrix} \{2, 4, 6\}$$

$$\hat{W} = (X^T X)^{-1} X^T y$$

X=

1	3	5
7	9	11
13	15	17

Y=

2	4	6

$$\left(\begin{pmatrix} 219 & 261 & 303 \\ 261 & 315 & 369 \\ 303 & 369 & 435 \end{pmatrix} \right)^{-1} (108, 132, 156)$$

$$\hat{W} = (X^T X)^{-1} X^T y$$

X=	1	3	5
	7	9	11
	13	15	17

Y=	2	4	6

$$\begin{pmatrix} \frac{1}{219} & \frac{1}{261} & \frac{1}{303} \\ \frac{1}{261} & \frac{1}{315} & \frac{1}{369} \\ \frac{1}{303} & \frac{1}{369} & \frac{1}{435} \end{pmatrix}$$

$$(108, 132, 156)$$

$$\hat{W} = (X^T X)^{-1} X^T y$$

X=

1	3	5
7	9	11
13	15	17

Y=

2	4	6

$$\{0.3984375, -0.609375, 0.5546875\}$$

Original file is located at
https://colab.research.google.com/drive/1-N3OoM1L6ZsTTQTD20YabcZi0z_XomID"""import numpy as np
import matplotlib.pyplot as

```
pltX=np.array([[1,3,5],[7,9,11],[13,15,17]])
print("X=")
print(X)

print("Xt=")
Xt=X.T
print(Xt)"""Invertida"""producto=Xt @ X
print("Producto")
print(producto)
invertida=np.linalg.inv(Xt @ X )
print(invertida)"""Segunda parte"""print("Xt=")
Xt=X.T
print(Xt)
y=np.array([2,4,6])
print("y=")
print(y)
segunda=Xt@y
print("segunda")
print(segunda)"""Total"""total=invertida@segunda
print(total)
```

Correlación y Pruebas

	Caracteristicas													
	0	1	2	3	4	5	6	7	8	9	10	11	12	R2
>=0.7						x							x	0.638561606
>=0.5						X					X		X	0.67862416
>=0.4			x		x	x				x	x		x	0.68102175
>=0.3	X	X	X		X	X	X		X	X	X	X	X	0.70627335
>=0	x	x	x	x	x	x	x	X	x	x	x	x	x	0.740642664
Baja relacion				x				x						0.10288072

Validación

Error Cuadrático Medio

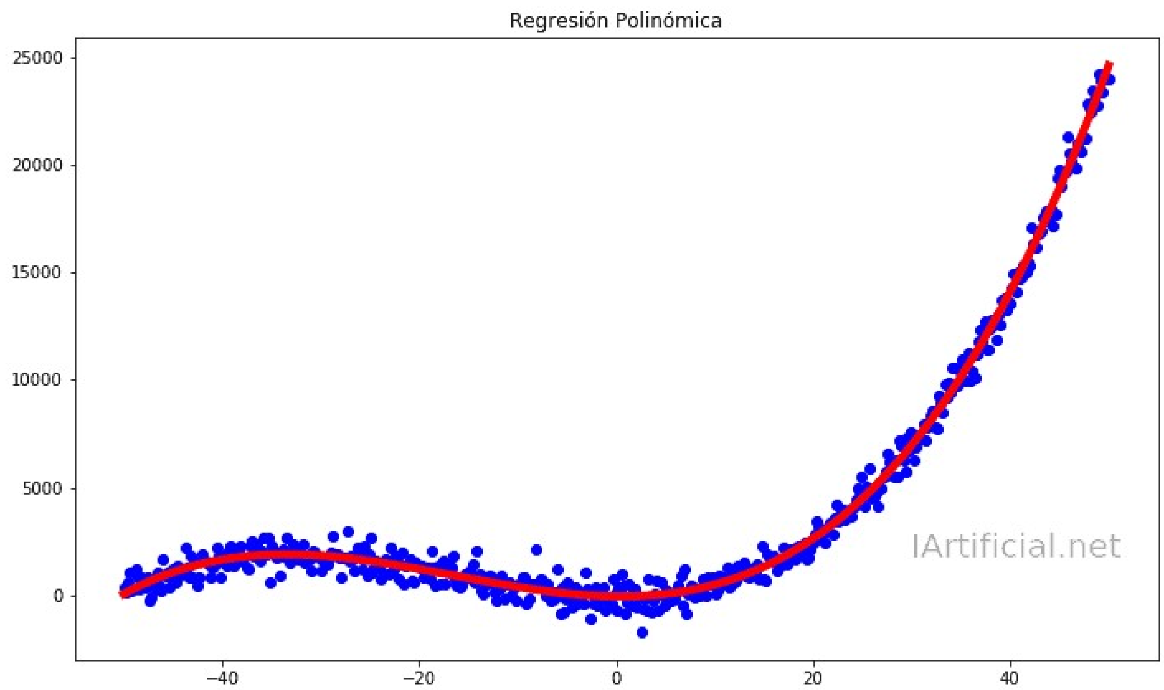
- Error Cuadrático Medio (MSE)
- Raíz del Error Cuadrático Medio (RMSE)

Validación de los resultados

Podemos evaluar la calidad del modelo midiendo el error cuadrático medio y el coeficiente de determinación R2.

El rango de R^2 está entre 0 y 1, siendo 1 lo mejor. Para medir el coeficiente de determinación R^2 de la regresión lineal usaremos el método ***score***.

Regresión Polinómica



Enlaces de interés

- https://www.iartificial.net/regresion-lineal-con-ejemplos-en-python/
- https://www.iartificial.net/error-cuadratico-medio-para-regresion/
- https://towardsdatascience.com/create-a-model-to-predict-house-prices-using-python-d34fe8fad88f
- https://towardsdatascience.com/how-to-begin-your-own-data-science-journey-2223caad8cee

- https://ichi.pro/es/clasificadores-generativos-vs-discriminativos-en-el-aprendizaje-automatico-174695206782366

FIN