

¿Cómo se calcula un error?

Ejemplo

Digamos que te contratan como empleado y te dan una relación de datos entre gastos de publicidad y ganancias en \$US.

| Gasto <i>K</i> \$ | Ganancia <i>K</i> \$ |
|-------------------|----------------------|
| 1.2 | 2 |
| 2 | 3 |
| 3.2 | 3.4 |
| 2.5 | 3.1 |
| 5 | 4 |
| 6 | 4.7 |
| 4 | 3.8 |
| 8 | 7 |

Despues me pide que elabore un modelo para predecir **¿cuanto voy a ganar o cuanto voy a gastar?**

```
In [ ]: import numpy as np
import matplotlib.pyplot as plt

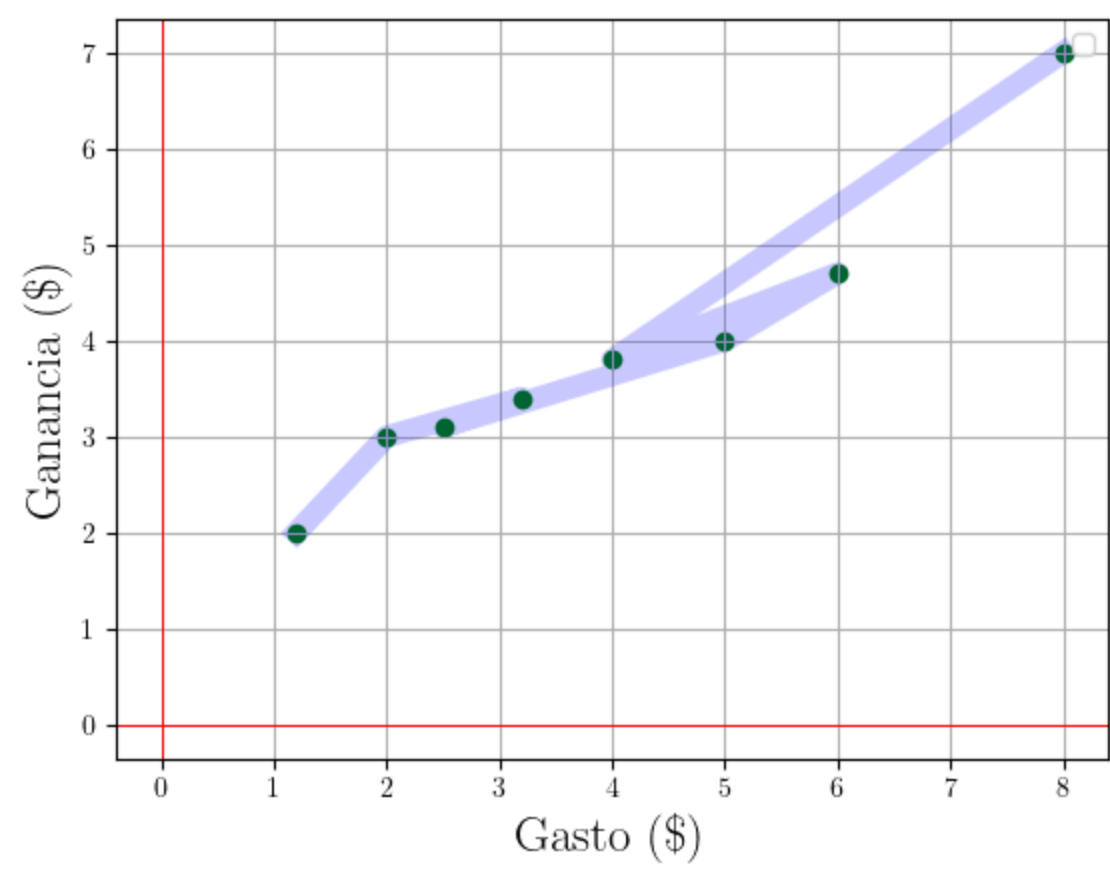
#Configurando Latex
# Configuración de Matplotlib para usar LaTeX
plt.rcParams.update({
    "text.usetex": True,
    "font.family": "serif",
    "font.serif": ["Computer Modern Roman"],
    "text.latex.preamble": r"\usepackage{amsmath}"
})
```

```
In [ ]: #Planteando el problema en código
gasto=np.array([1.2,2,3.2,2.5,5,6,4,8])
ganancia = np.array([2,3,3.4,3.1,4,4.7,3.8,7])

#Creando graficas
fig, ax = plt.subplots()
ax.plot(gasto, ganancia, color='blue', alpha=0.2, linewidth=8)
ax.scatter(gasto, ganancia, color='green')
ax.grid()
ax.axhline(y=0, color='r', linewidth=.7)
ax.axvline(x=0, color='r', linewidth=.7)
plt.title('Gasto vs Ganancia\n', fontsize=23)
plt.xlabel(r'Gasto  $(\$)$ ', fontsize=18)
plt.ylabel(r'Ganancia  $(\$)$ ', fontsize=18)
plt.legend()
plt.show()
```

No artists with labels found to put in legend. Note that artists whose label start with an underscore are ignored when legend() is called with no argument.

Gasto vs Ganancia



¿cómo se calcula un error?

Nosotros sabemos que lo que modelan en nuestros datos es una línea recta, pero vamos a llamar a la recta real con datos $y(x)$ (los puntos reales) y a nuestra predicción le llamaremos $\hat{y}(x)$.

La diferencia entre estos 2 es el error, es decir que tan alejado estoy del dato real.

$$\text{error} = \hat{y} - y$$

Pero en la anterior función en algunos puntos me darán errores negativos.

Podríamos usar el valor absoluto $|x|$

$$|\text{error}| = |\hat{y} - y|$$

Así siempre vamos a obtener valores positivos para nuestro error, pero aún hay otra forma de hacer los valores positivos, es mediante x^2 . Entonces también podemos calcular el error mediante:

$$\text{error}^2 = (\hat{y} - y)^2$$

¿Qué estaría pasando aquí y por qué esta función es buena para calcular errores? Digamos que con esta forma de calcular el error lo que se hace es que se pone énfasis en la diferencia de valores [real vs predicción], y cuando el error sea considerable se castiga con el término cuadrático x^2 , en caso contrario minimiza con errores pequeños. Entonces podemos hacer el error para todos los valores

$$\text{Error} = \sum_{i=1}^N (\hat{y} - y)^2$$

Para nuestro caso sería

$$\text{Error} = \sum_{i=1}^8 (\hat{y} - y)^2$$

Pero todo esto es el **Error total**

$$= \sum_{i=1}^N (\hat{y} - y)^2$$

Y para ver cómo cambia de manera individual nos apoyamos de la *media*. Entonces tenemos algo llamado **Error Cuadrático Medio = EMC**

$$\text{EMC} = \frac{1}{N} \sum_{i=1}^N (\hat{y} - y)^2$$

¿Qué es el EMC?

El EMC es una medida de cuán bien se ajusta la recta a los datos. O que tanto se está equivocando en cada uno de los puntos que ajusta la recta a los datos reales

NOTA

A las funciones que te dicen que tanto te equivocas en un modelo, se les llama **FUNCIONES DE COSTE** es un concepto muy útil, ya que nos dicen que tanto nos estamos equivocando.

Extras

Les recomiendo estos videos: -Regresión Lineal y Mínimos Cuadrados Ordinarios | DotCSV https://youtu.be/k964_uNn3l0

-IA NOTEBOOK #1 | Regresión Lineal y Mínimos Cuadrados Ordinarios | Programando IA <https://youtu.be/w2RJ1D6kz-o>