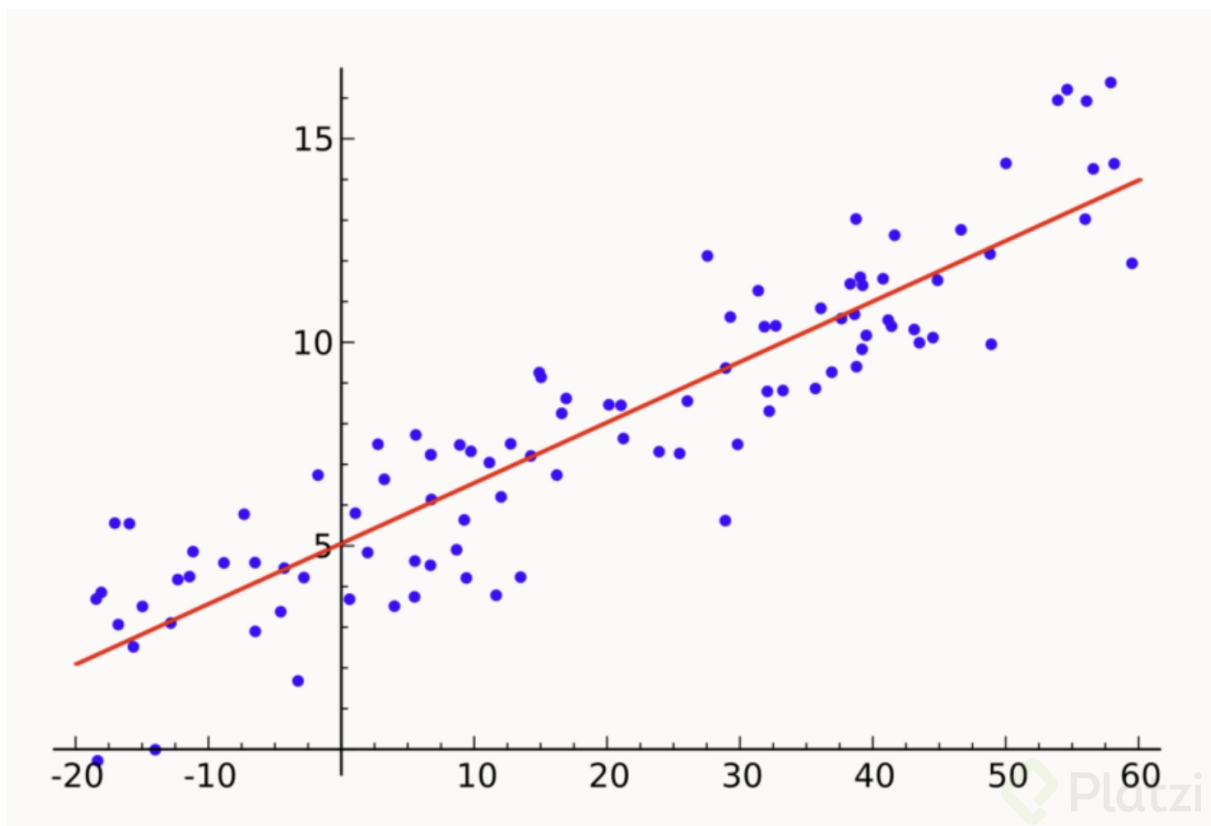


Regresión lineal

La regresión lineal es un modelo matemático que describe la relación entre varias variables. La representación gráfica de la regresión lineal simple es un gráfico de dispersión. Los modelos de regresión lineal son un procedimiento estadístico que ayuda a predecir el futuro. Se utiliza en los campos científicos y en los negocios, y en las últimas décadas se ha utilizado en el aprendizaje automático. Aunque la regresión lineal no sólo se utiliza con fines de predicción: también ha demostrado su eficacia para describir sistemas.

La tarea de la regresión en el aprendizaje automático consiste en predecir un parámetro (Y) a partir de un parámetro conocido X.



En una regresión lineal simple, se trata de establecer una relación entre una variable independiente y su correspondiente variable dependiente. Esta relación se expresa como una línea recta. No es posible trazar una línea recta que pase por todos los puntos de un gráfico si estos se encuentran ordenados de manera caótica. Por lo tanto, sólo se determina la ubicación óptima de esta línea mediante una regresión lineal. Algunos puntos seguirán distanciados de la

recta, pero esta distancia debe ser mínima. El cálculo de la distancia mínima de la recta a cada punto se denomina función de pérdida.

Hay dos ecuaciones que representan la regresión lineal (de forma independiente):

$$Y = \beta_0 + \beta_1 X + \varepsilon$$

donde:

- Y es la variable dependiente.
- X es la variable independiente.
- β_0 es el coeficiente de sesgo o la intersección de la línea de regresión con el eje Y cuando X es igual a cero.
- β_1 es el coeficiente de la variable X, que representa la pendiente de la línea de regresión.
- ε es el término de error o el error residual. .

$$Y = mx + b$$

Donde:

- y es la variable dependiente.
- x es la variable independiente.
- m es el coeficiente de la variable x, que representa la pendiente de la línea de regresión.
- b es el coeficiente de sesgo o la intersección de la línea de regresión con el eje y cuando x es igual a cero..

El término de error residual, ε , el cual solo aparece en la primera ecuación, que representa la discrepancia entre los valores reales de Y y los valores predichos por el modelo de regresión, no se incluye directamente en esta ecuación. Sin embargo, el error residual está implícitamente presente en la diferencia entre los valores reales de Y y los valores predichos por la ecuación " $Y = mx + b$ ".

```
import random
import matplotlib.pyplot as plt
```

```

import numpy as np

x = np.array([1.2, 2, 3.2, 2.5, 5, 6, 4, 8])
y = np.array([2, 3, 3.4, 3.1, 4, 4.7, 3.8, 7])

m = 1
b = 1
valid_err = 1.5
err_ant = 1000
best = [m, b]

def f(x):
    return m*x + b

def error(y_hat):
    return np.sum((y - y_hat)**2)

n = 100

for i in range(n):
    y_hat = f(x)
    err = error(y_hat)
    print(f"b: {round(b, 2)} m: {round(m, 2)} error: {round(e

    if err < err_ant:
        best = [m, b]
    else:
        m = best[0] + random.uniform(-0.1, 0.1)
        b = best[1] + random.uniform(-0.1, 0.1)
    err_ant = err

    if err < valid_err:
        print("r: ", round(m, 2), round(b, 2))
        break

```

```
m, b = best
y_hat = f(x)
err = error(y_hat)

print(f"b: {round(b, 2)} m: {round(m, 2)} error: {round(err, 2)}")

xo = x.copy()
xo.sort()
yo = f(xo)
plt.scatter(x, y)
plt.plot(xo, yo, color='red')
plt.show()
```