

Regresión Lineal Múltiple en Python

Jared Sandoval

March 22, 2025

1 Introducción

La regresión lineal múltiple es una extensión de la regresión lineal simple que permite modelar la relación entre una variable dependiente y dos o más variables independientes. Su ecuación general es:

$$Y = b + m_1X_1 + m_2X_2 + \cdots + m_nX_n \quad (1)$$

Esta técnica es útil para predecir valores y analizar la influencia de múltiples factores sobre una variable objetivo.

2 Metodología

Para llevar a cabo la regresión lineal múltiple en Python, seguimos los siguientes pasos:

2.1 Carga de datos y preprocesamiento

Se utilizaron datos con las siguientes variables:

- **Cantidad de palabras** en el artículo.
- **Suma de enlaces, comentarios e imágenes.**
- **Cantidad de veces compartido** (variable dependiente).

2.2 Implementación del Modelo

El siguiente código muestra la creación del modelo:

```
import numpy as np
import pandas as pd
from sklearn.linear_model import LinearRegression
from sklearn.metrics import mean_squared_error, r2_score

# Creación de la variable combinada
```

```

suma = (filtered_data["# of Links"] + filtered_data['# of
      comments'].fillna(0) + filtered_data['# Images video'])

dataX2 = pd.DataFrame()
dataX2["Word count"] = filtered_data["Word count"]
dataX2["suma"] = suma
XY_train = np.array(dataX2)
z_train = filtered_data['# Shares'].values

# Entrenamiento del modelo
regr2 = LinearRegression()
regr2.fit(XY_train, z_train)

# Resultados
print("Coefficients:", regr2.coef_)
print("Mean squared error: %.2f" % mean_squared_error(
      z_train, regr2.predict(XY_train)))
print("Variance score: %.2f" % r2_score(z_train, regr2.
      predict(XY_train)))

```

3 Resultados

Los coeficientes obtenidos fueron:

$$\text{Coeficientes} = [6.632, -483.408] \quad (2)$$

```

Coefficients:
[ 6.63216324 -483.40753769]
Mean squared error: 352122816.48
Variance score: 0.11

```

Figure 1: Resultados

El error cuadrático medio fue de 352122816.48 y la puntuación de varianza fue 0.11, indicando que el modelo tiene un bajo poder predictivo.

4 Visualización

Se generó un gráfico 3D para visualizar los resultados:

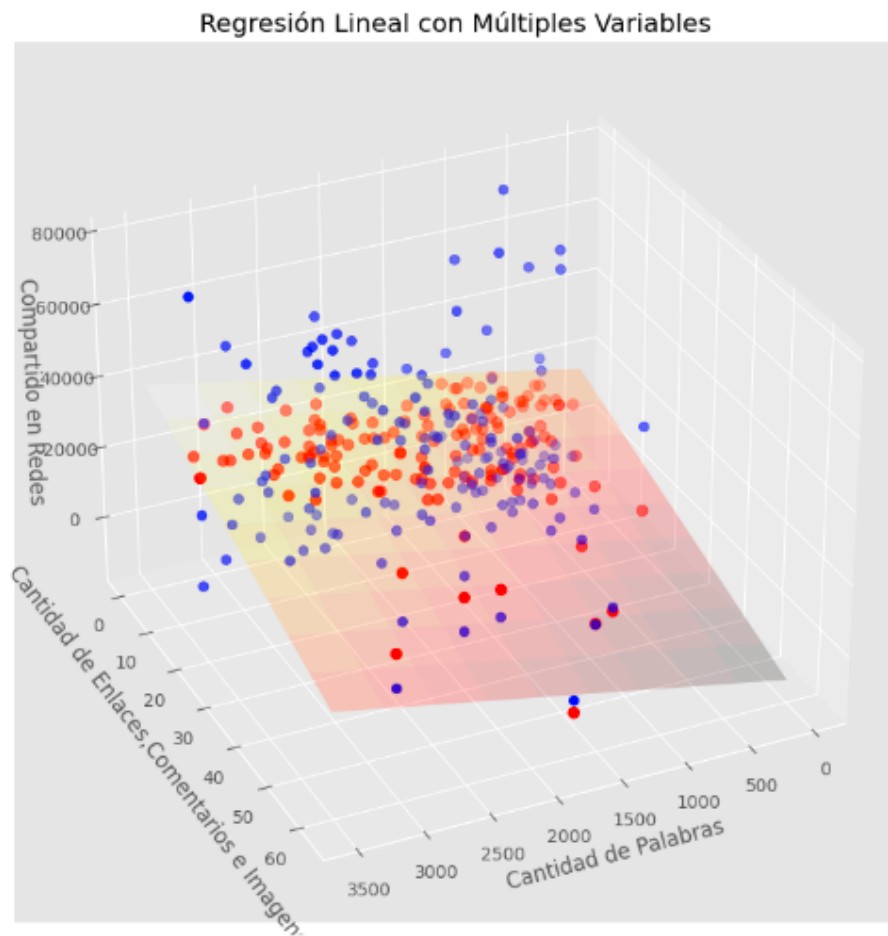


Figure 2: Regresión 3D

4.1 Predicción con el modelo de Múltiples Variables

Para evaluar el modelo, realizamos una predicción para un artículo con 2000 palabras, 10 enlaces, 4 comentarios y 6 imágenes. Utilizamos el siguiente código en Python:

```
# Si quiero predecir cuántos "Shares" voy a obtener por un artículo con:
# 2000 palabras y con enlaces: 10, comentarios: 4, imagenes: 6
# según nuestro modelo, hacemos:

z_Dosmil = regr2.predict([[2000, 10+4+6]])
print(int(z_Dosmil))
```

La predicción obtenida fue de 20,518, lo que indica una posible mejora con respecto a la regresión con una sola variable.

5 Conclusión

Este experimento demuestra la aplicación de la regresión lineal múltiple en Python. A pesar de la baja precisión del modelo, permite analizar el impacto de distintas variables sobre la cantidad de veces que un artículo es compartido. Para mejorar el modelo, se podrían incluir más variables predictivas o utilizar técnicas como PCA para la reducción de dimensionalidad.