

# Programando Regresion Lineal Multiple en Python

Julio Cesar Torres Marquez

marzo 2025

## 1 Introduccion

La regresion lineal multiple permite generar un modelo en el cual el valor de una variable dependiente se calcula en funcion de un conjunto de variables independientes. Es una extension de la regresion lineal simple. Estos modelos se pueden emplear para predecir valores o para evaluar la influencia de los factores predictivos.

## 2 Metodologia

Para implementar la regresion lineal en Python, es necesario instalar las siguientes librerias:

- pandas
- seaborn
- matplotlib
- scikit-learn

Para instalarlas, se ejecuta el siguiente comando en la consola:

```
pip install pandas seaborn matplotlib scikit-learn
```

Una vez instaladas, se procede con la programacion del modelo.

### 2.1 Importar las librerias

```
import numpy as np
import pandas as pd
import seaborn as sb
import matplotlib.pyplot as plt
from sklearn import linear_model
from sklearn.metrics import mean_squared_error, r2_score
```

## 2.2 Carga de los datos

```
data = pd.read_csv("ruta_del_archivo.csv")
```

## 2.3 Filtrado y procesamiento de los datos

Se filtran los datos para eliminar valores extremos y se genera una nueva variable que representa la suma de enlaces, comentarios e imagenes.

## 2.4 Creacion del modelo

Se define y entrena el modelo de regresion lineal multiple usando la libreria scikit-learn.

```
regr2 = linear_model.LinearRegression()  
regr2.fit(XY_train, z_train)  
z_pred = regr2.predict(XY_train)
```

## 3 Resultados

El modelo ajustado muestra dos coeficientes, cada uno representando una variable predictiva. Sin embargo, el error cuadratico medio es elevado, lo que indica que la prediccion no es muy precisa. Esto sugiere que los datos pueden requerir una mejor seleccion de variables o una transformacion para mejorar el ajuste del modelo.

## 4 Conclusion

La implementacion de la regresion lineal multiple en Python permite analizar la relacion entre multiples variables y predecir valores con base en los datos de entrenamiento. Aunque el modelo desarrollado en este caso no presento una alta precision, esta metodologia es util para comprender la influencia de diferentes factores en la variable dependiente. En el futuro, se podria mejorar el modelo mediante una seleccion mas rigurosa de datos o la aplicacion de tecnicas de optimizacion.

## 5 Referencias

IBM. (2020). Multiple Linear Regression. Recuperado de <https://www.ibm.com/docs/es/cognos-analytics/11.1.0?topic=tests-multiple-linear-regression>

Bagnato, J. (2020). Aprende machine learning. Leanpub.