

# UNIVERSIDAD DE GUADALAJARA



## CENTRO UNIVERSITARIO DE CIENCIAS EXACTAS E INGENIERÍAS

### Sistemas Inteligentes IV

#### Actividad 2. Regresión lineal y métricas de regresión

Pacheco Quintero Marco Antonio      213535019

19 de marzo de 2021

Semestre 2021A

Sección D01

## Objetivo

- Para la parte 1, realizar un programa en Python para aplicar una regresión lineal a los datos de los archivos adjuntos. Utilizar las herramientas sklearn para el modelo de regresión.
- Para la parte 2, realizar un programa en Python para aplicar una regresión lineal a los datos sobre la población del archivo adjunto. Utilizar las herramientas sklearn para el modelo de regresión. De igual forma predecir la población para el año 2020 con ayuda del modelo entrenado.

Para ambas partes utilizar las siguientes métricas de error para validar los modelos.

1. Error absoluto medio
2. Error cuadrático medio
3. Error mediano absoluto
4. Coeficiente  $R^2$
5. Coeficiente de explicación de la varianza

## Resultados

El resultado de la regresión lineal para los datos del archivo adjunto “df\_regresion\_lineal\_1.csv” se muestra en la Figura 1.

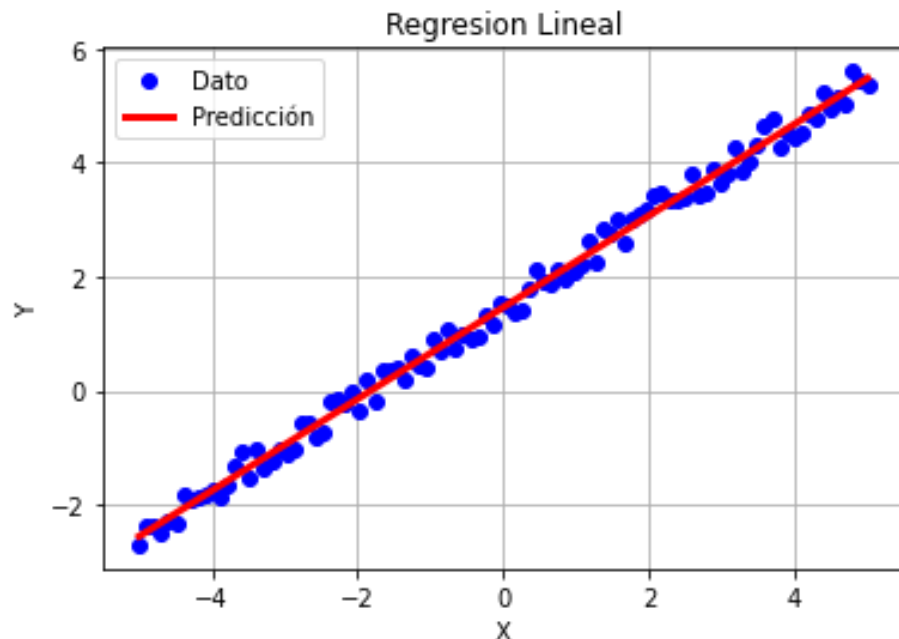


Figura 1. Regresión lineal de los datos del archivo adjunto 1.

Los resultados de las métricas de error para el modelo predicho fueron:

- Error absoluto medio

$$MAE(y, \hat{y}) = 0.16064031764688003$$

- Error cuadrático medio

$$MSE(y, \hat{y}) = 0.03307173080342715$$

- Error mediano absoluto

$$MedAE(y, \hat{y}) = 0.17157959685809088$$

- Coeficiente  $R^2$

$$R^2(y, \hat{y}) = 0.9940024504256648$$

- Coeficiente de explicación de la varianza

$$EVS(y, \hat{y}) = 0.9940024504256648$$

El resultado de la regresión lineal para los datos del archivo adjunto “df\_regresion\_lineal\_2.csv” se muestra en la Figura 2.

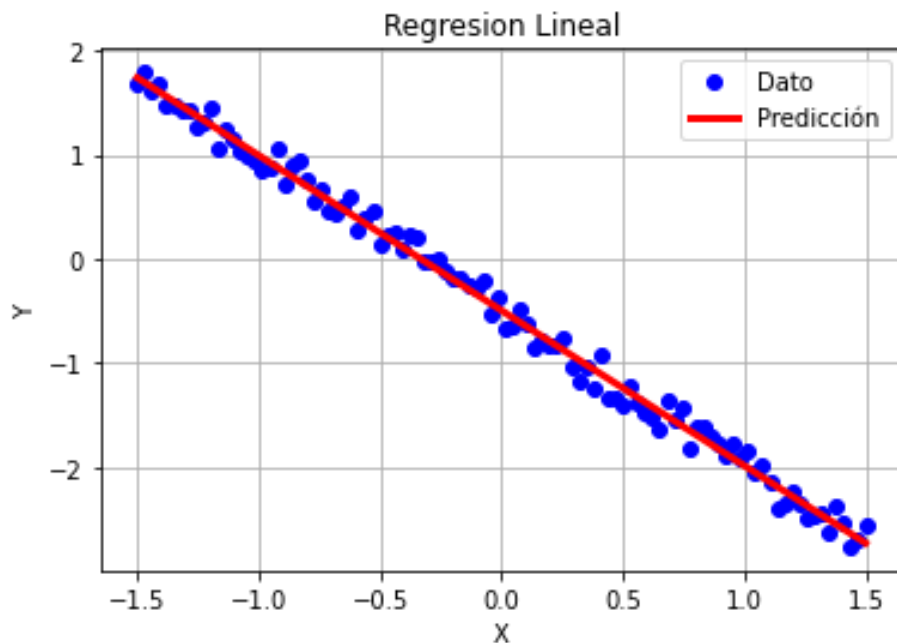


Figura 2. Regresión lineal de los datos del archivo adjunto 2.

Los resultados de las métricas de error para el modelo predicho fueron:

- Error absoluto medio

$$MAE(y, \hat{y}) = 0.09219580999368926$$

- Error cuadrático medio

$$MSE(y, \hat{y}) = 0.011747401311042304$$

- Error mediano absoluto

$$MedAE(y, \hat{y}) = 0.09279668280136555$$

- Coeficiente  $R^2$

$$R^2(y, \hat{y}) = 0.9931141742793483$$

- Coeficiente de explicación de la varianza

$$EVS(y, \hat{y}) = 0.9931141742793483$$

El resultado de la regresión lineal para los datos del archivo adjunto “df\_regresion\_lineal\_3.csv” se muestra en la Figura 3.

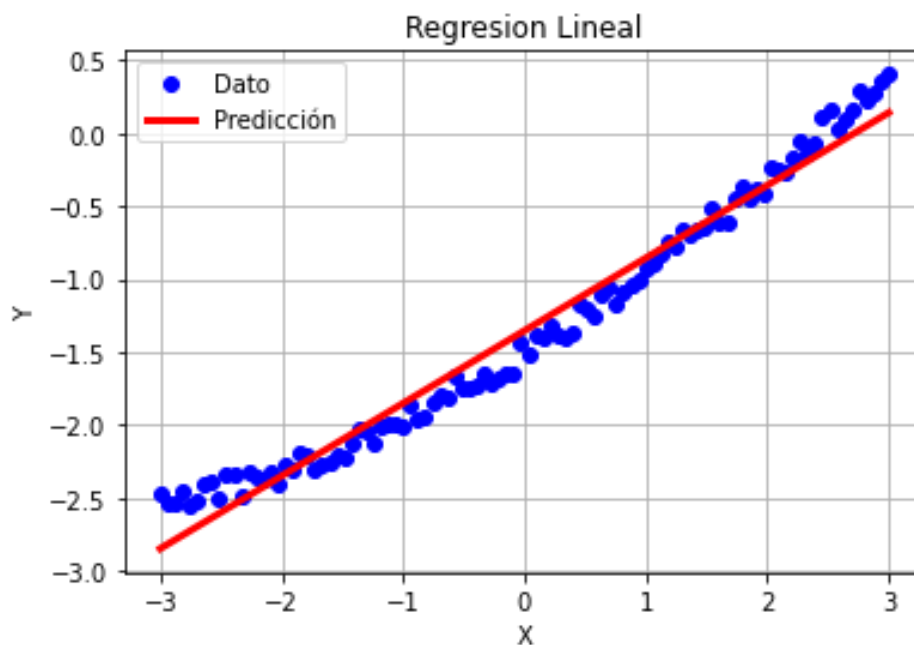


Figura 3. Regresión lineal de los datos del archivo adjunto 3.

Los resultados de las métricas de error para el modelo predicho fueron:

- Error absoluto medio

$$MAE(y, \hat{y}) = 0.12387517385319502$$

- Error cuadrático medio

$$MSE(y, \hat{y}) = 0.021754872061128278$$

- Error mediano absoluto

$$MedAE(y, \hat{y}) = 0.1016148385370356$$

- Coeficiente  $R^2$

$$R^2(y, \hat{y}) = 0.9721817814803592$$

- Coeficiente de explicación de la varianza

$$EVS(y, \hat{y}) = 0.9721817814803592$$

El resultado de la regresión lineal para los datos de población del archivo adjunto “df\_regresion\_lineal\_pop.csv” se muestra en la Figura 4.

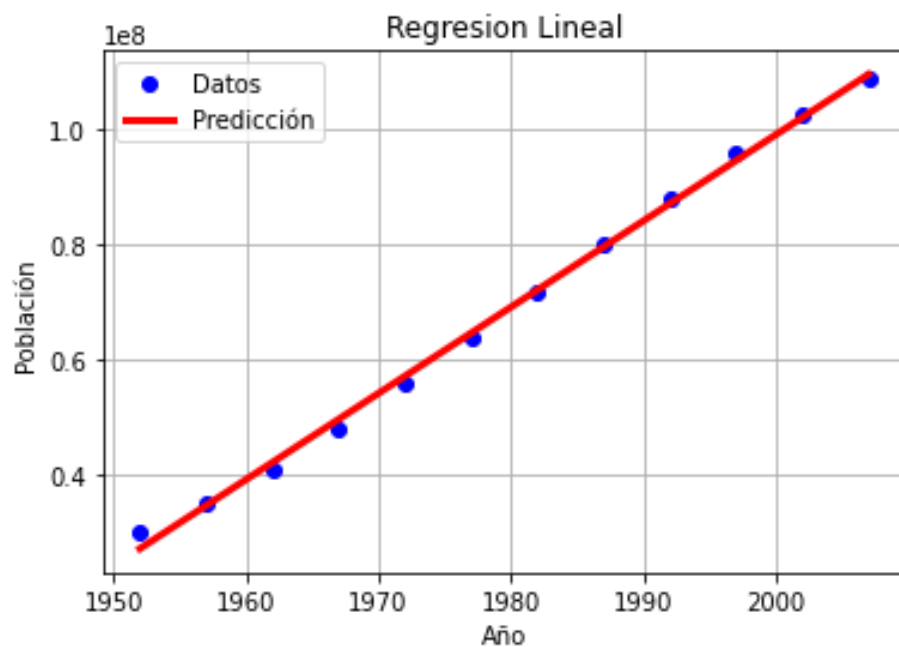


Figura 4. Regresión lineal de los datos de población del archivo adjunto.

Los resultados de las métricas de error para el modelo predicho fueron:

- Error absoluto medio

$$MAE(y, \hat{y}) = 1058079.682400902$$

- Error cuadrático medio

$$MSE(y, \hat{y}) = 1571199625216.2168$$

- Error mediano absoluto

$$MedAE(y, \hat{y}) = 974878.9458041191$$

- Coeficiente  $R^2$

$$R^2(y, \hat{y}) = 0.9976408962810307$$

- Coeficiente de explicación de la varianza

$$EVS(y, \hat{y}) = 0.9976408962810307$$

En cuanto a la población de México predicha para el año 2020 el resultado fue de **128,897,797** habitantes.

Al buscar la cifra oficial de población se tiene que en 2020 había **126,014,024** habitantes en México.

## Conclusión

Hay dos cosas a destacar que he observado en esta actividad: La primera es sobre las métricas de error obtenidas para el modelo de la población en México, para el MAE, MSE y MedAE se obtuvieron valores muy grandes que a simple vista no nos dan una idea tan clara de que tan buena fue la aproximación, esto debido a que se trabaja con datos tan grandes (de poco más de 1 millón) y por lo tanto tenemos errores muy grandes, no ocurre así con los coeficientes  $R^2$  y  $EVS$  los cuales nos dan una idea más directa. La segunda es sobre la predicción del modelo obtenido para la población, la cifra estimada fue bastante cercana a la cifra oficial, teniendo un error de casi 3 millones, lo cual es relativamente poco, teniendo en cuenta que hablamos de cientos de miles.