

Regresion Lineal en Python

Julio Cesar Torres Marquez

Marzo 2025

1 Introduccion

La regresion lineal es una tecnica utilizada en el analisis de datos para predecir valores desconocidos a partir de datos relacionados y conocidos. Se modela matematicamente como una ecuacion lineal entre una variable dependiente y una independiente.

2 Metodologia

2.1 Instalacion de librerias

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

- pandas
- seaborn
- matplotlib
- scikit-learn

El siguiente comando permite instalarlas:

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

2.2 Carga y visualizacion de datos

Se importan las librerias y se carga el conjunto de datos:

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

Cargar los datos

```
data = pd.read_csv("./articulos_ml.csv")
print("Forma de los datos:", data.shape)
print(data.head())
```

2.3 Graficos y analisis inicial

Para visualizar los datos, se genera un histograma:

```
data.drop(['Title', 'url', 'Elapsed days'], axis=1).hist()
plt.show()
```

Ademas, se crea un grafico de dispersion para analizar la relacion entre las variables:

```
plt.scatter(data['Word count'], data['# Shares'])
plt.show()
```

2.4 Regresion lineal con Scikit-Learn

Se implementa el modelo de regresion lineal:

```
X_train = np.array(data[['Word count']])
y_train = data['# Shares'].values
```

```
modelo = linear_model.LinearRegression()
modelo.fit(X_train, y_train)
```

Predicciones

```
y_pred = modelo.predict(X_train)
```

Los coeficientes obtenidos se muestran a continuacion:

```
print("Coeficiente:", modelo.coef_)
print("Interseccion:", modelo.intercept_)
```

3 Conclusion

La regresion lineal es una herramienta util para predecir tendencias, sin embargo, en este caso, el modelo presenta una varianza alta, lo que indica que se necesitan mas variables para mejorar la precision.

4 Referencias

- <https://aws.amazon.com/es/what-is/linear-regression/>
- Ignacio Bagnato, J. (2020). Aprende Machine Learning. Leanpub.