

Regresión Lineal en Python

1 Introducción

La regresión lineal es un método estadístico utilizado en aprendizaje automático para modelar la relación entre una variable dependiente y una o más variables independientes. Se basa en la ecuación de una línea recta:

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

En este documento, exploramos la implementación de la regresión lineal múltiple en Python para predecir la cantidad de veces que un artículo será compartido en redes sociales.

2 Metodología

El procedimiento se basó en los siguientes pasos:

1. Preprocesamiento de datos: Se cargaron los datos y se seleccionaron las variables relevantes.
2. Creación de variables: Se generó una nueva variable combinando enlaces, comentarios e imágenes.
3. Construcción del modelo: Se utilizó la librería **scikit-learn** para entrenar un modelo de regresión lineal múltiple.
4. Evaluación del modelo: Se calcularon los coeficientes, el error cuadrático medio y el puntaje de varianza.

2.1 Código Implementado

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

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

dataX2 = pd.DataFrame()
```

```

dataX2["Word count"] = filtered_data["Word count"]
dataX2["suma"] = suma

XY_train = np.array(dataX2)
z_train = filtered_data['# Shares'].values

regr2 = linear_model.LinearRegression()
regr2.fit(XY_train, z_train)

print('Coefficients:', regr2.coef_)
print("Mean squared error:", mean_squared_error(z_train, regr2.predict(XY_train)))
print('Variance score:', r2_score(z_train, regr2.predict(XY_train)))

```

3 Resultados

Los coeficientes obtenidos indican el impacto de cada variable en la cantidad de veces que un artículo es compartido. Se observó un error cuadrático medio elevado, lo que sugiere que el modelo puede no ser altamente predictivo.

4 Conclusión

A través de este experimento, se pudo implementar un modelo de regresión lineal múltiple en Python. Sin embargo, la precisión del modelo podría mejorar utilizando más características o técnicas avanzadas como la reducción de dimensionalidad.