

Inteligencia Artificial

Act 10: Programando Regresión Lineal Múltiple en Python

Estudiante: José Luis Calderón Galarza - 2132939

Docente: Luis Ángel Gutiérrez Rodríguez

1 Introducción

La regresión lineal múltiple es una extensión de la regresión lineal simple en la que se modela la relación entre una variable dependiente y y dos o más variables independientes X . En este caso, se busca ajustar un plano en un espacio tridimensional para minimizar el error cuadrático medio. Este modelo se emplea en diversas aplicaciones de análisis predictivo y tendencias.

2 Metodología

En esta actividad, se implementó un modelo de regresión lineal múltiple en Python para analizar la relación entre la cantidad de palabras de un artículo, el número de enlaces, comentarios e imágenes, y la cantidad de veces que es compartido en redes sociales. A continuación, se describe cada parte del código utilizado:

1. **Importación de librerías:** Se cargan las bibliotecas necesarias para la manipulación de datos, visualización y construcción del modelo.

```
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
plt.rcParams['figure.figsize'] = (16, 9)
plt.style.use('ggplot')
from sklearn import linear_model
from sklearn.metrics import mean_squared_error, r2_score
```

2. **Carga y exploración de datos:** Se carga el conjunto de datos desde un archivo CSV y se filtran los valores extremos.

```
data = pd.read_csv("./articulos_ml.csv")
filtered_data = data[(data['Word count'] <= 3500) & (data['# Shares'] <= 80000)]
```

3. **Creación de nueva variable:** Se suma el número de enlaces, comentarios e imágenes para crear una variable adicional.

```
suma = (filtered_data["# of Links"] + filtered_data['# of comments']).fillna(0) + filtered_data["# of images"]
dataX2 = pd.DataFrame()
dataX2["Word count"] = filtered_data["Word count"]
dataX2["suma"] = suma
```

4. **Entrenamiento del modelo:** Se seleccionan las variables independientes y dependientes y se entrena el modelo de regresión lineal múltiple.

```
XY_train = np.array(dataX2)
z_train = filtered_data['# Shares'].values
regr2 = linear_model.LinearRegression()
regr2.fit(XY_train, z_train)
```

5. **Predicciones y evaluación del modelo:** Se generan predicciones y se calculan los coeficientes, el error cuadrático medio y el puntaje de varianza.

```
z_pred = regr2.predict(XY_train)
print('Coefficients:', regr2.coef_)
print("Mean squared error: %.2f" % mean_squared_error(z_train, z_pred))
print('Variance score: %.2f' % r2_score(z_train, z_pred))
```

6. **Visualización del modelo:** Se genera una representación tridimensional de la regresión lineal múltiple.

```
fig = plt.figure()
ax = fig.add_subplot(111, projection='3d')
xx, yy = np.meshgrid(np.linspace(0, 3500, num=10), np.linspace(0, 60, num=10))
nuevoX = (regr2.coef_[0] * xx)
nuevoY = (regr2.coef_[1] * yy)
z = nuevoX + nuevoY + regr2.intercept_
z = z.reshape(xx.shape)
ax.plot_surface(xx, yy, z, alpha=0.2, cmap='hot')
ax.scatter(XY_train[:, 0], XY_train[:, 1], z_train, c='blue', s=30)
ax.scatter(XY_train[:, 0], XY_train[:, 1], z_pred, c='red', s=40)
ax.view_init(elev=30., azim=65)
ax.set_xlabel('Cantidad de Palabras')
ax.set_ylabel('Cantidad de Enlaces, Comentarios e Imágenes')
ax.set_zlabel('Compartido en Redes')
ax.set_title('Regresión Lineal con Múltiples Variables')
plt.show()
```

7. **Predicción para 2000 palabras:** Se usa el modelo entrenado para predecir cuántas veces será compartido un artículo con 2000 palabras, 10 enlaces, 4 comentarios y 6 imágenes.

```
z_Dosmil = regr2.predict([[2000, 10+4+6]])
print(int(z_Dosmil[0]))
```

3 Resultados

3.1 Visualización del modelo en 3D

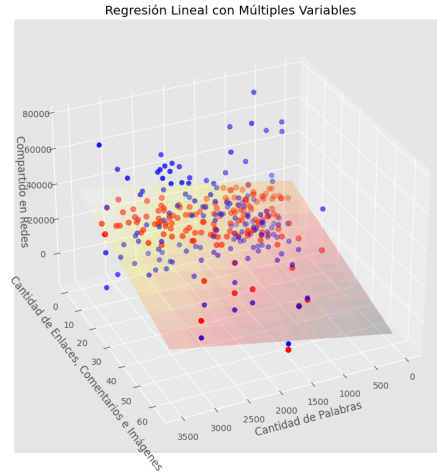


Figure 1: Plano de regresión lineal múltiple ajustado a los datos filtrados.

4 Conclusión

Se ha implementado con éxito un modelo de regresión lineal múltiple para predecir el número de veces que un artículo será compartido en redes sociales, considerando la cantidad de palabras y la cantidad de enlaces, comentarios e imágenes. Finalmente, se realizó una predicción para un artículo con 2000 palabras y ciertas características adicionales, obteniendo un valor estimado de 20518.