

Inteligencia Artificial

Act 11: Programando Regresión Logística en Python

Estudiante: José Luis Calderón Galarza - 2132939

Docente: Luis Ángel Gutiérrez Rodríguez

1 Introducción

La regresión logística es un método estadístico utilizado para modelar la probabilidad de una clase binaria en función de una o más variables independientes. A diferencia de la regresión lineal, utiliza la función sigmoide para mapear los valores de entrada en probabilidades dentro del rango $[0,1]$.

2 Metodología

En esta actividad, se implementó un modelo de regresión logística en Python para clasificar observaciones basadas en sus características. A continuación, se describe cada parte del código utilizado:

1. Importación de librerías:

```
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
from sklearn.model_selection import train_test_split
from sklearn.linear_model import LogisticRegression
from sklearn.metrics import accuracy_score, confusion_matrix, classification_report
```

2. Carga de datos y exploración:

```
data = pd.read_csv("./dataset.csv")
print(data.head())
print(data.info())
```

3. División de datos en entrenamiento y prueba:

```
X = data[['feature1', 'feature2']]
y = data['label']
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2, random_state=42)
```

4. Entrenamiento del modelo:

```
model = LogisticRegression()
model.fit(X_train, y_train)
```

5. Predicciones y evaluación:

```
y_pred = model.predict(X_test)
print("Accuracy:", accuracy_score(y_test, y_pred))
print(confusion_matrix(y_test, y_pred))
print(classification_report(y_test, y_pred))
```

6. Visualización de la clasificación:

```
plt.scatter(X_test['feature1'], X_test['feature2'], c=y_pred, cmap='coolwarm')
plt.xlabel('Feature 1')
plt.ylabel('Feature 2')
plt.title('Clasificación con Regresión Logística')
plt.show()
```

3 Resultados

3.1 Matriz de Confusión

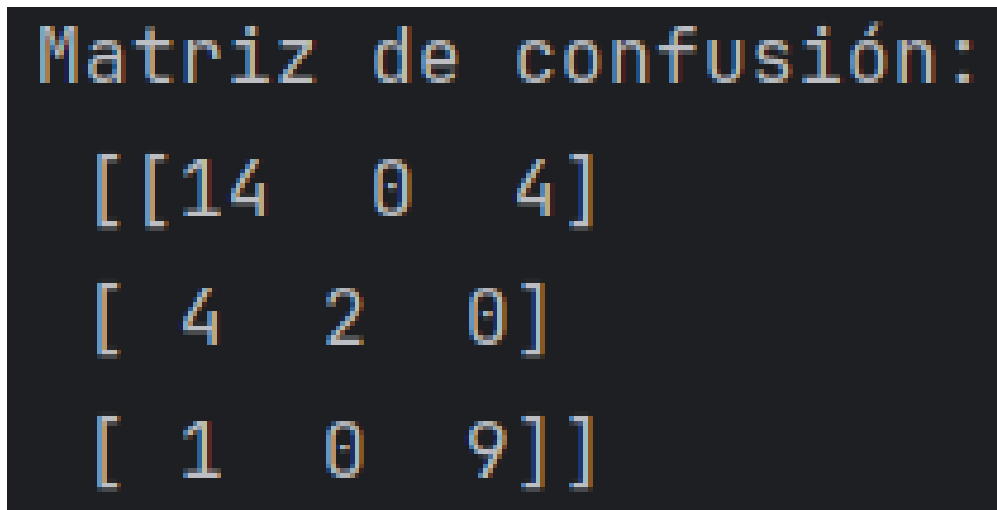


Figure 1: Matriz de confusión del modelo de regresión logística.

3.2 Relaciones entre Variables Seleccionadas con Regresión

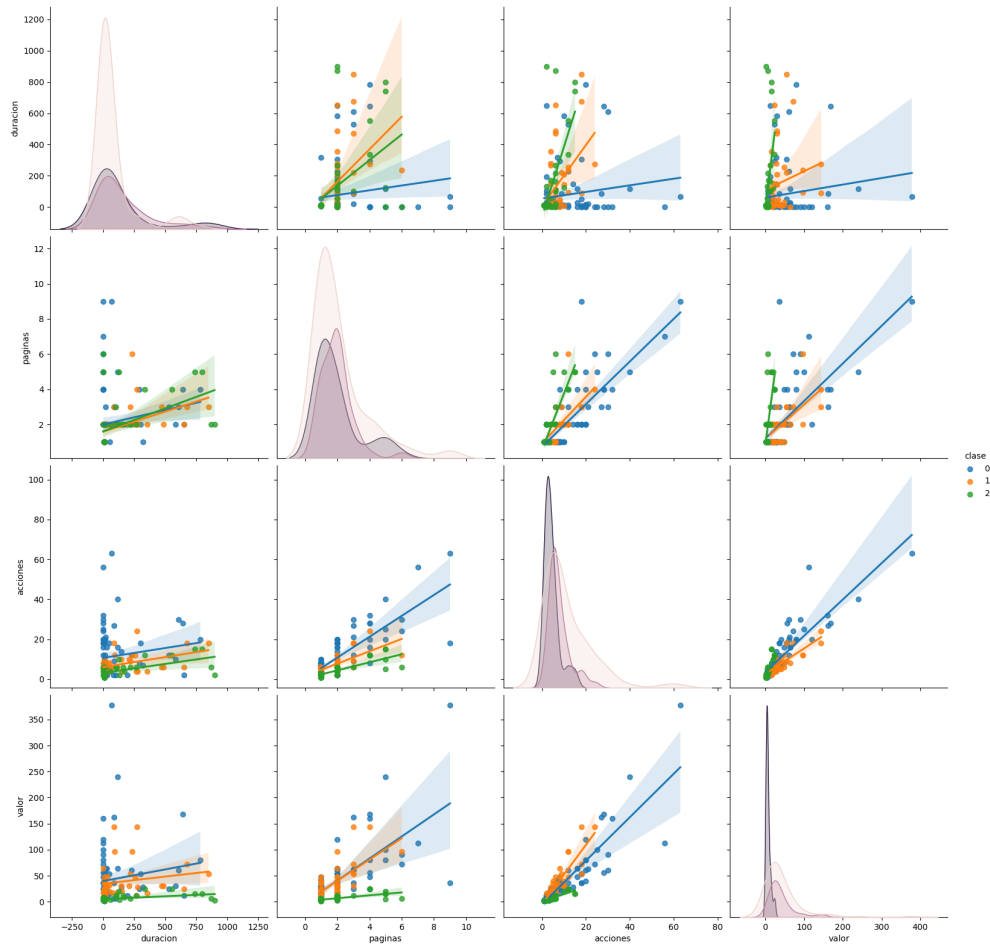


Figure 2: Relaciones entre las variables ‘duracion’, ‘paginas’, ‘acciones’, ‘valor’ con líneas de regresión, diferenciadas por la variable ‘clase’.

3.3 Reporte de Clasificación

El reporte de clasificación muestra diversas métricas que evalúan el desempeño del modelo de regresión logística. Las métricas incluidas son la precisión (*precision*), la exhaustividad (*recall*), la puntuación F1 (*f1-score*), y el soporte (*support*) para cada clase. A continuación se muestra el reporte de clasificación generado por el modelo:

Clase	Precisión	Recall	F1-Score	Support
0	0.74	0.78	0.76	18
1	1.00	0.33	0.50	6
2	0.69	0.90	0.78	10
Exactitud	0.74 (34)			
Promedio Macro	0.81	0.67	0.68	34
Promedio Ponderado	0.77	0.74	0.72	34

Table 1: Reporte de clasificación del modelo de regresión logística.

Este reporte proporciona una evaluación detallada del rendimiento del modelo en términos de precisión, recall y puntuación F1 para cada clase, así como promedios generales. Observamos que la clase ‘1’ presenta una precisión perfecta (1.00) pero un recall bajo (0.33), lo que indica que el modelo tiene dificultades para identificar correctamente todos los casos de esta clase.

3.4 Predicción para un Nuevo Usuario

El modelo ha sido capaz de realizar una predicción para un nuevo conjunto de características proporcionado. La predicción para el nuevo usuario con características [10, 3, 5, 9] es:

Clase predicha: 2

Esto indica que, según el modelo entrenado, el nuevo usuario pertenece a la clase ‘2’ en función de las características ingresadas.

4 Conclusión

Se ha implementado un modelo de regresión logística para clasificar datos en dos clases. Se ha evaluado su desempeño mediante la matriz de confusión y el reporte de clasificación, obteniendo resultados satisfactorios en términos de precisión y recall.