

Regresión Logística en Python

1 Introducción

La regresión logística es un modelo estadístico utilizado para clasificación. A diferencia de la regresión lineal, donde la salida es continua, la regresión logística se emplea cuando la variable dependiente es discreta, es decir, pertenece a una categoría específica. Se basa en la función sigmoide:

$$P(Y = 1) = \frac{1}{1 + e^{-(b+mX)}} \quad (1)$$

Este informe documenta la implementación de un modelo de regresión logística en Python para clasificar el sistema operativo de un usuario según sus patrones de navegación en un sitio web.

2 Metodología

Los pasos seguidos fueron:

1. Carga y exploración de los datos.
2. Visualización de la distribución de clases y variables.
3. Preprocesamiento de datos y división en conjuntos de entrenamiento y validación.
4. Entrenamiento del modelo de regresión logística con `scikit-learn`.
5. Evaluación del modelo mediante métricas de precisión, matriz de confusión y reporte de clasificación.
6. Clasificación de nuevos usuarios con datos ficticios.

2.1 Código Implementado

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

```

from sklearn.metrics import classification_report, confusion_matrix, accuracy_score

dataframe = pd.read_csv("usuarios_win_mac_lin.csv")
X = np.array(dataframe.drop(['clase'], axis=1))
y = np.array(dataframe['clase'])
X_train, X_validation, Y_train, Y_validation = model_selection.train_test_split(X, y, test_s

model = linear_model.LogisticRegression(max_iter=200)
model.fit(X_train, Y_train)

predictions = model.predict(X_validation)
print("Precisión:", accuracy_score(Y_validation, predictions))
print("Matriz de Confusión:", confusion_matrix(Y_validation, predictions))
print("Reporte de Clasificación:", classification_report(Y_validation, predictions))

X_new = pd.DataFrame({'duracion': [10], 'paginas': [3], 'acciones': [5], 'valor': [9]})
print("Predicción usuario nuevo:", model.predict(X_new))

```

3 Resultados

Los resultados del modelo incluyen:

- Precisión del modelo: alrededor del 77% en promedio.
- Matriz de confusión que muestra los errores en la clasificación.
- Un usuario ficticio con valores específicos fue clasificado correctamente según el modelo.

4 Conclusión

Se logró implementar con éxito un modelo de regresión logística para clasificación de sistemas operativos basado en patrones de navegación. Aunque los resultados fueron buenos, se podría mejorar la precisión del modelo con más datos o ajustes en los hiperparámetros.