

# Actividad 11: Programando Regresión Logística en Python

García Herrera Carlos Eduardo

March 2025

## 1 Introduction

La regresión lineal es un modelo estadístico que busca establecer una relación lineal entre una variable dependiente (o de salida) y una o más variables independientes (o explicativas). En términos simples, intenta encontrar la "mejor línea" que prediga el valor de la variable dependiente a partir de las variables independientes.

## 2 Metodología

### 2.1 Parte 1: Creación del Ambiente Virtual

Las siguientes líneas de código, se encargan de crear un ambiente virtual llamado VirtualEnv con las librerías necesarias para ejecutar el script

```
#Automatic creation of a virtual environment to run the script and install the libraries
import subprocess
import os
import venv
import sys
script_dir = os.path.dirname(os.path.realpath(__file__))
env_name = os.path.join(script_dir, "VirtualEnv")
if os.path.exists(os.path.join(script_dir, "VirtualEnv")):
    #Checks if the VirtualEnv is activated
    if sys.prefix == sys.base_prefix:
        print("Activating the Virtual Environment...")
        python_exe = os.path.join(env_name, "Scripts", "python")
        subprocess.run([python_exe, __file__])
    else:
        print("Installing the Required Libraries on a New Virtual Environment")
        venv.create(env_name, with_pip=True)

libraries = ["scikit-learn", "matplotlib", "seaborn", "pandas", "numpy"]
```

```

for lib in libraries:
    subprocess.run([os.path.join(env_name, "Scripts", "pip"), "install", lib],
                    check=True)

python_exe = os.path.join(env_name, "Scripts", "python")
subprocess.run([python_exe, __file__])

```

## 2.2 Parte 2: Analisis of the data

Se cargan los datos contenidos en articulos\_ml.csv y se analizan sus distintas propiedades

```

#cargamos los datos de entrada
dataframe = pd.read_csv("./usuarios_win_mac_lin.csv")

#vemos los primeros Registros
print(dataframe.head())

#vemos las caracteristicas del dataframe
print(dataframe.describe())

#analizaremos cuantos resultados tenemos de cada tipo usando la función groupby y vemos
#que tenemos 86 usuarios \Clase 0", es decir Windows, 40 usuarios Mac y 44 de Linux.
print(dataframe.groupby('clase').size())

dataframe.drop(['clase'],axis=1).hist()
plt.show()

sb.pairplot(dataframe.dropna(), hue='clase',height=4,vars=["duracion", "paginas","acciones"])
plt.show()

```

## 2.3 Parte 3: Regresion Lineal

A continuacion se hace el calculo de la regresion Lineal:

```

#Creacion del Modelo de Regresion Logistica
X = np.array(dataframe.drop(['clase'],axis=1))
y = np.array(dataframe['clase'])
X.shape

model = linear_model.LogisticRegression()
model.fit(X,y)

predictions = model.predict(X)
print("Predicciones:")
print(predictions[0:5])

```

```

print("Score:")
print(model.score(X,y))

#Validacion del Modelo
validation_size = 0.20
seed = 7
X_train, X_validation, Y_train, Y_validation = model_selection.train_test_split(X, y, test_s

name='Logistic Regression'
kfold = model_selection.KFold(n_splits=10, random_state=seed,shuffle=True)
cv_results = model_selection.cross_val_score(model, X_train,
Y_train, cv=kfold, scoring='accuracy')
msg = "%s: %f (%f)" % (name, cv_results.mean(), cv_results.std())
print(msg)

predictions = model.predict(X_validation)

print("Score")
print(accuracy_score(Y_validation, predictions))

print("Matriz Confusion")
print(confusion_matrix(Y_validation, predictions))

print("Reporte de Clasificacion")
print(classification_report(Y_validation, predictions))

##Clasificacion de Nuevos Valores
X_new = pd.DataFrame({'duracion': [10], 'paginas': [3],
'acciones': [5], 'valor': [9]})
new_predict=model.predict(X_new)
print("Nueva Prediccion:")
print(new_predict)

input("Press any key to Exit")

```

### 3 Resultados

Al ejecutar el script de pyhton la informacion obtenida es la siguiente:

#### 3.1 Analisis de los Datos

La informacion de los datos es la siguiente:

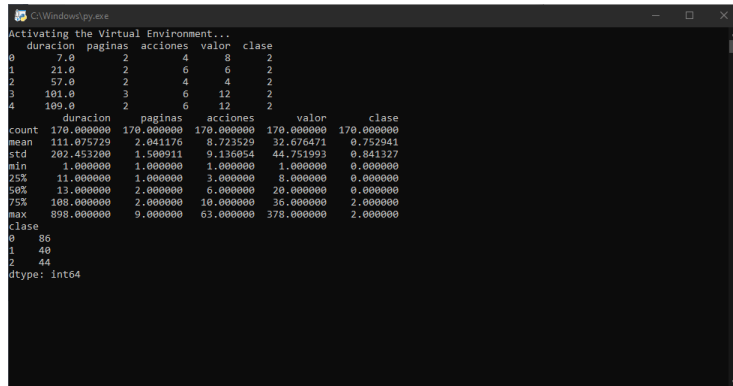


Figure 1: Informacion de los Datos

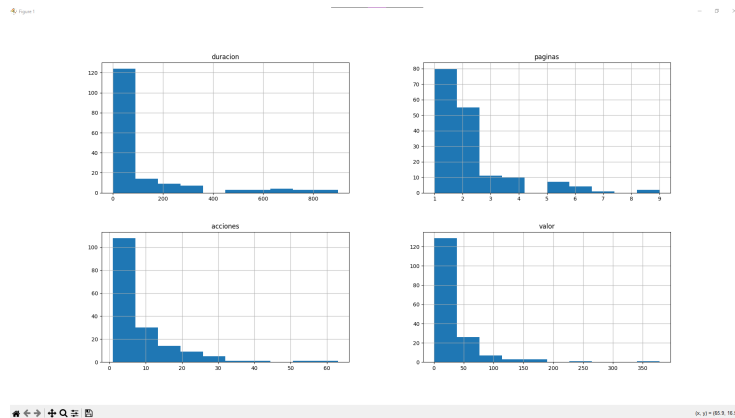


Figure 2: Numero de elementos por Post

### 3.2 Regresion Logistica

Los resultados de la regresion Logistica fueron los siguientes:

## 4 Conclusion

La regresión logística es una herramienta poderosa para abordar problemas de clasificación binaria y se ha aplicado con éxito en diversas áreas. Aunque es un modelo relativamente simple, su capacidad para predecir probabilidades y clasificar en función de características específicas la convierte en una técnica valiosa. Sin embargo, su rendimiento puede verse afectado por la calidad de los datos y la correcta selección de variables, por lo que es crucial realizar un análisis exhaustivo para obtener resultados confiables.



[illegible]

Figure 5: Datos calculados de la regresion Logistica con Validacion