

Act 11: Programando Regresión Logística en Python

Patricio Ricardí

March 2025

¿Qué es la regresión logística?

Es un método para predecir categorías (ej: "Windows/Mac/Linux") usando variables numéricas. A diferencia de la regresión lineal:

- Predice probabilidades (valores entre 0 y 1)
- Usa la función sigmoide:

$$\sigma(z) = \frac{1}{1 + e^{-z}}$$

donde $z = \beta_0 + \beta_1 x_1 + \dots + \beta_n x_n$

Pasos realizados

1. ****Cargar datos****: Información sobre usuarios y sus sistemas operativos
2. ****Preparar variables****:
 - X: Características de los usuarios (variables predictoras)
 - y: Sistema operativo (variable objetivo)
3. ****Entrenar modelo****: Usamos regresión logística multinomial con scikit-learn
4. ****Evaluar resultados****:
 - Matriz de confusión
 - Precisión, Recall y F1-score
 - Validación cruzada (5 divisiones)

Código clave

```
[language=Python] Entrenar modelo de regresión logística modelo = LogisticRegression( multi_class=' multinomial', solver =' lbfgs')modelo.fit(X_entrenamiento,y_entrenamiento)
Evaluar predicciones print("Precisión:", modelo.score(X_prueba,y_prueba))print(" Matriz de confusión :
", confusion_matrix(y_prueba,y_predicha))
```

Resultados importantes

- **Precisión general:** 77%
- **Validación cruzada:** 74% promedio
- **Mejor categoría:**
 - Precisión: 86%
 - Recall: 84%
 - F1-score: 84%
- Predijo correctamente el sistema operativo de un usuario ficticio

Conclusiones

- Funciona bien para clasificación con datos simples
- Ventajas:
 - Fácil de interpretar
 - Rápido de entrenar
- Mejoras futuras:
 - Usar regularización
 - Probar modelos más complejos (ej: redes neuronales)
 - Optimizar parámetros