

REGRESIÓN LOGÍSTICA

BY;

**JESUS PERAZA 20241578198
ANDRES ROJAS 20241578153**

¿QUE ES?

Es una tecnica o modelo estadistico que permite predecir la posibilidad de que pase un evento.

Esta es usada cuando la respuesta tiene dos opciones 0 o 1, si o no.

$$0 \leq p \leq 1$$

La Regresion logistica, esta basada en la relación entre las variables independientes y la probabilidad de que un evento no es lineal, sino sigmoidal.

A diferencia de la regresión lineal, que predice valores continuos, la regresión logística predice



¿POR QUE NO LINEAL?

01. Puede dar valores menores a 0 o mayores a 1

02. La relación entre X y la probabilidad no es lineal

03. Los errores no son normalmente distribuidos

04. El modelo se vuelve inestable

FORMULA

$$z = b_0 + b_1x_1 + b_2x_2 + \cdots + b_nx_n$$

Esta fórmula es una "combinación lineal"

Es una forma de mezclar varias variables x_1, x_2, \dots, x_n usando pesos (b_1, b_2, \dots)

Ejemplo

Si quieres saber si un estudiante aprobará, tomas en cuenta:

- Horas estudiadas (x_1)
- Nota en parciales (x_2)
- Asistencia (x_3)

Y cada una tiene un peso diferente en la decisión.

$$z = b_0 + b_1(\text{horas}) + b_2(\text{parcial}) + b_3(\text{asistencia})$$

LA REPRESENTACION DE Z

logit es una transformación matemática que convierte una probabilidad (p) en un número que puede ir desde $-\infty$ hasta $+\infty$.

Cuando tienes una probabilidad:

- $p = 0.30$
- $p = 0.75$
- $p = 0.15$

Eos valores solo van de 0 a 1.

Pero para hacer un modelo lineal necesitamos valores que puedan ir de $-\infty$ a $+\infty$.

El logit convierte la probabilidad en algo que sí puede recorrer todo ese rango.

$$\ln\left(\frac{p}{1-p}\right)$$

ODDS

son una forma de expresar qué tan probable es que ocurra un evento comparado con qué tan probable es que NO ocurra.

- p = probabilidad de que algo ocurra
- $1-p$ = probabilidad de que NO ocurra

supongamos que la probabilidad de aprobar un examen es:

$$p=0.75$$

Los odds = 3 significa que "es 3 veces más probable aprobar que no aprobar".

$$\text{odds} = \frac{p}{1 - p}$$

$$\text{odds} = \frac{0.75}{0.25} = 3$$

VENTAJAS Y DESVENTAJAS

Ventajas

- Probabilidades claras
- Fácil de interpretar
- Entrena rápido
- Funciona con datasets medianos
- Resistente a outliers
- Permite medir importancia de variables
- Muy utilizado en medicina, biología y ciencias sociales

Desventajas

- Solo maneja límites lineales en el espacio del logit
- No capta relaciones complejas (a menos que agregues polinomios)
- Sensible a la multicolinealidad
- Mal desempeño con clases muy desbalanceadas

PHYTON

Aquí se muestra donde se hace, por medio de la librería sklearn, el modelo predictivo de la Regresión logística.

- El segundo es un ejemplo con el modelo entrenado.

```
# 2. Inicializar y Entrenar el Modelo
modelo_logistico = LogisticRegression(random_state=42, max_iter=200)
modelo_logistico.fit(X_reshaped, Y)

# 3. Mostrar Coeficientes (para la interpretación)
print("---- Resultados del Entrenamiento ---")
print(f"Intercepto (B): {modelo_logistico.intercept_[0]:.3f}")
print(f"Coeficiente (B1 para Horas_Estudio): {modelo_logistico.coef_[0][0]:.3f}")

] ✓ 0.ls
```

RECURSOS USADOS



https://online.stat.psu.edu/stat462/node/207/?utm_source.com

https://en.wikipedia.org/wiki/Logistic_regression?utm_source.com

https://biz.libretexts.org/Workbench/OER_Textbook_for_Data_Analytics/06%3A_Logistic_Regression/6.11%3A_Appendix_-_Formulas_used_in_Logistic_Regression?utm_source.com

https://www.ugr.es/~mvargas/2.DiaNocioRegLogistica.pdf?utm_source.com

MUCHAS GRACIAS

Por ver esta presentación