

Análisis Predictivo y Gestión de Datos

Sesión 5: Regresión Logística

Oscar Leonardo Rincón León

24 de abril de 2025

Objetivo de la sesión

- Entender la lógica de la regresión logística y su función sigmoide.
- Usarla para clasificar de forma binaria: conectividad baja vs. alta.
- Evaluar su rendimiento con métricas como precisión, F1-score y curva ROC.

¿Por qué no usar regresión lineal para clasificación?

- La regresión lineal predice valores continuos sin límite.
- Para clasificación necesitamos probabilidades entre 0 y 1.
- La regresión logística lo soluciona usando la función sigmoide.

¿Qué es la regresión logística? (1/2)

- Modelo de clasificación binaria.
- Predice la probabilidad de que una observación pertenezca a la clase 1.
- Usa una combinación lineal de las variables:

$$z = \beta_0 + \beta_1 x_1 + \cdots + \beta_n x_n$$

¿Qué es la regresión logística? (2/2)

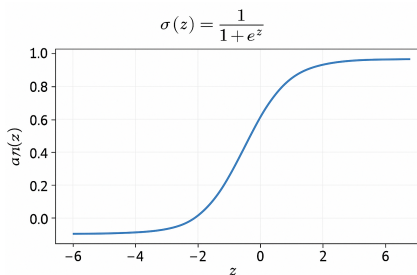
- Aplica la función sigmoide a z :

$$P(y = 1 \mid x) = \frac{1}{1 + e^{-z}}$$

- La salida es una probabilidad entre 0 y 1.
- Se predice clase 1 si $P \geq 0,5$ (o se ajusta el umbral).

Visualización de la función sigmoide

- Transforma cualquier valor real a una probabilidad.
- Representa la probabilidad de clase positiva.



¿Qué son los odds y el logit?

- **Odds:** cociente entre la probabilidad de éxito y de fracaso:

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

- **Logit:** logaritmo de los odds:

$$\log\left(\frac{p}{1 - p}\right) = z$$

Interpretación de los coeficientes

- Un aumento de 1 unidad en x_i incrementa los **log-odds** en β_i .
- En términos de odds:

$$\text{odds nuevos} = \text{odds actuales} \times e^{\beta_i}$$

- La relación con la probabilidad no es lineal (curva sigmoide).

- Métricas comunes:
 - Precisión, Recall, F1-score
 - Matriz de confusión
 - Curva ROC y AUC
- Se puede ajustar el umbral para mejorar el rendimiento en clase minoritaria.

Ventajas de la regresión logística

- Fácil de entrenar e interpretar.
- Permite estimar probabilidades.
- Buen punto de partida antes de usar modelos más complejos.

- La regresión logística modela probabilidades para clasificación binaria.
- Puede expresarse como función sigmoide o como logit.
- Es evaluada con precisión, F1 y curvas ROC.