

1. NORMALIZACIÓN DE DATOS

DEFINICIÓN 1: Normalización por el máximo.

Dado un conjunto de datos $X = \{x_1, x_2, \dots, x_n\}$, la normalización por el máximo transforma cada dato x_i en x'_i según la fórmula:

$$x'_i = \frac{x_i}{\max(X)}$$

DEFINICIÓN 2: Normalización Min-Max.

Dado un conjunto de datos $X = \{x_1, x_2, \dots, x_n\}$, la normalización Min-Max transforma cada dato x_i en x'_i según la fórmula:

$$x'_i = \frac{x_i - \min(X)}{\max(X) - \min(X)}$$

DEFINICIÓN 3: Estandarización (Z-score).

Dado un conjunto de datos $X = \{x_1, x_2, \dots, x_n\}$, la estandarización transforma cada dato x_i en x'_i según la fórmula:

$$x'_i = \frac{x_i - \mu}{\sigma}$$

donde μ es la media del conjunto de datos y σ es la desviación estándar.

2. DISCRETIZACIÓN DE DATOS

DEFINICIÓN 4: Discretización de igual amplitud.

Dado un conjunto de datos continuo $X = \{x_1, x_2, \dots, x_n\}$ y un número k de intervalos, la discretización de igual amplitud divide el rango de los datos en k intervalos de igual tamaño. Cada dato x_i se asigna al intervalo correspondiente. El tamaño de cada intervalo es:

$$\text{Tamaño del intervalo} = \frac{\max(X) - \min(X)}{k}$$

DEFINICIÓN 5: Discretización de igual frecuencia.

Dado un conjunto de datos continuo $X = \{x_1, x_2, \dots, x_n\}$ y un número k de intervalos, la discretización de igual frecuencia divide los datos en k intervalos de tal manera que cada intervalo contenga aproximadamente el mismo número de datos. Cada dato x_i se asigna al intervalo correspondiente.

3. DATOS CATEGÓRICOS

DEFINICIÓN 6: Codificación One-Hot.

Dado un conjunto de datos categóricos $C = \{c_1, c_2, \dots, c_m\}$ con m categorías distintas, la codificación One-Hot transforma cada categoría c_j en un vector binario de longitud m , donde el elemento correspondiente a la categoría es 1 y los demás son 0.

Ejemplo: En la siguiente tabla se aplica la codificación One-Hot a la variable categórica «Color» con las categorías «Rojo», «Verde» y «Azul».

Color		Rojo	Verde	Azul
Roj	$\xrightarrow{\text{One-Hot}}$	1	0	0
Verde		0	1	0
Azul		0	0	1
Verde		0	1	0
Roj		1	0	0