



# Escalamiento de variables

.....

Dr. Gaddiel Desirena López

Estandarización

Normalización basada en la media

Escalamiento de valores máximo y mínimo

Escalamiento de máximo absoluto

Escalamiento por cuantiles

- ▶ Muchos algoritmos de aprendizaje automático son sensibles a la escala y magnitud de las características.
  - ▶ Modelos lineales
  - ▶ Modelos que dependen de los cálculos de la distancia.

Las entidades con rangos de valores más grandes tienden a dominar las entidades con rangos más pequeños.

- ▶ Comparar la importancia de cada característica.
- ▶ Ayuda a que los algoritmos converjan más rápido.

- ▶ Centrar los datos en cero.
- ▶ Escalar los datos por medio de la desviación estándar

Estas manipulaciones se utilizan generalmente para mejorar la estabilidad numérica de algunos cálculos.

Para estandarizar las observaciones, se resta la media a cada una de ellas y se divide por la desviación estándar

$$Z = \frac{X - \bar{x}}{\sigma}$$

donde  $\bar{x}$  es la media aritmética y  $\sigma$  es la desviación estándar.

La desventaja de estas transformaciones es la pérdida de interpretabilidad de los valores individuales, ya que los datos ya no están en las unidades originales.

# Normalización basada en la media

- ▶ Centramos la media de la variable en cero.
- ▶ Escalamos de la distribución al rango de valores.

$$z = \frac{X - \bar{x}}{\text{máx}(X) - \text{mín}(X)}.$$

Esta transformación da como resultado una distribución centrada en 0, con sus valores mínimo y máximo dentro del rango de -1 a 1.

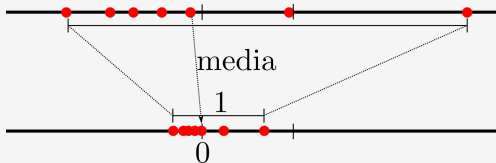


Figura 1: Normalización basada en la media.

# Escalamiento de valores máximo y mínimo

El escalamiento mínimo-máximo comprime (o estira) todos los valores de características para que estén dentro del intervalo cerrado de  $[0, 1]$ .

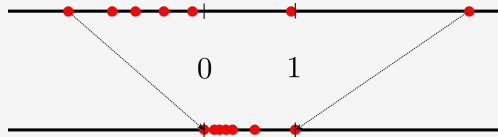


Figura 2: Escalamiento de valores máximo y mínimo.

# Escalamiento de valores máximo y mínimo

El escalamiento mínimo-máximo comprime (o estira) todos los valores de características para que estén dentro del intervalo cerrado de  $[0, 1]$ .

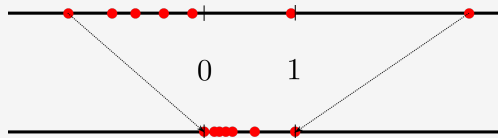


Figura 2: Escalamiento de valores máximo y mínimo.

$$z = \frac{X - \text{mín}(X)}{\text{máx}(X) - \text{mín}(X)}$$

# Escalamiento de máximo absoluto

- Esta transformación, mapea los valores de  $X$  hasta un máximo de 1.
- Conserva la posición relativa de todas las observaciones.

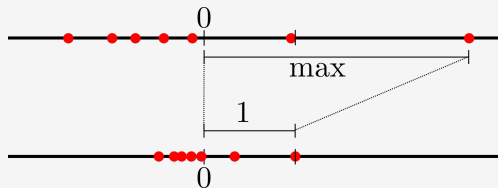


Figura 3: Escalamiento de máximo absoluto.



# Escalamiento de máximo absoluto

Se consigue dividiendo los datos entre el máximo

$$z = \frac{X}{\max |X|}$$

- ▶ Se recomienda usar esta transformación sobre datos centrados en cero o en un data-set con pocos datos.

# Escalamiento por cuantiles

Consiste en centrar las observaciones en cero usando la mediana y escalar el resultado por el rango intercuartílico (IQR)

$$z = \frac{X - \bar{x}}{Q_3 - Q_1}$$

donde  $\bar{x}$  es la mediana de  $X$  y  $Q_1$ ,  $Q_3$  corresponden a los cuartiles uno y tres.

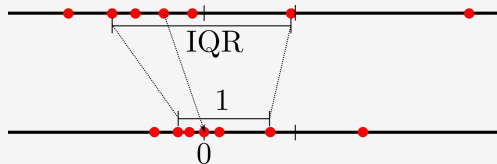


Figura 4: Escalamiento por cuantiles.

Este método se conoce como **escalamiento robusto** porque produce estimaciones más robustas para el centro y el rango de valores de la variable.

- ▶ Se recomienda si los datos contienen valores atípicos.