Andrés Merino • *Febrero 2026*

1. NORMALIZACIÓN DE DATOS

DEFINICIÓN 1: Normalización por el máximo.
Dado un conjunto de datos $X = \{x_1, x_2, \ldots, x_n\}$, la normalización por el máximo transforma cada dato x_i en x_i' según la fórmula:

$$x_i' = \frac{x_i}{\text{máx}(X)}$$

DEFINICIÓN 2: Normalización Min-Max.
Dado un conjunto de datos $X = \{x_1, x_2, \ldots, x_n\}$, la normalización Min-Max transforma cada dato x_i en x_i' según la fórmula:

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

DEFINICIÓN 3: Estandarización (Z-score).
Dado un conjunto de datos $X = \{x_1, x_2, \ldots, x_n\}$, 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 $X = \{x_1, x_2, \ldots, x_n\}$ 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:

$$\text{Tamaño del intervalo} = \frac{\text{máx}(X) - \text{mín}(X)}{k}$$

DEFINICIÓN 5: Discretización de igual frecuencia.
Dado un conjunto de datos continuo $X = \{x_1, x_2, \ldots, x_n\}$ 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
Verde
Rojo

$\xrightarrow{\text{One-Hot}}$

Rojo	Verde	Azul
1	0	0
0	1	0
0	0	1
0	1	0
1	0	0

4. Análisis de Componentes Principales (PCA)

Definición 7: Componente principal como dirección de máxima información.
Sea $X \in \mathbb{R}^{n\times p}$ una matriz de datos centrada (cada variable con media cero). La primera componente principal es el vector unitario $w \in \mathbb{R}^p$ que maximiza la información preservada, medida como la varianza de la proyección de los datos sobre dicha dirección:

$$w^* = \arg \max_{\|w\|=1} \operatorname{Var}(Xw)$$

Equivalente a:

$$w^* = \arg \max_{\|w\|=1} w^\top \Sigma w$$

donde $\Sigma = \frac{1}{n} X^\top X$ es la matriz de covarianza.

Teorema 1.
Sea Σ la matriz de covarianza de los datos centrados. El vector que define la primera componente principal es el vector propio asociado al mayor valor propio de Σ.
En general, si

$$\Sigma v_i = \lambda_i v_i, \quad \lambda_1 \geqslant \lambda_2 \geqslant \cdots \geqslant \lambda_p \geqslant 0,$$

entonces:

- La i-ésima componente principal es el vector propio v_i.
- La varianza explicada por dicha componente es λ_i.

Definición 8: Transformación PCA.
Sea $W_k = [v_1, \dots, v_k]$ la matriz cuyas columnas son los k vectores propios asociados a los mayores valores propios de la matriz de covarianza Σ.
Las proyecciones principales de los datos X sobre el subespacio generado por estas direcciones se obtienen como:

$$Z = XW_k$$

donde cada columna de $Z \in \mathbb{R}^{n\times k}$ contiene las coordenadas de los datos proyectados sobre una componente principal.

5. Descomposición en Valores Singulares (SVD)

Teorema 2.
Sea $A \in \mathbb{R}^{m\times n}$ una matriz de rango r. Entonces, existen únicas matrices $U \in \mathbb{R}^{m\times m}$, $V \in \mathbb{R}^{n\times n}$ y $\Sigma \in \mathbb{R}^{m\times n}$ tal que

$$A = U\Sigma V^T,$$

donde

- U y V son matrices ortogonales.
- Σ es una matriz diagonal con entradas no negativas en la diagonal.
- Las entradas de Σ están ordenadas de manera decreciente.

A los valores en la diagonal de Σ se les llama valores singulares de A, y las columnas de U y V se llaman vectores singulares izquierdos y derechos, respectivamente.

Demostración. Supongamos que la descomposición existe, es decir, $A = U\Sigma V^T$, nuestro objetivo es encontrar las matrices U, Σ y V.

Primero, consideremos la matriz A^TA, esta es simétrica; por el teorema espectral, también es diagonalizable y todos sus valores propios son no negativos. Sea P la matriz cuyas columnas son los vectores propios ortonormales de A^TA, y sea D la matriz diagonal cuyos elementos son los valores propios correspondientes. Entonces, podemos escribir

$$A^TA = PDP^T.$$

Por otro lado, como suponemos que $A = U\Sigma V^T$, tenemos

$$A^TA = (U\Sigma V^T)^T(U\Sigma V^T) = V\Sigma^TU^TU\Sigma V^T = V\Sigma^T\Sigma V^T.$$

Comparando ambas expresiones para A^TA, obtenemos

$$PDP^T = P\Sigma^T\Sigma P^T.$$

Así, podemos tomar

$$V = P, \quad \Sigma^T\Sigma = D.$$

Como tanto D y Σ son diagonales, tenemos que $\sigma_{ii}^2 = \lambda_{ii}$. Por otro lado, como los valores propios de $A^T A$ son no negativos, podemos tomar

$$\sigma_{ii} = \sqrt{\lambda_{ii}}.$$

Ahora, tomemos v_i las columnas de V y u_i las columnas de U. Observemos que, como V es ortogonal, tenemos que $V^T v_i = e_i$, donde e_i es el vector canónico. Entonces, de $A = U\Sigma V^T$, tenemos que

$$Av_i = U\Sigma V^T v_i = U\Sigma e_i = U\sigma_{ii} e_i = \sigma_{ii} u_i,$$

así, podemos tomar

$$u_i = \frac{1}{\sigma_{ii}} A v_i.$$

De esta manera, tenemos la descomposición en valores singulares. □