Actividad 1 Recuperación: Caracterización de Matrices Euclídeas

Juan Rubio Cobeta

15 de diciembre de 2025

Índice

1. Introducción y Definiciones Previas

El objetivo de este documento es formalizar las condiciones necesarias y suficientes para que una matriz de disimilaridades pueda ser representada como una configuración de puntos en un espacio euclídeo.

Sean $n, K \in \mathbb{N}$. Consideremos el espacio de matrices reales $\mathbb{R}^{n\times n}$. Denotamos por $\mathbf{1} \in \mathbb{R}^n$ al vector columna cuyos elementos son todos iguales a la unidad.

Definimos la **matriz de centrado** $H \in \mathbb{R}^{n\times n}$ como el proyector ortogonal sobre el complemento ortogonal del subespacio generado por $\mathbf{1}$:

$$H := I_n - \frac{1}{n}\mathbf{1}\mathbf{1}^T. \tag{1}$$

La matriz H satisface las siguientes propiedades fundamentales:

1. **Simetría:** $H^T = H$.
2. **Idempotencia:** $H^2 = H$.
3. **Centrado:** $H\mathbf{1} = \mathbf{0}$ (y por simetría, $\mathbf{1}^T H = \mathbf{0}^T$).

2. Teorema Principal

Teorema 1: Caracterización de Matrices Euclídeas

Sea $D = (d_{rs}) \in \mathbb{R}^{n\times n}$ una matriz de distancias (tal que $d_{rs} \geq 0$, $d_{rr} = 0$, $d_{rs} = d_{sr}$). Sea $A \in \mathbb{R}^{n\times n}$ la matriz cuyos elementos están definidos por $a_{rs} = -\frac{1}{2}d_{rs}^2$.

La matriz D es una **Matriz de Distancia Euclídea** correspondientes a una configuración de puntos en $\mathbb{R}^K$ si y solo si la matriz doblemente centrada

$$B = HAH$$

es semidefinida positiva ($B \succeq 0$) y rango$(B) \leq K$.

Además, se verifican las siguientes propiedades estructurales:

1. Si D es generada por una configuración $Z \in \mathbb{R}^{n\times K}$, entonces B es la matriz de Gram de la configuración centrada HZ.
2. Recíprocamente, si $B \succeq 0$, las coordenadas de la configuración X pueden recuperarse mediante la descomposición espectral de B.

3. Demostración

La demostración se divide en dos implicaciones lógicas.

Parte I: Necesidad ($\implies$)

Supongamos que D es una matriz de distancias euclídeas. Esto implica que existen puntos $z_1, \ldots, z_n \in \mathbb{R}^K$ tales que $d_{rs} = \|z_r - z_s\|$. Sea $Z = [z_1, \ldots, z_n]^T \in \mathbb{R}^{n\times K}$ la matriz de la configuración.

Utilizando la expansión del producto escalar usual, desarrollamos el cuadrado de la distancia:

$$d_{rs}^2 = \|z_r - z_s\|^2 = \langle z_r, z_r\rangle + \langle z_s, z_s\rangle - 2\langle z_r, z_s\rangle. \tag{2}$$

Sea $G = ZZ^T$ la matriz de Gram asociada, donde $g_{rs} = \langle z_r, z_s\rangle$, y definamos el vector $\mathbf{g} = \mathrm{diag}(G)$. La ecuación anterior puede expresarse matricialmente como:

$$D^{(2)} = \mathbf{g}\mathbf{1}^T + \mathbf{1}\mathbf{g}^T - 2G, \tag{3}$$

donde $D^{(2)}$ es la matriz de distancias al cuadrado. Dado que $A = -\frac{1}{2}D^{(2)}$, tenemos:

$$A = G - \frac{1}{2}\mathbf{g}\mathbf{1}^T - \frac{1}{2}\mathbf{1}\mathbf{g}^T. \tag{4}$$

Aplicamos ahora el operador de doble centrado $B = HAH$. Recordando que $H\mathbf{1} = \mathbf{0}$, los términos que involucran al vector $\mathbf{1}$ se anulan:

$$\begin{aligned} B &= H\left(G - \frac{1}{2}\mathbf{g}\mathbf{1}^T - \frac{1}{2}\mathbf{1}\mathbf{g}^T\right)H \\ &= HGH - \frac{1}{2}H\mathbf{g}(\mathbf{1}^T H) - \frac{1}{2}(H\mathbf{1})\mathbf{g}^T H \\ &= HGH. \end{aligned}$$

Sustituyendo $G = ZZ^T$:

$$B = H(ZZ^T)H = (HZ)(HZ)^T. \tag{5}$$

Sea $M = HZ$. La expresión $B = MM^T$ demuestra inmediatamente que B es una matriz de Gram y, por consiguiente, es **simétrica y semidefinida positiva** ($B \succeq 0$). Adicionalmente, $\text{rango}(B) = \text{rango}(HZ) \leq \text{mín}(n-1, K)$.

Parte II: Suficiencia ($\Longleftarrow$)

Supongamos ahora que B es semidefinida positiva con $\text{rango}(B) = K$. Por el **Teorema Espectral** para matrices simétricas reales, existe una descomposición ortogonal:

$$B = U\Lambda U^T, \tag{6}$$

donde $\Lambda = \text{diag}(\lambda_1, \ldots, \lambda_n)$ con autovalores $\lambda_1 \geq \cdots \geq \lambda_K > 0 = \cdots = 0$.

Construimos la matriz de coordenadas $X \in \mathbb{R}^{n\times K}$ como:

$$X = U_K \Lambda_K^{1/2}, \tag{7}$$

donde U_K contiene los primeros K autovectores y $\Lambda_K = \text{diag}(\lambda_1, \ldots, \lambda_K)$. Entonces $B = XX^T$, lo que implica que $b_{rs} = \langle x_r, x_s\rangle$.

Debemos verificar que la distancia euclídea inducida por esta configuración X recupera la matriz original D. Calculamos la distancia al cuadrado entre las filas x_r y x_s:

$$\|x_r - x_s\|^2 = b_{rr} + b_{ss} - 2b_{rs}. \tag{8}$$

Para relacionar esto con A, expandimos la relación $B = HAH$ elemento a elemento. Si denotamos las medias por fila, columna y global de A como $\bar{a}_{r\cdot}$, $\bar{a}_{\cdot s}$ y $\bar{a}_{\cdot\cdot}$ respectivamente:

$$b_{rs} = a_{rs} - \bar{a}_{r\cdot} - \bar{a}_{\cdot s} + \bar{a}_{\cdot\cdot}. \tag{9}$$

Sustituyendo en (8):

$$\begin{aligned} \|x_r - x_s\|^2 &= (a_{rr} - \bar{a}_{r\cdot} - \bar{a}_{\cdot r} + \bar{a}_{\cdot\cdot}) \\ &\quad + (a_{ss} - \bar{a}_{s\cdot} - \bar{a}_{\cdot s} + \bar{a}_{\cdot\cdot}) \\ &\quad - 2(a_{rs} - \bar{a}_{r\cdot} - \bar{a}_{\cdot s} + \bar{a}_{\cdot\cdot}). \end{aligned}$$

Al simplificar algebraicamente, todos los términos de medias se cancelan. Además, dado que $d_{rr} = 0 \implies a_{rr} = 0$, obtenemos:

$$\|x_r - x_s\|^2 = -2a_{rs} = -2\left(-\frac{1}{2}d_{rs}^2\right) = d_{rs}^2. \tag{10}$$

Por lo tanto, $\|x_r - x_s\| = d_{rs}$. Esto demuestra que existe una configuración X en $\mathbb{R}^K$ que genera exactamente las distancias D.

4. Observaciones Finales

- **Centrado de la configuración:** De la identidad $B\mathbf{1} = HAH\mathbf{1} = HA\mathbf{0} = \mathbf{0}$, y sabiendo que $B = XX^T$, se deduce que $XX^T\mathbf{1} = \mathbf{0}$. Multiplicando por la izquierda por $(X^TX)^{-1}X^T$ (asumiendo rango completo), obtenemos $X^T\mathbf{1} = \mathbf{0}$. Esto implica que la suma de las filas de X es el vector nulo; es decir, la configuración recuperada tiene su baricentro en el origen.
- **Unicidad (Isometría):** Si existiese otra configuración Y tal que $YY^T = B = XX^T$, entonces existe una matriz ortogonal $Q \in O(K)$ tal que $Y = XQ$. La solución es única salvo rotaciones y reflexiones.