

Reconstrucción Exacta de Matrices de Datos mediante Valores y Vectores Propios

1 Reconstrucción Exacta de una Matriz de Datos usando Valores y Vectores Propios

En el análisis multivariado, y en particular en métodos como el Análisis de Componentes Principales (ACP), es posible reconstruir exactamente una matriz de datos utilizando sus valores y vectores propios gracias a la Descomposición en Valores Singulares (DVS).

Este procedimiento descompone una matriz en el producto de tres matrices con características específicas, lo que no solo facilita su interpretación geométrica, sino que también resulta muy útil para realizar reducción de dimensionalidad.

1.1 La Descomposición en Valores Singulares (DVS)

La DVS muestra que cualquier matriz C de tamaño $n \times p$ y rango r puede factorizarse de la siguiente forma:

$$C = VLU'$$

Donde:

- L : matriz diagonal $r \times r$ que contiene los **valores propios** de C , es decir, $\sqrt{\lambda_j}$, donde λ_j es el j -ésimo valor propio de $C'C$
- U : matriz $p \times r$ cuyas columnas son los **vectores propios de $C'C$**
- V : matriz $n \times r$ cuyas columnas son los **vectores propios de CC'**
- Ambas matrices U y V son **ortonormales**: $U'U = V'V = I_r$

Luego, cuando aplicamos la DVS a la matriz X centrada y estandarizada, obtenemos que:

$$X = VLU'$$

Donde:

- L contiene $\sqrt{\lambda_\alpha}$, con λ_α siendo los valores propios de $X'X$
- Las columnas de U son los vectores propios de $X'X$
- Las columnas de V son los vectores propios de XX'

Ahora bien, Se puede escribir X en su forma expandida

$$X = [v_1 \ v_2 \ \dots \ v_p] \begin{bmatrix} \sqrt{\lambda_1} & 0 & \dots & 0 \\ 0 & \sqrt{\lambda_2} & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \sqrt{\lambda_p} \end{bmatrix} \begin{bmatrix} u'_1 \\ u'_2 \\ \vdots \\ u'_p \end{bmatrix}$$

Luego, al multiplicar (V) por (L) tenemos que:

$$VL = [\sqrt{\lambda_1}v_1, \sqrt{\lambda_2}v_2, \dots, \sqrt{\lambda_p}v_p]$$

y al multiplicar por (U'):

$$X = [\sqrt{\lambda_1}v_1, \sqrt{\lambda_2}v_2, \dots, \sqrt{\lambda_p}v_p] \begin{bmatrix} u'_1 \\ u'_2 \\ \vdots \\ u'_p \end{bmatrix}$$

$$X = \sqrt{\lambda_1}v_1u'_1 + \sqrt{\lambda_2}v_2u'_2 + \dots + \sqrt{\lambda_p}v_pu'_p$$

$$X = \sum_{\alpha=1}^p \sqrt{\lambda_\alpha} v_\alpha u'_\alpha$$

Esto nos muestra que toda la información de la matriz original está contenida en sus valores y vectores propios.

Por otro lado, un resultado clave que surge de la DVS es la **equivalencia entre los valores propios de (X'X) y (XX')**. Ambos comparten los mismos valores propios no nulos (λ_α). Esto es:

- Si (u_α) es un vector propio de (X'X), entonces $\frac{1}{\sqrt{\lambda_\alpha}} X u_\alpha$ es un vector propio de (XX').
- Análogamente, si (v_α) es un vector propio de (XX'), entonces $\frac{1}{\sqrt{\lambda_\alpha}} X' v_\alpha$ es un vector propio de (X'X).

Esto implica que **solo es necesario calcular los valores y vectores propios de una de estas matrices** para obtener también los de la otra.

1.2 Interpretación y utilidad en ACP

Así, la factorización $X = VLU'$ muestra una transformación diferente de los datos:

- **Las Componentes Principales (Scores de los Individuos):**

La matriz \$ Z = V L \$ contiene las **coordenadas de los individuos**, es decir, cada columna de Z

representa la proyección de todas las observaciones sobre la α -ésima componente. Esto muestra cómo se ubican mis muestras en el nuevo espacio de características reducido

- **Las Cargas Factoriales (Contribuciones de las Variables):**

La matriz U (o, en su versión escalada, U) reúne las cargas factoriales, que muestran el peso y la dirección que tiene cada variable original en la construcción de las componentes principales. Cuando una carga es alta en valor absoluto, significa que esa variable tiene una influencia importante sobre la componente correspondiente. Es decir, la matriz U nos muestra qué variables originales son las que realmente están impulsando cada componente

En general, La reconstrucción exacta de la matriz X a través de la DVS no solo es un resultado teórico interesante, sino que también constituye el fundamento computacional del ACP. Gracias a este enfoque es posible reducir la dimensionalidad sin perder la estructura esencial de los datos, facilitar la visualización de las relaciones entre individuos y variables, y comprender mejor la geometría que hay detrás del conjunto de datos.