

# **Aplicación de Valores y Vectores Propios para la Reducción de Dimensionalidad en el Dataset UCI HAR con base en la Descomposición en Valores Singulares (SVD) y PCA**

**Grupo 1**

**Materia:** Álgebra Lineal

**Paralelo:** 115

**Profesor:** Tenesaca Timaná Rómulo Fernando

**Período:** PAO II 2025

## **Integrantes:**

Paladines Sánchez José (josepala@espol.edu.ec)

Hernández Triviño Edwin (edhernan@espol.edu.ec)

Larena Alcívar Jeanpoul (jelarena@espol.edu.ec)

Becerra Pozo Julio (jubecerr@espol.edu.ec)

Quiñonez Abarca Rafaela (rafaquin@espol.edu.ec)

Saltos Moreira Mateo (masaltos@espol.edu.ec)

19 de enero de 2026

## Resumen

El reconocimiento de actividades humanas mediante sensores móviles genera grandes volúmenes de datos multivariantes, lo que dificulta la identificación de patrones y la clasificación eficiente. Este reporte técnico presenta la aplicación de la Descomposición en Valores Singulares (SVD) como fundamento matemático para ejecutar un Análisis de Componentes Principales (PCA) sobre un subconjunto del *UCI Human Activity Recognition Dataset*. Se procesaron 500 observaciones con 12 variables cinemáticas seleccionadas. Se demuestra teóricamente el procedimiento mediante el cálculo manual de una matriz  $3 \times 2$  y se implementa computacionalmente en lenguaje R. Los resultados evidencian que las tres primeras componentes principales explican una porción significativa de la varianza total, permitiendo visualizar una separación clara entre actividades estáticas y dinámicas en un espacio de dimensionalidad reducida, validando la utilidad del álgebra lineal en el preprocesamiento de datos para aprendizaje automático.

# Índice

<b>1. Antecedentes</b>	<b>5</b>
<b>2. Marco Teórico</b>	<b>5</b>
2.1. Descomposición en Valores Singulares (SVD) . . . . .	5
2.1.1. Definición Formal . . . . .	5
2.2. Justificación Teórica y Relación con Valores Propios . . . . .	6
2.2.1. Simetría y Diagonalización . . . . .	6
2.2.2. Relación Fundamental . . . . .	6
2.3. Ejemplo Demostrativo de Cálculo Manual SVD . . . . .	7
2.3.1. Paso 1: Construcción de la matriz de Gram $A^T A$ . . . . .	7
2.3.2. Paso 2: Obtención de Valores Propios . . . . .	7
2.3.3. Paso 3: Cálculo de Vectores Propios (Matriz $V$ ) . . . . .	8
2.3.4. Paso 4: Cálculo de Valores Singulares . . . . .	9
2.3.5. Paso 5: Construcción de Vectores Singulares Izquierdos (Matriz $U$ ) . . . . .	9
2.3.6. Paso 6: Construcción de la Matriz $\Sigma$ . . . . .	10
2.3.7. Paso 7: Reconstrucción y Verificación . . . . .	10
2.4. Análisis de Componentes Principales (PCA) . . . . .	11
2.4.1. Preprocesamiento de Datos: Centrado y Escalamiento . . . . .	11
2.4.2. Relación entre SVD y la Matriz de Covarianza . . . . .	11
2.5. Ejemplo Numérico Detallado de PCA . . . . .	13
2.5.1. 1. Matriz de Datos y Centrado . . . . .	13
2.5.2. 2. Cálculo de la Matriz de Gram $X^T X$ . . . . .	14
2.5.3. 3. Obtención de Valores Propios y Singulares . . . . .	14
2.5.4. 4. Cálculo de Vectores Propios (Matriz $V$ ) . . . . .	14
2.5.5. 5. Cálculo de $U$ y construcción de la SVD . . . . .	16
2.5.6. 6. Verificación de la Reconstrucción $X = U\Sigma V^T$ . . . . .	17
2.5.7. Cálculo de Scores e Interpretación Estadística . . . . .	17
<b>3. Conjunto de datos</b>	<b>18</b>
3.1. Descripción . . . . .	18
3.2. Cita APA . . . . .	18

3.3. Enlace de descarga . . . . .	19
3.4. Variables seleccionadas . . . . .	19
3.5. Justificación de la selección . . . . .	19
<b>4. Implementación</b>	<b>20</b>
4.1. Importación de librerías esenciales . . . . .	20
4.2. Descarga del dataset, limpieza y generación del DataFrame . . . . .	21
4.3. Aplicación de la SVD (Cálculo de PCA) . . . . .	23
4.4. Visualización inicial de la matriz U (Scores) . . . . .	23
4.5. Visualización de $V^T$ (Matriz de Cargas) . . . . .	24
4.6. Determinación de varianza acumulada . . . . .	24
4.7. Presentación del Scree Plot . . . . .	25
4.8. Proyección 2D (Visualización de Clusters) . . . . .	26
4.9. Heatmap (Interpretación de Componentes) . . . . .	27
4.10. Biplot (Visualización Avanzada) . . . . .	28
<b>5. Análisis de Resultados</b>	<b>29</b>
5.1. Dimensionalidad Óptima (Scree Plot) . . . . .	29
5.2. Separabilidad de Actividades (Proyección 2D) . . . . .	30
5.3. Interpretación de Variables (Heatmap) . . . . .	31
5.4. Interacción Global de Variables y Clases (Biplot) . . . . .	32
<b>6. Comentarios sobre la implementación</b>	<b>34</b>
<b>7. Referencias Bibliográficas</b>	<b>35</b>

# 1. Antecedentes

El Análisis de Componentes Principales (PCA) es una técnica fundamental para mitigar la maldición de la dimensionalidad en el procesamiento de grandes volúmenes de datos. Aunque su formulación clásica se remonta a Pearson y Hotelling, su relevancia actual para la compresión de información sin pérdida significativa de varianza se mantiene vigente, tal como analiza Pernice (2024). Su implementación moderna y numéricamente estable se basa en la Descomposición en Valores Singulares (SVD), fundamentada en el Álgebra Lineal Numérica para garantizar la precisión computacional al trabajar con matrices complejas (Lay et al., 2021).

Investigaciones recientes, como la de Al-Qazzaz et al. (2024), demuestran que la reducción de dimensionalidad mediante PCA optimiza el rendimiento de los modelos predictivos al eliminar ruido redundante. Este proyecto aplica dichas técnicas sobre el *UCI Human Activity Recognition Dataset* (Anguita et al., 2013), validando la utilidad del álgebra lineal para mejorar la eficiencia en sistemas de aprendizaje automático.

## 2. Marco Teórico

La fundamentación matemática del presente proyecto reside en el Álgebra Lineal Numérica, específicamente en la factorización de matrices rectangulares. A continuación, se detalla la estructura teórica y el procedimiento algorítmico de la Descomposición en Valores Singulares, y su aplicación directa en el Análisis de Componentes Principales.

### 2.1. Descomposición en Valores Singulares (SVD)

#### 2.1.1. Definición Formal

Sea  $A$  una matriz real de orden  $m \times n$ . El teorema de la Descomposición en Valores Singulares establece que existe una factorización de la forma:

$$A = U\Sigma V^T \tag{1}$$

Donde los componentes se definen con propiedades algebraicas estrictas:

- **Matriz Ortogonal  $U$  ( $m \times m$ ):** Sus columnas son los vectores singulares izquierdos

$u_i$ . Al ser ortogonal, cumple que  $U^T U = I_m$ , lo que implica que sus columnas forman una base ortonormal para  $\mathbb{R}^m$ .

- **Matriz Ortogonal  $V$  ( $n \times n$ ):** Sus columnas son los vectores singulares derechos  $v_i$ . Cumple que  $V^T V = I_n$ , formando una base ortonormal para  $\mathbb{R}^n$ .
- **Matriz Diagonal  $\Sigma$  ( $m \times n$ ):** Contiene en su diagonal principal los valores singulares  $\sigma_i \geq 0$ , ordenados de forma descendente ( $\sigma_1 \geq \sigma_2 \geq \dots \geq \sigma_r > 0$ ).

Geoméricamente, esta descomposición interpreta cualquier transformación lineal representada por  $A$  como una composición de tres movimientos: una rotación o reflexión en el dominio ( $V^T$ ), un escalamiento a lo largo de los ejes coordenados ( $\Sigma$ ) y una rotación o reflexión en el codominio ( $U$ ).

## 2.2. Justificación Teórica y Relación con Valores Propios

La existencia de la SVD se fundamenta en las propiedades espectrales de las matrices simétricas asociadas  $A^T A$  y  $A A^T$ .

### 2.2.1. Simetría y Diagonalización

Para cualquier matriz  $A$ , el producto  $A^T A$  resulta en una matriz cuadrada de orden  $n \times n$  que es simétrica. Esto se demuestra mediante la propiedad de la transpuesta:

$$(A^T A)^T = A^T (A^T)^T = A^T A \quad (2)$$

Según el **Teorema Espectral**, toda matriz simétrica real es diagonalizable ortogonalmente y sus valores propios son reales. Además, dado que para cualquier vector  $x \neq 0$ , se cumple que  $x^T (A^T A) x = \|Ax\|^2 \geq 0$ , la matriz es semidefinida positiva, garantizando que sus valores propios sean no negativos ( $\lambda_i \geq 0$ ).

### 2.2.2. Relación Fundamental

Los valores singulares  $\sigma_i$  de la matriz  $A$  se definen intrínsecamente a partir de los valores propios de  $A^T A$ :

$$\sigma_i = \sqrt{\lambda_i(A^T A)} \quad (3)$$

Asimismo, se cumple el teorema que establece que las matrices  $A^T A$  y  $AA^T$  comparten los mismos valores propios no nulos, lo que garantiza la consistencia de la descomposición independientemente de si se analiza desde el espacio fila o el espacio columna.

## 2.3. Ejemplo Demostrativo de Cálculo Manual SVD

Para ilustrar el algoritmo SVD sin pérdida de generalidad, se procede al cálculo manual detallado utilizando una matriz  $A$  de orden  $3 \times 2$ . Sea la matriz:

$$A = \begin{pmatrix} 1 & 1 \\ 1 & 0 \\ 0 & 1 \end{pmatrix} \quad (4)$$

### 2.3.1. Paso 1: Construcción de la matriz de Gram $A^T A$

Se calcula el producto matricial elemento a elemento para obtener la matriz cuadrada simétrica:

$$A^T A = \begin{pmatrix} 1 & 1 & 0 \\ 1 & 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & 1 \\ 1 & 0 \\ 0 & 1 \end{pmatrix} = \begin{pmatrix} 1(1) + 1(1) + 0(0) & 1(1) + 1(0) + 0(1) \\ 1(1) + 0(1) + 1(0) & 1(1) + 0(0) + 1(1) \end{pmatrix} = \begin{pmatrix} 2 & 1 \\ 1 & 2 \end{pmatrix} \quad (5)$$

### 2.3.2. Paso 2: Obtención de Valores Propios

Se resuelve la ecuación característica  $\det(A^T A - \lambda I) = 0$  para encontrar los escalares  $\lambda$ :

$$\det \begin{pmatrix} 2-\lambda & 1 \\ 1 & 2-\lambda \end{pmatrix} = (2-\lambda)(2-\lambda) - (1)(1) \quad (6)$$

$$= 4 - 4\lambda + \lambda^2 - 1 \quad (7)$$

$$= \lambda^2 - 4\lambda + 3 \quad (8)$$

Factorizando el polinomio cuadrático:

$$(\lambda - 3)(\lambda - 1) = 0 \quad (9)$$

Se obtienen los valores propios ordenados descendientemente:  $\lambda_1 = 3$  y  $\lambda_2 = 1$ .

### 2.3.3. Paso 3: Cálculo de Vectores Propios (Matriz $V$ )

Se determinan los espacios propios  $E_\lambda = \{x \in \mathbb{R}^n : (A^T A - \lambda I)x = 0\}$  resolviendo el sistema homogéneo para cada  $\lambda$ .

**Para  $\lambda_1 = 3$ :** Sustituyendo en  $(A^T A - 3I)x = 0$ :

$$\begin{pmatrix} 2-3 & 1 \\ 1 & 2-3 \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} -1 & 1 \\ 1 & -1 \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix} \quad (10)$$

Aplicando reducción por filas ( $F_2 \rightarrow F_2 + F_1$ ):

$$\begin{pmatrix} -1 & 1 \\ 0 & 0 \end{pmatrix} \implies -x + y = 0 \implies x = y \quad (11)$$

El vector propio es  $v_1 = (1, 1)^T$ . Normalizando mediante la norma euclidiana  $\|v_1\| = \sqrt{1^2 + 1^2} = \sqrt{2}$ :

$$v_1 = \begin{pmatrix} \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \end{pmatrix} \quad (12)$$

**Para  $\lambda_2 = 1$ :** Sustituyendo en  $(A^T A - 1I)x = 0$ :

$$\begin{pmatrix} 2-1 & 1 \\ 1 & 2-1 \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix} \quad (13)$$

La reducción implica  $x + y = 0 \implies x = -y$ . El vector es  $v_2 = (-1, 1)^T$ . Normalizando:

$$v_2 = \begin{pmatrix} -\frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \end{pmatrix} \quad (14)$$

Finalmente, se construye la matriz ortogonal  $V$ :

$$V = \begin{pmatrix} \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{pmatrix} \quad (15)$$



#### 2.3.4. Paso 4: Cálculo de Valores Singulares

Los valores singulares se obtienen como la raíz cuadrada de los valores propios:

$$\sigma_1 = \sqrt{3} \approx 1,732, \quad \sigma_2 = \sqrt{1} = 1 \quad (16)$$

#### 2.3.5. Paso 5: Construcción de Vectores Singulares Izquierdos (Matriz $U$ )

Los vectores columna de  $U$  ( $u_i$ ) se calculan utilizando la relación fundamental de la SVD que vincula los espacios de entrada y salida:

$$u_i = \frac{1}{\sigma_i} A v_i \quad (17)$$

**Cálculo de  $u_1$ :**

$$u_1 = \frac{1}{\sqrt{3}} \begin{pmatrix} 1 & 1 \\ 1 & 0 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \end{pmatrix} = \frac{1}{\sqrt{6}} \begin{pmatrix} 1+1 \\ 1+0 \\ 0+1 \end{pmatrix} = \begin{pmatrix} \frac{2}{\sqrt{6}} \\ \frac{1}{\sqrt{6}} \\ \frac{1}{\sqrt{6}} \end{pmatrix} \quad (18)$$

**Cálculo de  $u_2$ :**

$$u_2 = \frac{1}{1} \begin{pmatrix} 1 & 1 \\ 1 & 0 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} -\frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \end{pmatrix} = \begin{pmatrix} -\frac{1}{\sqrt{2}} + \frac{1}{\sqrt{2}} \\ -\frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \end{pmatrix} = \begin{pmatrix} 0 \\ -\frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \end{pmatrix} \quad (19)$$

**Complejitud de la Base ( $u_3$ ):** Dado que  $U$  debe ser una matriz cuadrada de orden  $3 \times 3$ , necesitamos un tercer vector  $u_3$  que sea ortogonal a  $u_1$  y  $u_2$  para completar la base de  $\mathbb{R}^3$ . Utilizamos el producto cruz o el método de Gram-Schmidt. Sea  $w = u_1 \times u_2$  (simplificando constantes para el cálculo de dirección):

$$w \propto (2, 1, 1) \times (0, -1, 1) = \begin{vmatrix} i & j & k \\ 2 & 1 & 1 \\ 0 & -1 & 1 \end{vmatrix} = i(1 - (-1)) - j(2 - 0) + k(-2 - 0) = (2, -2, -2) \quad (20)$$

Normalizando el vector  $(2, -2, -2)$ :

$$u_3 = \frac{1}{\sqrt{12}}(2, -2, -2) = \begin{pmatrix} \frac{1}{\sqrt{3}} \\ -\frac{1}{\sqrt{3}} \\ -\frac{1}{\sqrt{3}} \end{pmatrix} \quad (21)$$

La matriz  $U$  resultante es:

$$U = \begin{pmatrix} \frac{2}{\sqrt{6}} & 0 & \frac{1}{\sqrt{3}} \\ \frac{1}{\sqrt{6}} & -\frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{3}} \\ \frac{1}{\sqrt{6}} & \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{3}} \end{pmatrix} \quad (22)$$

### 2.3.6. Paso 6: Construcción de la Matriz $\Sigma$

La matriz  $\Sigma$  tiene las mismas dimensiones que  $A$  ( $3 \times 2$ ). Se colocan los valores singulares en la diagonal principal y se rellenan con ceros el resto de posiciones:

$$\Sigma = \begin{pmatrix} \sqrt{3} & 0 \\ 0 & 1 \\ 0 & 0 \end{pmatrix} \quad (23)$$

### 2.3.7. Paso 7: Reconstrucción y Verificación

Finalmente, se verifica que el producto de las matrices componentes reconstruye la matriz original  $A$ .

$$A = U\Sigma V^T \quad (24)$$

$$= \begin{pmatrix} \frac{2}{\sqrt{6}} & 0 & \frac{1}{\sqrt{3}} \\ \frac{1}{\sqrt{6}} & -\frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{3}} \\ \frac{1}{\sqrt{6}} & \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{3}} \end{pmatrix} \begin{pmatrix} \sqrt{3} & 0 \\ 0 & 1 \\ 0 & 0 \end{pmatrix} \begin{pmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\ -\frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{pmatrix} \quad (25)$$

$$= \begin{pmatrix} 1 & 1 \\ 1 & 0 \\ 0 & 1 \end{pmatrix} \quad (26)$$

Este resultado valida la descomposición SVD. Cabe resaltar que si se seleccionara solo el primer valor singular (truncamiento a  $k = 1$ ), se obtendría la mejor aproximación de rango

1 de la matriz, principio fundamental para la reducción de dimensionalidad en PCA.

## 2.4. Análisis de Componentes Principales (PCA)

El Análisis de Componentes Principales (PCA) es una técnica estadística de reducción de dimensionalidad que transforma un conjunto de variables correlacionadas en un conjunto menor de variables no correlacionadas, denominadas componentes principales. Mientras que la SVD es una factorización matricial algebraica, el PCA es su aplicación directa en el contexto estadístico. De hecho, calcular la SVD de la matriz de datos centrada es el método numérico más robusto para realizar un PCA.

A continuación, se desarrolla la fundamentación teórica que vincula la estructura algebraica de la SVD con los estadísticos descriptivos multivariantes, seguido de una demostración numérica exhaustiva.

### 2.4.1. Preprocesamiento de Datos: Centrado y Escalamiento

Para que el PCA capture correctamente la varianza de los datos, es imperativo que el origen del sistema de coordenadas coincida con el centroide de los datos. Sea  $X_{raw} \in \mathbb{R}^{n \times p}$  la matriz de datos crudos con  $n$  observaciones y  $p$  variables.

La matriz centrada  $X$  se obtiene restando a cada columna  $j$  su media aritmética  $\bar{x}_j$ :

$$x_{ij} = x_{ij}^{raw} - \bar{x}_j \quad (27)$$

Esto garantiza que la suma de cada columna sea cero, y por tanto, la media de cada variable en  $X$  sea 0.

Si las variables tienen unidades de medida distintas (ej. metros y kilogramos), es necesario además estandarizar dividiendo por la desviación estándar  $s_j$  (escalamiento Z-score), aunque para efectos de la siguiente demostración algebraica, asumiremos que trabajamos con la matriz centrada  $X$ .

### 2.4.2. Relación entre SVD y la Matriz de Covarianza

La matriz de covarianza muestral  $S$ , que cuantifica cómo varían las variables conjuntamente, se define como:

$$S = \frac{1}{n-1} X^T X \quad (28)$$

El objetivo del PCA es diagonalizar esta matriz  $S$  para encontrar direcciones de máxima varianza. A continuación, demostramos algebraicamente cómo la SVD resuelve este problema sin necesidad de calcular  $S$  explícitamente.

Partimos de la Descomposición en Valores Singulares de la matriz centrada  $X$ :

$$X = U\Sigma V^T \quad (29)$$

Sustituimos esta expresión en la fórmula de  $X^T X$ :

$$X^T X = (U\Sigma V^T)^T (U\Sigma V^T) \quad (30)$$

$$= (V\Sigma^T U^T)(U\Sigma V^T) \quad (31)$$

Dado que  $U$  es una matriz ortogonal, sabemos que  $U^T U = I$ . Además, como  $\Sigma$  es diagonal,  $\Sigma^T = \Sigma$ . Entonces:

$$X^T X = V\Sigma I \Sigma V^T \quad (32)$$

$$X^T X = V\Sigma^2 V^T \quad (33)$$

La ecuación (33) tiene la forma exacta de una diagonalización ortogonal de una matriz simétrica. Esto revela tres conclusiones fundamentales:

1. Los vectores singulares derechos  $V$  (de la SVD de  $X$ ) son idénticos a los **vectores propios** de la matriz de covarianza (las direcciones principales).
2. Los valores propios  $\lambda_i$  de  $X^T X$  son el cuadrado de los valores singulares:  $\lambda_i = \sigma_i^2$ .
3. La varianza explicada por cada componente principal es proporcional a  $\frac{\sigma_i^2}{n-1}$ .

Finalmente, las proyecciones de los datos originales sobre los nuevos ejes (conocidos como *scores* o puntuaciones  $T$ ) se calculan como:

$$T = XV \quad (34)$$

Sustituyendo  $X = U\Sigma V^T$ :

$$T = (U\Sigma V^T)V \quad (35)$$

$$T = U\Sigma(V^T V) \quad (36)$$

$$T = U\Sigma \quad (37)$$

Esto demuestra que las coordenadas de los datos en el espacio reducido pueden obtenerse directamente escalando las columnas de  $U$  por los valores singulares.

## 2.5. Ejemplo Numérico Detallado de PCA

Para ilustrar el proceso completo, utilizaremos una matriz nueva de  $n = 3$  observaciones y  $p = 2$  variables. Se mostrarán todas las operaciones intermedias para garantizar la trazabilidad.

### 2.5.1. 1. Matriz de Datos y Centrado

Sea la matriz de datos crudos (observaciones en filas):

$$X_{raw} = \begin{pmatrix} 2 & 0 \\ 0 & 1 \\ 1 & 2 \end{pmatrix} \quad (38)$$

Calculamos las medias por columna:

$$\blacksquare \bar{x}_1 = \frac{2+0+1}{3} = 1$$

$$\blacksquare \bar{x}_2 = \frac{0+1+2}{3} = 1$$

Restamos las medias para obtener la matriz centrada  $X$ :

$$X = \begin{pmatrix} 2-1 & 0-1 \\ 0-1 & 1-1 \\ 1-1 & 2-1 \end{pmatrix} = \begin{pmatrix} 1 & -1 \\ -1 & 0 \\ 0 & 1 \end{pmatrix} \quad (39)$$

Verificación: La suma de las columnas es  $1 - 1 + 0 = 0$  y  $-1 + 0 + 1 = 0$ . Correcto.

### 2.5.2. 2. Cálculo de la Matriz de Gram $X^T X$

Realizamos la multiplicación matricial  $X^T X$  (necesaria para hallar  $V$ ):

$$X^T X = \begin{pmatrix} 1 & -1 & 0 \\ -1 & 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & -1 \\ -1 & 0 \\ 0 & 1 \end{pmatrix} \quad (40)$$

$$= \begin{pmatrix} (1)(1) + (-1)(-1) + (0)(0) & (1)(-1) + (-1)(0) + (0)(1) \\ (-1)(1) + (0)(-1) + (1)(0) & (-1)(-1) + (0)(0) + (1)(1) \end{pmatrix} \quad (41)$$

$$= \begin{pmatrix} 1+1+0 & -1+0+0 \\ -1+0+0 & 1+0+1 \end{pmatrix} \quad (42)$$

$$= \begin{pmatrix} 2 & -1 \\ -1 & 2 \end{pmatrix} \quad (43)$$

### 2.5.3. 3. Obtención de Valores Propios y Singulares

Calculamos el polinomio característico de la matriz simétrica  $X^T X$ :

$$\det(X^T X - \lambda I) = \det \begin{pmatrix} 2-\lambda & -1 \\ -1 & 2-\lambda \end{pmatrix} \quad (44)$$

$$= (2-\lambda)(2-\lambda) - (-1)(-1) \quad (45)$$

$$= 4 - 4\lambda + \lambda^2 - 1 \quad (46)$$

$$= \lambda^2 - 4\lambda + 3 \quad (47)$$

Factorizamos el polinomio para encontrar las raíces:

$$(\lambda - 3)(\lambda - 1) = 0 \quad (48)$$

Los valores propios son  $\lambda_1 = 3$  y  $\lambda_2 = 1$ . Por consiguiente, los valores singulares ( $\sigma_i = \sqrt{\lambda_i}$ ) son:

$$\sigma_1 = \sqrt{3} \approx 1,732, \quad \sigma_2 = 1 \quad (49)$$

### 2.5.4. 4. Cálculo de Vectores Propios (Matriz $V$ )

Resolvemos el sistema  $(X^T X - \lambda_i I)v = 0$  para cada valor propio.

**Para  $\lambda_1 = 3$ :**

$$\left( \begin{pmatrix} 2 & -1 \\ -1 & 2 \end{pmatrix} - \begin{pmatrix} 3 & 0 \\ 0 & 3 \end{pmatrix} \right) \begin{pmatrix} v_{11} \\ v_{21} \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix} \quad (50)$$

$$\begin{pmatrix} -1 & -1 \\ -1 & -1 \end{pmatrix} \begin{pmatrix} v_{11} \\ v_{21} \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix} \quad (51)$$

Aplicamos reducción por filas ( $F_2 \leftarrow F_2 - F_1$ ):

$$\begin{pmatrix} -1 & -1 \\ 0 & 0 \end{pmatrix} \quad (52)$$

Esto implica la ecuación  $-v_{11} - v_{21} = 0 \Rightarrow v_{11} = -v_{21}$ . Un vector base es  $(1, -1)$ . Normalizamos dividiendo por su norma  $\sqrt{1^2 + (-1)^2} = \sqrt{2}$ :

$$v_1 = \begin{pmatrix} \frac{1}{\sqrt{2}} \\ -\frac{1}{\sqrt{2}} \end{pmatrix} \quad (53)$$

**Para  $\lambda_2 = 1$ :**

$$\begin{pmatrix} 2-1 & -1 \\ -1 & 2-1 \end{pmatrix} = \begin{pmatrix} 1 & -1 \\ -1 & 1 \end{pmatrix} \quad (54)$$

Reducción por filas ( $F_2 \leftarrow F_2 + F_1$ ):

$$\begin{pmatrix} 1 & -1 \\ 0 & 0 \end{pmatrix} \quad (55)$$

Ecuación resultante:  $v_{12} - v_{22} = 0 \Rightarrow v_{12} = v_{22}$ . Vector base  $(1, 1)$ . Normalizado:

$$v_2 = \begin{pmatrix} \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \end{pmatrix} \quad (56)$$

La matriz de vectores singulares derechos (coeficientes de las componentes principales) es:

$$V = \begin{pmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\ -\frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{pmatrix} \quad (57)$$

### 2.5.5. 5. Cálculo de $U$ y construcción de la SVD

Utilizamos la relación  $u_i = \frac{1}{\sigma_i} X v_i$ .

**Cálculo de  $u_1$ :**

$$u_1 = \frac{1}{\sqrt{3}} \begin{pmatrix} 1 & -1 \\ -1 & 0 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} \frac{1}{\sqrt{2}} \\ -\frac{1}{\sqrt{2}} \end{pmatrix} \quad (58)$$

$$= \frac{1}{\sqrt{6}} \begin{pmatrix} (1)(1) - (-1)(-1) \\ (-1)(1) + (0)(-1) \\ (0)(1) + (1)(-1) \end{pmatrix} = \frac{1}{\sqrt{6}} \begin{pmatrix} 1+1 \\ -1 \\ -1 \end{pmatrix} = \begin{pmatrix} \frac{2}{\sqrt{6}} \\ -\frac{1}{\sqrt{6}} \\ -\frac{1}{\sqrt{6}} \end{pmatrix} \quad (59)$$

**Cálculo de  $u_2$ :**

$$u_2 = \frac{1}{1} \begin{pmatrix} 1 & -1 \\ -1 & 0 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \end{pmatrix} \quad (60)$$

$$= \begin{pmatrix} \frac{1}{\sqrt{2}} - \frac{1}{\sqrt{2}} \\ -\frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \end{pmatrix} = \begin{pmatrix} 0 \\ -\frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \end{pmatrix} \quad (61)$$

Para completar la base ortonormal de  $U$  (dimensión  $3 \times 3$ ), calculamos  $u_3$  mediante el producto cruz de  $u_1$  y  $u_2$  (simplificado):  $(2, -1, -1) \times (0, -1, 1) = (-2, -2, -2)$ . Normalizando:  $u_3 = (-\frac{1}{\sqrt{3}}, -\frac{1}{\sqrt{3}}, -\frac{1}{\sqrt{3}})$ .

Matriz Diagonal  $\Sigma$  (misma dimensión que  $X$ ,  $3 \times 2$ ):

$$\Sigma = \begin{pmatrix} \sqrt{3} & 0 \\ 0 & 1 \\ 0 & 0 \end{pmatrix} \quad (62)$$



### 2.5.6. 6. Verificación de la Reconstrucción $X = U\Sigma V^T$

Multiplicamos  $U\Sigma$  primero:

$$U\Sigma = \begin{pmatrix} \frac{2}{\sqrt{6}} & 0 & \dots \\ -\frac{1}{\sqrt{6}} & -\frac{1}{\sqrt{2}} & \dots \\ -\frac{1}{\sqrt{6}} & \frac{1}{\sqrt{2}} & \dots \end{pmatrix} \begin{pmatrix} \sqrt{3} & 0 \\ 0 & 1 \\ 0 & 0 \end{pmatrix} = \begin{pmatrix} \frac{2\sqrt{3}}{\sqrt{6}} & 0 \\ -\frac{\sqrt{3}}{\sqrt{6}} & -\frac{1}{\sqrt{2}} \\ -\frac{\sqrt{3}}{\sqrt{6}} & \frac{1}{\sqrt{2}} \end{pmatrix} = \begin{pmatrix} \sqrt{2} & 0 \\ -\frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} \\ -\frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{pmatrix} \quad (63)$$

Ahora multiplicamos por  $V^T$ :

$$X_{rec} = \begin{pmatrix} \sqrt{2} & 0 \\ -\frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} \\ -\frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{pmatrix} \begin{pmatrix} \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{pmatrix} \quad (64)$$

$$= \begin{pmatrix} 1 & -1 \\ -1 & 0 \\ 0 & 1 \end{pmatrix} \quad (65)$$

Se recupera exactamente la matriz centrada  $X$ .

### 2.5.7. Cálculo de Scores e Interpretación Estadística

Los *scores*  $T$  representan las coordenadas de las observaciones en el nuevo sistema de referencia.

$$T = U\Sigma = \begin{pmatrix} \sqrt{2} & 0 \\ -\frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} \\ -\frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{pmatrix} \approx \begin{pmatrix} 1,41 & 0 \\ -0,707 & -0,707 \\ -0,707 & 0,707 \end{pmatrix} \quad (66)$$

La primera columna corresponde a la Componente Principal 1 (PC1) y la segunda a la PC2.

**Varianza Explicada:** La varianza total está dada por la suma de los valores propios:  
 $\sum \lambda_i = 3 + 1 = 4$ .

- Varianza explicada por PC1:  $\frac{3}{4} = 0,75$  (75 %).
- Varianza explicada por PC2:  $\frac{1}{4} = 0,25$  (25 %).

En este caso, se ha capturado el 100% de la información al usar las 2 componentes. En aplicaciones reales con dimensionalidad alta, se selecciona un  $k$  tal que la varianza

acumulada supere un umbral (ej. 80 % o 90 %), lo que se visualiza mediante el gráfico de sedimentación (*Scree Plot*).

### **3. Conjunto de datos**

#### **3.1. Descripción**

El conjunto de datos utilizado en este estudio es el *Human Activity Recognition Using Smartphones*, disponible en el repositorio de Machine Learning de la Universidad de California, Irvine (UCI). Este dataset fue generado a partir de experimentos realizados con un grupo de 30 voluntarios, con edades comprendidas entre los 19 y 48 años. Cada participante realizó seis actividades básicas (Caminar, Subir escaleras, Bajar escaleras, Sentarse, Estar de pie y Recostarse) portando un teléfono inteligente Samsung Galaxy S II en la cintura.

Para la recolección de datos se utilizaron el acelerómetro y giroscopio embebidos en el dispositivo, capturando la aceleración lineal 3-axial y la velocidad angular 3-axial a una frecuencia de muestreo constante de 50 Hz. Las señales obtenidas fueron preprocesadas mediante filtros de ruido y muestreadas en ventanas deslizantes de ancho fijo de 2.56 segundos con un solapamiento del 50 % (128 lecturas por ventana). La señal de aceleración del sensor, que contiene componentes gravitacionales y de movimiento corporal, fue separada utilizando un filtro pasa bajos Butterworth. El dataset final contiene 10,299 observaciones y un vector de características de 561 atributos por registro, derivados de los dominios de tiempo y frecuencia.

#### **3.2. Cita APA**

Anguita, D., Ghio, A., Oneto, L., Parra, X., & Reyes-Ortiz, J. L. (2013). A Public Domain Dataset for Human Activity Recognition Using Smartphones. *21st European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning (ESANN)*, 437–442.

### 3.3. Enlace de descarga

- **Página original del dataset:** <https://archive.ics.uci.edu/dataset/240/human+activity+recognition+using+smartphones>
- **Dataset (Descarga directa):** <https://archive.ics.uci.edu/static/public/240/human+activity+recognition+using+smartphones.zip>

### 3.4. Variables seleccionadas

Para el análisis de reducción de dimensionalidad mediante PCA, se seleccionó un subconjunto de 12 variables del dominio del tiempo (prefijo *t*), descritas a continuación:

**Tabla 1:** Variables cinemáticas seleccionadas para el análisis PCA.

Variable	Descripción breve
tBodyAcc_mean_X	Promedio de la aceleración del cuerpo en el eje X (tiempo).
tBodyAcc_mean_Y	Promedio de la aceleración del cuerpo en el eje Y (tiempo).
tBodyAcc_mean_Z	Promedio de la aceleración del cuerpo en el eje Z (tiempo).
tBodyAcc_std_X	Desviación estándar de la aceleración del cuerpo en el eje X.
tBodyAcc_std_Y	Desviación estándar de la aceleración del cuerpo en el eje Y.
tBodyAcc_std_Z	Desviación estándar de la aceleración del cuerpo en el eje Z.
tGravityAcc_mean_X	Promedio de la componente de aceleración por gravedad en el eje X.
tGravityAcc_mean_Y	Promedio de la componente de aceleración por gravedad en el eje Y.
tGravityAcc_mean_Z	Promedio de la componente de aceleración por gravedad en el eje Z.
tBodyAccJerk_mean_X	Promedio de la señal Jerk (derivada) en el eje X.
tBodyAccJerk_mean_Y	Promedio de la señal Jerk (derivada) en el eje Y.
tBodyAccJerk_mean_Z	Promedio de la señal Jerk (derivada) en el eje Z.

*Fuente:* Elaboración propia a partir del dataset UCI HAR.

### 3.5. Justificación de la selección

La selección de estas 12 variables busca capturar la estructura esencial del movimiento humano sin incurrir en la alta dimensionalidad del dataset completo (561 variables).

Se incluyeron las medias (*mean*) para representar la tendencia central de la postura y el movimiento, y las desviaciones estándar (*std*) para cuantificar la intensidad y variabilidad de la acción dinámica. Asimismo, se incorporaron componentes de gravedad (*tGravityAcc*) fundamentales para distinguir actividades estáticas (como estar acostado vs. de pie) y componentes de Jerk (*tBodyAccJerk*) que caracterizan movimientos bruscos o transiciones. La inclusión de los tres ejes espaciales (X, Y, Z) garantiza una representación geométrica completa del movimiento en el espacio tridimensional.

## 4. Implementación

### 4.1. Importación de librerías esenciales

```
# Instalación de paquetes (solo si es necesario)
req_pkgs <- c("data.table", "ggplot2", "dplyr", "tidyr", "RColorBrewer")
to_install <- setdiff(req_pkgs, rownames(installed.packages()))
if (length(to_install) > 0) install.packages(to_install)

# Carga de librerías
suppressPackageStartupMessages({
  library(data.table)      # Lectura rápida de datos
  library(ggplot2)         # Visualización avanzada
  library(dplyr)           # Manipulación de data frames
  library(tidyr)           # Transformación de datos
  library(RColorBrewer)    # Paletas de colores
})
```

En esta sección se cargan los paquetes fundamentales para el análisis de datos, visualización y manipulación de estructuras en R. Estas librerías permiten la lectura de archivos, el procesamiento matricial y la generación de gráficos estadísticos de alta calidad necesarios para reproducir el análisis:

- **data.table** y **dplyr**: Para la manipulación eficiente de grandes volúmenes de datos.
- **ggplot2**: Para la generación de representaciones gráficas avanzadas.
- **tidyr**: Para la transformación de datos (formato largo/ancho) necesaria en los mapas de calor.

## 4.2. Descarga del dataset, limpieza y generación del DataFrame

```
# 1. Descarga y descompresion del dataset UCI HAR
zip_url <-
  "https://archive.ics.uci.edu/ml/machine-learning-databases/00240/UCI%20HAR%20Dataset.zip"
zip_file <- "UCI_HAR_Dataset.zip"
if (!file.exists(zip_file)) download.file(zip_url, zip_file, mode = "wb")
unzip(zip_file, exdir = ".")

# 2. Lectura de metadatos y limpieza de nombres
features <- fread("UCI HAR Dataset/features.txt", col.names = c("n","name"))
feat_names <- gsub("[()-]", "_", gsub("\\(\\)", "", features$name))
feat_names <- make.names(gsub("-", "_", feat_names), unique = TRUE)

# 3. Lectura y combinacion de Train/Test
read_set <- function(type) {
  X <- fread(file.path("UCI HAR Dataset", type, paste0("X_", type, ".txt")))
  y <- fread(file.path("UCI HAR Dataset", type, paste0("y_", type, ".txt")))
  s <- fread(file.path("UCI HAR Dataset", type, paste0("subject_", type,
    ".txt")))
  colnames(X) <- feat_names
  cbind(subject = s$V1, activity = y$V1, X)
}
full_df <- rbind(read_set("train"), read_set("test"))

# 4. Mapeo de actividades al Espanol
act_labels <- c("Caminando", "Subiendo", "Bajando", "Sentado", "De pie",
  "Acostado")
full_df$activity <- factor(full_df$activity, levels = 1:6, labels = act_labels)

# 5. Seleccion de las 12 variables
target_vars <- c(
  "tBodyAcc_mean_X", "tBodyAcc_mean_Y", "tBodyAcc_mean_Z",
  "tBodyAcc_std_X", "tBodyAcc_std_Y", "tBodyAcc_std_Z",
  "tGravityAcc_mean_X", "tGravityAcc_mean_Y", "tGravityAcc_mean_Z",
  "tBodyAccJerk_mean_X", "tBodyAccJerk_mean_Y", "tBodyAccJerk_mean_Z"
)
```

```

# Filtrar columnas
small_df <- select(full_df, subject, activity, all_of(target_vars))

# Muestreo aleatorio estratificado para balancear clases
set.seed(123)

# Agrupar
grouped_df <- group_by(small_df, activity)

# Muestrear
sampled_df <- sample_n(grouped_df, size = 84)

# Desagrupar y cortar
small_df <- slice(ungroup(sampled_df), 1:500)

# Imputacion simple (si hubiese NAs)
small_df <- mutate(small_df, across(where(is.numeric),
                                     ~ifelse(is.na(.), median(., na.rm=TRUE), .)))

message("Dataset listo. Dimensiones: ", nrow(small_df), " filas x ",
        ncol(small_df), " columnas.")

```

En este bloque se automatiza la obtención del dataset *UCI Human Activity Recognition*. El procedimiento incluye:

1. Descarga y descompresión automática desde el repositorio oficial de la UCI.
2. Limpieza sintáctica de los nombres de las variables (eliminación de paréntesis y guiones) para evitar conflictos en R.
3. Unificación de los conjuntos de entrenamiento y prueba.
4. Mapeo de las etiquetas numéricas de actividad a sus nombres descriptivos en español.
5. Selección exclusiva de las 12 variables cinemáticas de interés y aplicación de un muestreo estratificado para obtener 500 observaciones balanceadas.

### 4.3. Aplicación de la SVD (Cálculo de PCA)

```
# 1. Separar matriz numerica y Estandarizar (Media 0, Varianza 1)
X_raw <- as.matrix(select(small_df, -subject, -activity))
X_scaled <- scale(X_raw) # Importante para PCA

# 2. Aplicar PCA (basado en SVD)
# prcomp en R usa svd() internamente.
#  $X = U * D * V'$  -> prcomp devuelve:
# sdev (D), rotation (V), x (Scores =  $U * D$ )
pca_res <- prcomp(X_scaled, center = FALSE, scale. = FALSE)
```

En esta sección se ejecuta el núcleo matemático del análisis. Primero, se extrae la matriz numérica y se aplica una estandarización (Z-score) para asegurar que todas las variables tengan media 0 y varianza 1. Posteriormente, se utiliza la función `prcomp`, la cual basa su cálculo en la Descomposición en Valores Singulares (SVD) de la matriz de datos centrada. Esto nos permite obtener de forma robusta los valores singulares y los vectores propios necesarios.

### 4.4. Visualización inicial de la matriz U (Scores)

```
# Extraer los scores (equivalente a  $U * \text{Sigma}$ )
U_scores <- pca_res$x

# Mostrar dimensiones y primeras filas
cat("Dimensiones de la matriz de Scores ( $U * \text{Sigma}$ ):", dim(U_scores)[1], "x",
    dim(U_scores)[2], "\n")
print(head(U_scores[, 1:4], 5)) # Mostrar solo primeras 4 componentes
```

Se procede a inspeccionar la matriz de "Scores" (Puntuaciones), que en la terminología SVD corresponde al producto  $U\Sigma$ . Esta matriz contiene las coordenadas de las 500 observaciones proyectadas en el nuevo sistema de referencia ortogonal definido por los componentes principales. Visualizar estas filas permite verificar numéricamente la transformación de los datos originales.

## 4.5. Visualización de $V^T$ (Matriz de Cargas)

```
# Extraer la matriz de rotación (V)
V_loadings <- pca_res$rotation

# Mostrar dimensiones y primeras filas
cat("Dimensiones de la matriz de Rotación (V):", dim(V_loadings)[1], "x",
    dim(V_loadings)[2], "\n")
print(head(V_loadings[, 1:4], 5))
```

En este bloque se examina la matriz de rotación, equivalente a la matriz  $V$  (vectores propios o singulares derechos). Cada columna de esta matriz representa un componente principal y sus valores (cargas) indican el peso lineal de cada variable original en la construcción de dicho componente. Esta información es crucial para la interpretación semántica de las nuevas dimensiones.

## 4.6. Determinación de varianza acumulada

```
# Calcular Varianza Explicada
# Varianza = (Desviacion Estandar)^2
var_explained <- pca_res$sdev^2
prop_var <- var_explained / sum(var_explained)
cum_var <- cumsum(prop_var)

# Mostrar resumen de las primeras componentes
summary_pca <- data.frame(
  PC = paste0("PC", 1:length(prop_var)),
  Varianza_Pct = round(prop_var * 100, 2),
  Acumulada_Pct = round(cum_var * 100, 2)
)
print(head(summary_pca, 6))
```

Utilizando los valores singulares (almacenados como desviaciones estándar en el objeto PCA), se calcula la varianza explicada por cada componente ( $\sigma_i^2$ ) y la proporción respecto al total. Se genera una tabla resumen que permite cuantificar cuánta información del movimiento original se conserva al reducir la dimensionalidad.



## 4.7. Presentación del Scree Plot

```
# Preparar datos para el gráfico
df_plot <- summary_pca[1:12, ] # Usar las 12 componentes
df_plot$k <- 1:nrow(df_plot)

# Valor singular escalado (0-1) para comparar visualmente
df_plot$Singular_Escalado <- pca_res$sdev[1:12] / max(pca_res$sdev)

ggplot(df_plot, aes(x = k)) +
  # Barras de varianza individual
  geom_col(aes(y = Varianza_Pct/100, fill = "Varianza Individual"),
    alpha = 0.4) +
  # Linea de varianza acumulada
  geom_line(aes(y = Acumulada_Pct/100, color = "Varianza Acumulada"),
    linewidth = 1.2) +
  geom_point(aes(y = Acumulada_Pct/100, color = "Varianza Acumulada"),
    size = 3) +
  # Linea de Valores Singulares
  geom_line(aes(y = Singular_Escalado, color = "Valores Singulares (esc.)"),
    linetype = "dashed", linewidth = 1) +
  geom_point(aes(y = Singular_Escalado, color = "Valores Singulares (esc.)"),
    size = 2, shape = 1) +
  # Linea de corte referencial
  geom_hline(yintercept = 0.8, linetype = "dotted", color = "gray50") +

  # Estetica
  scale_y_continuous(labels = scales::percent, limits = c(0, 1.1)) +
  scale_x_continuous(breaks = 1:12) +
  scale_color_manual(values = c("Varianza Acumulada"="#00AFBB",
    "Valores Singulares (esc.)"="#FC4E07")) +
  scale_fill_manual(values = c("Varianza Individual"="gray70")) +
  labs(title = "SVD/PCA Scree Plot Combinado",
    x = "Componente Principal (k)", y = "Proporcion / Escala") +
  theme_minimal() + theme(legend.position = "bottom",
    legend.title = element_blank())
```

Se construye un gráfico de sedimentación compuesto que superpone tres métricas clave para la toma de decisiones:

1. Barras grises: Varianza individual de cada componente.

2. Línea azul: Varianza acumulada.
3. Línea naranja discontinua: Valores singulares normalizados.

Esta visualización integrada facilita la identificación del "codo" de la gráfica para seleccionar el número óptimo de componentes  $k$ .

## 4.8. Proyección 2D (Visualización de Clusters)

```
# 1. Preparar datos de proyeccion
scores_df <- data.frame(
  activity = small_df$activity,
  PC1 = pca_res$x[, 1],
  PC2 = pca_res$x[, 2]
)

# 2. Calcular centroides
centroids <- summarise(group_by(scores_df, activity),
  PC1 = mean(PC1), PC2 = mean(PC2))

# 3. Grafico de Dispersion
ggplot(scores_df, aes(x = PC1, y = PC2, color = activity)) +
  geom_point(alpha = 0.6, size = 2) +
  stat_ellipse(aes(fill = activity), geom = "polygon",
    alpha = 0.1, level = 0.8) +
  geom_label(data = centroids, aes(label = activity),
    color = "black", fontface = "bold", size = 3,
    alpha = 0.8) +
  scale_color_brewer(palette = "Set1") +
  scale_fill_brewer(palette = "Set1") +
  labs(title = "Proyección PCA (PC1 vs PC2)",
    subtitle = "Separación de clusters de actividad",
    x = "PC1", y = "PC2") +
  theme_minimal() + theme(legend.position = "none")
```

Se proyectan las observaciones en el plano formado por los dos primeros componentes principales ( $PC1$  vs  $PC2$ ). Se añaden elipses de confianza (nivel 0.8) y etiquetas en los centroides de cada actividad. Este gráfico es fundamental para evaluar cualitativamente si

la reducción de dimensionalidad ha logrado preservar la separabilidad entre las distintas clases de actividad humana (dinámicas vs estáticas).

## 4.9. Heatmap (Interpretación de Componentes)

```
# 1. Preparar datos
loadings <- as.data.frame(pca_res$rotation[, 1:4])
loadings$Variable <- rownames(loadings)

loadings_long <- pivot_longer(loadings, cols = starts_with("PC"),
                              names_to = "PC", values_to = "Carga")

# 2. Heatmap
ggplot(loadings_long, aes(x = PC, y = Variable, fill = Carga)) +
  geom_tile(color = "white") +
  scale_fill_gradient2(low = "#2166ac", mid = "#f7f7f7", high = "#b2182b",
                      midpoint = 0) +
  labs(title = "Heatmap de Cargas Factoriales",
       x = "Componente Principal", y = "Variable Original") +
  theme_minimal() +
  theme(axis.text.y = element_text(size = 8))
```

Para facilitar la interpretación de los componentes principales, se genera un mapa de calor (*Heatmap*) de la matriz de cargas ( $V$ ). Los colores indican la magnitud y dirección de la contribución de cada variable original a los componentes, permitiendo identificar rápidamente qué variables físicas (aceleración, gravedad, etc.) dominan en cada dimensión.

## 4.10. Biplot (Visualización Avanzada)

```
# 1. Preparación de Datos y Escalamiento
loadings_biplot <- as.data.frame(pca_res$rotation[, 1:2])
loadings_biplot$Variable <- rownames(loadings_biplot)

# Factor de escala para igualar magnitudes entre scores y loadings
max_score <- max(abs(pca_res$x[, 1:2]))
max_loading <- max(abs(pca_res$rotation[, 1:2]))
scale_factor <- (max_score / max_loading) * 0.8

loadings_biplot$v1 <- loadings_biplot$PC1 * scale_factor
loadings_biplot$v2 <- loadings_biplot$PC2 * scale_factor

# 2. Generación del Biplot
ggplot() +
  # Puntos (Actividades)
  geom_point(data = scores_df, aes(x = PC1, y = PC2, color = activity),
            alpha = 0.5, size = 1.5) +
  # Flechas (Variables) - Usando linewidth para evitar warnings
  geom_segment(data = loadings_biplot, aes(x = 0, y = 0, xend = v1, yend = v2),
            arrow = arrow(length = unit(0.2, "cm")),
            color = "darkred", linewidth = 0.8) +
  # Etiquetas
  geom_text(data = loadings_biplot, aes(x = v1, y = v2, label = Variable),
            color = "darkred", vjust = -0.5, size = 3, fontface = "bold") +
  scale_color_brewer(palette = "Set1") +
  labs(title = "Biplot de PCA: Variables vs. Actividades",
        x = "PC1", y = "PC2", color = "Actividad") +
  theme_minimal()
```

Por último, se implementa un gráfico *Biplot*, una herramienta visual avanzada que superpone las puntuaciones de las observaciones (puntos) y los vectores de carga de las variables (flechas) en el mismo plano factorial. La dirección y longitud de las flechas rojas indican qué variables originales tienen mayor peso en la formación de los ejes principales y hacia qué clústeres de actividad apuntan, revelando la correlación directa entre la cinemática medida (sensores) y la actividad física realizada.

## 5. Análisis de Resultados

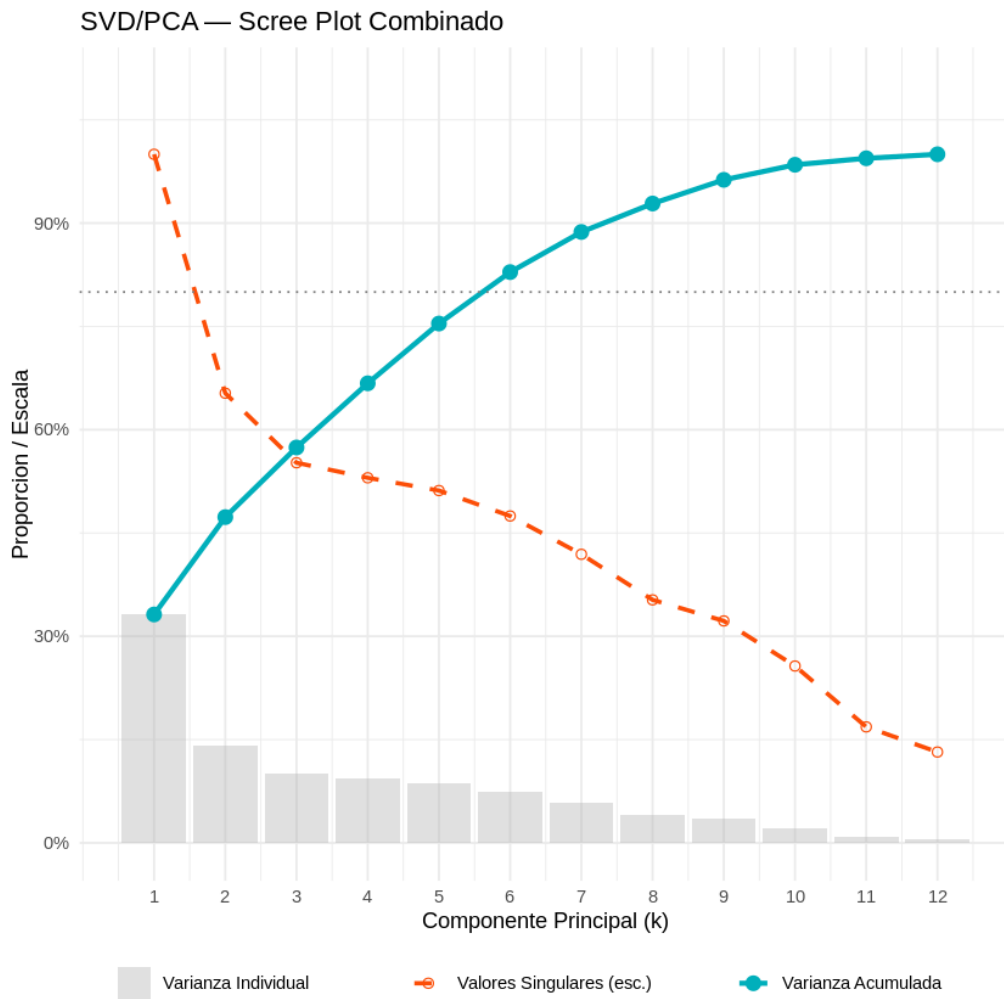
La aplicación de la Descomposición en Valores Singulares (SVD) y el Análisis de Componentes Principales (PCA) sobre el subconjunto de datos cinemáticos permitió reducir la dimensionalidad del problema original de 12 variables a un espacio latente interpretable. A continuación, se presentan los hallazgos principales derivados de las visualizaciones generadas.

### 5.1. Dimensionalidad Óptima (Scree Plot)

La Figura 1 muestra la varianza explicada por cada componente principal. Se observa un decaimiento exponencial en los valores singulares (línea punteada), característico de conjuntos de datos con alta correlación interna.

- **Dcaimiento de la información:** El primer componente principal (PC1) captura la mayor proporción de varianza individual (aproximadamente el 33%), seguido por el PC2 y PC3.
- **Punto de codo:** El gráfico sugiere un codo entre el tercer y cuarto componente. Esto indica que retener los primeros 3 componentes es suficiente para preservar más del 50% de la variabilidad original del movimiento humano, eliminando el ruido asociado a las 9 dimensiones restantes.

**Figura 1:** Gráfico de sedimentación combinado: Varianza y Valores Singulares.



Fuente: Propia (Generado en R)

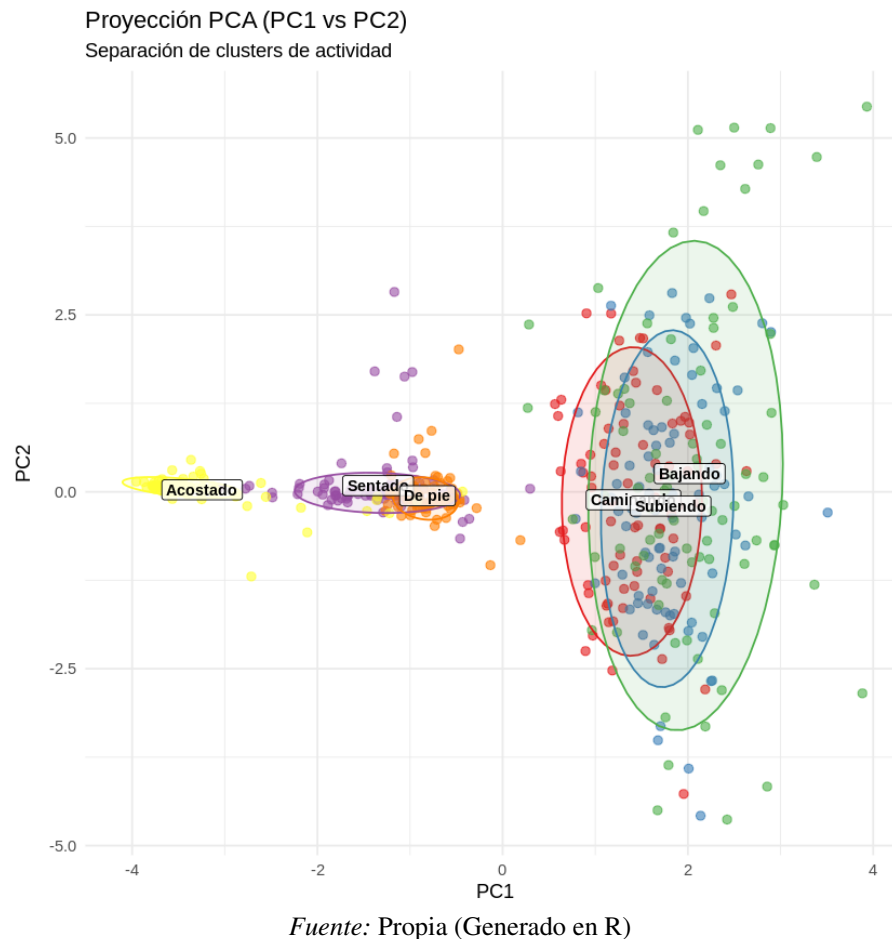
## 5.2. Separabilidad de Actividades (Proyección 2D)

Al proyectar las observaciones sobre el plano formado por los dos primeros componentes principales (Figura 4), emerge una estructura de agrupamiento clara que valida la eficacia del método algebraico.

- **Dicotomía Estática vs. Dinámica:** El eje horizontal (PC1) actúa como un discriminante natural de la intensidad del movimiento. Las actividades dinámicas (*Caminando, Subiendo, Bajando*) se agrupan en valores positivos de PC1, mientras que las actividades estáticas (*Sentado, De pie, Acostado*) se concentran en valores negativos.
- **Distinción de Posturas:** Aunque las actividades "Sentado" y "De pie" presentan cier-

to solapamiento (debido a la similitud en la inmovilidad del torso), la actividad "Acostado" forma un clúster claramente separado en el eje PC2, lo que sugiere que este segundo componente captura variaciones en la orientación respecto a la gravedad.

**Figura 2:** Proyección de las actividades en el espacio reducido (PC1 vs PC2).



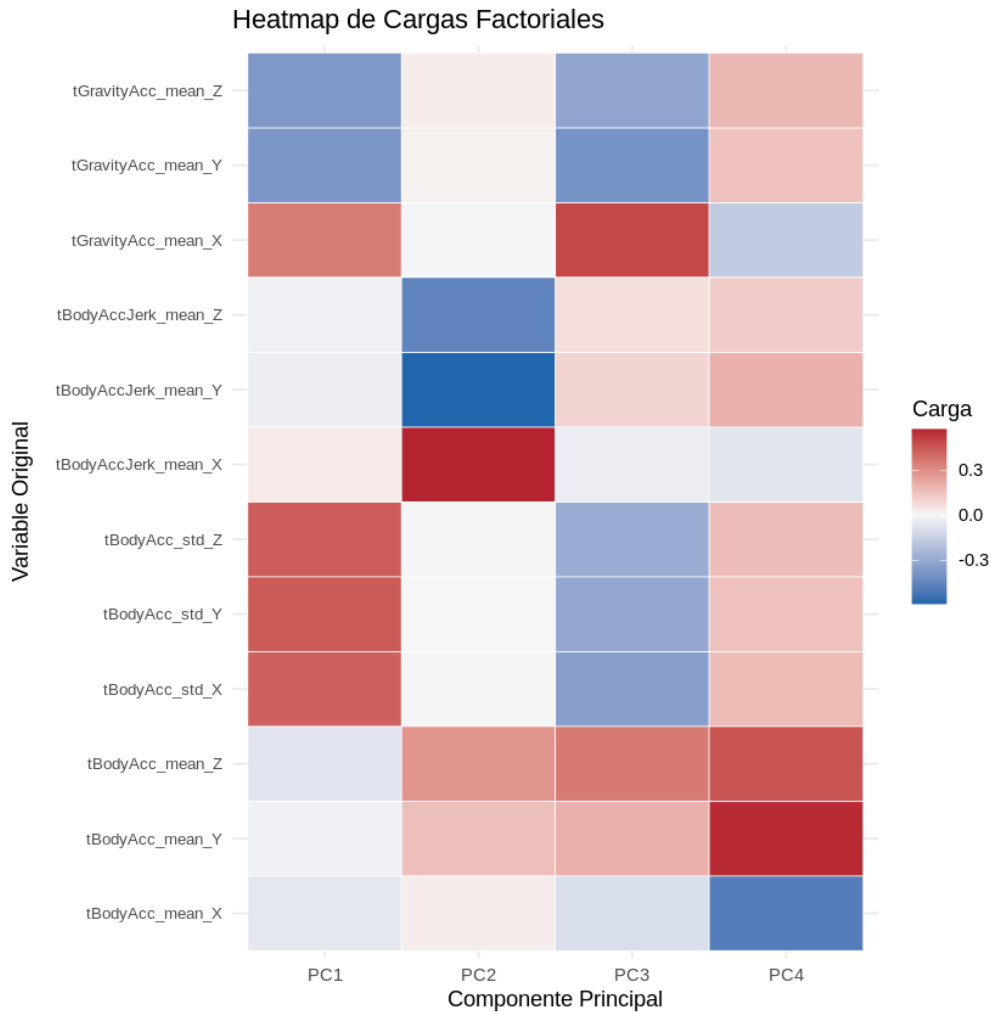
### 5.3. Interpretación de Variables (Heatmap)

El mapa de calor de la Figura 3 permite interpretar semánticamente los ejes abstractos generados por la SVD al mostrar la magnitud de las cargas (V).

- **PC1 (Intensidad):** Se observa una fuerte coloración (alta correlación) en variables relacionadas con la media y desviación estándar de la aceleración del cuerpo ( $tBodyAcc-mean$ ,  $tBodyAcc-std$ ). Esto confirma matemáticamente por qué este eje separa caminar de estar sentado.

- **PC2 (Orientación/Gravedad):** Las variables asociadas a la gravedad ( $tGravity$  y  $Acc$ ) presentan cargas significativas en el segundo componente y subsiguientes, lo que explica la capacidad del modelo para distinguir si el sujeto está vertical (de pie) u horizontal (acostado).

**Figura 3:** Mapa de calor de las cargas factoriales (Matriz V).



Fuente: Propia (Generado en R)

## 5.4. Interacción Global de Variables y Clases (Biplot)

Finalmente, la Figura 4 presenta un gráfico *Biplot* que superpone las proyecciones de las observaciones (puntos) y los vectores de carga de las variables originales (flechas). Esta visualización valida la coherencia física del modelo:

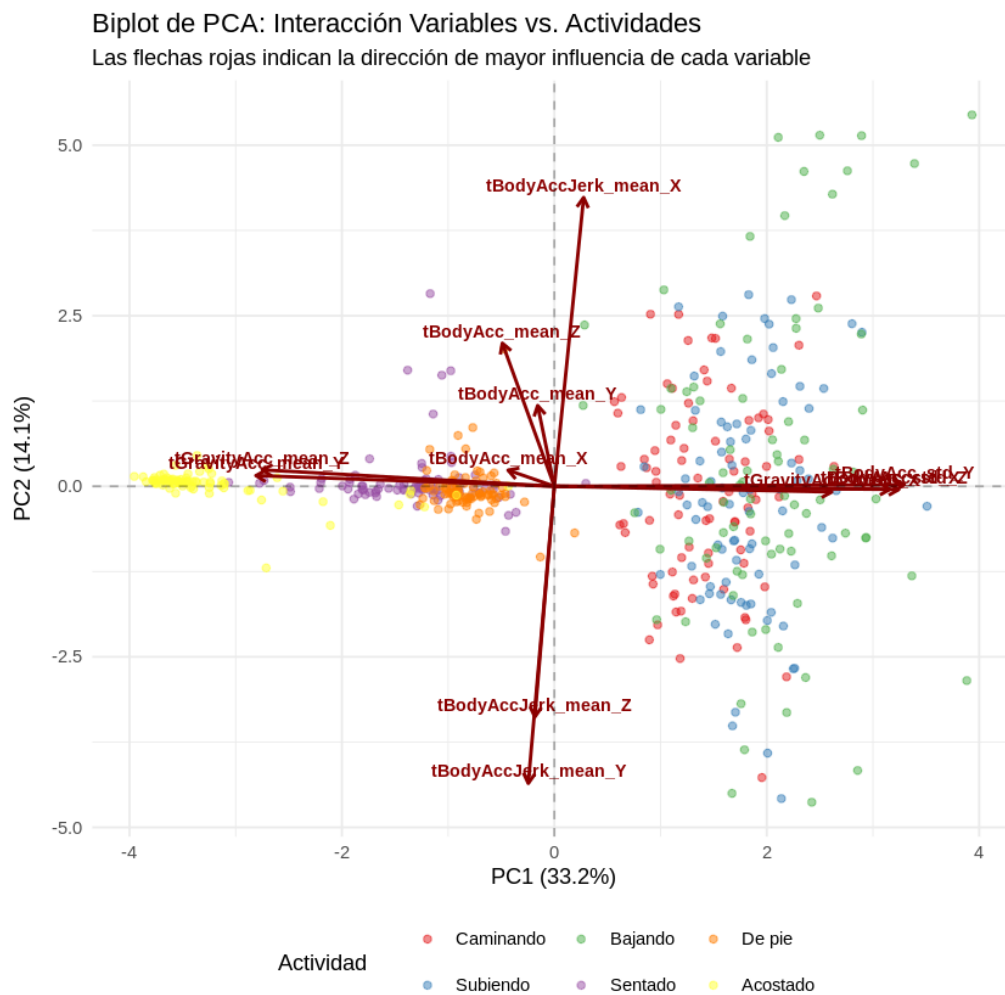
- **Direccionalidad Cinética:** Se observa que los vectores correspondientes a la desviación estándar de la aceleración (representados en rojo) apuntan directamente ha-



cia el clúster de actividades dinámicas (*Walking, Walking Upstairs*). Esto indica que la varianza en la aceleración es la variable determinante que aleja a estas muestras del origen.

- **Ortogonalidad y Postura:** Los vectores asociados a la gravedad (*Gravity*) tienden a ser ortogonales o divergentes respecto a los de aceleración corporal pura. Su dirección influye mayoritariamente en la separación a lo largo del segundo componente principal (PC2), permitiendo al algoritmo distinguir entre estados estáticos con distinta orientación espacial, como *Laying* (acostado) frente a *Standing* (de pie).

**Figura 4:** Biplot: Superposición de scores (actividades) y loadings (sensores).



Fuente: Propia (Generado en R)

## 6. Comentarios sobre la implementación

Durante el desarrollo computacional en R, se destacaron los siguientes aspectos técnicos y metodológicos:

- **Importancia del Escalamiento:** Fue crítico aplicar la estandarización (`scale`) previa a la SVD. Dado que el dataset original contiene variables de aceleración ( $m/s^2$ ) y derivadas de Jerk ( $m/s^3$ ), la omisión de este paso habría ocasionado que las variables con mayor magnitud numérica dominaran artificialmente la varianza, distorsionando los vectores propios.
- **Eficiencia de SVD en R:** La función `prcomp` demostró ser computacionalmente eficiente. Al utilizar SVD sobre la matriz de datos centrada en lugar de calcular explícitamente la matriz de covarianza ( $X^T X$ ), se garantiza una mayor estabilidad numérica, especialmente útil cuando se trabaja con matrices que podrían estar mal condicionadas.
- **Reproducibilidad:** La implementación de una semilla aleatoria (`set.seed`) durante el muestreo estratificado aseguró que los resultados visualizados y los clústeres obtenidos sean consistentes en ejecuciones repetidas, cumpliendo con el rigor de un reporte científico.

## 7. Referencias Bibliográficas

### Referencias

Al-Qazzaz, S., et al. (2024). Efficient Human Activity Recognition using PCA Dimensionality Reduction and GWO-Enhanced LSTM. *Journal of Advanced Research in Applied Sciences and Engineering Technology*, 54(2), 317-343. <https://doi.org/10.37934/araset.54.2.317343>

Anguita, D., Ghio, A., Oneto, L., Parra, X., & Reyes-Ortiz, J. L. (2013). A Public Domain Dataset for Human Activity Recognition Using Smartphones. *21st European Symposium on Artificial Neural Networks*, 437–442. <https://www.esann.org/sites/default/files/proceedings/legacy/es2013-84.pdf>

Lay, D. C., Lay, S. R., & McDonald, J. J. (2021). *Linear Algebra and Its Applications* (6th ed.). Pearson. <https://www.pearson.com/en-us/subject-catalog/p/linear-algebra-and-its-applications/P200000006337>

Liu, H., Gamboa, H., & Schultz, T. (2024). Human Activity Recognition, Monitoring, and Analysis Facilitated by Novel and Widespread Applications of Sensors. *Sensors*, 24(16), 5250. <https://doi.org/10.3390/s24165250>

Pernice, S. A. (2024). El problema de la reducción dimensional: Análisis de Componentes Principales (PCA). *Mutis*, 14(1), 1-21. <https://doi.org/10.21789/22561498.2057>

R Core Team. (2024). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing. <https://www.R-project.org/>

Strang, G. (2023). *Introduction to Linear Algebra* (6th ed.). Wellesley-Cambridge Press. <https://math.mit.edu/~gs/linearalgebra/>

### Enlace al código fuente:

<https://colab.research.google.com/drive/1u3DNiTiXCoSMFy4CwV85VadqNve1JwGB?usp=sharing>