

1 Descomposición en Valores Singulares

La **descomposición en valores singulares** (SVD: Singular Value Decomposition) es una factorización de matrices, la cual se aplica en varios ámbitos del álgebra lineal computacional.

1.1 Observación Geométrica

La imagen de la esfera unitaria bajo cualquier matriz $m \times n$ es una hiperelipse. El término “hiperelipse”, es la generalización de una elipse en el espacio m -dimensional. Se puede definir una hiperelipse en \mathbb{R}^m como la superficie obtenida, “alargando” la esfera unitaria en \mathbb{R}^m por factores $\sigma_1, \sigma_2, \dots, \sigma_m$ (posiblemente nulos) en direcciones ortogonales $\mathbf{u}_1, \mathbf{u}_2, \dots, \mathbf{u}_m \in \mathbb{R}^m$. Por conveniencia, sean los vectores \mathbf{u}_i unitarios, i.e., $\|\mathbf{u}_i\|_2 = 1$. Los vectores $\{\sigma_i \mathbf{u}_i\}$ son los *semiejes principales* de la hiperelipse, con longitud $\sigma_1, \sigma_2, \dots, \sigma_m$. Si A tiene rango r , exactamente r de las m longitudes son no nulas, y en particular, si $m \geq n$, al menos n longitudes son no nulas.

Cuando se denomina la esfera unitaria, es más que nada la esfera unitaria en el espacio n -dimensional en norma 2, la cual denominaremos S . Entonces AS , la imagen de S bajo el mapeo de A , es una hiperelipse (Figura 1).

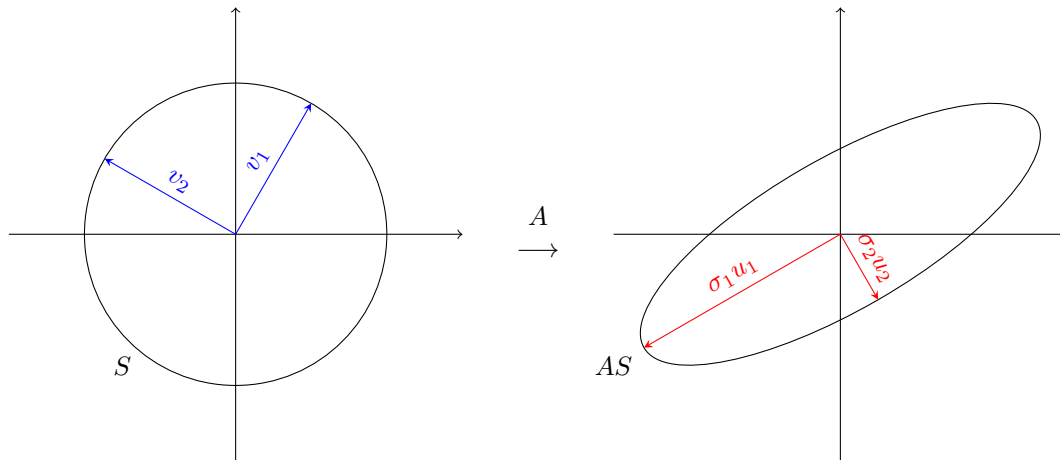


Figure 1: SVD de una matriz 2×2

Se definen algunas propiedades de A en términos de la forma de AS , suponer que $A \in \mathbb{R}^{m \times n}$ con $m \geq n$ y es full rank.

1. Los n **valores singulares** de A , son las longitudes de los n semiejes principales de AS , denominados $\sigma_1, \sigma_2, \dots, \sigma_n$. Se asume que los valores singulares son enumerados en orden descendiente $\sigma_1 \geq \sigma_2 \geq \dots \geq \sigma_n > 0$.
2. Los n **vectores singulares izquierdos** de A , son los vectores unitarios $\{\mathbf{u}_1, \mathbf{u}_2, \dots, \mathbf{u}_n\}$ orientados en la dirección de los semiejes principales de AS .
3. Los n **vectores singulares derechos** de A , son los vectores unitarios $\{\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_n\} \in S$, que son las pre - imágenes de los semiejes principales de AS , luego $A\mathbf{v}_j = \sigma_j \mathbf{u}_j$.

1.2 SVD Reducida y SVD Completa

Las ecuaciones que relacionan los vectores singulares izquierdos y derechos pueden ser escritas como:

$$A\mathbf{v}_j = \sigma_j \mathbf{u}_j, \quad 1 \leq j \leq n.$$

Matricialmente, puede expresarse como:

$$\left[\begin{array}{c} A \end{array} \right] \left[\begin{array}{c|c|c|c} \mathbf{v}_1 & \mathbf{v}_2 & \cdots & \mathbf{v}_n \end{array} \right] = \left[\begin{array}{c|c|c|c} \mathbf{u}_1 & \mathbf{u}_2 & \cdots & \mathbf{u}_n \end{array} \right] \left[\begin{array}{cccc} \sigma_1 & & & \\ & \sigma_2 & & \\ & & \ddots & \\ & & & \sigma_n \end{array} \right]$$

más compacto, $AV = \hat{U}\hat{\Sigma}$. $\hat{\Sigma}$ es una matriz diagonal de dimensión $n \times n$ con elementos positivos (asumiendo que A es full rank n), \hat{U} es una matriz de dimensión $m \times n$, con columnas ortonormales y V es una matriz de dimensión $n \times n$ con columnas ortonormales. Luego V es unitaria, por lo tanto $V^{-1} = V^*$, se obtiene entonces:

$$A = \hat{U}\hat{\Sigma}V^*,$$

esta factorización se denomina *descomposición en valores singulares reducida* de A (Figura 2).

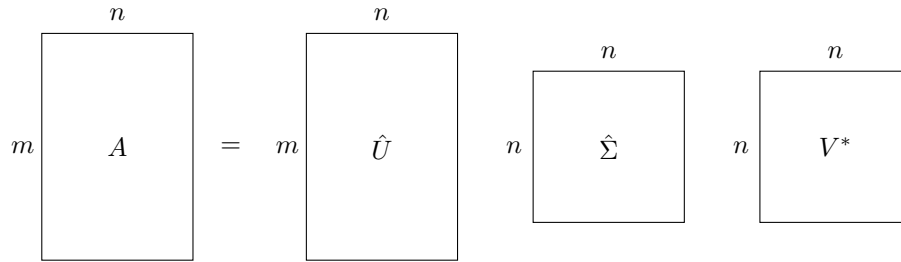


Figure 2: Factorización SVD reducida.

Las columnas de \hat{U} son n vectores ortonormales en el espacio \mathbb{C}^m . A menos que $m = n$, este conjunto de vectores no forman una base para \mathbb{C}^m , es más, \hat{U} no es unitaria. Aún así, si se adjuntan los $m - n$ vectores ortonormales (V.O.) que faltan, \hat{U} puede convertirse en una matriz unitaria. Por lo tanto, ahora se cambia la matriz \hat{U} , por una matriz unitaria $U \in \mathbb{C}^{m \times m}$. A raíz de este cambio, la matriz $\hat{\Sigma}$ también debe ser reemplazada, tomando la misma forma que la matriz A ; para lograr esto, simplemente se agregan elementos nulos para completar la dimensión $m \times n$, dando paso a una matriz Σ . Ahora entonces, se tiene una nueva factorización:

$$A = \underbrace{U}_{\substack{m \times m \\ \text{unitaria}}} \underbrace{\Sigma}_{\substack{m \times n \\ \text{diagonal}}} \underbrace{V^*}_{\substack{n \times n \\ \text{unitaria}}}$$

esta factorización corresponde a la *descomposición en valores singulares completa* de A (Figura 3).

Es preciso notar que la SVD es muy similar a la descomposición en valores propios:

$$AX = X\Lambda$$

Donde matricialmente se tiene lo siguiente:

$$\left[\begin{array}{c} A \end{array} \right] \left[\begin{array}{c|c|c|c} \mathbf{x}_1 & \mathbf{x}_2 & \cdots & \mathbf{x}_n \end{array} \right] = \left[\begin{array}{c|c|c|c} \mathbf{x}_1 & \mathbf{x}_2 & \cdots & \mathbf{x}_n \end{array} \right] \left[\begin{array}{cccc} \lambda_1 & & & \\ & \lambda_2 & & \\ & & \ddots & \\ & & & \lambda_n \end{array} \right]$$

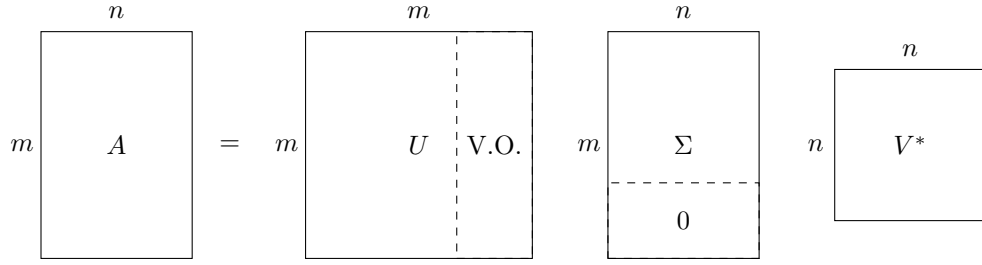


Figure 3: Factorización SVD completa.

Cada valor propio λ_i se encuentra relacionado con un vector propio asociado \mathbf{x}_i y cumple con la relación:

$$A\mathbf{x}_i = \lambda_i\mathbf{x}_i,$$

Generalmente, ambas descomposiciones son distintas, aunque podría darse el caso de que sean iguales. Otro aspecto a considerar es que toda matriz tiene una descomposición en valores singulares, pero no todas las matrices tienen una descomposición en valores propios. La matriz Λ es una matriz diagonal, y X es una matriz no singular. Debido a esto $X^* \neq X^{-1}$ y la matriz A está dada por:

$$A = X\Lambda X^{-1}$$

En este caso, se dice entonces que las matrices A y Λ son similares.

1.3 Algunas Propiedades

Asumir que $A \in \mathbb{C}^{m \times n}$, además $p = \min\{m, n\}$, $r \leq p$ denota el número de valores singulares no nulos y $\langle \mathbf{x}, \mathbf{y}, \dots, \mathbf{z} \rangle$ denota el espacio generado por los vectores $\mathbf{x}, \mathbf{y}, \dots, \mathbf{z}$.

Teorema 1. *El rango de A es r , la cantidad de valores singulares no nulos.*

Demostración: Como U y V^* son matrices unitarias, $\text{range}\{A\} = \text{range}\{\Sigma\}$, donde la última corresponde al número de valores diagonales distintos de cero.

Teorema 2. $\text{range}\{A\} = \langle \mathbf{u}_1, \mathbf{u}_2, \dots, \mathbf{u}_r \rangle$ y $\text{null}\{A\} = \langle \mathbf{v}_{r+1}, \dots, \mathbf{v}_n \rangle$

Demostración: Ya que r es el número de valores singulares no nulos. En general, se cumple lo siguiente:

$$\left[\begin{array}{c|c|c|c|c|c|c} \overbrace{\text{col}\{A\}} & & & & \overbrace{\text{null}\{A^*\}} & & \\ \hline \mathbf{u}_1 & \dots & \mathbf{u}_r & \mathbf{u}_{r+1} & \dots & \mathbf{u}_m & \end{array} \right] \left[\begin{array}{ccccccc} \sigma_1 & & & & & & \\ & \ddots & & & & & \\ & & \sigma_r & & & & \\ & & & 0 & & & \\ & & & & \ddots & & \\ & & & & & 0 & \end{array} \right] \left[\begin{array}{c} \mathbf{v}_1^* \\ \vdots \\ \mathbf{v}_r^* \\ \mathbf{v}_{r+1}^* \\ \vdots \\ \mathbf{v}_n^* \end{array} \right] \left\{ \begin{array}{l} \text{row}\{A\} \\ \text{null}\{A\} \end{array} \right.$$

Teorema 3. $\|A\|_2 = \sigma_1$ y $\|A\|_F = \sqrt{\sigma_1^2 + \dots + \sigma_r^2}$

Demostración: Las normas son invariantes frente a la multiplicación de matrices unitarias, por lo tanto:

$$\|A\|_2 = \|U\Sigma V^*\|_2 = \|\Sigma\|_2 = \max_{1 \leq i \leq m} |d_i| = \sigma_1$$

$$\|A\|_F = \|U\Sigma V^*\|_F = \|\Sigma\|_F = \sqrt{\sigma_1^2 + \dots + \sigma_r^2}$$

Teorema 4. *Los valores singulares no nulos de A , son la raíz cuadrada de los valores propios no nulos de A^*A o AA^* (Estas matrices tienen los mismos valores propios no nulos).*

Demostración: A^*A es una matriz cuadrada, simétrica y semidefinida positiva que acepta una diagonalización del tipo $X\Lambda X^{-1}$, con Λ la matriz diagonal de los valores propios y X la matriz de vectores propios. Además:

$$A^*A = X\Lambda X^{-1} = (U\Sigma V^*)^*(U\Sigma V^*) = V\Sigma^*U^*U\Sigma V^* = V\Sigma^*\Sigma V^*$$

Se cumple que:

$$\begin{aligned} \Lambda &= \Sigma\Sigma^* \\ \begin{bmatrix} \lambda_1 & & & \\ & \lambda_2 & & \\ & & \ddots & \\ & & & \lambda_n \end{bmatrix} &= \begin{bmatrix} \sigma_1 & & & \\ & \sigma_2 & & \\ & & \ddots & \\ & & & \sigma_n \end{bmatrix} \begin{bmatrix} \sigma_1 & & & \\ & \sigma_2 & & \\ & & \ddots & \\ & & & \sigma_n \end{bmatrix} \\ \begin{bmatrix} \lambda_1 & & & \\ & \lambda_2 & & \\ & & \ddots & \\ & & & \lambda_n \end{bmatrix} &= \begin{bmatrix} \sigma_1^2 & & & \\ & \sigma_2^2 & & \\ & & \ddots & \\ & & & \sigma_n^2 \end{bmatrix} \end{aligned}$$

Donde $\lambda_i = \sigma_i^2$. Este procedimiento es análogo con la matriz AA^* .

Teorema 5. A es la suma de r matrices de rango 1

$$A = \sum_{k=1}^r \sigma_k \mathbf{u}_k \mathbf{v}_k^*,$$

Donde r es el rango de A y \mathbf{u}_i y \mathbf{v}_i son las i -ésimas columnas de las matrices U y V , respectivamente.

Demostración

$$\begin{aligned} A &= U\Sigma V^* \\ &= U \begin{bmatrix} \sigma_1 & & & \\ & \ddots & & \\ & & \sigma_r & \end{bmatrix} V^* \\ &= U \left(\begin{bmatrix} \sigma_1 & & & \\ & & & \end{bmatrix} + \begin{bmatrix} & \sigma_2 & & \\ & & & \end{bmatrix} + \dots + \begin{bmatrix} & & & \sigma_r \end{bmatrix} \right) V^* \\ &= \sigma_1 \mathbf{u}_1 \mathbf{v}_1^* + \sigma_2 \mathbf{u}_2 \mathbf{v}_2^* + \dots + \sigma_r \mathbf{u}_r \mathbf{v}_r^* \end{aligned}$$

1.4 Cálculo de SVD

Para obtener la factorización SVD con una matriz $A \in \mathbb{C}^{m \times n}$ se debe seguir el siguiente procedimiento:

1. Calcular el producto A^*A .

$$\begin{aligned} A^*A &= (U\Sigma V^*)^*(U\Sigma V^*) \\ &= V\Sigma^*U^*U\Sigma V^* \\ &= V\Sigma^*\Sigma V^* \end{aligned}$$

2. La matriz $\Sigma^*\Sigma$ es la matriz diagonal:

$$\begin{bmatrix} \sigma_1^2 & & & \\ & \sigma_2^2 & & \\ & & \ddots & \\ & & & \sigma_n^2 \end{bmatrix}$$

Donde los valores de la diagonal corresponden a los valores propios de la matriz A^*A . Se calculan los vectores propios $\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_n$ asociados cada valor propio. Estos corresponden a los valores para construir la matriz V . La matriz Σ se construye con los valores $\sigma_1, \sigma_2, \dots, \sigma_n$.

3. Como ya se tienen las matrices V y Σ , se calcula la matriz U despejando.

Es preciso hacer notar que computacionalmente este método no es preferido para calcular el *SVD* cuando el número de condición de la matriz A es muy grande. En dichos casos, es preferible formar la matriz B :

$$B = \begin{bmatrix} 0 & A^* \\ A & 0 \end{bmatrix}$$

Esta matriz B es una matriz simétrica de $(m+n)$ filas y $(m+n)$ columnas. De esta forma, la matriz tiene valores propios reales y una base de vectores propios. Sea λ un valor propio y $[\mathbf{u}, \mathbf{v}]$ un $(m+n)$ -vector que es un vector propio de la matriz B , entonces según la ecuación 1.2:

$$\begin{aligned} B\mathbf{x} &= \lambda\mathbf{x} \\ \begin{bmatrix} 0 & A^* \\ A & 0 \end{bmatrix} \begin{bmatrix} \mathbf{u} \\ \mathbf{v} \end{bmatrix} &= \lambda \begin{bmatrix} \mathbf{u} \\ \mathbf{v} \end{bmatrix} \\ \begin{bmatrix} A^*\mathbf{v} \\ A\mathbf{u} \end{bmatrix} &= \lambda \begin{bmatrix} \mathbf{u} \\ \mathbf{v} \end{bmatrix} \end{aligned}$$

Se tiene un problema de valores propios, donde se tienen las ecuaciones:

$$A^*\mathbf{v} = \lambda\mathbf{u} \tag{1}$$

$$A\mathbf{u} = \lambda\mathbf{v} \tag{2}$$

Al multiplicar por la izquierda la ecuación 2 por A^T resulta:

$$A^*A\mathbf{u} = \lambda A^T\mathbf{v}$$

Ahora, utilizando la ecuación 1:

$$A^*A\mathbf{u} = \lambda^2\mathbf{u}$$

Donde \mathbf{u} es el vector propio asociado a A^*A con su correspondiente valor propio λ^2 . Se pueden determinar los valores propios y vectores propios sin calcular la matriz A^*A . Luego se determinan las matrices U , V y Σ de manera similar al otro método.

1.5 Ejemplo

Considerar la matriz A .

$$A = \begin{bmatrix} 1 & -1 \\ 1 & 1 \\ 1 & 1 \end{bmatrix}$$

La matriz A no tiene grandes dimensiones y está bien condicionada, pues su número de condición es un valor pequeño. Por lo tanto se seguirá el primer método explicado para obtener los valores propios. En primer lugar, se calcula el producto A^*A :

$$A^*A = \begin{bmatrix} 1 & 1 & 1 \\ -1 & 1 & 1 \end{bmatrix} \begin{bmatrix} 1 & -1 \\ 1 & 1 \\ 1 & 1 \end{bmatrix} = \begin{bmatrix} 3 & 1 \\ 1 & 3 \end{bmatrix}$$

A continuación, se determinan los valores propios de la matriz A^*A , resultando $\lambda_1 = \sigma_1^2 = 4$ y $\lambda_2 = \sigma_2^2 = 2$. Estos valores corresponden a la diagonal de la matriz $\Sigma^*\Sigma$. Por lo tanto, la matriz $\hat{\Sigma}$ está dada por:

$$\hat{\Sigma} = \begin{bmatrix} \sigma_1 & 0 \\ 0 & \sigma_2 \end{bmatrix} = \begin{bmatrix} 2 & 0 \\ 0 & \sqrt{2} \end{bmatrix}$$

Luego, se calculan los valores de \mathbf{v}_1 y \mathbf{v}_2 , los vectores propios asociados a λ_1 y λ_2 , respectivamente:

$$\mathbf{v}_1 = \begin{bmatrix} 1 \\ 1 \end{bmatrix}, \mathbf{v}_2 = \begin{bmatrix} 1 \\ -1 \end{bmatrix}$$

Todos los vectores propios representan las columnas de la matriz V . Sin embargo, debido a que la matriz V presenta columnas ortonormales, cada vector propio pasa por un proceso previo de normalización.

$$\mathbf{v}_1 = \begin{bmatrix} \sqrt{2}/2 \\ \sqrt{2}/2 \end{bmatrix}, \mathbf{v}_2 = \begin{bmatrix} \sqrt{2}/2 \\ -\sqrt{2}/2 \end{bmatrix}$$

Cada vector normalizado corresponde a una columna para la matriz V :

$$V = [\mathbf{v}_1 | \mathbf{v}_2] = \begin{bmatrix} \sqrt{2}/2 & \sqrt{2}/2 \\ \sqrt{2}/2 & -\sqrt{2}/2 \end{bmatrix}$$

En este caso $V = V^*$ (la matriz es simétrica). Ya que se tienen las matrices $\hat{\Sigma}$ (1.5) y V^* (1.5) se calcula la matriz \hat{U} , al resolver el sistema de ecuaciones $AV = \hat{U}\hat{\Sigma}$ (1.2):

$$\begin{bmatrix} 1 & -1 \\ 1 & 1 \\ 1 & 1 \end{bmatrix} \begin{bmatrix} \sqrt{2}/2 & \sqrt{2}/2 \\ \sqrt{2}/2 & -\sqrt{2}/2 \end{bmatrix} = \hat{U} \begin{bmatrix} 2 & 0 \\ 0 & \sqrt{2} \end{bmatrix}$$

De esta forma, se obtiene la matriz \hat{U} :

$$U = \begin{bmatrix} 0 & 1 \\ \sqrt{2}/2 & 0 \\ \sqrt{2}/2 & 0 \end{bmatrix}$$

Finalmente, se tiene la *descomposición en valores singulares reducida* de la matriz A :

$$A = \hat{U}\hat{\Sigma}V^* = \begin{bmatrix} 0 & 1 \\ \sqrt{2}/2 & 0 \\ \sqrt{2}/2 & 0 \end{bmatrix} \begin{bmatrix} 2 & 0 \\ 0 & \sqrt{2} \end{bmatrix} \begin{bmatrix} \sqrt{2}/2 & \sqrt{2}/2 \\ \sqrt{2}/2 & -\sqrt{2}/2 \end{bmatrix}$$

Para obtener la *descomposición en valores singulares completa* de A , debe agregarse vector columna ortonormal a la matriz \hat{U} y un correspondiente vector fila nulo en la matriz $\hat{\Sigma}$:

$$A = U\Sigma V^* = \begin{bmatrix} 0 & 1 & 0 \\ \sqrt{2}/2 & 0 & -\sqrt{2}/2 \\ \sqrt{2}/2 & 0 & \sqrt{2}/2 \end{bmatrix} \begin{bmatrix} 2 & 0 \\ 0 & \sqrt{2} \\ 0 & 0 \end{bmatrix} \begin{bmatrix} \sqrt{2}/2 & \sqrt{2}/2 \\ \sqrt{2}/2 & -\sqrt{2}/2 \end{bmatrix}$$

2 Principal Component Analysis

El *Principal Component Analysis* o Análisis de Componentes Principales es una técnica utilizada para la reducción dimensional de conjuntos de datos. El objetivo es encontrar una nueva base vectorial para expresar el conjunto de datos basados en la distribución de ellos mismos.

2.1 Algunos Conceptos Previos

Dado un vector $\mathbf{x} \in \mathbb{R}^n$, el promedio \bar{x} de \mathbf{x} está dado por (3)

$$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i \quad (3)$$

La varianza es una medida de dispersión de datos alrededor de la media. En el caso de un conjunto x con n datos, la varianza está dada por:

$$\text{var}(x) = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2 \quad (4)$$

Se utiliza el cociente $\frac{1}{n-1}$ en lugar de $\frac{1}{n}$, para eliminar el sesgo. Al considerar un segundo conjunto de n datos y , se define la covarianza, que mide el grado de relación lineal entre dos variables, como:

$$\text{cov}(x,y) = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y}) \quad (5)$$

Al convertir x e y a vectores fila con n valores, se puede definir la covarianza (6) como un producto punto entre \mathbf{x} y \mathbf{y}^T :

$$\text{cov}(x,y) = \frac{1}{n-1} \mathbf{x} \mathbf{y}^T \quad (6)$$

Notar que cuando el conjunto de datos es el mismo, la covarianza entre estos conjuntos de datos es igual a la varianza de los datos. Esto quiere decir que $\text{cov}(x, x) = \text{var}(x)$.

2.2 Cambio de base

Suponer que se cuenta con una muestra de n puntos $\mathbf{w}_1, \mathbf{w}_2, \dots, \mathbf{w}_n$, con $\mathbf{w}_i \in \mathbb{R}^d$. De esta manera, cada punto puede verse como un vector de d filas o características. Entonces, se forma una matriz de n filas y d columnas con todos los datos de la muestra, con \mathbf{w}_i en la i -ésima fila y considerando $1 \leq i \leq n$. Cada columna de la matriz representaría una misma característica en el conjunto de datos.

$$W = \begin{bmatrix} \mathbf{w}_1 \\ \mathbf{w}_2 \\ \vdots \\ \mathbf{w}_n \end{bmatrix} \quad (7)$$

Una base ingenua que sirve para representar el conjunto de datos sería la base canónica. Por ejemplo, en \mathbb{R}^2 la base canónica sería $B_2 = \{(1,0), (0,1)\}$. Esto es fácilmente escalable hasta d dimensiones. Entonces, cada dato \mathbf{w}_i puede ser expresado en función de la base canónica. El problema con este tipo de base es que probablemente no sea capaz de expresar de buena manera el conjunto de datos. Una mejor expresión del conjunto de datos está relacionado con un cambio de base y con dos conceptos: el ruido y la redundancia.

Un punto \mathbf{w}_i expresado en la base canónica se expresa como:

$$\mathbf{w}_i = \sum_{j=1}^d w_{ij} \mathbf{e}_j$$

Donde \mathbf{e}_j es la base canónica. De manera análoga, con una base distinta, el punto \mathbf{w}_i queda expresado como:

$$\mathbf{w}_i = \sum_{j=1}^d y_{ij} \mathbf{p}_j$$

En este caso, el punto \mathbf{w}_i expresado en la nueva base \mathbf{p}_j utiliza valores y_{ij} . Un cambio de base es una transformación lineal representada por una matriz de cambio de base B . Matricialmente un cambio de base se expresa de la siguiente manera:

$$Y = WB \quad (8)$$

La matriz W posee los valores w_{ij} , que corresponde a \mathbf{w}_i expresado con la base canónica. Por otro lado, la matriz Y posee los valores y_{ij} , \mathbf{w}_i expresado con la base \mathbf{p}_j . Geométricamente, un cambio de base representa una rotación o un reajuste que transforma W en Y .

El ruido en un conjunto de datos corresponde a interferencias en la medición que ocasiona que los datos registrados no sean los correctos. Un conjunto de datos con mucho ruido no permitiría extraer información. Una forma de medir el ruido en la muestra de datos es por medio de la Relación Señal/Ruido (SNR en inglés), que establece una proporción entre la señal que se transmite y el ruido que corrompe la medición. Valores mucho mayores a uno representan alta precisión, mientras que valores cercanos a cero corresponden a información contaminada.

En el caso de la redundancia, al tener un sistema dinámico y realizar mediciones en algún momento es posible que se puedan registrar las mismas mediciones una gran cantidad de veces. Esto puede deberse porque las fuentes de medición están muy cercanas o se está midiendo una misma característica en distintas unidades. Cuando ocurre esto es posible establecer relaciones entre características, lo que daría paso a la eliminación de una de las características.

¿Cómo relacionar el cambio de base con el ruido y la redundancia? Considerar la matriz W (7), se define la matriz de covarianzas de W como C_W :

$$C_W = \frac{1}{n-1} W^T W$$

Esta matriz tiene varias propiedades:

- C_W es una matriz simétrica de $d \times d$.
- Los términos de la diagonal corresponden a la varianza de característica en particular.
- Los términos fuera de la diagonal corresponden a la covarianza entre características.

C_W captura la correlación entre todas las características y refleja el ruido y la redundancia. Los valores diagonales mayores son características interesantes, mientras que los menores valores diagonales corresponden al ruido. Fuera de la diagonal, valores altos corresponden a alta redundancia y valores bajos a baja redundancia. El objetivo de PCA es minimizar la redundancia, medida por la covarianza y maximizar la señal, medida por la varianza. Entonces, la matriz de covarianza de Y (8), llamada C_Y , debe ser una matriz diagonal.

2.3 PCA y SVD

Sea la Z la matriz formada por la matriz X al restarle a cada columna el promedio de la característica respectiva:

$$Z = X - \mu$$

Donde μ es un vector fila con valores μ_1, \dots, μ_d promedios de las características. Se utilizará esta matriz que tiene la propiedad de tener media cero en cada una de las características y covarianza positiva entre las características. PCA está relacionado con SVD, ya que es posible obtener un cambio de base para la matriz Z , consiguiendo una diagonalización de la matriz de covarianzas, con el objetivo de maximizar la señal y minimizar el ruido:

$$\begin{aligned}
Z &= U\Sigma V^* \\
ZV &= U\Sigma \\
ZV &= Y
\end{aligned}$$

De (2.3) se desprende que el cambio de base buscado en (8) corresponde a $Y = ZV$, donde la matriz V de la SVD es la matriz de cambio de base. Las columnas de V , representan las **componentes principales** de Z . Por otro lado, se tiene la igualdad $Y = U\Sigma$, con la cual se demuestra que la matriz de covarianzas C_Y es una matriz diagonal con los valores propios de Z :

$$\begin{aligned}
C_Y &= \frac{1}{n-1} Y^T Y \\
&= \frac{1}{n-1} (U\Sigma)^T U\Sigma \\
&= \frac{1}{n-1} \Sigma^T U^T U \Sigma \\
&= \frac{1}{n-1} \Lambda
\end{aligned}$$

Es imperante mencionar que para realizar PCA podemos hacer la SVD de Z como realizar la descomposición en valores propios de Z^*Z , ambos procedimientos entregan la matriz V , más los valores singulares y valores propios respectivamente.