



PONTIFICIA  
UNIVERSIDAD  
CATÓLICA  
DE CHILE




PONTIFICIA  
UNIVERSIDAD  
CATÓLICA  
DE CHILE

# Algebra Lineal Numérica

Clase SVD

**Manuel A. Sánchez**  
**2024.04.10**

## Algebra Lineal

- 1 Descomposición en valores singulares
  - 2 Aproximación matricial
  - 3 Análisis de Componentes Principales (PCA)
  - 4 Mínimos cuadrados ordinarios
  - 5 Problema de mínimos cuadrados
  - 6 Las ecuaciones normales
  - 7 Resolver las ecuaciones normales
  - 8 Mínimos cuadrados y data fitting
  - 9 Mínimos cuadrados multi-objetivo
  - 10 Mínimos cuadrados: solución de mínima norma
- 

# Descomposición en valores singulares

## ■ Introducción y motivación

- La descomposición en valores singulares (SVD) y el análisis de componentes principales (PCA) son herramientas esenciales en álgebra lineal y análisis de datos.
- Ambos tienen aplicaciones extensas en diversas áreas, incluyendo procesamiento de imágenes, aprendizaje automático, estadísticas y más.
- Comprender estos conceptos nos permite entender la estructura subyacente y reducir la dimensionalidad de los datos.

### **Objetivos de la clase**

- Introducir la descomposición en valores singulares (SVD) como una herramienta fundamental en álgebra lineal.
- Explorar el análisis de componentes principales (PCA) y su aplicación en la reducción de la dimensionalidad.

## ■ Introducción a la SVD

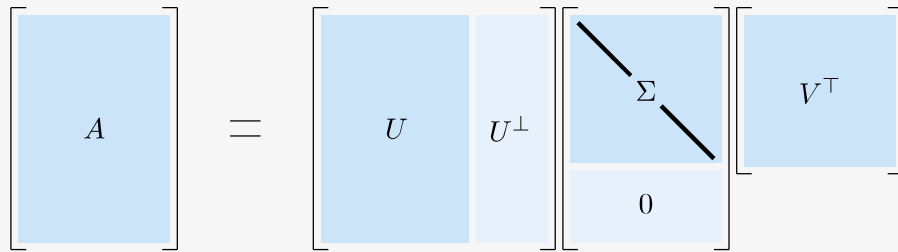
La **descomposición en valores singulares** (SVD) es una factorización matricial importante. Dada una matriz  $A$ , se puede expresar como:

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

donde:

- $U$  y  $V$  son matrices ortogonales.
- $\Sigma$  es una matriz diagonal con los valores singulares.

## ■ SVD completa

$$\begin{bmatrix} A \end{bmatrix} = \begin{bmatrix} U & U^\perp \end{bmatrix} \begin{bmatrix} \Sigma \\ 0 \end{bmatrix} \begin{bmatrix} V^\top \end{bmatrix}$$
The diagram illustrates the complete Singular Value Decomposition (SVD) of a matrix A. Matrix A is represented by a light blue square. It is equal to the product of three matrices. The first matrix is [U, U^perp], where U is a light blue square and U^perp is a lighter blue vertical rectangle. The second matrix is a block matrix consisting of a top-left light blue square labeled Σ with a diagonal line through it, and a bottom light blue rectangle labeled 0. The third matrix is V^top, a light blue square. The entire equation is set within a light gray rectangular frame with a blue border.

## ■ SVD reducida

$$\begin{bmatrix} A \end{bmatrix} = \begin{bmatrix} U \end{bmatrix} \begin{bmatrix} \Sigma \end{bmatrix} \begin{bmatrix} V^T \end{bmatrix}$$



## Teorema

Sea  $A \in \mathbb{R}^{m \times n}$ . Entonces, existen matrices ortogonales

$$U = \begin{bmatrix} u_1 & \cdots & u_m \end{bmatrix} \in \mathbb{R}^{m \times m} \quad \text{y} \quad V = \begin{bmatrix} v_1 & \cdots & v_n \end{bmatrix} \in \mathbb{R}^{n \times n}$$

tales que

$$U^T A V = \Sigma = \text{diag}(\sigma_1, \dots, \sigma_p) \in \mathbb{R}^{m \times n}, \quad p = \min\{m, n\}$$

y con  $\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_p \geq 0$ .

## Corolario

Los escalares  $\sigma_i$ ,  $1 \leq i \leq p$  son los **valores singulares**, los vectores  $u_i$ ,  $1 \leq i \leq m$  son los **vectores singulares izquierdos**, los vectores  $v_i$ ,  $1 \leq i \leq n$  son los **vectores singulares derechos**.

$$A = U\Sigma V^\top$$

- Si  $m \geq n$  se tiene que:  $Av_i = \sigma_i u_i$  y  $A^\top u_i = \sigma_i v_i$ , para  $1 \leq i \leq n$ .
- Si  $m \leq n$  se tiene que:  $Av_i = \sigma_i u_i$  y  $A^\top u_i = \sigma_i v_i$ , para  $1 \leq i \leq m$ .

## Corolario

Los escalares  $\sigma_i$ ,  $1 \leq i \leq p$  son los **valores singulares**, los vectores  $u_i$ ,  $1 \leq i \leq m$  son los **vectores singulares izquierdos**, los vectores  $v_i$ ,  $1 \leq i \leq n$  son los **vectores singulares derechos**.

$$A = \begin{bmatrix} | & \cdots & | \\ u_1 & \cdots & u_m \\ | & \cdots & | \end{bmatrix} \begin{bmatrix} \sigma_1 & 0 & 0 \\ 0 & \ddots & 0 \\ 0 & 0 & \sigma_n \\ 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} | & \cdots & | \\ v_1 & \cdots & v_n \\ | & \cdots & | \end{bmatrix}^{\top}, \quad m \geq n$$

- ▣ Si  $m \geq n$  se tiene que:  $Av_i = \sigma_i u_i$  y  $A^{\top} u_i = \sigma_i v_i$ , para  $1 \leq i \leq n$ .
- ▣ Si  $m \leq n$  se tiene que:  $Av_i = \sigma_i u_i$  y  $A^{\top} u_i = \sigma_i v_i$ , para  $1 \leq i \leq m$ .

## Corolario

Los escalares  $\sigma_i$ ,  $1 \leq i \leq p$  son los **valores singulares**, los vectores  $u_i$ ,  $1 \leq i \leq m$  son los **vectores singulares izquierdos**, los vectores  $v_i$ ,  $1 \leq i \leq n$  son los **vectores singulares derechos**.

$$A = \begin{bmatrix} | & \cdots & | \\ u_1 & \cdots & u_m \\ | & \cdots & | \end{bmatrix} \begin{bmatrix} \sigma_1 & 0 & 0 \\ 0 & \ddots & 0 \\ 0 & 0 & \sigma_n \\ 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} \text{---} & v_1 & \text{---} \\ \vdots & \vdots & \vdots \\ \text{---} & v_n & \text{---} \end{bmatrix}, \quad m \geq n$$

- ▣ Si  $m \geq n$  se tiene que:  $Av_i = \sigma_i u_i$  y  $A^\top u_i = \sigma_i v_i$ , para  $1 \leq i \leq n$ .
- ▣ Si  $m \leq n$  se tiene que:  $Av_i = \sigma_i u_i$  y  $A^\top u_i = \sigma_i v_i$ , para  $1 \leq i \leq m$ .

## Corolario

Los escalares  $\sigma_i$ ,  $1 \leq i \leq p$  son los **valores singulares**, los vectores  $u_i$ ,  $1 \leq i \leq m$  son los **vectores singulares izquierdos**, los vectores  $v_i$ ,  $1 \leq i \leq n$  son los **vectores singulares derechos**.

$$A = \begin{bmatrix} | & \cdots & | \\ u_1 & \cdots & u_m \\ | & \cdots & | \end{bmatrix} \begin{bmatrix} \sigma_1 & 0 & 0 & 0 \\ 0 & \ddots & 0 & 0 \\ 0 & 0 & \sigma_n & 0 \end{bmatrix} \begin{bmatrix} - & v_1 & - \\ \vdots & \vdots & \vdots \\ - & v_n & - \end{bmatrix}, \quad m \leq n$$

- Si  $m \geq n$  se tiene que:  $Av_i = \sigma_i u_i$  y  $A^\top u_i = \sigma_i v_i$ , para  $1 \leq i \leq n$ .
- Si  $m \leq n$  se tiene que:  $Av_i = \sigma_i u_i$  y  $A^\top u_i = \sigma_i v_i$ , para  $1 \leq i \leq m$ .

## ■ Descomposición en Valores Singulares

Dada la matriz  $A = \begin{bmatrix} 3 & 0 \\ 4 & 5 \end{bmatrix}$ , queremos calcular su descomposición en valores singulares

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

## ■ Estrategia del cálculo

- Calcular  $AA^T$  y  $A^T A$ .
- Calcular valores y vectores propios de  $AA^T \rightarrow \lambda_i, u_i$ .
- Calcular valores y vectores propios de  $A^T A \rightarrow \lambda_i, v_i$ .
- Valores singulares  $\sigma_i = \sqrt{\lambda_i}$ .

## ■ Cálculo de $A^T A$ y $AA^T$

Primero, calculemos  $A^T A$  y  $AA^T$ :

$$AA^T = \begin{bmatrix} 3 & 0 \\ 4 & 5 \end{bmatrix} \begin{bmatrix} 3 & 0 \\ 4 & 5 \end{bmatrix}^T = \begin{bmatrix} 3 & 0 \\ 4 & 5 \end{bmatrix} \begin{bmatrix} 3 & 4 \\ 0 & 5 \end{bmatrix} = \begin{bmatrix} 9 & 12 \\ 12 & 41 \end{bmatrix}$$

$$A^T A = \begin{bmatrix} 3 & 0 \\ 4 & 5 \end{bmatrix}^T \begin{bmatrix} 3 & 0 \\ 4 & 5 \end{bmatrix} = \begin{bmatrix} 3 & 4 \\ 0 & 5 \end{bmatrix} \begin{bmatrix} 3 & 0 \\ 4 & 5 \end{bmatrix} = \begin{bmatrix} 25 & 12 \\ 12 & 25 \end{bmatrix}$$



## ■ Cálculo de valores y vectores propios de $AA^\top$

$$\det(AA^\top - \lambda I) = 0 \implies (9 - \lambda)(41 - \lambda) - 12^2 = 0 \implies \lambda_1 = 45, \lambda_2 = 5$$

$$(AA^\top - \lambda_1 I)u_1 \implies \begin{bmatrix} 9 - 45 & 12 \\ 12 & 41 - 45 \end{bmatrix} u_1 = 0 \implies u_1 = \frac{1}{\sqrt{10}} \begin{bmatrix} 1 \\ 3 \end{bmatrix}$$

$$(AA^\top - \lambda_2 I)u_2 \implies \begin{bmatrix} 9 - 5 & 12 \\ 12 & 41 - 5 \end{bmatrix} u_2 = 0 \implies u_2 = \frac{1}{\sqrt{10}} \begin{bmatrix} -3 \\ 1 \end{bmatrix}$$

$$\text{Así } U = [u_1, u_2].$$

## ■ Cálculo de valores y vectores propios de $AA^\top$

$$\det(A^\top A - \lambda I) = 0 \implies (25 - \lambda)(25 - \lambda) - 12^2 = 0 \implies \lambda_1 = 45, \lambda_2 = 5$$

$$(A^\top A - \lambda_1 I)v_1 \implies \begin{bmatrix} 25 - 45 & 12 \\ 12 & 25 - 45 \end{bmatrix} v_1 = 0 \implies v_1 = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 \\ 1 \end{bmatrix}$$

$$(A^\top A - \lambda_2 I)v_2 \implies \begin{bmatrix} 25 - 5 & 12 \\ 12 & 25 - 5 \end{bmatrix} v_2 = 0 \implies v_2 = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 \\ -1 \end{bmatrix}$$

$$\text{Así } V = \begin{bmatrix} v_1 & v_2 \end{bmatrix}.$$

## Construcción de la SVD

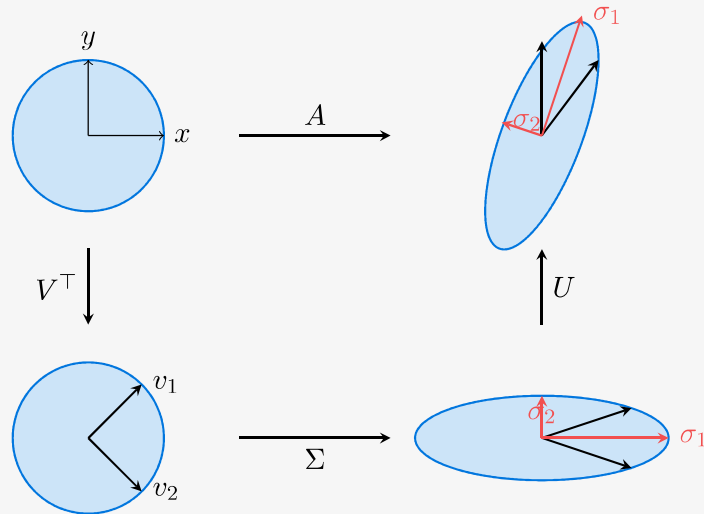
Finalmente, obtenemos los valores singulares

$$\sigma_1 = \sqrt{\lambda_1} = \sqrt{45}, \quad \sigma_2 = \sqrt{\lambda_2} = \sqrt{5},$$

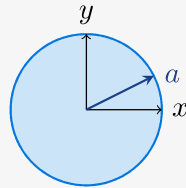
y construimos la descomposición en valores singulares  $A = U\Sigma V^T$ :

$$\underbrace{\begin{bmatrix} 3 & 0 \\ 4 & 5 \end{bmatrix}}_A = \underbrace{\frac{1}{\sqrt{10}} \begin{bmatrix} 1 & -3 \\ 3 & 1 \end{bmatrix}}_U \underbrace{\begin{bmatrix} \sqrt{45} & 0 \\ 0 & \sqrt{5} \end{bmatrix}}_\Sigma \underbrace{\frac{1}{\sqrt{2}} \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix}^T}_{V^T}$$

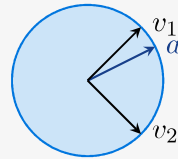
## ■ Interpretación geométrica



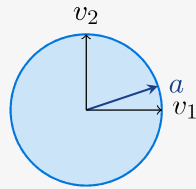
## ■ Interpretación geométrica



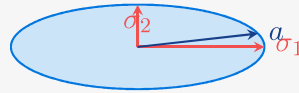
## ■ Interpretación geométrica



## ■ Interpretación geométrica

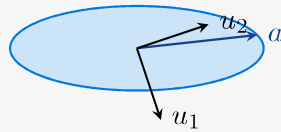


## ■ Interpretación geométrica

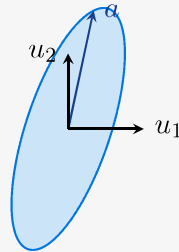




## ■ Interpretación geométrica



## ■ Interpretación geométrica



# Aproximación matricial

## Observación: aproximación de bajo rango óptima

Descomposición de  $A$  en matrices de rango 1

$$A = \begin{bmatrix} | & | & & | & & | \\ u_1 & u_2 & \cdots & u_n & \cdots & u_m \\ | & | & & | & & | \end{bmatrix} \begin{bmatrix} \sigma_1 & 0 & \cdots & 0 \\ 0 & \sigma_2 & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & \sigma_n \\ 0 & \cdots & 0 & 0 \end{bmatrix} \begin{bmatrix} \text{---} & v_1 & \text{---} \\ \text{---} & v_2 & \text{---} \\ \vdots & \vdots & \vdots \\ \text{---} & v_n & \text{---} \end{bmatrix}$$

## Observación: aproximación de bajo rango óptima

Descomposición de  $A$  en matrices de rango 1

$$A = \begin{bmatrix} | & | & & | & \cdots & | \\ \mathbf{u}_1 & u_2 & \cdots & u_n & \cdots & u_m \\ | & | & & | & & | \end{bmatrix} \begin{bmatrix} \sigma_1 & 0 & \cdots & 0 \\ 0 & \sigma_2 & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & \sigma_n \\ 0 & \cdots & 0 & 0 \end{bmatrix} \begin{bmatrix} \text{---} & \mathbf{v}_1 & \text{---} \\ \text{---} & v_2 & \text{---} \\ \vdots & \vdots & \vdots \\ \text{---} & v_n & \text{---} \end{bmatrix}$$
$$= \sigma_1 \mathbf{u}_1 \mathbf{v}_1^\top$$

## Observación: aproximación de bajo rango óptima

Descomposición de  $A$  en matrices de rango 1

$$A = \begin{bmatrix} | & | & & | & & | \\ \textcolor{blue}{u_1} & \textcolor{red}{u_2} & \cdots & u_n & \cdots & u_m \\ | & | & & | & & | \end{bmatrix} \begin{bmatrix} \textcolor{blue}{\sigma_1} & 0 & \cdots & 0 \\ 0 & \textcolor{red}{\sigma_2} & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & \sigma_n \\ 0 & \cdots & 0 & 0 \end{bmatrix} \begin{bmatrix} \text{---} & \textcolor{blue}{v_1} & \text{---} \\ \text{---} & \textcolor{red}{v_2} & \text{---} \\ \vdots & \vdots & \vdots \\ \text{---} & v_n & \text{---} \end{bmatrix}$$
$$= \textcolor{blue}{\sigma_1 u_1 v_1^\top} + \textcolor{red}{\sigma_2 u_2 v_2^\top}$$

## Observación: aproximación de bajo rango óptima

Descomposición de  $A$  en matrices de rango 1

$$A = \begin{bmatrix} | & | & & | & & | \\ \textcolor{blue}{u}_1 & \textcolor{red}{u}_2 & \cdots & \textcolor{green}{u}_n & \cdots & u_m \\ | & | & & | & & | \end{bmatrix} \begin{bmatrix} \textcolor{blue}{\sigma}_1 & 0 & \cdots & 0 \\ 0 & \textcolor{red}{\sigma}_2 & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & \textcolor{green}{\sigma}_n \\ 0 & \cdots & 0 & 0 \end{bmatrix} \begin{bmatrix} \text{---} & \textcolor{blue}{v}_1 & \text{---} \\ \text{---} & \textcolor{red}{v}_2 & \text{---} \\ \vdots & \vdots & \vdots \\ \text{---} & \textcolor{green}{v}_n & \text{---} \end{bmatrix}$$
$$= \textcolor{blue}{\sigma}_1 \textcolor{blue}{u}_1 \textcolor{blue}{v}_1^\top + \textcolor{red}{\sigma}_2 \textcolor{red}{u}_2 \textcolor{red}{v}_2^\top + \cdots + \textcolor{green}{\sigma}_n \textcolor{green}{u}_n \textcolor{green}{v}_n^\top$$

## Teorema de Eckart-Young

Considere la matriz  $A \in \mathbb{R}^{m \times n}$  y su descomposición en valores singulares  $A = U\Sigma V^\top$ .  
Sea la matriz

$$A_k = \sigma_1 u_1 v_1^\top + \cdots + \sigma_k u_k v_k^\top$$

para  $k \leq \min\{m, n\}$ . Entonces,  $(\dim(\text{im}(A_k)) =) \text{rank}(A_k) = k$ . Además si  $B \in \mathbb{R}^{m \times n}$  es una matriz con  $\text{rank}(B) = k$ , entonces

$$\|A - A_k\|_F \leq \|A - B\|_F$$

Además,

$$\|A - A_k\|_F^2 = \sum_{j=k+1}^p \sigma_j^2$$



## Teorema de Eckart-Young

Considere la matriz  $A \in \mathbb{R}^{m \times n}$  y su descomposición en valores singulares  $A = U\Sigma V$ .  
Sea la matriz

$$A_k = \sigma_1 u_1 v_1^\top + \cdots + \sigma_k u_k v_k^\top$$

para  $k \leq \min\{m, n\}$ . Entonces,  $(\dim(\text{im}(A_k)) =) \text{rank}(A_k) = k$ . Además si  $B \in \mathbb{R}^{m \times n}$  es una matriz con  $\text{rank}(B) = k$ , entonces

$$\|A - A_k\|_2 \leq \|A - B\|_2$$

Además,

$$\|A - A_k\|_2^2 = \sigma_{k+1}^2$$

## Ejemplo de compresión de imágenes



Figure: Imagen original

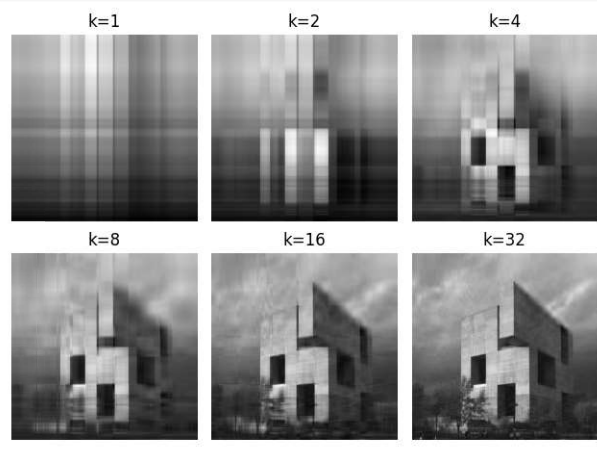


Figure: Sucesión de imágenes creadas con la SVD, matrices truncadas

## Ejemplo de compresión de imágenes

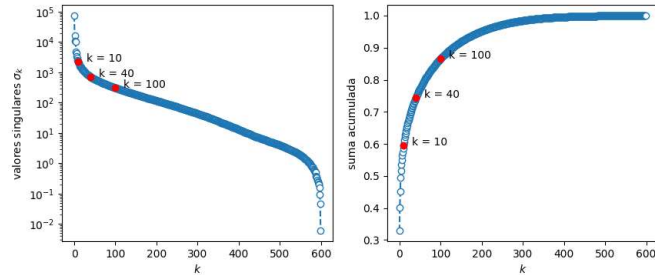


Figure: Valores singulares  $\sigma_k$  y suma acumulada  $\sum_{j=1}^k \sigma_j / \sum_{j=1}^p \sigma_j$  de los primeros  $k$  valores singulares.

## ■ Interpretaciones

La SVD de una matriz  $A$  esta estrechamente relacionada a un problema de valores propios que involucra a las matrices de correlación  $AA^T$  y  $A^TA$ .

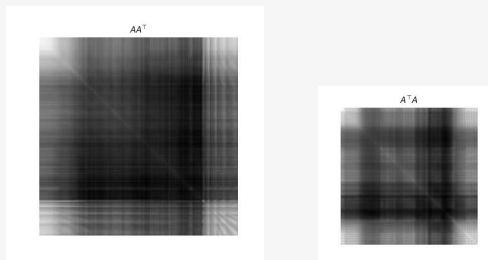


Figure: Matrices de correlación asociadas a la figura

## ■ Interpretaciones

La SVD de una matriz  $A$  esta estrechamente relacionada a un problema de valores propios que involucra a las matrices de correlación  $AA^\top$  y  $A^\top A$ .

$$A = U\Sigma V^\top \Rightarrow \begin{cases} AA^\top = U \begin{bmatrix} \Sigma \\ 0 \end{bmatrix} V^\top V \begin{bmatrix} \Sigma & 0 \end{bmatrix} U^\top = U \begin{bmatrix} \Sigma^2 & 0 \\ 0 & 0 \end{bmatrix} U^\top \\ A^\top A = V \begin{bmatrix} \Sigma & 0 \end{bmatrix} U^\top U \begin{bmatrix} \Sigma \\ 0 \end{bmatrix} V^\top = V \Sigma^2 V^\top \end{cases}$$

## Interpretación

Recordemos que  $U$  y  $V$  son matrices ortogonales, entonces  $U$ ,  $\Sigma$  y  $V$  son soluciones de los problemas de valores propios:

$$AA^T = U \begin{bmatrix} \Sigma^2 & 0 \\ 0 & 0 \end{bmatrix}$$
$$A^T A V = V \Sigma^2$$

**Cada valor singular es la raíz cuadrada de un valor propio de  $AA^T$  y  $AA^T$ .**

**Observación:** Si  $A = A^T$  entonces los valores singulares de  $A$  son iguales al valor absoluto de los valores propios de  $A$ .

## Método de Snapshots

Calcular la matriz  $AA^T$  puede resultar **impráctico** para muchos datos (número de filas  $m \gg n$ ). Para estos casos, es posible calcular las primeras  $k$  columnas de  $U$  usando el **método de Snapshots**.

En lugar de calcular la descomposición de valores propios de  $AA^T$  que nos da los vectores singulares izquierdos  $U$ , calculamos la descomposición de  $A^T A$  y obtenemos los vectores singulares izquierdo  $V$  y los valores singulares  $\Sigma$ .

Luego, seleccionamos los  $k$  valores singulares no cero en  $\tilde{\Sigma}$  y los correspondientes vectores  $\tilde{V}$ . Así aproximamos las primeras  $k$  columnas de  $U$  por:

$$\tilde{U} = A\tilde{V}\tilde{\Sigma}^{-1}$$

## ■ Aplicaciones de la SVD

- Reducción de la dimensionalidad
- Eliminación de ruido
- Obtener subespacios, espacio columna, espacio fila, espacio nulo derecho y espacio nulo izquierdo
- Obtener la Pseudoinversa Moore-Penrose
- Resolver ecuaciones lineales y regresión lineal
- Preprocesamiento de atributos y blanqueamiento en aprendizaje de máquina
- Detección de outliers
- Ingeniería de atributos



## Moore-Penrose pseudo inversa

Sistemas de ecuaciones lineales:  $Ax = b$ ,  $A \in \mathbb{R}^{m \times n}$ ,  $b \in \mathbb{R}^m$

■  $m \ll n$  sistema **subdeterminado**

$$\begin{bmatrix} A \end{bmatrix} \begin{bmatrix} x \end{bmatrix} = \begin{bmatrix} b \end{bmatrix}$$

■  $m \gg n$  sistema **sobredeterminado**

$$\begin{bmatrix} A \end{bmatrix} \begin{bmatrix} x \end{bmatrix} = \begin{bmatrix} b \end{bmatrix}$$

## ■ Moore-Penrose pseudo inversa

Sistemas de ecuaciones lineales:  $Ax = b$ ,  $A \in \mathbb{R}^{m \times n}$ ,  $b \in \mathbb{R}^m$

■  $m \ll n$  sistema **subdeterminado**

Solución de norma mínima: minimizar  $\|x\|_2$ , tal que  $Ax = b$

■  $m \gg n$  sistema **sobredeterminado**

Solución de mínimos cuadrados: minimizar  $\|Ax - b\|_2^2$

## ■ Moore-Penrose pseudo inversa

Sea la SVD truncada exacta  $A = U\Sigma V^\top$ , (valores singulares no cero), entonces

$$A^\dagger = V\Sigma^{-1}U^\top \Rightarrow A^\dagger A = VV^\top \quad (\neq I_{n \times n})$$

es la **Moore-Penrose pseudo inversa por la izquierda**

$$A^\dagger Ax = A^\dagger b \Rightarrow x = V\Sigma^{-1}U^\top b \Rightarrow Ax = U\Sigma V^\top V\Sigma^{-1}U^\top b = UU^\top b$$

Así  $x$  es la solución exacta sólo cuando  $b$  está en el espacio columna de  $U$

## Regresión lineal unidimensional

Buscamos la pendiente  $x$ :  $Ax = b$  donde  $A \in \mathbb{R}^{n \times 1}$ .

La solución por la pseudo inversa es:  $x = V\Sigma^{-1}U^T b$

Aquí,  $\tilde{\Sigma} = \|a\|_2$ ,  $\tilde{V} = 1$ ,  $\tilde{U} = a_1/\|a_1\|_2$

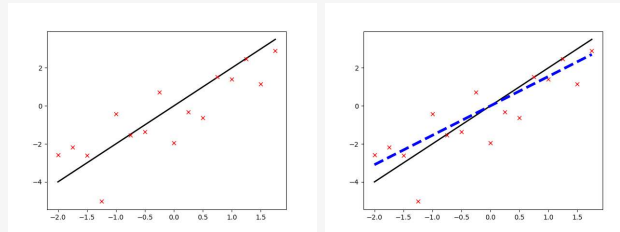


Figure: Línea original y datos con ruidos (izquierda). Línea de regresión (azul derecha)

## Regresión lineal multivariada

describe la generación de calor de 13 mezclas de cemento que están compuestas de 4 ingredientes básicos. El objetivo es determinar el peso  $x$  que relaciona las proporciones de los 4 ingredientes con la generación de calor.

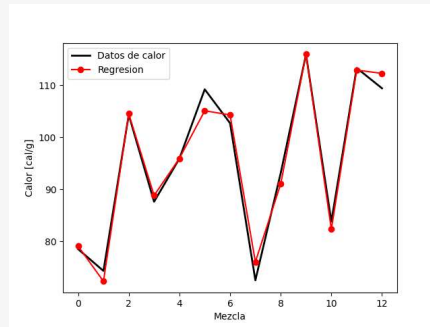


Figure: Datos de generación de calor para mezclas de cemento

## Boston Housing data

Conjunto de datos: 13 atributos correlacionados con el precio de 506 casas. El mejor ajuste se grafica versus el valor verdadero de las casas, primero desordenados y luego ordenados

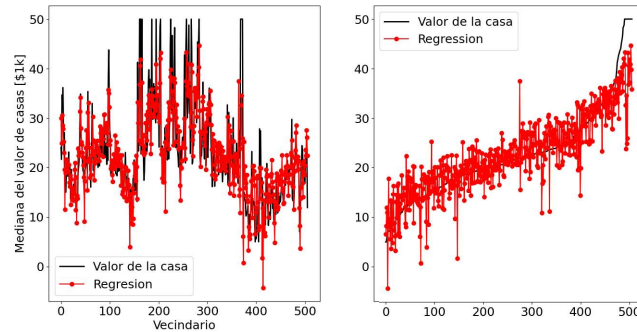


Figure: Regresion de precios de casa, datos desordenados (izquierda) y ordenados (derecha)

## ■ Boston Housing Data

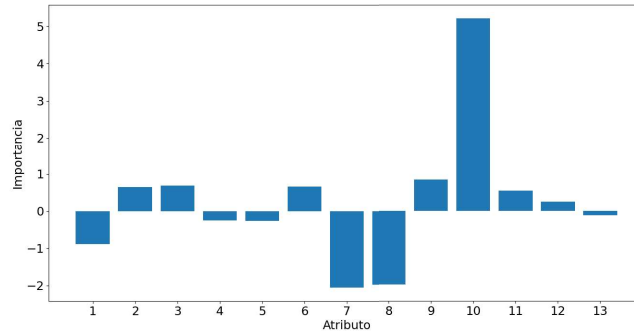


Figure: Importancia de los atributos en la regresion

# Análisis de Componentes Principales (PCA)



## ■ Motivación - PCA

Análisis de componentes principales es una de las aplicaciones centrales de la SVD. Provee una interpretación estadística del sistema de coordenadas **jerárquico** impulsado por los datos utilizado para representar datos correlacionados de alta dimensionalidad. Este sistema de coordenadas implica las matrices de correlación. Es importante destacar que el PCA **preprocesa los datos** mediante la sustracción de la media y estableciendo la varianza en unidad antes de realizar la SVD.

- Encuentra los ejes principales (componentes) en los cuales los datos tienen la mayor varianza.
- Proyecta los datos en estos nuevos ejes para obtener una representación de menor dimensión.

## Problema

Frecuentemente en estadística un número de datos son recopiladas en un experimento. Estas medidas son organizadas en un vector fila. Cada uno de estos es organizado como fila en un matriz  $X \in \mathbb{R}^{N \times m}$ , donde  $m$  es el número de atributos y  $N$  es el número de experimentos. Buscamos reducir la dimensionalidad de estos conservando sus propiedades principales. Esto resulta ser relevante para muchas aplicaciones. Por ejemplo, donde queremos

- Pre-procesar los datos.
- Visualizar datos en menor dimensión.
- Extraer los features mas relevantes.

En resumen, queremos reducir la dimensión de los datos  $X$  a  $\tilde{X} \in \mathbb{R}^{N \times n}$ , donde  $n < m$  y  $\tilde{X}$  sea cercano a  $X$ .

## Cómputo de PCA

- 1 Calculamos la fila promedio  $\bar{x}$  (el promedio de todas las filas) y lo restamos de  $X$

$$\bar{x}_j = \frac{1}{n} \sum_{i=1}^n X_{ij}, \quad B = X - \begin{bmatrix} 1 \\ \vdots \\ 1 \end{bmatrix} \bar{x} \quad (\text{datos centrados})$$

- 2 Calcular la matriz de covarianza de  $B$

$$C = \frac{1}{n-1} B^\top B, \quad (\text{corrección de Bessel } \frac{1}{n-1} \text{ en lugar de } \frac{1}{n})$$

- 3 Calculas las componentes principales de  $C$ , los vectores propios,

$$CV = VD \Rightarrow C = VDV^\top \Rightarrow D = V^\top CV$$

$D_{ii}$  son las varianzas de los datos respecto a la dirección de la columna  $i$  de  $V$ .

## ■ Cómputo de PCA

**Observación:** La matriz de componentes principales  $V$  es también la matriz de vectores singulares de  $B$ .

$$B = U\Sigma V^{\top} \Rightarrow C = \frac{1}{n-1} B^{\top} B = \frac{1}{n-1} V \Sigma^2 V^{\top} \Rightarrow D = \frac{1}{n-1} \Sigma^2$$

Por lo tanto la varianza de los datos en estas coordenadas, los elementos de la diagonal de  $D$ , están relacionados con los valores singulares por

$$D_{ii} = \frac{\sigma_i^2}{n-1}$$

## Ejemplo 1 PCA

### Generación de datos:

$[x, y]^T = X \sim N_2(0, I)$ . Luego escalamos en las direcciones  $x$  e  $y$  por 2 y  $1/2$  y rotamos en un ángulo de  $\pi/3$  radianes de una distribución normal bidimensional con media cero. Finalmente trasladamos al centro  $x_C = [2, 1]^T$

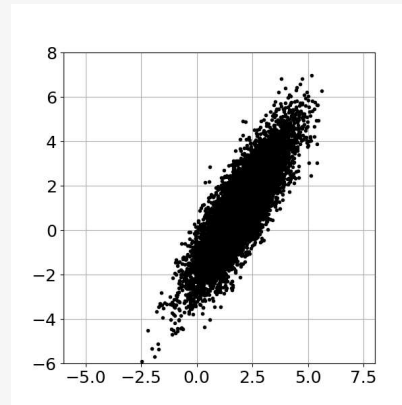


Figure: Datos generados

## Visualización de PCA

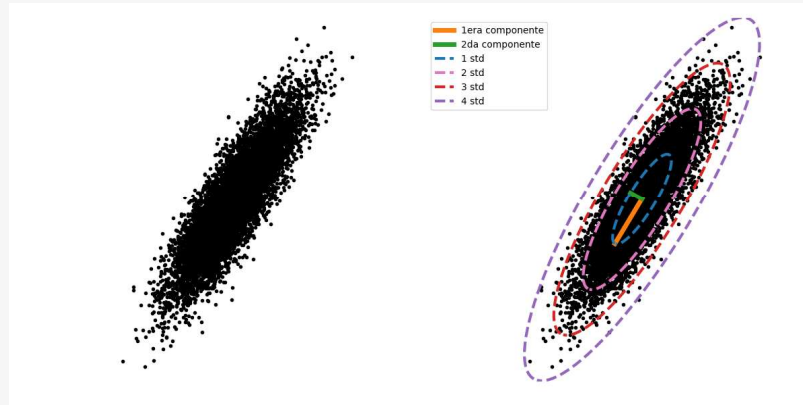


Figure: Distribucion de datos e interpretacion de PCA. Valores singulares son:  $\sigma_1 = 1.9820579$  y  $\sigma_2 = 0.49922381$

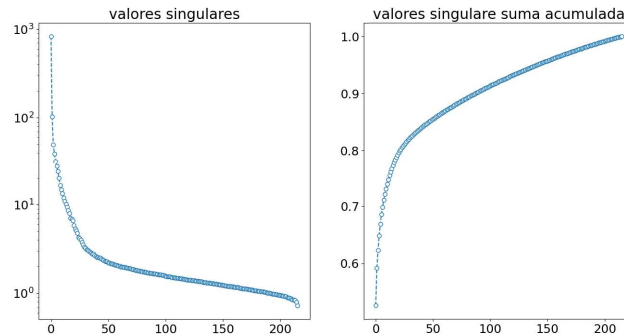
## ■ Ejemplo 1 PCA

Además observamos que la matriz  $U$  de la SVD aproxima la matriz de rotación por el ángulo de  $\pi/3$ :

$$U = \begin{bmatrix} -0.5011 & -0.8654 \\ -0.8654 & 0.5011 \end{bmatrix}, \quad Q_{\pi/3} = \begin{bmatrix} \cos(\pi/3) & -\sin(\pi/3) \\ \sin(\pi/3) & \cos(\pi/3) \end{bmatrix} = \begin{bmatrix} 0.5 & 0.8660 \\ -0.8660 & 0.5 \end{bmatrix}$$

## Ejemplo 2 PCA

**Conjunto de datos:** 216 pacientes, 121 que tienen cáncer de ovarios y 95 que no tienen. Para cada paciente se tiene un vector de datos que contiene la expresión de 4000 genes (alta dimensión de los atributos).





## Ejemplo 2 PCA

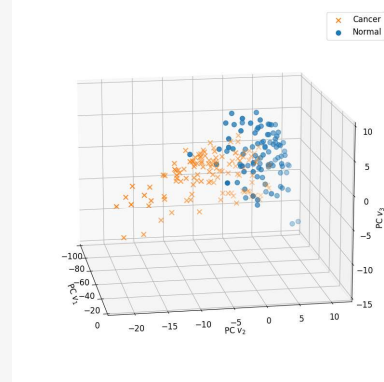
Una de las razones para usar PCA es la capacidad de visualizar patrones y correlaciones

Graficamos en el espacio generado por las tres primeras componentes principales, proyectando las coordenadas de cada dato

$$x = v_1^\top X^\top$$

$$y = v_2^\top X^\top$$

$$z = v_3^\top X^\top$$



### Ejemplo 3 PCA

**Eigenfaces.** 38 individuos y 64 fotografías ( $192 \times 168$  píxeles) de cada uno



## Ejemplo 3 PCA

**Eigenfaces.** 38 individuos y 64 fotografías ( $192 \times 168$  pixeles) de cada uno  
Cada imagen de la base de datos es transformada en un vector columna de largo 32356  
( $=192 \times 168$ ).

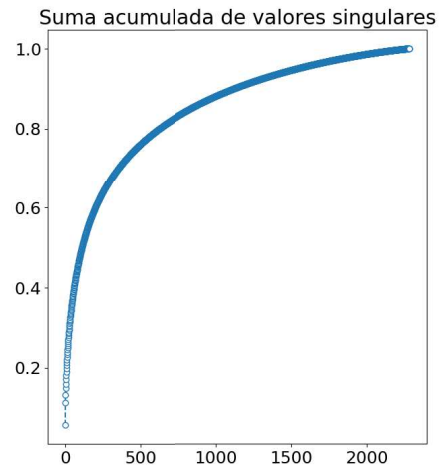
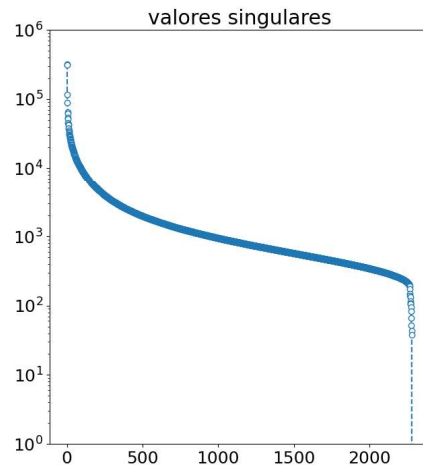
Vector media,  
rostro medio



Rostros  
menos  
media

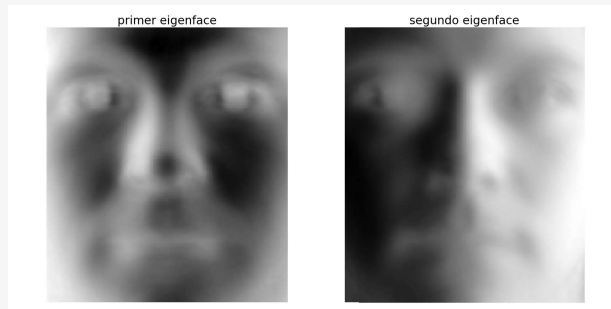
la

### ■ Ejemplo 3 PCA



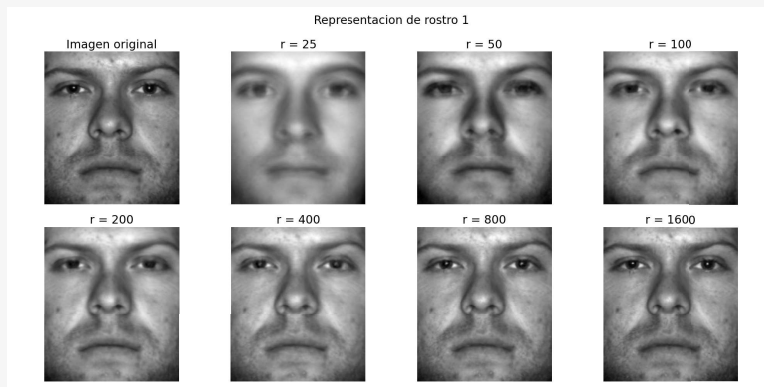
## ■ Ejemplo 3 PCA

Los vectores columnas de la matriz  $U$  son los eigenfaces o rostros propios



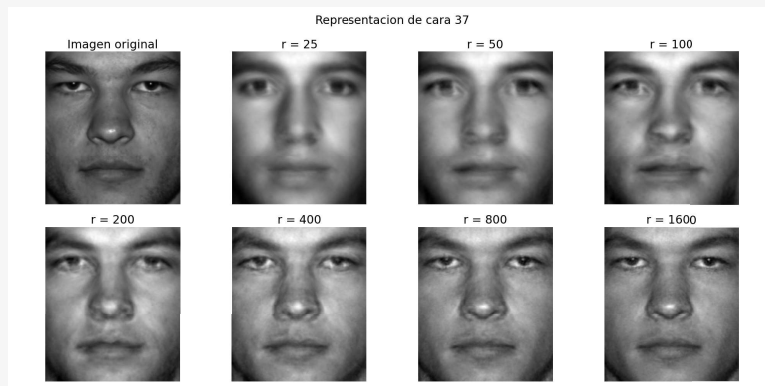
### Ejemplo 3 PCA

Usando las eigenfaces podemos tratar de aproximar cada imagen de entrenamiento con base la SVD de rango  $r$ .



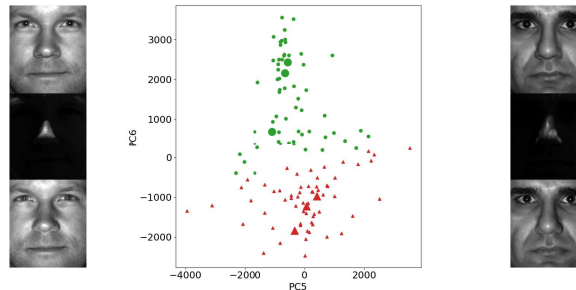
### Ejemplo 3 PCA

Usando las eigenfaces podemos tratar de aproximar una imagen de test con base la SVD de rango  $r$ .  $\hat{x} = \tilde{U}\tilde{U}^\top x$



## Ejemplo 3 PCA

**Sistema de coordenadas** Proyectamos una imagen  $x$  sobre los  $r$  primeros PCA modos, así obtenemos un conjunto de coordenadas en este espacio  $\tilde{x} = \tilde{U}^\top x$ . Algunas componentes principales pueden capturar los atributos más comunes compartidos por todos los rostros humanos, mientras que otros serán más útiles para distinguir entre individuos. Otras también pueden capturar diferencias en el ángulo de luz.





## ■ Conclusiones

- Hemos introducido la SVD como una herramienta poderosa en álgebra lineal y análisis de datos.
- PCA es útil para reducir la dimensionalidad y preservar la información relevante en los datos.
- Profundizaremos en la aplicación computacional de estos en python en el tutorial.