UC | Chile

Algebra Lineal Aplicada para Ciencia de Datos

Clase 8. Descomposición en valores singulares

1 Introducción

2 Tema 1: Descomposición en valores singulares

3 Tema 2: Teorema de Eckart-Young

4 Tema 3: Análisis de Componentes principales

Introducción

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.

UC | Chile

Tema 1: Descomposición en valores singulares

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.
- Σ es una matriz diagonal con los valores singulares.

Teorema

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

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

tales que

$$U^\top AV = \Sigma = \mathrm{diag}(\sigma_1, \dots, \sigma_p) \in \mathbb{R}^{m\times n}, \quad p = \min\{m, n\}$$

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

Corolario

Los escalares σ_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 σ_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 σ_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**.

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

- 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 σ_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

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

Estrategia del cálculo

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

Cálculo de $A^T A$ y AA^T

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

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

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

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

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

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

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

Así $U = \begin{bmatrix} u_1, u_2 \end{bmatrix}$.

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

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

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

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

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

Construcción de la Descomposición en Valores Singulares

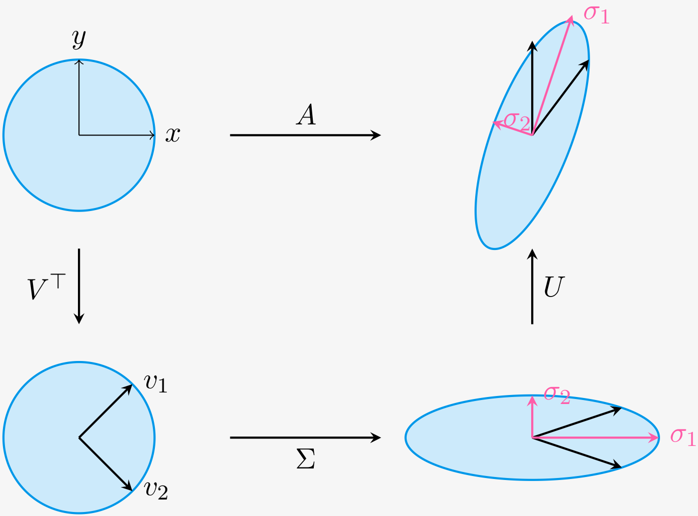
Finalmente, obtenemos los valores singulares

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

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

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

Interpretación geométrica



Tema 2: Teorema de Eckart-Young

Observación

Descomposición de A en matrices de rango 1.

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

$$=$$

Observación

Descomposición de A en matrices de rango 1.

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

$$= \quad \boldsymbol{\sigma_1 u_1 v_1^\top}$$

Observación

Descomposición de A en matrices de rango 1.

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

$$= \quad \boldsymbol{\sigma_1 u_1 v_1^\top} \; + \; \boldsymbol{\sigma_2 u_2 v_2^\top}$$

Observación

Descomposición de A en matrices de rango 1.

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

$$= \boldsymbol{\sigma_1 u_1 v_1^\top} + \boldsymbol{\sigma_2 u_2 v_2^\top} + \cdots + \boldsymbol{\sigma_n u_n v_n^\top}$$

Teorema de Eckart-Young

Considere la matriz $A \in \mathbb{R}^{m \times n}$ y su descomposición en valores singulares $A = U\Sigma V$. 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(**im**(A_k)) $=$)rank$(A_k) = k$. Además si $B \in \mathbb{R}^{m \times n}$ es una matriz con rank$(B) = k$, entonces

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

Ejemplo de compresión de imágenes

Figure: Imagen original

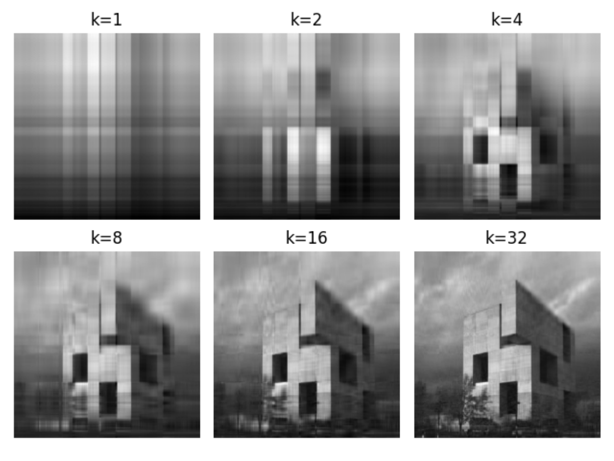


Figure: Sucesion de imagenes creadas con la SVD

Tema 3: Análisis de Componentes principales

Problema

Dado un conjunto de datos $X \in \mathbb{R}^{n \times N}$, donde m es el número de atributos, 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 $\widetilde{X} \in \mathbb{R}^{n \times N}$, donde $n < m$ y $\widetilde{X}$ sea cercano a X.

Introducción a PCA

El Análisis de Componentes Principales (PCA) es una técnica estadística que utiliza la SVD para reducir la dimensionalidad de los datos.

- 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.

Pasos Básicos de PCA

1. Centrar los datos.
2. Calcular la matriz de covarianza.
3. Calcular valores y vectores propios de la matriz de covarianza.
4. Seleccionar los componentes principales.
5. Proyectar los datos en el nuevo espacio.

Pasos Básicos de PCA

1. Centrar los datos.
2. Calcular valores y vectores singulares de los datos centrados.
3. Seleccionar los componentes principales.
4. Proyectar los datos en el nuevo espacio.

Visualización de PCA

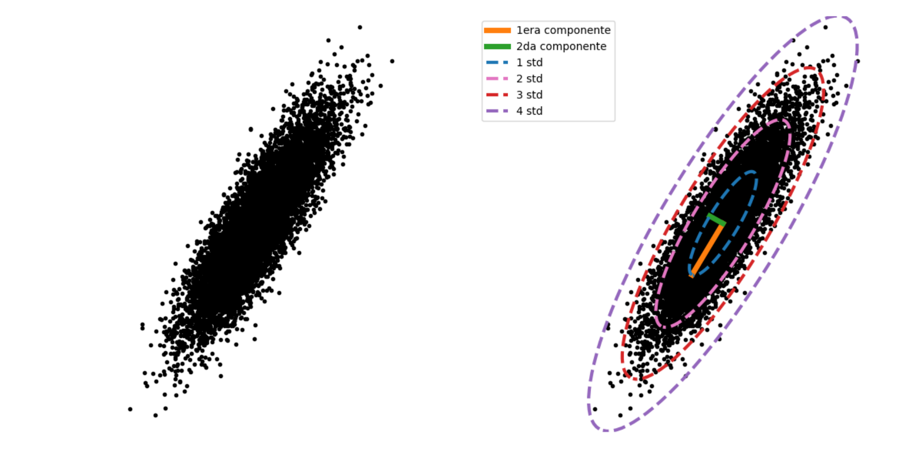


Figure: Distribucion de datos e interpretacion de PCA.

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 aplicacion computacional de estos en python en el tutorial.

PONTIFICIA UNIVERSIDAD CATÓLICA DE CHILE

UC | Chile