PONTIFICIA
UNIVERSIDAD
CATÓLICA
DE CHILE

PONTIFICIA
UNIVERSIDAD
CATÓLICA
DE CHILE

Fundamentos Matemáticos de Inteligencia Artificial

Clase 3

Manuel A. Sánchez

2024.04.01

Algebra Lineal

Introducción

Introducción y motivación

Factorización $A = QR$ es una de los conceptos mas importantes en Algebra Lineal. En esta clase responderemos primero la pregunta fundamental de como decidir si un conjunto de vectores en linealmente independiente. Luego avanzaremos a mostrar la factorización y sus aplicaciones para resolver sistemas lineales, proyecciones ortogonales, pseudoinversa, y mas adelante problemas de mínimos cuadrados.

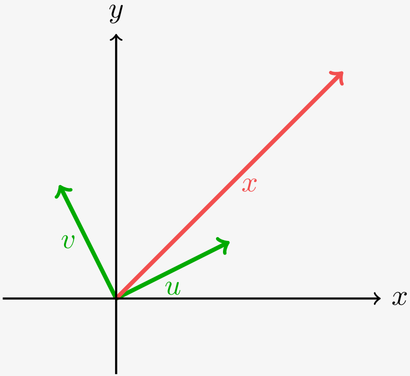
Ortogonalización/Ortonormalización

Bases ortonormales

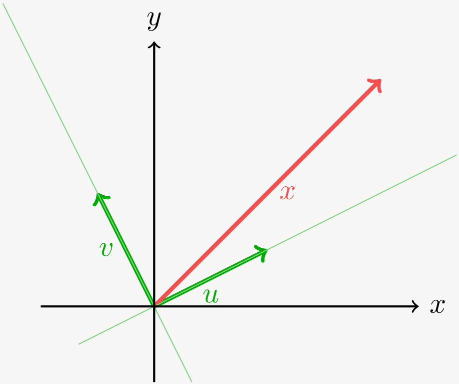
¿Cómo calculamos las coordenadas de un vector x en una base $a_1, \ldots, a_n$?

Existe un tipo de base para la cual es simple determinar las coordenadas de un vector cualquiera

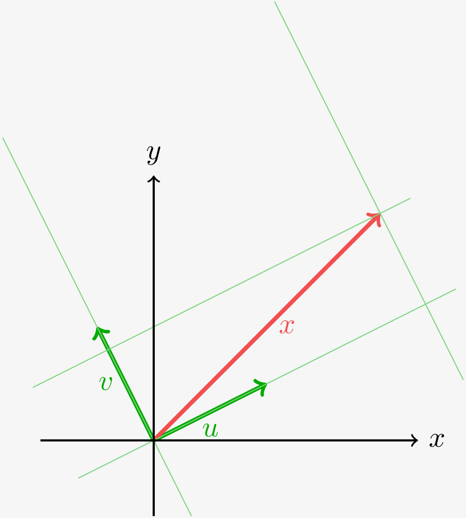
Bases ortonormales



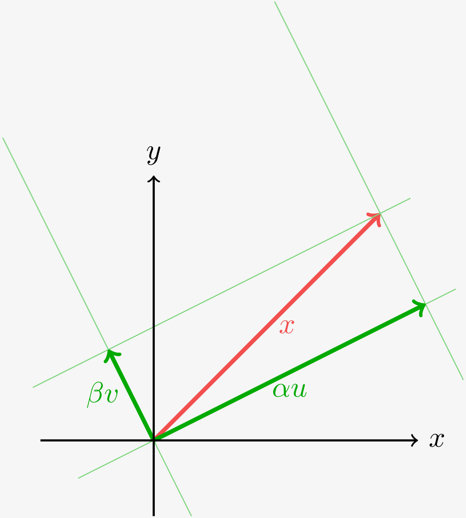
Bases ortonormales



Bases ortonormales

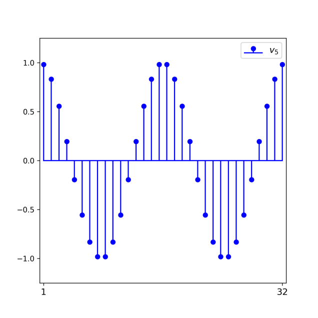


Bases ortonormales

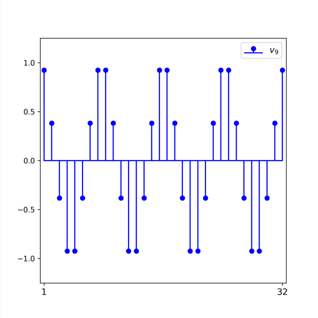


Bases ortogonales

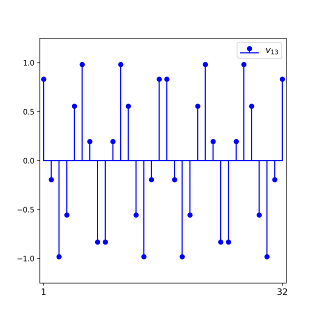
La base del coseno discreto es **ortogonal** para cualquier d



v_5



v_9



v_{13}

Matrices ortogonales

Definición: Una colección de vectores $a_1, ..., a_k$ es ortogonal o mutuamente ortogonal si

$$a_i \cdot a_j = a_i^\top a_j = 0, \quad \text{para } 1 \leq i, j \leq k,\ i \neq j.$$

Si además tenemos que $\|a_i\|_2^2 = a_i^\top a_i = 1$, entonces los vectores se dicen ortonormales, es decir

$$a_i \cdot a_j = a_i^\top a_j = \begin{cases} 1 & \text{si } i = j, \\ 0 & \text{si } i \neq j. \end{cases}$$

Ejemplos: de vectores ortogonales

- Los vectores canónicos unitarios son ortonormales

$$e_1 = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}, \quad e_2 = \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}, \quad e_3 = \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}$$

- Los siguientes vectores son ortonormales

$$a_1 = \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}, \quad a_2 = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 \\ -1 \\ 0 \end{bmatrix}, \quad a_3 = \frac{1}{\sqrt{2}} \begin{bmatrix} -1 \\ -1 \\ 0 \end{bmatrix}$$

Observación

- Si tenemos un conjunto de vectores **ortonormales** entonces tenemos que estos vectores son inmediatamente **linealmente independientes**.
  En efecto, si $a_1, ..., a_k$ son ortonormales y existen constantes $\beta_1, ..., \beta_k$ tales que

$$\beta_1 a_1 + ... + \beta_k a_k = 0 \quad \Longrightarrow \quad 0 = a_i^\top(\beta_1 a_1 + ... + \beta_k a_k) = \beta_i \|a_i\|^2 \quad \Longrightarrow a_i = 0.$$

- Si $a_1, ..., a_k$ son vectores ortonormales y $x = \beta_1 a_1 + ... + \beta_k a_k$, entonces las constantes están dadas por $\beta_i = a_i^\top x$, $1 \leq i \leq k$.
- Un conjunto de $n-$ vectores ortonormales $a_1, ...a_n$ forman una base y se dicen base ortonormal.

Ejemplo base ortonormal

Sea el vector $x \in \mathbb{R}^3$, dado por $x = \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix}$, y la base ortonormal

$$a_1 = \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}, \quad a_2 = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 \\ -1 \\ 0 \end{bmatrix}, \quad a_3 = \frac{1}{\sqrt{2}} \begin{bmatrix} -1 \\ -1 \\ 0 \end{bmatrix}$$

Entonces, describimos x en términos de la base ortonormal por

$$x = (x^\top a_1) \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} + (x^\top a_2) \frac{1}{\sqrt{2}} \begin{bmatrix} 1 \\ -1 \\ 0 \end{bmatrix} + (x^\top a_3) \frac{1}{\sqrt{2}} \begin{bmatrix} -1 \\ -1 \\ 0 \end{bmatrix}$$

Ejemplo base ortonormal

Sea el vector $x \in \mathbb{R}^3$, dado por $x = \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix}$, y la base ortonormal

$$a_1 = \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}, \quad a_2 = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 \\ -1 \\ 0 \end{bmatrix}, \quad a_3 = \frac{1}{\sqrt{2}} \begin{bmatrix} -1 \\ -1 \\ 0 \end{bmatrix}$$

Entonces, describimos x en términos de la base ortonormal por

$$x = (-3) \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} + (\frac{3}{\sqrt{2}}) \frac{1}{\sqrt{2}} \begin{bmatrix} 1 \\ -1 \\ 0 \end{bmatrix} + (-\frac{1}{\sqrt{2}}) \frac{1}{\sqrt{2}} \begin{bmatrix} -1 \\ -1 \\ 0 \end{bmatrix}$$

Pregunta

¿Como podemos saber si una lista de vectores son linealmente independientes?

Proceso de ortogonalización de Gram-Schmidt

Algorithm Ortogonalización de Gram-Schmidt

Input: Vectores $\{a_1, a_2, ..., a_k\}$, $a_i \in \mathbb{R}^n$, $1 \leq i \leq k$.
Output: Vectores $\{q_1, ..., q_l\}$, $q_i \in \mathbb{R}^n$, $1 \leq i \leq l \leq k$.

for i=1:k **do**
  $\tilde{q}_i = a_i - (q_1^\top a_i)q_1 - ... - (q_{i-1}^\top a_i)q_{i-1}$
  if $\tilde{q}_i = 0$ **then**
    $l = i - 1$ break
  else
    $q_i = \tilde{q}_i / \|q_i\|_2$
  end if
end for

Ejemplo de ortogonalización

Considere los vectores $$a_1 = \begin{bmatrix} 1/2 \\ 2 \end{bmatrix}, \quad a_2 = \begin{bmatrix} 2 \\ 1 \end{bmatrix}.$$

El proceso de ortogonalización de Gram-Schmidt entrega los vectores ortonormales

$$q_1 = \begin{bmatrix} 0.2425 \\ 0.9701 \end{bmatrix}, \quad q_2 = \begin{bmatrix} 0.9701 \\ -0.2425 \end{bmatrix}.$$

Ilustración gráfica

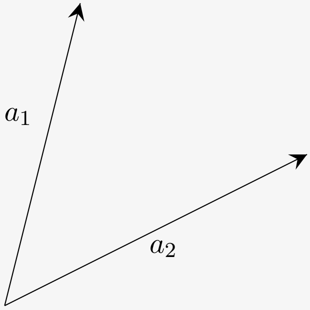


Ilustración gráfica

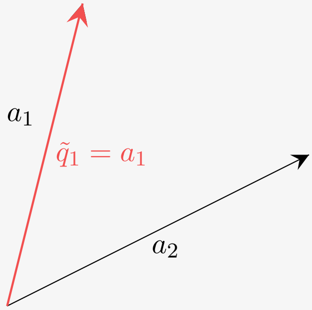


Ilustración gráfica

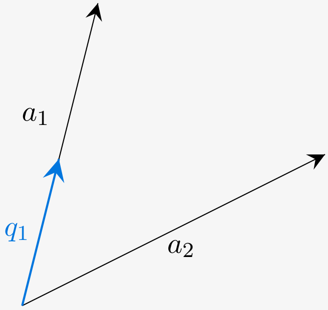


Ilustración gráfica

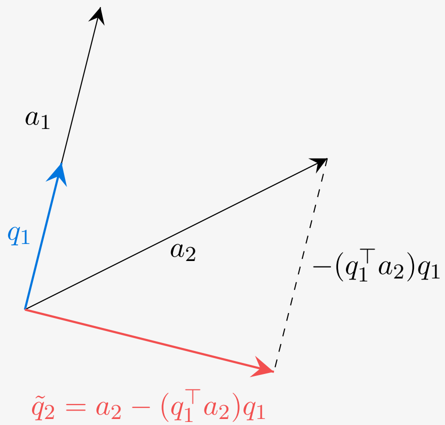
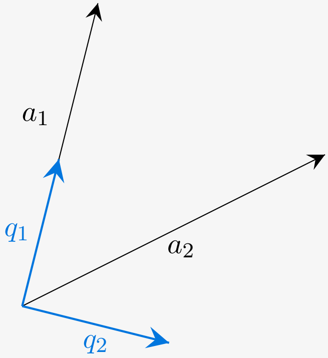


Ilustración gráfica



Ejemplo de ortogonalización

Considere los vectores $\quad a_1 = \begin{bmatrix} 1 \\ 1 \\ -1 \end{bmatrix}, \quad a_2 = \begin{bmatrix} 1 \\ 0 \\ 2 \end{bmatrix}, \quad a_3 = \begin{bmatrix} 2 \\ -2 \\ 3 \end{bmatrix}.$

El proceso de ortogonalización de Gram-Schmidt entrega los vectores ortonormales

$$q_1 = \frac{1}{\sqrt{3}} \begin{bmatrix} 1 \\ 1 \\ -1 \end{bmatrix}, \quad q_2 = \frac{1}{\sqrt{42}} \begin{bmatrix} 4 \\ 1 \\ 5 \end{bmatrix}, \quad q_3 = \sqrt{14} \begin{bmatrix} 2 \\ -3 \\ -1 \end{bmatrix}$$

Ejemplo de ortogonalización

$$a_1 = \begin{bmatrix} 0 \\ -2 \\ -1 \\ -4 \\ 1 \end{bmatrix}, \quad a_2 = \begin{bmatrix} 2 \\ -2 \\ -4 \\ 0 \\ -2 \end{bmatrix}, \quad a_3 = \begin{bmatrix} -4 \\ -1 \\ 0 \\ 1 \\ 0 \end{bmatrix}, \quad a_4 = \begin{bmatrix} 0 \\ 3 \\ 1 \\ -4 \\ 0 \end{bmatrix}, \quad a_5 = \begin{bmatrix} 8 \\ -2 \\ 6 \\ 1 \\ 5 \end{bmatrix}$$

Ejemplo de ortogonalización

$$a_1 = \begin{bmatrix} 0 \\ -2 \\ -1 \\ -4 \\ 1 \end{bmatrix}, \quad a_2 = \begin{bmatrix} 2 \\ -2 \\ -4 \\ 0 \\ -2 \end{bmatrix}, \quad a_3 = \begin{bmatrix} -4 \\ -1 \\ 0 \\ 1 \\ 0 \end{bmatrix}, \quad a_4 = \begin{bmatrix} 0 \\ 3 \\ 1 \\ -4 \\ 0 \end{bmatrix}, \quad a_5 = \begin{bmatrix} 8 \\ -2 \\ 6 \\ 1 \\ 5 \end{bmatrix}$$

$$\tilde{q}_1 = \begin{bmatrix} 0 \\ -2 \\ -1 \\ -4 \\ 1 \end{bmatrix},$$

Ejemplo de ortogonalización

$$a_1 = \begin{bmatrix} 0 \\ -2 \\ -1 \\ -4 \\ 1 \end{bmatrix}, \quad a_2 = \begin{bmatrix} 2 \\ -2 \\ -4 \\ 0 \\ -2 \end{bmatrix}, \quad a_3 = \begin{bmatrix} -4 \\ -1 \\ 0 \\ 1 \\ 0 \end{bmatrix}, \quad a_4 = \begin{bmatrix} 0 \\ 3 \\ 1 \\ -4 \\ 0 \end{bmatrix}, \quad a_5 = \begin{bmatrix} 8 \\ -2 \\ 6 \\ 1 \\ 5 \end{bmatrix}$$

$$\tilde{q}_1 = \begin{bmatrix} 0 \\ -2 \\ -1 \\ -4 \\ 1 \end{bmatrix}, \quad \tilde{q}_2 = \begin{bmatrix} 2. \\ -1.454 \\ -3.727 \\ 1.090 \\ -2.272 \end{bmatrix},$$

Ejemplo de ortogonalización

$$a_1 = \begin{bmatrix} 0 \\ -2 \\ -1 \\ -4 \\ 1 \end{bmatrix}, \quad a_2 = \begin{bmatrix} 2 \\ -2 \\ -4 \\ 0 \\ -2 \end{bmatrix}, \quad a_3 = \begin{bmatrix} -4 \\ -1 \\ 0 \\ 1 \\ 0 \end{bmatrix}, \quad a_4 = \begin{bmatrix} 0 \\ 3 \\ 1 \\ -4 \\ 0 \end{bmatrix}, \quad a_5 = \begin{bmatrix} 8 \\ -2 \\ 6 \\ 1 \\ 5 \end{bmatrix}$$

$$\tilde{q}_1 = \begin{bmatrix} 0 \\ -2 \\ -1 \\ -4 \\ 1 \end{bmatrix}, \quad \tilde{q}_2 = \begin{bmatrix} 2. \\ -1.454 \\ -3.727 \\ 1.090 \\ -2.272 \end{bmatrix}, \quad \tilde{q}_3 = \begin{bmatrix} -3.58 \\ -1.482 \\ -0.862 \\ 0.862 \\ -0.379 \end{bmatrix},$$

Ejemplo de ortogonalización

$$a_1 = \begin{bmatrix} 0 \\ -2 \\ -1 \\ -4 \\ 1 \end{bmatrix}, \quad a_2 = \begin{bmatrix} 2 \\ -2 \\ -4 \\ 0 \\ -2 \end{bmatrix}, \quad a_3 = \begin{bmatrix} -4 \\ -1 \\ 0 \\ 1 \\ 0 \end{bmatrix}, \quad a_4 = \begin{bmatrix} 0 \\ 3 \\ 1 \\ -4 \\ 0 \end{bmatrix}, \quad a_5 = \begin{bmatrix} 8 \\ -2 \\ 6 \\ 1 \\ 5 \end{bmatrix}$$

$$\tilde{q}_1 = \begin{bmatrix} 0 \\ -2 \\ -1 \\ -4 \\ 1 \end{bmatrix}, \quad \tilde{q}_2 = \begin{bmatrix} 2. \\ -1.454 \\ -3.727 \\ 1.090 \\ -2.272 \end{bmatrix}, \quad \tilde{q}_3 = \begin{bmatrix} -3.58 \\ -1.482 \\ -0.862 \\ 0.862 \\ -0.379 \end{bmatrix}, \quad \tilde{q}_4 = \begin{bmatrix} -0.937 \\ 2.352 \\ -0.804 \\ -1.395 \\ -1.681 \end{bmatrix},$$

Ejemplo de ortogonalización

$$a_1 = \begin{bmatrix} 0 \\ -2 \\ -1 \\ -4 \\ 1 \end{bmatrix}, \quad a_2 = \begin{bmatrix} 2 \\ -2 \\ -4 \\ 0 \\ -2 \end{bmatrix}, \quad a_3 = \begin{bmatrix} -4 \\ -1 \\ 0 \\ 1 \\ 0 \end{bmatrix}, \quad a_4 = \begin{bmatrix} 0 \\ 3 \\ 1 \\ -4 \\ 0 \end{bmatrix}, \quad a_5 = \begin{bmatrix} 8 \\ -2 \\ 6 \\ 1 \\ 5 \end{bmatrix}$$

$$\tilde{q}_1 = \begin{bmatrix} 0 \\ -2 \\ -1 \\ -4 \\ 1 \end{bmatrix}, \quad \tilde{q}_2 = \begin{bmatrix} 2. \\ -1.454 \\ -3.727 \\ 1.090 \\ -2.272 \end{bmatrix}, \quad \tilde{q}_3 = \begin{bmatrix} -3.58 \\ -1.482 \\ -0.862 \\ 0.862 \\ -0.379 \end{bmatrix}, \quad \tilde{q}_4 = \begin{bmatrix} -0.937 \\ 2.352 \\ -0.804 \\ -1.395 \\ -1.681 \end{bmatrix}, \quad \tilde{q}_5 = \begin{bmatrix} -1.2e-15 \\ -8.8e-16 \\ 3.3e-16 \\ 2.2e-16 \\ 2.2e-15 \end{bmatrix}$$

Observaciones

1. Si dado un conjunto de vectores $a_1, ..., a_k$ el proceso de Gram-Schmidt se completa entonces los vectores son linealmente independientes.
2. Si dado un conjunto de vectores $a_1, ..., a_k$ el proceso de Gram-Schmidt no se completa, o termina prematuramente en la iteración j, entonces el vector a_j es una combinación lineal de los vectores $q_1, ..., q_{j-1}$ (o también de $a_1, ..., a_{j-1}$).
3. **Algoritmo de Gram-Schmidt Modificado**. En implementaciones una version modificada del algoritmo, pero equivalente matemáticamente, se considera debido a que posee propiedades de errores de redondeo superiores (**ver Jupyter Notebook**). Ver también Algoritmo de Houselholder y Rotaciones de Givens.

Factorización QR

Definición de matriz ortogonal

Una matrix $A \in \mathbb{R}^{m \times n}$ se dice **ortogonal** si satisface que

$$A^\top A = I \in \mathbb{R}^{m \times m}.$$

Observe que las columnas de matrices ortogonales forman una base ortonormal.

Factorización QR

Dado un conjunto de vectores linealmente independientes $a_1, ..., a_n$, el proceso de Gram-Schmidt nos entrega los vectores ortonormales $q_1, ..., q_n$ y podemos escribir

$$\begin{aligned} a_i &= (q_1^\top a_i)q_1 + ... + (q_{i-1}^\top a_i)_{i-1} + +(q_i^\top a_i)q_i \\ &= R_{1i}q_1 + ... + R_{i-1i}q_{i-1} + R_{ii}q_i \end{aligned}$$

Entonces, si $A \in \mathbb{R}^{m\times n}$ con columnas linealmente independientes, usamos el proceso de ortogonalización de Gram-Schmidt para las columnas de A

$$A = \begin{bmatrix} a_1 & a_2 & \cdots & a_n \end{bmatrix} = \begin{bmatrix} q_1 & q_2 & \dots & q_n \end{bmatrix} R = QR$$

donde $R \in \mathbb{R}^{n\times n}$ es triangular superior con elementos en la diagonal no cero y $Q \in \mathbb{R}^{m\times n}$ es ortogonal.

Teorema: factorización QR

Sea $A \in \mathbb{R}^{m\times n}$, $m \geq n$ una matriz con rango completo n. Entonces, existe $Q \in \mathbb{R}^{m\times n}$, ortogonal y $R \in \mathbb{R}^{n\times n}$ triangular superior, tales que:

$$A = QR \qquad \text{factorización reducida}$$

Una factorización QR completa es una matriz ortogonal $\tilde{Q} \in \mathbb{R}^{m\times m}$ y una matriz $\tilde{R} \in \mathbb{R}^{m\times n}$, con ceros bajo la diagonal

$$A = \tilde{Q}\tilde{R}$$

Ejemplo de factorización QR

Se la matriz

$$A = \begin{bmatrix} 1 & 1 & 2 \\ 1 & 0 & -2 \\ -1 & 2 & 3 \end{bmatrix}$$

Entonces, la factorización QR de A es

$$A = \underbrace{\begin{bmatrix} 1/\sqrt{3} & 4/\sqrt{42} & 2/\sqrt{14} \\ 1/\sqrt{3} & 1/\sqrt{42} & -3/\sqrt{14} \\ -1/\sqrt{3} & 5/\sqrt{42} & -1/\sqrt{14} \end{bmatrix}}_{Q} \underbrace{\begin{bmatrix} \sqrt{3} & -1/\sqrt{3} & -\sqrt{3} \\ 0 & \sqrt{14}/\sqrt{3} & \sqrt{21}/\sqrt{2} \\ 0 & 0 & \sqrt{7}/\sqrt{2} \end{bmatrix}}_{R}$$

Resolver ecuaciones lineales con QR

Sea $A \in \mathbb{R}^{n \times n}$ una matriz no singular y sea $n \in \mathbb{R}^n$. Entonces

$$Ax = b \quad \Longleftrightarrow \quad (QR)x = b \quad \Longleftrightarrow \quad Rx = Q^\top b.$$

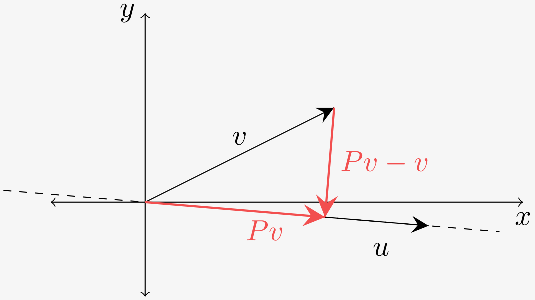
Algorithm Solver de ecuaciones lineales con factorización QR

Input: $A \in \mathbb{R}^{n \times n}$, $b \in \mathbb{R}^n$
Output: $x \in \mathbb{R}^n$.

Calcular factorización $A = QR$
Calcular $Q^\top b$
Resolver sistema triangular $Rx = Q^\top b$

Matriz de proyección

Definición: Una proyecccion es una matriz cuadrada P que satisface $P^2 = P$. Además, decimos que una proyección es ortogonal si $P^\top = P$



Proyección ortogonal

Observación: Si Q es una matriz ortogonal alta, entonces $P = QQ^\top$ satisface que:

$$P^2 = (QQ^\top)(QQ^\top) = QQ^\top = P$$
$$P^\top = (QQ^\top)^\top = QQ^\top = P$$

Proyección

Dado un vector arbitrario v y $\{q_1, q_2, ..., q_n\}$ un conjunto ortonormal

$$
\begin{aligned}
v &= r - (q_1^\top v)q_1 - (q_n^\top v)q_n - ... - (q_n^\top v)q_n \\
&= r - (q_1 q_1^\top)v - (q_2 q_2^\top)v - ... - (q_n q_n^\top)v
\end{aligned}
$$

donde r es la parte de v ortogonal a $\{q_1, q_2, ..., q_n\}$.
La transformación sobre el rango de $Q = [q_1, q_2, ..., q_n]$

$$
v \mapsto (q_1 q_1^\top)v - (q_2 q_2^\top)v - ... - (q_n q_n^\top)v
$$

se escribe

$$
y = QQ^\top v
$$

Proyección para matriz no ortogonal

Es posible definir una proyección sobre una colección de vectores $\{a_1, ..., a_n\}$ columnas de una matriz A. Este es:

$$P = A(A^\top A)^{-1} A^\top$$

Proyección

Considere las matrices

$$A = \begin{bmatrix} 1 & 0 \\ 0 & 1 \\ 1 & 0 \end{bmatrix}, \quad B = \begin{bmatrix} 1 & 2 \\ 0 & 1 \\ 1 & 0 \end{bmatrix}$$

Preguntas:

- Cual es el proyector ortogonal P_A sobre el espacio imagen de A?
- Cual es la imagen sobre P_A del vector $b = [1, 2, 3]^\top$?
- Cual es el proyector ortogonal P_B sobre el espacio imagen de B?
- Cual es la imagen sobre P_B del vector $b = [1, 2, 3]^\top$?

Pseudoinversa

Matriz de Gram

Definición. Sea $A \in \mathbb{R}^{m\times n}$. se define la matriz de Gramm asociada como la matriz cuadrada $A^\top A$.
Observación. Note que A tiene columnas linealmente independientes si y sólo sí su matriz de Gram es invertible. En efecto,

$$\begin{aligned}
(A^\top A)x = 0 \quad &\longrightarrow \quad 0 = x^\top (A^\top A)x = x^\top A^\top Ax = \|Ax\|^2 \\
&\longrightarrow \quad Ax = 0 \\
&\longrightarrow \quad x = 0
\end{aligned}$$

Para el recíproco observe que si existe $x \neq 0$ tal que $Ax = 0$, entonces $A^\top Ax = 0$, lo que implica que la matriz de Gram no es invertible.

Definición de Pseudoinversa

Sea $A \in \mathbb{R}^{m \times n}$.

- Si A tiene columnas linealmente independientes entonces la matriz $(A^\top A)^{-1}A^\top$ es una inversa por la izquierda de A (matrices altas o cuadradas.) La peudoinversa se define entonces

$$A^\dagger = (A^\top A)^{-1}A^\top$$

- Si A tiene filas linealmente independientes entonces la matriz $A^\top(AA^\top)^{-1}$ es una inversa por la derecha de A (matrices anchas o cuadradas.) La peudoinversa se define entonces

$$A^\dagger = A^\top(AA^\top)^{-1}$$

Pseudoinvera por factorización QR

Suponga que $A \in \mathbb{R}^{m \times n}$, y sus columnas son linealmente independientes, y sea su factorización $A = QR$. Entonces:

$$A^\top A = (QR)^\top (QR) = R^\top R$$

y así su pseudoinversa queda:

$$A^\dagger = (A^\top A)^{-1} A = (R^\top R)^{-1} (QR)^\top = R^{-1} Q^\top$$

Resolviendo sistemas sobredeterminados

Considere el sistema sobre determinado

$$Ax = b \quad \iff \quad \begin{bmatrix} -3 & -4 \\ 4 & 6 \\ 1 & 1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 1 \\ -2 \\ 0 \end{bmatrix}$$

La factorización QR de A es

$$A = \begin{bmatrix} -3 & -4 \\ 4 & 6 \\ 1 & 1 \end{bmatrix} = \begin{bmatrix} -0.5883 & 0.4576 \\ 0.7845 & 0.5230 \\ 0.1961 & -0.7191 \end{bmatrix} \begin{bmatrix} 5.0990 & 7.2563 \\ 0 & 0.5883 \end{bmatrix}$$

La solución usando la pseudoinversa es (no necesariamente solución del sistema lineal)

$$x = A^{\dagger} b = R^{-1} Q^{\top} = \begin{bmatrix} -1.2222 & -1.1111 & 1.7778 \\ 0.7778 & 0.8889 & -1.2222 \end{bmatrix} \begin{bmatrix} 1 \\ -2 \\ 0 \end{bmatrix} = \begin{bmatrix} 1 \\ -1 \end{bmatrix}$$

Valores y vectores propios

Definición de valor y vector propio

Sea $A \in \mathbb{R}^{n\times n}$. Un número λ se dice **valor propio** de A si existe un vector llamado **vector propio** $v \in \mathbb{R}^n \backslash \{0\}$ tal que:

$$Av = \lambda v$$

Ejemplo: Considere la siguiente matriz $A = \begin{bmatrix} 3 & 1 \\ 1 & 3 \end{bmatrix}$.

Entonces $\lambda = 2$ y $v = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$, en efecto

$$Av = \begin{bmatrix} 3 & 1 \\ 1 & 3 \end{bmatrix} \begin{bmatrix} 1 \\ 1 \end{bmatrix} = \begin{bmatrix} 4 \\ 4 \end{bmatrix} = 4 \begin{bmatrix} 1 \\ 1 \end{bmatrix} = \lambda v$$

Polinomio característico

Definición: Sea $A = (A_{ij}) \in \mathbb{R}^{n \times n}$ y sea $A_{[i,j]}$ la matriz formada al quitar de A la fila i-ésima y la columna j-ésima. Entonces, el **determinante** de A, denotado por det(A), se define recursivamente por:

1. Si $n = 1$, entonces det$(A) = A_{11}$, la única componente de A.
2. Si $n \geq 2$, entonces para un valor fijo $j \in \{1, ..., n\}$

$$\text{det}(A) = \sum_{i=1}^{n} (-1)^{(i+j)} A_{ij} \text{det}(A_{[i,j]})$$

Definición: Sea $A \in \mathbb{R}^{n \times n}$. El **polinomio característico** asociado a A es el polinomio de grado n en λ dado por

$$p_A(\lambda) = \text{det}(A - \lambda I), \quad \lambda \in \mathbb{C}.$$

Teorema

Un número λ es un valor propio de $A \in \mathbb{R}^{n\times n}$ si y solo si la matriz $(A - \lambda I)$ es singular, es decir, $\mathrm{dim}(\mathbf{im}(A)) =: \mathrm{rank}(A) < n$. Además, λ satisface la ecuación característica

$$p_A(\lambda) = \det(A - \lambda I) = 0.$$

Ejemplo

Sea la siguiente matriz $A = \begin{bmatrix} 0 & -1 & -1 \\ 1 & 2 & 1 \\ 1 & 1 & 2 \end{bmatrix}$

- Compruebe que el polinomio característico es $p_A(\lambda) = (\lambda - 2)(\lambda - 1)^2$.
- Compruebe que el valor propio $\lambda_1 = 2$ y el vector propio asociado es

$$v_1 = \frac{1}{\sqrt{3}} \begin{bmatrix} -1 \\ 1 \\ 1 \end{bmatrix}.$$

- Compruebe que el valor propio $\lambda_2 = 1$ tiene dos vectores propios asociados es

$$v_2 = \frac{1}{\sqrt{2}} \begin{bmatrix} -1 \\ 1 \\ 0 \end{bmatrix} \quad \text{y} \quad v_3 = \frac{1}{\sqrt{2}} \begin{bmatrix} -1 \\ 0 \\ 1 \end{bmatrix}.$$

Propiedades

Sea $A \in \mathbb{R}^{n\times n}$

- La matriz A possee al menos 1 y a lo más n valores propios, los cuales pueden ser complejos.
- La matriz $A^\top$ tiene los mismos valores propios que la matri A.
- Si $\lambda_1, ..., \lambda_k$ sin valores propios distintos de A entonces lo vectores propios correspondientes $v_1, ..., v_k$ son linealmente independientes.
- Si A tiene n valores propios distintos y reales entonces los vectores propios asociados forman una base de $\mathbb{R}^n$.
- La matriz A es singular si y solo si $\lambda = 0$ es valor propio de A.

Matrices diagonalizables

Definición: Una matriz $A \in \mathbb{R}^{n\times n}$ es **diagonalizable** si existe una matriz no singular V y una matriz diagonal $\Lambda = \text{diag}(\lambda_1, ..., \lambda_n)$ tales que

$$A = V\Lambda V^{-1},$$

es decir, A es semejante a la matriz diagonal Λ.

Teorema. Sea $A \in \mathbb{R}^{n\times n}$. Si los valores propios $\lambda_1, ..., \lambda_n$ de A son todos distintos, entonces los vectores propios asociados $v_1, ..., v_n$ son linealmente indpendientes y la matriz A es diagonalizable

$$A = V\Lambda V^{-1} = \begin{bmatrix} v_1 & \dots & v_n \end{bmatrix} \text{diag}(\lambda_1, ..., \lambda_n) \begin{bmatrix} v_1 & \dots & v_n \end{bmatrix}^{-1}$$

Matrices diagonalizables

Teorema. Sea $A \in \mathbb{R}^{n \times n}$ una matriz simétrica $(A = A^\top)$. Entonces A existe una matriz Q ortogonal y una matriz diagonal $\Lambda = \text{diag}(\lambda_1, ..., \lambda_n)$ tales que

$$A = Q\Lambda Q^{-1} = Q\Lambda Q^\top.$$

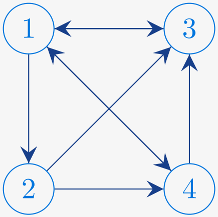
Propiedad. Si λ y v son valor y vector propio de una matriz $A \in \mathbb{R}^{n \times n}$, entonces, λ^2 y v sin valor y vector propio de $A^2 = AA$.

$$Av = \lambda v \quad \Longrightarrow \quad A^2 v = A(Av) = A(\lambda v) = \lambda Av = \lambda\lambda v = \lambda^2 v.$$

PageRank

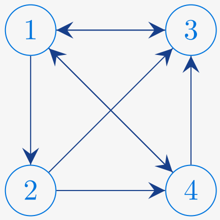
Descripción del problema

Nos interesa una web d n-páginas, donde cada página está indexada con $1 \leq k \leq n$. Denotamos por x_k el puntaje de relevancia de la página k en la web. Así si para dos paginas i y j tenemos que $x_j > x_i$ entonces concluímos que la página j es mas relevante que la página i.



Descripción del problema

Nos interesa una web d n-páginas, donde cada página está indexada con $1 \leq k \leq n$. Denotamos por x_k el puntaje de relevancia de la página k en la web. Así si para dos páginas i y j tenemos que $x_j > x_i$ entonces concluímos que la página j es mas relevante que la página i.

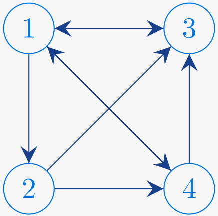


Matriz de conectividad

$$A = \begin{bmatrix} 0 & 0 & 1 & 1 \\ 1 & 0 & 0 & 0 \\ 1 & 1 & 0 & 1 \\ 1 & 1 & 0 & 0 \end{bmatrix}$$

Descripción del problema

Nos interesa una web d n-páginas, donde cada página está indexada con $1 \leq k \leq n$. Denotamos por x_k el puntaje de relevancia de la página k en la web. Así si para dos páginas i y j tenemos que $x_j > x_i$ entonces concluímos que la página j es mas relevante que la página i.



Matriz de conectividad

$$A = \begin{bmatrix} 0 & 0 & 1 & 1 \\ 1 & 0 & 0 & 0 \\ 1 & 1 & 0 & 1 \\ 1 & 1 & 0 & 0 \end{bmatrix}$$

Vector de relevancia

$$x = \begin{bmatrix} 2 \\ 1 \\ 3 \\ 2 \end{bmatrix}$$

Problema

La solución anterior solo considera el número de páginas que llegan a una determinada, sin considerar la relevancia de estas.

Problema

La solución anterior solo considera el número de páginas que llegan a una determinada, sin considerar la relevancia de estas.

Cómo incluir la relevancia de las páginas?

Problema

La solución anterior solo considera el número de páginas que llegan a una determinada, sin considerar la relevancia de estas.

Cómo incluir la relevancia de las páginas?

Bryan, K., & Leise, T. (2006). The $25,000,000,000 eigenvector: The linear algebra behind Google. *SIAM review*, 48(3), 569-581.

Algortimo PageRank

Calcularemos el puntaje de relevancia de la página j como la suma de los puntajes de relevancia de las páginas que tienen link a j, es decir:

$$x_1 = x_3 + x_4, \quad x_2 = x_1, \quad x_3 = x_1 + x_2 + x_4, \quad x_4 = x_1 + x_2$$

Pero obviamente x_3 y x_4 dependen de x_1.
Si la página j contiene n_j links, uno de ellos a k, entonces el puntaje x_k se incrementará por x_j/n_j.
Ejemplo: El nodo 1 apunta a los nodos $2, 3, 4$, entonces contiene 3 links, así aportará acada uno $x_1/3$.

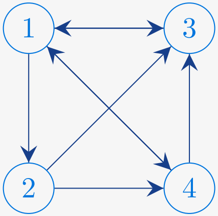
Algortimo PageRank

Calcularemos el puntaje de relevancia de la página j como la suma de los puntajes de relevancia de las páginas que tienen link a j, es decir:

$$x_1 = x_3 + x_4, \quad x_2 = x_1, \quad x_3 = x_1 + x_2 + x_4, \quad x_4 = x_1 + x_2$$

Pero obviamente x_3 y x_4 dependen de x_1.
Si la página j contiene n_j links, uno de ellos a k, entonces el puntaje x_k se incrementará por x_j/n_j.
Ejemplo: El nodo 1 apunta a los nodos $2, 3, 4$, entonces contiene 3 links, así aportará acada uno $x_1/3$.

Algortimo PageRank

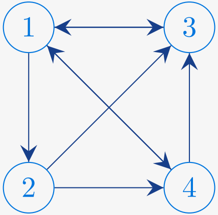


Matriz link

$$A = \begin{bmatrix} 0 & 0 & 1 & 1/2 \\ 1/3 & 0 & 0 & 0 \\ 1/3 & 1/2 & 0 & 1/2 \\ 1/3 & 1/2 & 0 & 0 \end{bmatrix}$$

Buscamos $Ax = x$

Algortimo PageRank



Matriz link

$$A = \begin{bmatrix} 0 & 0 & 1 & 1/2 \\ 1/3 & 0 & 0 & 0 \\ 1/3 & 1/2 & 0 & 1/2 \\ 1/3 & 1/2 & 0 & 0 \end{bmatrix}$$

Buscamos $Ax = x$

Solución es el vector:

$$x = \begin{bmatrix} 12/31 \\ 4/31 \\ 9/31 \\ 6/31 \end{bmatrix}$$

Como calculamos este vector?

Iteración de potencia

Algoritmos para el cálculo de valores propios

Iteración de potencia: genera una aproximación del vector propio correspondiente al valor propio de mayor magnitud de A.

$$\begin{aligned} \text{Inicializar } x_0 : &\quad \|x_0\|_1 = 1, \\ \text{iteración } k : &\quad x^{(k)} = Ax^{(k-1)} \end{aligned}$$

Ejemplo: Considere la matriz de link:

$$x^{(0)} = \frac{1}{8}\begin{bmatrix} 2 \\ 1 \\ 3 \\ 2 \end{bmatrix}, \quad x^{(1)} = \frac{1}{8}\begin{bmatrix} 0 & 0 & 1 & 1/2 \\ 1/3 & 0 & 0 & 0 \\ 1/3 & 1/2 & 0 & 1/2 \\ 1/3 & 1/2 & 0 & 0 \end{bmatrix}\begin{bmatrix} 2 \\ 1 \\ 3 \\ 2 \end{bmatrix} = \begin{bmatrix} 1/2 \\ 1/12 \\ 13/48 \\ 7/48 \end{bmatrix}, \quad x^{(2)} = \begin{bmatrix} 11/32 \\ 1/6 \\ 9/32 \\ 5/24 \end{bmatrix}$$

Descomposición en valores singulares

1. Introducción
2. Ortogonalización/Ortonormalización
3. Factorización QR
4. Pseudoinversa
5. Valores y vectores propios
6. PageRank
7. Iteración de potencia
8. Teorema de Eckart-Young
9. Análisis de Componentes Principales (PCA)

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.
- 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

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

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 $AA^\top \to \lambda_i$, u_i.
- Calcular valores y vectores propios de $A^\top A \to \lambda_i, v_i$.
- 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 $U = \left[u_1, u_2\right]$.

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 SVD

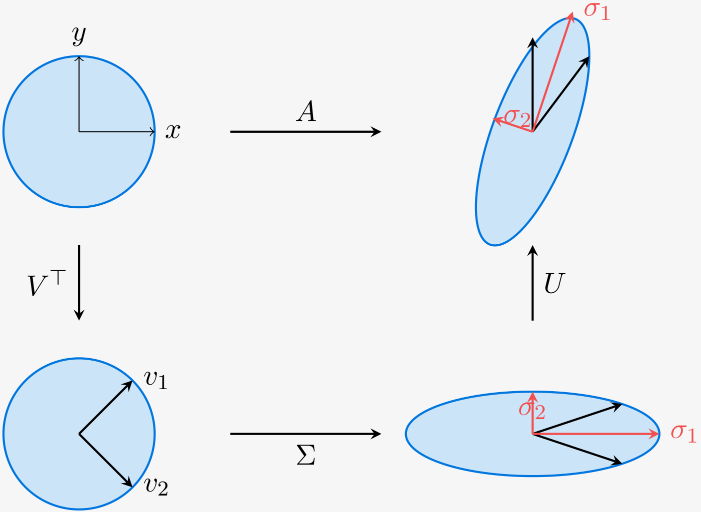
Finalmente, obtenemos los valores singulares

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

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

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

Interpretación geométrica



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

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

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

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

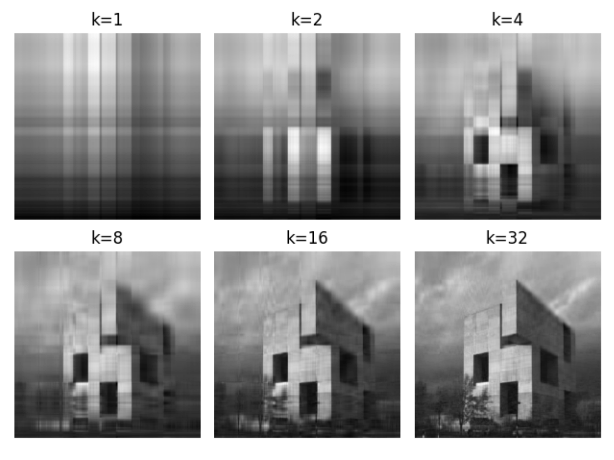


Figure: Sucesion de imagenes creadas con la SVD

Análisis de Componentes Principales (PCA)

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

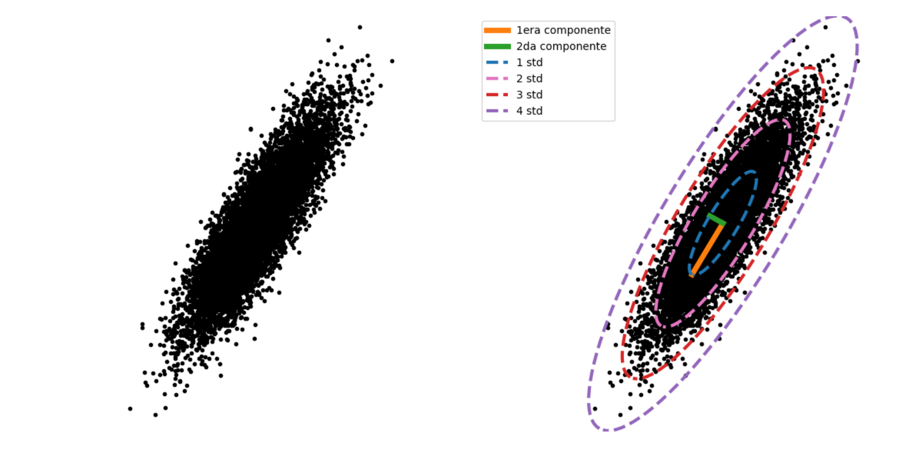


Figure: Distribucion de datos e interpretacion de PCA.

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