Análisis Matemático para Inteligencia Artificial

Verónica Pastor (vpastor@fi.uba.ar),
Martín Errázquin (merrazquin@fi.uba.ar)

Especialización en Inteligencia Artificial

16/9/2022

Se acuerdan cuando vimos formas cuadráticas:

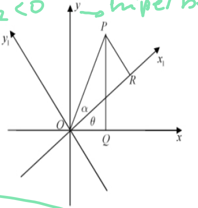
$$\underbrace{\begin{bmatrix} x & y \end{bmatrix}}^{x^T} \overbrace{\begin{bmatrix} 2 & 3 \\ 3 & 9 \end{bmatrix}}^{A} \overbrace{\begin{bmatrix} x \\ y \end{bmatrix}}^{x} = 2x^2 + 6xy + 9y^2 > 0 \quad \forall (x,y) \neq (0,0)$$

$\lambda_1 = \lambda_2 \to$ ec. de una ○

$\lambda_1 \neq \lambda_2,\ \lambda_1, \lambda_2 > 0 \to$ una ⬭

$\lambda_1 > 0,\ \lambda_2 < 0 \to$ hipérbolas

Si rotamos un ángulo θ ✓
la figura dada por: $2x^2 + 6xy + 9y^2 = 1$
a la forma canónica $\boxed{\lambda_1 x_1^2 + \lambda_2 x_2^2 = 1}$
Esto nos lleva a
una matriz de transición $S = \begin{bmatrix} cos(\theta) & -sen(\theta) \\ sen(\theta) & cos(\theta) \end{bmatrix}$



$\tilde{x}^T S^T$

$$x^T A x = (S\tilde{x})^T A (S\tilde{x}) = \tilde{x}^T (S^T A S) \tilde{x} \Rightarrow \boxed{x^T A x = \lambda_1 x_1^2 + \lambda_2 x_2^2}$$

$SS^{-1} = SS^T$

Obs.: S resulta una matriz ortogonal, $S^T A S = S^{-1} A S = \begin{bmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{bmatrix} = D$

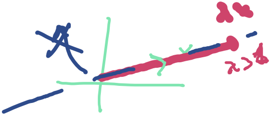
Autovalores y Autovectores: definición

Sea $A \in \mathbb{K}^{n\times n}, \mathbb{R}$ *o* $\mathbb{C}$ diremos que $\lambda \in \mathbb{K}$ es un autovalor (ava) de A y $x \in \mathbb{K}^n, x \neq 0$ es un autovector (ave) asociado a λ si:

$n \times n \quad n\times 1 \quad 1\times 1 \quad 1 \times n$

$$Ax = \lambda x$$

$n\times 1 = n\times 1$



→ Interpretación geométrica: A cada vector que se encuentre en la dirección de x, la transformación $T(x) = Ax$ lo contrae (o expende) por un factor λ.

$0 < \lambda < 1$ $\quad \lambda > 1$ $\quad$ son [illegible] si $\lambda < 0$

Importante: un autovalor puede ser nulo, un autovector no.

Resolución $\quad Ax - \lambda x = 0$

$$\underbrace{(\underset{n\times n}{A} - \underset{\in \mathbb{R}}{\lambda} \underset{n\times n}{I})}_{M} \underset{n\times 1}{x} = \underset{n\times 1}{0}$$

$$M \; x = 0$$

→ seguro es solución $x = 0$ ↳ y NO aporta información

sist. de ec. con n ec. y $n+1$ incógnitas

Queremos que el sist. tenga solución que sea compatible indeterminado

$\det M = 0$

Autovalores y Autovectores: algunas aplicaciones

1. Geometría: curvas planas o superficies.
2. Sistemas dinámicos.
3. Análisis del comportamiento de sistemas de ecuaciones diferenciales.
4. Análisis de estabilidad.
5. Cadenas de Markov.
6. Grafos.
7. Reducción de dimensiones.
8. Cálculo de resonancias del sistema.
9. PageRank.

Hallando los autovalores

Dada $A \in \mathbb{K}^{n \times n}, \mathbb{R}$ *o* $\mathbb{C}$, $\lambda \in \mathbb{K}$ es un autovalor de A sii λ es un cero del polinomio característico, $p(\lambda)$, de A:

$$p(\lambda) = det(A - \lambda I)$$

$p(\lambda) = (\lambda - 1)^2 = 0$

$\lambda = 1$ tiene mult 2

Observación: Puede ocurrir que algún λ sea raíz múltiple de $p(\lambda)$.

Se llama multiplicidad algebraica, m_a, a la cantidad de veces que λ aparece como raíz. $m_a = 2$, $\lambda = 1$

Se llama multiplicidad geométrica, m_g, del autovalor λ, a la cantidad de autovectores l.i. asociados a λ. $m_a \geq m_g \geq 1$

Propiedades de los autovectores

Dada $A \in \mathbb{K}^{n \times n}$, el conjunto de todos los autovectores asociados al autovalor λ generan un subespacio de $\mathbb{K}^n$ (autoespacio de A respecto de λ). El conjunto de todos los autovectores de A se llama autoespectro.

$\det(\underbrace{A - \lambda I})$ tiene que ser cuadrada

Si λ es un autovalor de $A \in \mathbb{K}^{n \times n}$, el autoespacio E_λ asociado es la solución al sistema homogéneo $(A - \lambda I)x = 0$.

Teorema: Los autovectores $x_1, ..., x_n$ de una matriz $A \in \mathbb{R}^{n \times n}$ con n autovalores distintos de $\lambda_1, ..., \lambda_n$ son linealmente independientes. Es decir, forman una base de $\mathbb{R}^n$.

Importante
ahorro de cálculos

$P(\lambda)$ puede tener raíces $\mathbb{R}$ ó $\mathbb{C}$

$P(\lambda) = \lambda^2 + 1 = 0$

Teorema: Si $A \in \mathbb{R}^{n \times n}$ es simétrica, entonces existe una base ortonormal correspondiente a un espacio vectorial formado por autovectores de A y además $\lambda_i \in \mathbb{R}$, $\forall i = 1, \ldots, n$.

Ejemplo

Sea $A = \begin{bmatrix} 2 & -3 \\ 2 & -5 \end{bmatrix}$, calculamos $\det(A - \lambda I) = 0$, $\left| \begin{pmatrix} 2 & -3 \\ 2 & -5 \end{pmatrix} - \begin{pmatrix} \lambda & 0 \\ 0 & \lambda \end{pmatrix} \right|$

$$\begin{vmatrix} 2-\lambda & -3 \\ 2 & -5-\lambda \end{vmatrix} = (2-\lambda)(-5-\lambda) - 2\cdot(-3) = -10 - 2\lambda + 5\lambda + \lambda^2 + 6 =$$

$$= \lambda^2 + 3\lambda - 4 = 0 \iff \lambda_1 = 1, \ \lambda_2 = -4$$

Autoesp. $\{1, -4\}$ son distintos

- Buscamos $E_{\lambda=1}$, $\begin{pmatrix} 2 & -3 \\ 2 & -5 \end{pmatrix}\begin{pmatrix} x \\ y \end{pmatrix} = 1\begin{pmatrix} x \\ y \end{pmatrix} \to \begin{cases} 2x - 3y = x \\ 2x - 5y = y \end{cases}$

$AX = \lambda X$

$(A - \lambda I)X = 0$

ó equiv. $\begin{pmatrix} 1 & -3 \\ 2 & -6 \end{pmatrix}\begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix} \iff x - 3y = 0$, $2x - 6y = 0$

$\boxed{x = 3y}$ → y libre

$E_{\lambda=1} = \left\{ \begin{pmatrix} x \\ y \end{pmatrix} \in \mathbb{R}^2 : x = 3y \right\} = \text{gen}\left\{ \begin{pmatrix} 3 \\ 1 \end{pmatrix} \right\}$ ✓

$E_{\lambda=-4} = \text{gen}\left\{ \begin{pmatrix} 1 \\ 2 \end{pmatrix} \right\}$

- Buscamos $E_{\lambda=-4}$, $\begin{pmatrix} 6 & -3 \\ 2 & -1 \end{pmatrix}\begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$ $\quad 2x - y = 0$

$\boxed{y = 2x}$ x libre

Puedo definir $S = \begin{bmatrix} 3 & 1 \\ 1 & 2 \end{bmatrix}$ (que λ_1, que λ_2)

$S^{-1} A S = D = \begin{bmatrix} 1 & 0 \\ 0 & -4 \end{bmatrix}$ ← diagonal

$(A = S D S^{-1})$

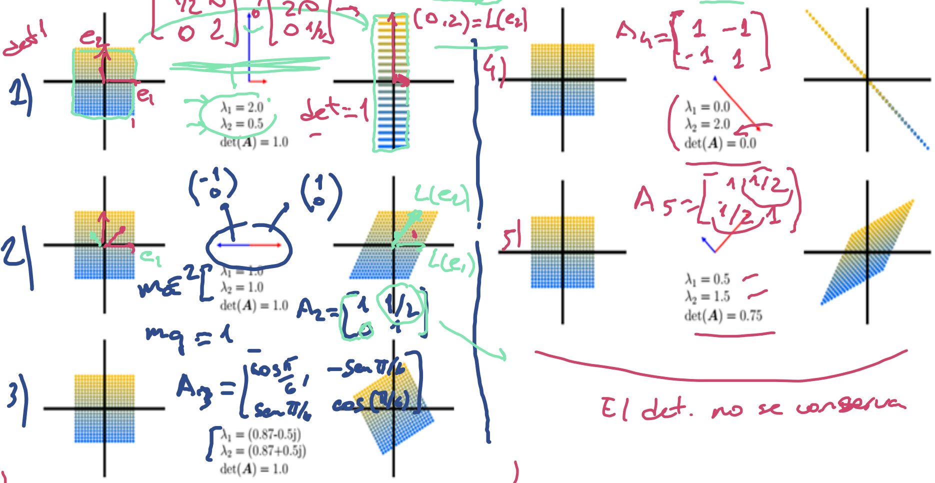
$S^{-1} = \frac{1}{|S|}\begin{bmatrix} a_{22} & -a_{12} \\ -a_{21} & a_{11} \end{bmatrix}$

Importante el orden de los aut... y aut...

Propiedades

$det(A) = det(SDS^{-1}) = det\,S \cdot det\,D \cdot det\,S^{-1}$

Sea $A \in \mathbb{K}^{n\times n}$: Área: $det(A) = \prod_{i=1}^{n} \lambda_i$ y Perímetro: $Tr(A) = \sum_{i=1}^{n} \lambda_i$



El det. no se conserva

det en todos los casos se conserva

Diagonalización

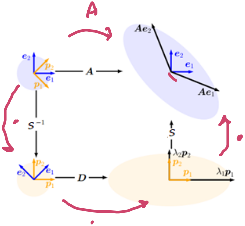
Diagonalización

Definición:

Una matriz $A \in \mathbb{K}^{n\times n}$ es diagonalizable si $\exists S \in \mathbb{K}^{n\times n}$ no singular (es decir, tiene inversa) tal que:

$$\boxed{S^{-1}AS = D} \rightarrow AS = SD$$

donde D es una matriz *diagonal*.



Teorema: Sea $A \in \mathbb{K}^{n\times n}$ es diagonalizable sii A tiene n autovectores linealmente independientes.

1. A es diagonalizable $\Rightarrow$ los vectores columna de la matriz de diagonalización S son los autovectores de A y los elementos de D son los autovalores de A. respetando un orden.
2. ~~B~~ S no es única, se pueden reordenar columnas, etc $\Rightarrow$ D será distinta.
3. A tiene n autovalores distintos $\Rightarrow$ los autovectores correspondientes son l.i. $\Rightarrow$ A es diagonalizable.
4. A tiene menos de n autovectores l.i. $\Rightarrow$ A no es diagonalizable.
5. Si hay autovalores repetidos y si:
   - $m_a = m_g$ (en los ava repetidos) $\Rightarrow$ los ave son l.i.
   - $m_a > m_g$ (en un ava repetido) $\Rightarrow$ A no es diagonalizable.

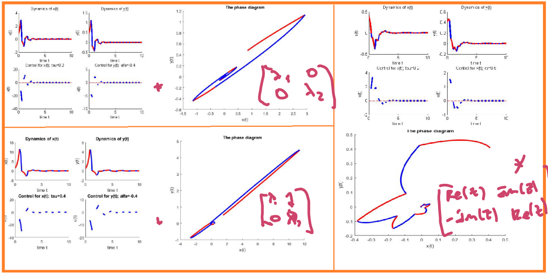
   → equivale a simétrica en $\mathbb{R}$
6. Si $A \in \mathbb{C}^{n\times n}$ es una matriz hermítica ($A = \overline{A^T} = A^H$) $\Rightarrow$ A es diagonalizable y las columnas forman una BON.

Sea $L : \mathbb{V} \to \mathbb{V}$ endomorfismo sobre $\mathbb{K} - e.v.$ con $dim(\mathbb{V}) = n$. Si $p(\lambda)$ se factoriza en $\mathbb{K}$ entonces existe una base que viene dada por la matriz en bloques $(m < n)$.

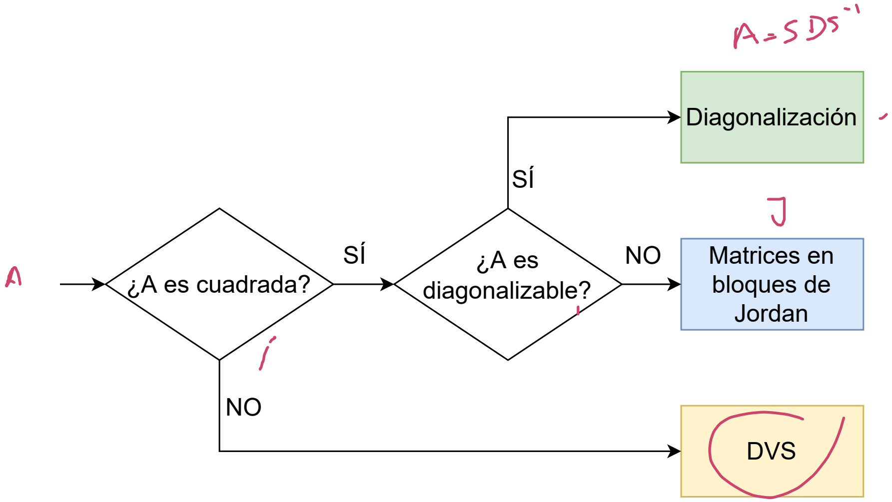
$$J = \begin{bmatrix} A_1 & . & . & . & 0 \\ 0 & . & & & . \\ . & & . & & . \\ . & & & . & . \\ 0 & . & . & . & A_m \end{bmatrix} \in \mathbb{K}^{n\times n} \quad A_k = \begin{bmatrix} \lambda_k & 1 & 0 & \dots & 0 \\ 0 & . & . & & . \\ . & & . & . & . \\ . & & & . & 1 \\ 0 & . & . & . & \lambda_k \end{bmatrix}, \quad \sum_{k=1}^{m} \lambda_k = n.$$

Obs.: Cuando es diagonalizable $m = n$.

Aplicación en Control



¿Qué puede pasar al intentar diagonalizar?



Descomposición en Valores Singulares (DVS)

$\mathbb{R}^{m\times n}$ simétrica

Teorema: Dada $A \in \mathbb{C}^{m\times n}$, siempre se puede obtener una matriz hermítica, semidefinida positiva $S \in \mathbb{C}^{n\times n}$ como:

S es siempre (herm o simétrica)

$$S \doteq A^H A$$

$$S = A^T A$$

En $\mathbb{R}$, ¿$S = S^T$?

$S = A^T A \Rightarrow S^T = (A^T A)^T = A^T (A^T)^T =$

$= A^T A = S$

Obs.: $\tilde{S} \in \mathbb{C}^{m\times m}$ definida como $\tilde{S} \doteq AA^H$ también resulta hermítica y semidefinida positiva.

Obs. (II): Como $S, \tilde{S}$ son semidef. positivas, todos sus autovalores son reales **no negativos**.

$$\underset{A}{\begin{bmatrix} \lambda_1 & \\ & \lambda_2 \end{bmatrix}} \underset{A^T}{\begin{bmatrix} \lambda_1 & \\ & \lambda_2 \end{bmatrix}} \rightarrow \begin{matrix} \lambda_1^2 & - \\ & \lambda_2^2 \end{matrix}$$

Teorema: Sea $A \in \mathbb{R}^{m\times n}$ una matriz de rango $r \in \{0, \dots, \min\{m, n\}\}$. La **descomposición en valores singulares** (DVS, SVD en inglés) de A se define como:

simil la diagonal.

$$\underbrace{A}_{m\times n} = \underbrace{U}_{m\times m} \cdot \underbrace{\Sigma}_{m\times n} \cdot \underbrace{V^T}_{n\times n}$$

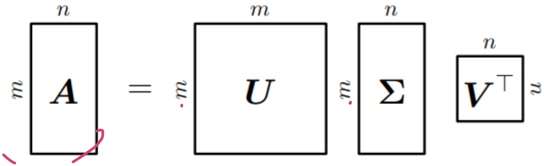
donde:

m×n n×m

- $U \in \mathbb{R}^{m\times m}$ es la matriz ortogonal formada por los aves de AA^T.
- $\Sigma \in \mathbb{R}^{m\times n}$ es la matriz "diagonal" de valores singulares $\boxed{\sigma_i} = \sqrt{\lambda_i}$, con λ_i los primeros $p = \min\{m, n\}$ avas de AA^T y $A^T A$.

n×m m×n

- $V \in \mathbb{R}^{n\times n}$ es la matriz ortogonal formada por los aves de $A^T A$.



Ejemplo: Hallar una DVS

Sea $A = \begin{pmatrix} 1 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}$

Calculamos $A^TA = \begin{pmatrix} 1 & 0 \\ 1 & 0 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} = \begin{pmatrix} 1 & 1 & 0 \\ 1 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}$

1) Observemos que A^TA la puedo ver con 2 bloques $\Rightarrow \lambda = 1$ es un ava

2) Como A^TA tiene 2 filas iguales $\Rightarrow \det(A^TA) = 0 \Rightarrow \lambda = 0$ es un ava

3) Ya tenemos $\lambda_1 = 1$, $\lambda_2 = 0$ y sabemos que $\lambda_1 + \lambda_2 + \lambda_3 = tr(A) = 3$ entonces $\lambda_3 = 2$.

4) Ahora buscamos los autoespacios asociados a c/ava, nos dá:
$E_0 = gen\left\{\begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}\right\}$; $E_1 = gen\left\{\begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix}\right\}$ y $E_2 = gen\left\{\begin{pmatrix} -1 \\ 1 \\ 0 \end{pmatrix}\right\}$

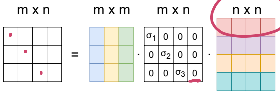
5) Como los ave provienen de ava distintos, son l.i. Además por ser A^TA simétrica los ave son ortogonales, entonces si normalizamos podemos construir $V^T = \begin{bmatrix} -1/\sqrt{2} & 1/\sqrt{2} & 0 \\ 1/\sqrt{2} & 1/\sqrt{2} & 0 \\ 0 & 0 & 1 \end{bmatrix}$ con $\Sigma = \begin{bmatrix} \sqrt{2} & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix}$

6) Análogo para AA^T tiene ava 2 y 1

y finalmente, verificar $A = U \Sigma V^T$

DVS Compacta, Reducida y Truncada

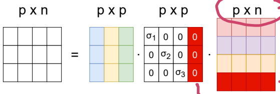
Sea $A \in \mathbb{R}^{m \times n}$ con DVS $A = U\Sigma V^T$:



→ sí es cuadrada

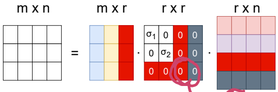
- Si $p = min\{m, n\}$, la DVS compacta es $A = U_p \Sigma_p V_p^T$

→ saco 0's



¿la volamos? ;)
¿No guardo ceros

- Si $\sigma_1, \ldots, \sigma_r > 0, \sigma_{r+1}, \ldots, \sigma_p = 0$, la DVS reducida es $A = U_r \Sigma_r V_r^T$



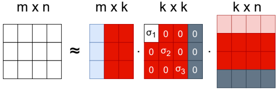
Lo rojo no lo guardo, son cero.

$\sigma_3 = 0$

- Sea $k \in \{1, \ldots, r-1\}$, la DVS truncada es $A \approx U_k \Sigma_k V_k^T$

aproximación.

$\sigma_1 >> \sigma_7$ (por ej.)
puedo despreciar σ_7
y los que sigan



Pseudoinversa de Moore-Penrose

Sea $A \in \mathbb{R}^{m \times n}$ y el sistema lineal $Ax = y$. La inversa de A no está definida, pero si $m > n$ una posible solución (cuadrados mínimos) era

$$\hat{x} = (A^T A)^{-1} A^T y$$

¿Pero qué ocurre si $A^T A$ no es invertible? Se define la pseudoinversa de Moore-Penrose como

$$A^\dagger = V_r \Sigma_r^{-1} U_r^T$$ → se refiere a reducida

y $\hat{x} = \underline{A^\dagger y}$ es (si $m < n$) la solución de mínima norma euclídea o (si $m > n$) la aproximación de mínimo error cuadrático/distancia euclídea.

Obs.: Todo lo anteriormente visto también vale para $A \in \mathbb{C}^{m \times n}$, utilizando $\cdot^H$ en vez de $\cdot^T$. Se utilizaron matrices reales para facilitar la lectura.

Colab

Una aplicación obvia para la DVS es, como vimos recién, la "solución"de sistemas de ecuaciones para los cuales el método de cuadrados mínimos no es factible.

¿Pero es ese el único uso que le podemos dar? Las descomposiciones siempre nos brindan un criterio bajo el cual obtener la *mejor* aproximación de *orden k* para el objeto a descomponer, y eso podría llegar a ser muy útil.