Análisis Matemático para Inteligencia Artificial

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

Especialización en Inteligencia Artificial

11/10/2022

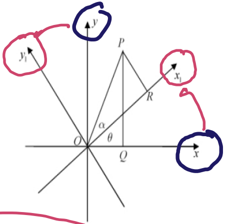
Autovalores y Autovectores: Motivación

Se acuerdan cuando vimos formas cuadráticas:

$$\underbrace{\begin{bmatrix} x & y \end{bmatrix} \begin{bmatrix} 2 & 3 \\ 3 & 9 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix}} = 2x^2 + 6xy + 9y^2$$

me deshago de este término

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



$(BC)^T = C^T B^T$

Diagonal.

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

asociativa

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

$S\,S^{-1} AS = SD$

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:

$$\boxed{Ax = \lambda x}$$

invariante por una transf.
↳ conserva la dirección, (no el sentido, ni la norma)

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.

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

Para determinar la matriz S, $S = \begin{bmatrix} | & | \\ s_1 & s_2 \\ | & | \end{bmatrix}$ tales q

$$AS = SD$$

para el caso s_1: $\underbrace{\overset{n\times n}{A}\ \overset{n\times 1}{s_1}} = \overset{n\times 1}{s_1}\ \overset{1\times 1}{\lambda} = \underbrace{\overset{n\times n}{\lambda I}\ \overset{n\times 1}{s_1}}$ invariante. (no cambia la dirección)

$$\begin{bmatrix} 2 & 3 \\ 3 & 9 \end{bmatrix}\begin{bmatrix} s_{1(1)} \\ s_{1(2)} \end{bmatrix} = \lambda \begin{bmatrix} s_{1(1)} \\ s_{1(2)} \end{bmatrix} \Rightarrow \begin{cases} 2s_{1(1)} + 3s_{1(2)} = \lambda s_{1(1)} \\ 3s_{1(1)} + 9s_{1(2)} = \lambda s_{1(2)} \end{cases}$$ (despejar)

o bien $\underbrace{A s_1}_{\text{matriz}} - \underbrace{\lambda s_1}_{\text{vector}} = \vec{0} \Rightarrow (A - \lambda I) s_1 = \vec{0}$ $\quad \det(A - \lambda I) = 0$ para sol. $x \neq 0$

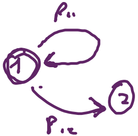
Autovalores y Autovectores: algunas aplicaciones

$\lambda_1 = \lambda_2$

$\lambda_1 \neq \lambda_2,\ \lambda_i > 0$

$\lambda_1 > 0,\ \lambda_2 < 0$ o vicev.

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. Aplic. Blockchain
6. Grafos.
7. Reducción de dimensiones.
8. Cálculo de resonancias del sistema.
9. PageRank.



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:

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

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

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

$\lambda = 1$

Se llama multiplicidad algebraica, m_a, a la cantidad de veces que λ aparece como raíz.

$\lambda = 1$ es una raiz de multi alg 2

Se llama multiplicidad geométrica, m_g, del autovalor λ, a la cantidad de autovectores l.i. asociados a λ.

↪ linealmente independientes

En el ej: $A = \begin{bmatrix} 2 & 3 \\ 3 & 9 \end{bmatrix}$ $\quad p(\lambda) = det(A - \lambda I) = \begin{vmatrix} 2-\lambda & 3 \\ 3 & 9-\lambda \end{vmatrix} = (2-\lambda)(9-\lambda) - 9$

$= 18 - 2\lambda - 9\lambda + \lambda^2 - 9 = \lambda^2 - 11\lambda + 9 = 0$

Si: $\lambda_1 \neq \lambda_2$, $A \in \mathbb{R}^{2\times 2}$

Busco $\lambda_{1,2} = \dfrac{11 \pm \sqrt{11^2 - 4.1.9}}{2.1}$ $\begin{cases} \lambda_1 = \\ \lambda_2 = \end{cases}$

ent v_{λ_1} y v_{λ_2} son autovect. l.i

En gral $1 \leq m_g \leq m_a$

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.

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 $\mathbb{R}^n$.

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 $\lambda_i \in \mathbb{R}$, $\forall i = 1, \ldots, n$.

En el ejemplo $v_{\lambda_1} \perp v_{\lambda_2}$

áng $(v_{\lambda_1}, v_{\lambda_2}) = \pi/2$

Ejemplo

Sea $A = \begin{bmatrix} 2 & -3 \\ 2 & -5 \end{bmatrix}$ no es simétrica $(A \neq A^T)$

① Calculo $p(\lambda) = \det(A - \lambda I) = \begin{vmatrix} 2-\lambda & -3 \\ 2 & -5-\lambda \end{vmatrix}$

$= (2-\lambda)(-5-\lambda) + 6 = \lambda^2 + 3\lambda - 4 \underset{\text{pido}}{=} 0$ Resuelvo $\lambda_1 = 1$, $\lambda_2 = -4$

Autoespectro $= \{1, -4\}$ son distintos ent los ave son l.i.

Busco los autoesp, $\lambda_1 = 1$

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

$\det(A) = 1(-4)$

$Tr(A) = 1 + (-4)$

$E_1 = \left\{ \begin{pmatrix} x \\ y \end{pmatrix} : \begin{pmatrix} 3y \\ y \end{pmatrix} = \begin{pmatrix} 3 \\ 1 \end{pmatrix} \cdot y,\ y \in \mathbb{R} \right\} = \text{gen}\left\{ \begin{pmatrix} 3 \\ 1 \end{pmatrix} \right\}$

subesp

Análog E_{-4}, $\underbrace{\begin{pmatrix} 2 & -3 \\ 2 & -5 \end{pmatrix}\begin{pmatrix} x \\ y \end{pmatrix}} = -4\begin{pmatrix} x \\ y \end{pmatrix}$ resuelvo el sist.

$2y = -4x + 4y$

$y = 2x$

$E_{-4} = \left\{ \begin{pmatrix} x \\ y \end{pmatrix} : \begin{pmatrix} x \\ 2x \end{pmatrix} = \begin{pmatrix} 1 \\ 2 \end{pmatrix} \cdot x,\ x \in \mathbb{R} \right\}$

$= \text{gen}\left\{ \begin{pmatrix} 1 \\ 2 \end{pmatrix} \right\}$

voy a poder escribir $S^{-1}AS = D = \begin{bmatrix} 1 & 0 \\ 0 & -4 \end{bmatrix}$

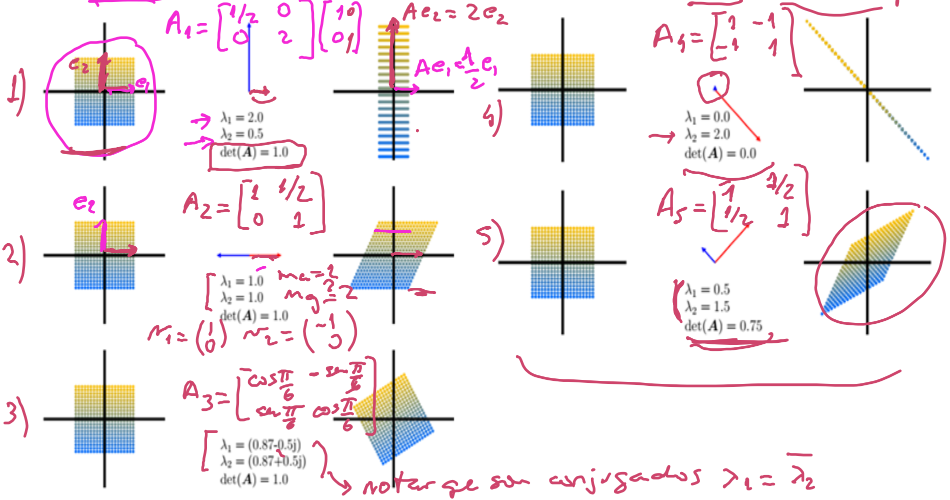
$S = \begin{bmatrix} 3 & 1 \\ 1 & 2 \end{bmatrix}$ $\det(S) \neq 0$

$AS = SD$

Interpretación gráfica (Mathematics for Machine Learning)

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

La suma de los elem. de la diag principal



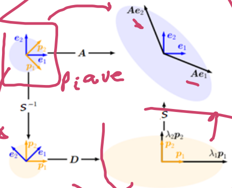
notar que son conjugados $\lambda_1 = \overline{\lambda_2}$

Diagonalización

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix}\begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} f \\ g \end{pmatrix} \longrightarrow \text{muchas cuentas}$$

$$\begin{pmatrix} a & 0 \\ 0 & d \end{pmatrix}\begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} f \\ g \end{pmatrix} \longrightarrow \begin{array}{l} x = f/a \,, \quad a \neq 0 \\ y = g/d \,, \quad d \neq 0 \end{array}$$

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:

$$S^{-1}AS = D$$

donde D es una matriz *diagonal*.

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

dem) ($\Leftarrow$) Sea x_i ave asociado a a λ_i , $i=1,n$ l.i

Para c/ i $\;A x_i = x_i \lambda_i$, $A\vec{X} = A[x_1 \dots x_i \dots x_n] = [\lambda x_1 \dots \lambda x_n] = XD$

Si tengo $AX = XD \Rightarrow X^{-1}AX = D$ → x_i son l.i

(similar a lo ya visto, construyo)

($\Rightarrow$) Si A es diagonalizable; $\exists X$ tq X sea no sing ($\det X \neq 0$) y las columnas de X son los que llamo los ave l.i.

Conclusiones para $A \in \mathbb{K}^{n \times n}$

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.

   conservar el orden de ave con ava

2. P 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.
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 $(m < n)$.

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

λ_1

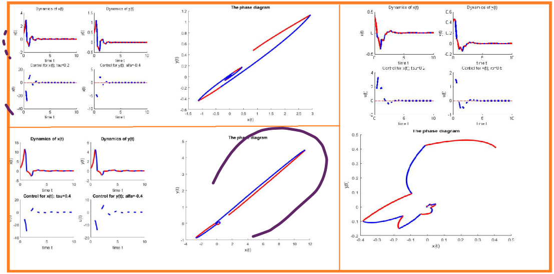
Obs.: Cuando es diagonalizable $m = n$.

$D = \begin{bmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{bmatrix}$

$\lambda_1 \neq \lambda_2$

y reales ?

Aplicación en Control



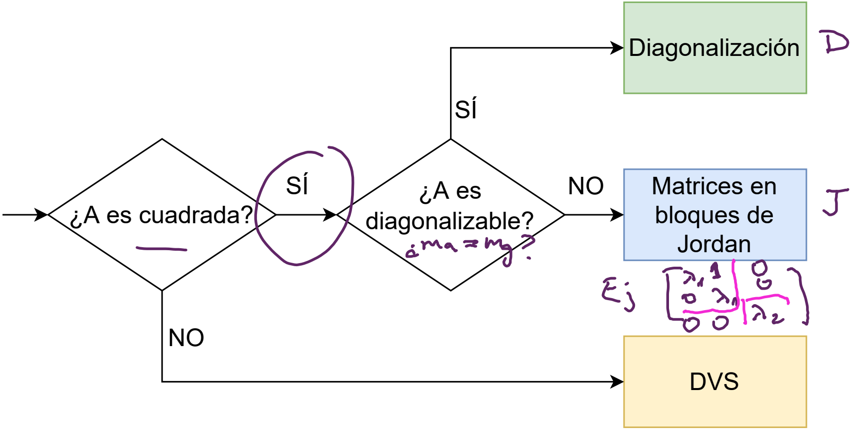
$\lambda_1, \lambda_2 \in \mathbb{C}$

$\lambda_1 = \overline{\lambda_2}$

$\lambda_1 = \lambda_2 \to$

$J = \begin{bmatrix} \lambda_1 & 1 \\ 0 & \lambda_1 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix}$

¿Qué puede pasar al intentar diagonalizar?



Descomposición en Valores Singulares (DVS)

¿Qué pasa cuando A no es cuadrada?

$\mathbb{R}^{m\times n}$ 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:

$x^T A x \geq 0$

$$\underset{n\times n}{S} \doteq \underbrace{A^{H}}_{n\times m}\underbrace{A}_{m\times n}$$

T

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**.

$a_{11}^2 = \lambda_1$

DVS: Definición

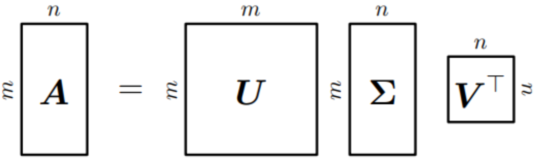
Teorema: Sea $A \in \mathbb{R}^{m\times n}$ una matriz de rango $r \in \{0, \ldots, \text{mín}\{m, n\}\}$. La descomposición en valores singulares (DVS, SVD en inglés) de A se define como:

$$\underbrace{A}_{m\times n} = \underbrace{U}_{m\times m} \cdot \overset{\text{"D"}}{\underbrace{\boxed{\Sigma}}_{(m\times n)}} \cdot \underbrace{V^T}_{n\times n}$$

donde:

- $U \in \mathbb{R}^{m\times m}$ 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 $p = \text{mín}\{m, n\}$ avas de AA^T y A^TA.

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



$m=2,\ n=3$

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

1) Calculo $A^TA = \begin{pmatrix} 1 & 0 \\ 1 & 0 \\ 0 & 1 \end{pmatrix}\begin{pmatrix} 1 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} = \begin{pmatrix} 1 & 1 & 0 \\ 1 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}$ son

2) Busco ava

(i) Como A^TA es en bloques un $\lambda_3 = 1$, me falta $\lambda_{1,2}$ $(\det(A_1 - \lambda I))$

(ii) Si $tr(A^TA) = \sum_{i=1}^{3} \lambda_i = \underset{0}{\lambda_1} + \lambda_2 + \overset{=1}{\lambda_3} = 3$ ent $\boxed{\lambda_2 = 2}$

$\det(A^TA) = \prod_{i=1}^{3} \lambda_i = \lambda_1 \cdot \lambda_2 = 0$

$\lambda_1 = 0 \vee \lambda_2 = 0$

Ver que $\lambda_i \geq 0$ $\sigma_i = \sqrt{\lambda_i}$

3) Busco ave (A revisar) $E_0 = gen\left\{\begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}\right\}$, $E_1 = gen\left\{\begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix}\right\}$ $\lambda=1$, $\lambda = 2$

$E_2 = gen\left\{\begin{pmatrix} -1 \\ 1 \\ 0 \end{pmatrix}\right\}$ $\lambda = 0$

$V = \begin{bmatrix} 1/\sqrt{2} & 0 & -1/\sqrt{2} \\ 1/\sqrt{2} & 0 & 1/\sqrt{2} \\ 0 & 1 & 0 \end{bmatrix} \Rightarrow \Sigma = \begin{bmatrix} \sqrt{2} & 0 & 0 \\ 0 & \sqrt{1} & 0 \end{bmatrix}$

conviene ordenar de mayor a menor

4) Busco U → (i) AA^T y buscar ava(ii) ave

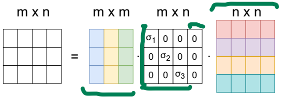
↳ o bien $A = U \underbrace{\Sigma\, V^T}_{\text{de } (3)}$ (U: 2×2, V^T: 3×3) y despejo U

$$u_i = \frac{A v_i}{\sigma_i}$$

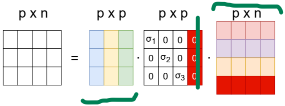
Otra opción. Convenía empezar por 4)

DVS Compacta, Reducida y Truncada

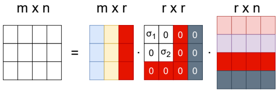
Sea $A \in \mathbb{R}^{m \times n}$ con DVS $A = U\Sigma V^T$:



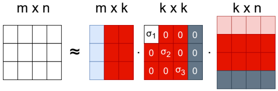
- Si $p = min\{m, n\}$, la DVS compacta es $A = U_p \Sigma_p V_p^T$



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



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



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

$$A^\dagger = V_r \Sigma_r^{-1} U_r^T$$

$A = V_r \cdot \Sigma_r \cdot V_r^T$

y $\hat{x} = A^\dagger y$ 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.

a ver en
clase 5