Análisis Matemático para Inteligencia Artificial

Martín Errázquin (merrazquin@fi.uba.ar)

Especialización en Inteligencia Artificial

Clase 3

$\Pi(\Pi(v)) = P(P\cdot v) = P^2 \cdot v$

$v \xrightarrow{\Pi} ? \xrightarrow{\Pi} ? \xrightarrow{\Pi} ?$

Proyección

Matriz de proyeccion

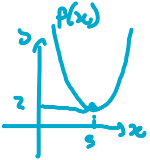
$\Pi(v) = \underbrace{P}_{n\times n} \cdot v$

Sea $\mathbb{V}$ un EV y $S \subset \mathbb{V}$ un SEV. Una transformación lineal $\Pi : \underbrace{\mathbb{V}}_{n} \rightarrow \underbrace{S}_{n}$ es una proyección si $\Pi^2 = \Pi \circ \Pi = \Pi$
Esto significa que la matriz de transformación asociada a la proyección cumple con la propiedad de idempotencia: $[\Pi]^2 = [\Pi]$. $P^2 = P$

$x \in S ? \xrightarrow{\checkmark} x$, $\xrightarrow{\times} y$

Proyección Ortogonal

$x \to \boxed{f} \to y$

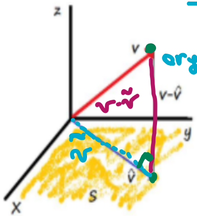


Dado $\mathbb{V}$ un EV con p.i. y $S \subset \mathbb{V}$ un SEV, el objetivo es dado $v \in \mathbb{V}$ hallar $\tilde{v} \in S$ que sea "lo más parecido posible" a v.

$$\tilde{v} \in S : \tilde{v} = arg\ min_{s\in S} \overbrace{||v - s||}^{d(v,s)}$$

$(v-\tilde{v}) \in S^{\perp}$

Además vale que $\langle v - \tilde{v}, s\rangle = 0, \ \forall s \in S$



$\underbrace{arg\ mín\ f(x)}_{2} = s$

Obs: La ortogonalidad de la proyección tiene que ver con el p.i. que se use.

Teorema de proyección

Sea $\mathbb{V}$ un EV de dimensión finita con p.i. $\langle .,. \rangle$, S un SEV. Dado $v \in \mathbb{V}$ existe un único $\tilde{v} \in S$ tal que

$\exists$ $!$

$$\underbrace{||v - \tilde{v}|| \leq ||v - u||}_{\min}, \quad \forall u \in \mathbb{S}$$

¿Cómo hallar la proyección?

Sea $\mathbb{V}$ un EV de dimensión n con p.i. $\langle .,. \rangle$, y $S \subset \mathbb{V}$ un SEV, $dim(S) = m \geq 1$, y sea $B = \{s_1, ..., s_m\}$ una BON de S. Buscamos encontrar la proyección de $\tilde{v} \in S$ de $v \in \mathbb{V}$ ($\tilde{v} = \Pi_S(v)$). Proyección ortgonal de v sobre s

Como $\tilde{v} \in S$, $\tilde{v} = \sum_{i=1}^{m} \alpha_i s_i \Rightarrow$ busco los coeficientes que minimizan $||v - \sum_{i=1}^{m} \alpha_i s_i||$. El problema puede escribirse como:

$||v - \tilde{v}||$

$$\Pi_S(v) = \sum_{i=1}^{m} \alpha_i s_i = B\alpha, \quad B = [s_1, ..., s_m], \; \alpha = [\alpha_1, ..., \alpha_m]^T \in \mathbb{R}^m$$

$= \tilde{v}$ $\quad = \begin{pmatrix} | & & | \\ s_1 & \cdots & s_m \\ | & & | \end{pmatrix} \in \mathbb{R}^{n \times m}$ $\quad = \begin{pmatrix} \alpha_1 \\ \vdots \\ \alpha_m \end{pmatrix}$

$v - \tilde{v} \perp s \quad \forall s \in S$

Como por definición $\langle v - \Pi_S(v), s\rangle = 0$, $\forall s \in S$, debo resolver el siguiente sistema de ecuaciones:

$\langle x, y\rangle = x^T y = y^T x$ producto escalar

$$\begin{pmatrix} | & & | \\ s_1 & \cdots & s_m \\ | & & | \end{pmatrix}$$

$$\begin{cases} \langle v - \overbrace{\Pi_S(v)}^{B\alpha}, s_1\rangle = 0 \\ \vdots \\ \langle v - \Pi_S(v), s_m\rangle = 0 \end{cases} \Rightarrow \begin{cases} s_1^T(v - B\alpha) = 0 \\ \vdots \\ s_m^T(v - B\alpha) = 0 \end{cases} \Rightarrow \underbrace{\begin{pmatrix} -s_1^T- \\ \vdots \\ -s_m^T- \end{pmatrix}}_{\to B^T} (v - B\alpha) = \vec{0}$$

$$B^T(v - B\alpha) = 0 \Leftrightarrow B^T v = \underbrace{B^T B}\alpha \Leftrightarrow \alpha = (B^T B)^{-1} B^T v$$

$\tilde{v} = B \cdot \alpha$
$= \underbrace{B(B^T B)^{-1} B^T}_{} v$
$= P \cdot v$

$$\Pi_S = \underline{B(B^T B)^{-1} B^T} = P$$

Observación: Si B es una BON entonces $P_{\Pi_S} = BB^T$.

Aplicación: Cuadrados Mínimos

$\hat{A} \overset{?}{x} = \hat{y}$ $\quad$ si: $A \in \mathbb{R}^{n \times n} \wedge \det(A) \neq 0 \Rightarrow x^* = A^{-1} \cdot y$

Supongamos que tenemos un sistema sobredeterminado de la forma:

subdeterminado $m < n$

$m > n$

$$Ab = y, \; A \in \mathbb{R}^{m \times n}, \; b \in \mathbb{R}^n, \; y \in \mathbb{R}^m, \; m > n.$$

b es la solución de cuadrados mínimos

Como $m > n$, puede que no exista b que satisfaga todas las m ecuaciones, entonces busco la solución que más se acerque (busco $Proy_{Col(A)} y$)

$$b = (A^T A)^{-1} A^T y \;\Rightarrow\; P = A(A^T A)^{-1} A^T$$

MULTICOLINEALIDAD

$\beta_1, \dots, \beta_n \in \mathbb{R} /$

$S = Col(A)$

si: $A = \begin{pmatrix} | & & | \\ A_1 & \dots & A_n \\ | & & | \end{pmatrix} \Rightarrow Col(A) = \langle \{ A_1, \dots, A_n \} \rangle$

$\underbrace{\beta_1 \cdot \begin{pmatrix} | \\ A_1 \\ | \end{pmatrix} + \dots \beta_n \begin{pmatrix} | \\ A_n \\ | \end{pmatrix}}_{= \hat{y}} - \begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix}$ $\quad \|\cdot\|^2$

Colab

Ya vimos que el ajuste por cuadrados mínimos funciona en la teoría, ahora veámoslo funcionar en la práctica!

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:

(λ, x)
ava ave → dirección

$$Ax = \lambda x$$

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.

Sea (λ, x) par ava-ave de A, $y = k \cdot x$ con $k \in \mathbb{K} \neq 0$

$$A \cdot y = A \cdot (k \cdot x) = k \cdot A \cdot x = \underbrace{k \cdot \lambda}_{\text{producto entre } \mathbb{K}\text{'s}} \cdot x = \lambda (k \cdot x) = \lambda \cdot y$$

↑ escalar

si (λ, x) es ava-ave de $A \Rightarrow \forall k_{\neq 0}$ $(\lambda, k \cdot x)$ también

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:

$\exists x \neq \vec{0} \,/\, Ax = \lambda x$

$Ax - \lambda x = \vec{0}$

$(A - \lambda . I) . x = \vec{0}$

$\Updownarrow$

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

$n = \sum_i m_a(\lambda_i)$

$R \to R$

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

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.

$1 \leq m_g(\lambda) \leq m_a(\lambda) \leq n$

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

$A = \begin{pmatrix} 4 & -4 \\ 1 & 0 \end{pmatrix} \to p(\lambda) = \begin{vmatrix} 4-\lambda & -4 \\ 1 & -\lambda \end{vmatrix} = -4\lambda + \lambda^2 + 4 = (\lambda - 2)^2 = 0 \Rightarrow \lambda_1 = 2$

$\lambda_2 = 2$

$m_a(2) = 2$

ave?

ave?

$(A - 2I)\vec{x} = \vec{0}$

$\begin{pmatrix} 2 & -4 \\ 1 & -2 \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix} \sim \begin{cases} 2x - 4y = 0 \\ x - 2y = 0 \end{cases}$

$0 = 0 \quad \forall y$

$x = 2y$

$\vec{x} = \begin{pmatrix} 2y \\ y \end{pmatrix} = y \begin{pmatrix} 2 \\ 1 \end{pmatrix}$ $\forall y$

$m_g(2) = 1$

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.

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 $\mathbb{R}^n$.

(λ_1, x_1)
$\vdots$
(λ_n, x_n)

con todos $\lambda_i \neq \lambda_j \Rightarrow B = \{x_1, \dots, x_n\}$ es 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 $\lambda_i \in \mathbb{R}$, $\forall i = 1, \dots, n$.

$p(\lambda) = \lambda^2 + 1 = 0 \rightarrow \lambda_1 = i$
$\lambda_2 = -i$

Ejemplo

$A=\begin{pmatrix}2&-3\\2&5\end{pmatrix}$ $\quad\dots\quad \lambda_1=1$

$\lambda_2=-4$

$(A-I)\vec{x}=\vec{0}$ $\quad\dots\quad \vec{x}=y\begin{pmatrix}3\\1\end{pmatrix}\Rightarrow E_1=\left\langle\begin{pmatrix}3\\1\end{pmatrix}\right\rangle$

$(A+4I)\vec{x}=\vec{0}$ $\quad\dots\quad \vec{x}=x\begin{pmatrix}1\\2\end{pmatrix}\Rightarrow E_{-4}=\left\langle\begin{pmatrix}1\\2\end{pmatrix}\right\rangle$

$m_g(1)=m_a(1)$

$m_g(-4)=m_a(-4)$

$\Downarrow$

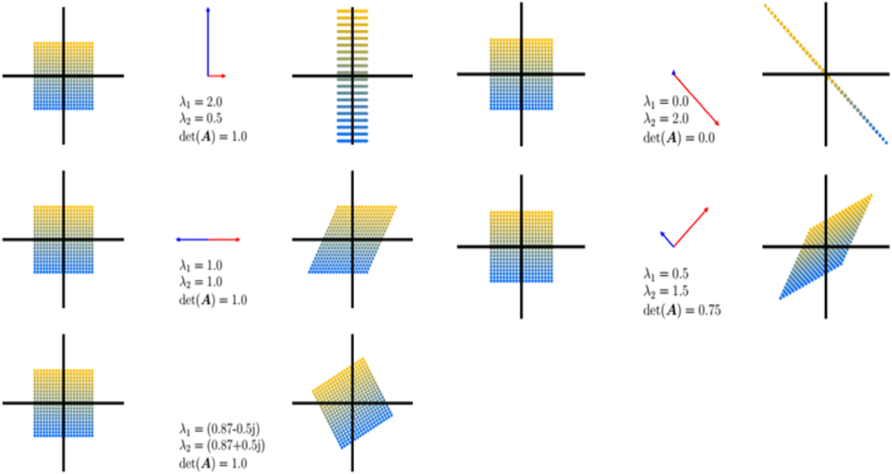
$B=\left\{\begin{pmatrix}3\\1\end{pmatrix},\begin{pmatrix}1\\2\end{pmatrix}\right\}$

$\Downarrow$

$D=\begin{pmatrix}1&0\\0&-4\end{pmatrix}\quad S=\begin{pmatrix}3&1\\1&2\end{pmatrix}\Leftarrow$ A es diag.

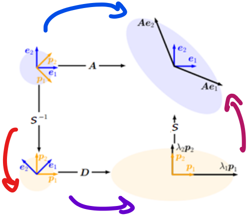
Interpretación gráfica (Mathematics for Machine Learning)

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



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:



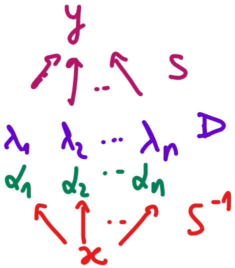
$$S^{-1}AS = D \iff A = S \cdot D \cdot S^{-1}$$

donde D es una matriz *diagonal*.

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

$$A \quad \begin{matrix}(\lambda_1, v_1)\\ \vdots \\ (\lambda_n, v_n)\end{matrix} \Rightarrow S = \begin{pmatrix} | & & | \\ v_1 & \cdots & v_n \\ | & & | \end{pmatrix} \qquad D = \begin{pmatrix} \lambda_1 & & 0 \\ & \ddots & \\ 0 & & \lambda_n \end{pmatrix}$$

$$B = \{v_1, \dots, v_n\}$$



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.
2. 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. $\Leftrightarrow$ A es diagonalizable. $m_a(\lambda_i)=1 \ \forall i \Rightarrow m_a(\lambda_i)=1=m_g(\lambda_i) \ \forall i$
4. A tiene menos de n autovectores l.i. $\Leftrightarrow$ 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.
7. A es diagonalizable $\Rightarrow$ $A^n = SD\cancel{S^{-1}S}DS^{-1}\cdots = SD^nS^{-1}$

$A^{1000} = \underbrace{A\cdot A\cdots A}_{1000}$ $\quad D^k = \begin{pmatrix}\lambda_1^k & & 0\\ & \ddots & \\ 0 & & \lambda_n^k\end{pmatrix}$ $\quad = S\cdot D\cdots D\cdot S^{-1} =$

Recordatorio: Encuesta de Clase

Recordar que está disponible la encuesta de clase! Completarla es cortito y sirve para ir monitoreando el estado del curso.

¿Dónde encontrarla? En la hoja de notas (e.g. "Notas CEIA 10Co2024"), abajo de todo, junto al link de las grabaciones de las clases.