

Análisis Matemático para Inteligencia Artificial

Martín Errázquin (merrazquin@fi.uba.ar)

Especialización en Inteligencia Artificial

Clase 3

Proyección Ortogonal

$$T(\alpha v + \beta w) = \alpha T(v) + \beta T(w)$$

Proyección

$$v \in S \xrightarrow{\Pi} v$$

$$v \notin S \xrightarrow{\Pi} \tilde{v}$$

Sea \mathbb{V} un EV y $S \subset \mathbb{V}$ un SEV. Una transformación lineal $\Pi : \mathbb{V} \rightarrow S$ 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: $[\Pi]^2 = [\Pi]$.



$$y = Ax \quad Ay = y$$

Proyección Ortogonal

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 .

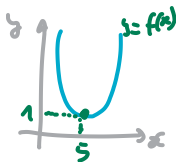
$$v - \tilde{v} \in S^\perp$$

$$\tilde{v} \in S : \tilde{v} = \arg \min_{s \in S} \|v - s\|$$

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

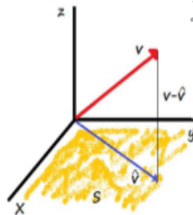
Obs: La ortogonalidad de la proyección tiene que ver con el p.i. que se use.

$$y = A \cdot A \cdot x$$



$$1 = \min f(x)$$

$$s = \arg \min f(x)$$



Teorema de proyección

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

$$\|v - \tilde{v}\| \leq \|v - u\|, \quad \forall u \in S$$

¿Cómo hallar la proyección?

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

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:

$\underbrace{v - \tilde{v} \in S^\perp}_{\tilde{v} - v \in S^\perp} \Leftrightarrow v - \tilde{v} \perp s_i \quad \forall s_i \in B$

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

$\tilde{v} = \Pi \cdot v = B \cdot \alpha$

$B = \begin{pmatrix} | & & | \\ s_1 & \cdots & s_m \\ | & & | \end{pmatrix} \in \mathbb{R}^{n \times m}$

¿Cómo hallar la proyección?

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

Ad *ni p.i. es el mod. menor* $\langle x, y \rangle = x^T y = y^T x$ $B = \begin{pmatrix} | & & | \\ s_1 & \dots & s_m \\ | & & | \end{pmatrix}$

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

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

$$\rightarrow \Pi_S = B(B^T B)^{-1} B^T \quad \tilde{v} = B \alpha$$

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


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

Aplicación: Cuadrados Mínimos

$$y = \begin{pmatrix} y_1 \\ y_2 \\ y_3 \end{pmatrix} = \alpha \cdot \begin{pmatrix} x_1 \\ 1 \end{pmatrix} + \beta \cdot \begin{pmatrix} x_2 \\ 1 \end{pmatrix} + \dots = \underbrace{\begin{pmatrix} x_1 & 1 & \dots \end{pmatrix}}_A \underbrace{\begin{pmatrix} \alpha \\ \beta \\ \vdots \end{pmatrix}}_b$$

$$\min \|y - \hat{y}\|$$

Supongamos que tenemos un sistema sobredeterminado de la forma: $\hat{y} = A \cdot \hat{b}$


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

b es la solución de cuadrados mínimos

$$\text{Col}(A) = \left\langle \begin{pmatrix} x_1 \\ 1 \end{pmatrix}, \dots, \begin{pmatrix} x_n \\ 1 \end{pmatrix} \right\rangle$$

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

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

multicolinealidad

↓
eliminar las cols. "malas"

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

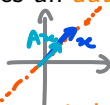
Autovalores y Autovectores: definición

$(2, (5, 5))$

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:

$(2, (1, 1))$

$(2, (-1, -1))$



$$Ax = \lambda x$$

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

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

sea A con (λ, v) ave-ave y sea $k \in \mathbb{K}$, $k \neq 0$

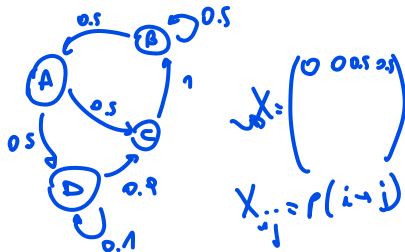
sea $w = k \cdot v$

$$\boxed{A \cdot w = A \cdot k \cdot v = k \cdot A \cdot v = k \cdot \lambda \cdot v = \lambda \cdot \underbrace{k \cdot v}_w = \lambda \cdot w}$$

si (λ, v) es un ave-ave de A , $(\lambda, k \cdot v)$ también $\forall k \neq 0$

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 si λ es un cero del polinomio característico, $p(\lambda)$, de A :

$$\sum_{\lambda_i} m_a(\lambda_i) = n$$

$$|\mathbb{K}| = |\mathbb{K}|$$

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

$$\det(A - \lambda I) = 0 \Rightarrow 3x + 0 / (x - \lambda)x = 0$$

$$Ax - \lambda x = 0$$

$$Ax = \lambda x$$

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.

$$\forall \lambda: 1 \leq m_g(\lambda_i) \leq m_a(\lambda_i) \leq n$$

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

$$A: \begin{pmatrix} 4 & -1 \\ 1 & 2 \end{pmatrix} \leadsto p(\lambda) = \det \left(\begin{pmatrix} 4 & -1 \\ 1 & 2 \end{pmatrix} - \lambda \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \right) = \begin{vmatrix} 4-\lambda & -1 \\ 1 & 2-\lambda \end{vmatrix} = \lambda^2 - 4\lambda + 4 = (\lambda - 2)^2 = 0$$

$$\left\{ \begin{array}{l} \lambda_1 = 2 \\ \lambda_2 = 2 \end{array} \right\}$$

$$m_a(2) = 2$$

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

$$\begin{pmatrix} 2 & -1 \\ 1 & -2 \end{pmatrix} \begin{pmatrix} a \\ b \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix} \leadsto \begin{cases} 2a - b = 0 \\ a - 2b = 0 \end{cases} \begin{matrix} 4b - b = 0 \\ 6a = 2b \end{matrix} \begin{matrix} \forall b \\ a = \frac{1}{3}b \end{matrix} \begin{pmatrix} 2b \\ b \end{pmatrix} = b \begin{pmatrix} 2 \\ 1 \end{pmatrix} \quad m_g(2) = 1$$

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 $(\lambda, v_1), (\lambda, v_2)$ son eve-eve $\Rightarrow (\lambda, \alpha v_1 + \beta v_2)$ También $\forall \alpha, \beta \neq 0, 0$

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, \dots, x_n de una matriz $A \in \mathbb{R}^{n \times n}$ con n autovalores distintos de $\lambda_1, \dots, \lambda_n$ 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 $\lambda_i \in \mathbb{R}, \forall i = 1, \dots, n$.

Ejemplo

$$A = \begin{pmatrix} 1 & -3 \\ 1 & 1 \end{pmatrix}$$

2×2

$$\left. \begin{array}{l} \lambda_1 = 1 \\ \lambda_2 = -4 \end{array} \right\} 2 \text{ ova}$$

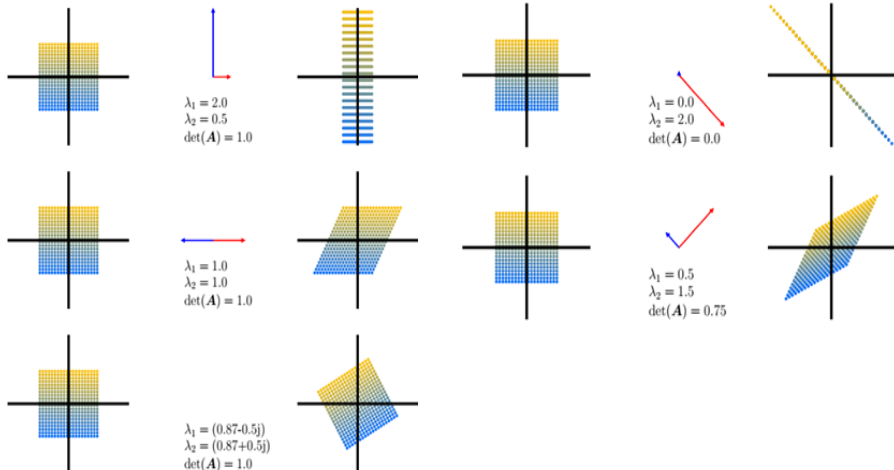
$$(A - I\lambda) \vec{x} = \vec{0} \rightarrow \lambda = 1 \Rightarrow \vec{x} = a \begin{pmatrix} 1 \\ 1 \end{pmatrix} \rightarrow m_a(1) = 1 = m_g(1)$$

$$\downarrow \lambda = -4 \Rightarrow \vec{x} = b \begin{pmatrix} 1 \\ 2 \end{pmatrix} \rightarrow m_b(-4) = 1 = m_g(-4)$$

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

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: $Tr(A) = \sum_{i=1}^n \lambda_i$



Diagonalización

Definición:

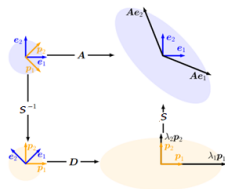
Una matriz $A \in \mathbb{K}^{n \times n}$ es diagonalizable si $\exists S \in \mathbb{K}^{n \times n}$ no singular (es decir, tiene inversa) tal que:

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

donde D es una matriz *diagonal*.

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

$$(\lambda_1, v_1) \dots (\lambda_n, v_n) \Rightarrow S = \begin{pmatrix} | & & | \\ v_1 & \dots & v_n \\ | & & | \end{pmatrix} \quad D = \begin{pmatrix} \lambda_1 & & 0 \\ & \ddots & \\ 0 & & \lambda_n \end{pmatrix}$$



$$y = Ax$$

$$[x]_{\mathcal{B}} = S^{-1} \cdot x$$

$$[y]_{\mathcal{B}} = D \cdot [x]_{\mathcal{B}}$$

$$y = S \cdot [y]_{\mathcal{B}}$$

$$y = S \cdot D \cdot S^{-1} \cdot x$$

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

$$S: \mathcal{B} \rightarrow \mathcal{E}$$

$$S^{-1}: \mathcal{E} \rightarrow \mathcal{B}$$

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

$$D = \begin{pmatrix} \lambda_1 & 0 \\ 0 & \lambda_n \end{pmatrix} \quad D' = \begin{pmatrix} \lambda_n & 0 \\ 0 & \lambda_1 \end{pmatrix}$$

- 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.
- 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 $\Rightarrow A^n = SDS^{-1}SDS^{-1} \dots = SD^nS^{-1}$

$$D^k = \begin{pmatrix} \lambda_1^k & & 0 \\ & \ddots & \\ 0 & & \lambda_n^k \end{pmatrix}$$

$$\underbrace{I}$$

$$A \text{ es diag} \Leftrightarrow \sum m_g(\lambda_i) = n$$

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.