Análisis Matemático para Inteligencia Artificial

Martín Errázquin (merrazquin@fi.uba.ar)

Especialización en Inteligencia Artificial

Clase 3

$v \to \boxed{\Pi} \to \tilde{v} \to \boxed{\Pi} \to \tilde{v}$

si $v \in S \to \Pi(v) = v$

$v \notin S \to \Pi(v) = \tilde{v}$

Proyección

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

$y = A \cdot A \cdot x$

$= A^2$

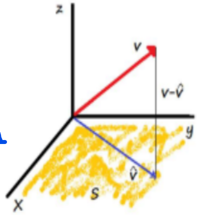
Proyección Ortogonal

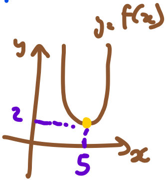
Dado $\mathbb{V}$ un EV con p.i. y $S \subset \mathbb{V}$ un SEV, el objetivo es dado $v \in \mathbb{V}$ hallar $\tilde{v} \in S$ que sea "lo más parecido posible" a v.

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

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

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





$\min f(x) = 2$

$\operatorname{argmin} f(x) = 5$

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

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

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

¿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 $v \in \mathbb{V}$ ($\tilde{v} = \Pi_S(v)$).

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:

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

$$B = \begin{pmatrix} | & & | \\ s_1 & \cdots & s_m \\ | & & | \end{pmatrix} \in \mathbb{R}^{n \times m} \qquad \alpha = \begin{pmatrix} \alpha_1 \\ \vdots \\ \alpha_m \end{pmatrix} \in \mathbb{R}^m$$

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

$v - \tilde{v} \in S^\perp \Leftrightarrow v - \tilde{v} \perp s_i \quad \forall i = 1, \dots, m$

$\begin{pmatrix} | & & | \\ s_1 & \cdots & s_m \\ | & & | \end{pmatrix}^T = B^T$

p.i. canónico

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

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

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

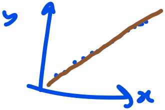
$= \underbrace{B \cdot (B^T B)^{-1} \cdot B^T}_{\Pi_S} \cdot v$

$= \Pi_S \cdot v$

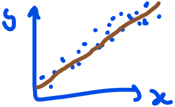
$$\Pi_S = B(B^T B)^{-1}B^T$$

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

Aplicación: Cuadrados Mínimos



$y = ax + b$



Supongamos que tenemos un sistema sobredeterminado de la forma:

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

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 $Proy_{Col(A)}y$)

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

β

(x_1, x_2, x_3) $\qquad$ $\hat{y} = \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3$

$(\beta_1, \beta_2, \beta_3) \in \mathbb{R}^3$

Colab

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

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:

(λ, x) ava-ave de A

$$Ax = \lambda x$$

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) un par ava-ave de A, sea $k \in \mathbb{K}$, $k \neq 0$

sea $y = k \cdot x$

$$\boxed{A \cdot y} = A \cdot k \cdot x = k \cdot A \cdot x = k \cdot \lambda \cdot x = \lambda \cdot \underbrace{k \cdot x} = \boxed{\lambda \cdot y}$$

$\Rightarrow (\lambda, k \cdot x)$ es par ava-ave de A

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.

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:

$p(\lambda)=0 \Leftrightarrow \det(A-\lambda I)=0 \Leftrightarrow \exists x \neq \vec{0} \,/\, Ax=\lambda\cdot x$

$\sum_i m_a(\lambda_i) = n$

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

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.

$A = \begin{pmatrix} 4 & -4 \\ 1 & 0 \end{pmatrix} \rightsquigarrow p(\lambda) = \begin{vmatrix} 4-\lambda & -4 \\ 1 & -\lambda \end{vmatrix} = -4\lambda + \lambda^2 + 4 = \underset{=0}{(\lambda-2)^2} \rightsquigarrow \begin{matrix} \lambda_1 = 2 \\ \lambda_2 = 2 \end{matrix}$ $\quad m_a(2) = 2$

resuelvo $(A - \underset{\lambda=2}{2}\cdot I)x = \vec{0}$:

$\begin{pmatrix} 2 & -4 \\ 1 & -2 \end{pmatrix}\begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix} \rightsquigarrow \begin{cases} 2x - 4y = 0 \\ x - 2y = 0 \end{cases}$ $\quad 0=0\ \forall y$ $\quad \rightarrow x = 2y$

$x = \begin{pmatrix} 2y \\ y \end{pmatrix} = y\begin{pmatrix} 2 \\ 1 \end{pmatrix} \ \forall y \quad m_g(2) = 1$

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.

$I = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$ $\quad \left(1, \begin{pmatrix} 1 \\ 0 \end{pmatrix}\right)$ $\quad \left(1, \begin{pmatrix} 0 \\ 1 \end{pmatrix}\right)$

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 $\mathbb{R}^n$. $B = \{x_1, ..., x_n\}$ forman una base

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

$A = \begin{pmatrix} 2 & -3 \\ 2 & 5 \end{pmatrix}$ $\quad \left.\begin{matrix} \lambda_1 = 1 \\ \lambda_2 = -4 \end{matrix}\right\}$ avo

2×2

$\downarrow$

$n = 2$

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

n avo's distintos

resuelvo $(A - \lambda I)\vec{x} = \vec{0} \xrightarrow{\lambda = 1} \vec{x} = a\begin{pmatrix} 3 \\ 1 \end{pmatrix} \rightarrow m_g(1) = 1$

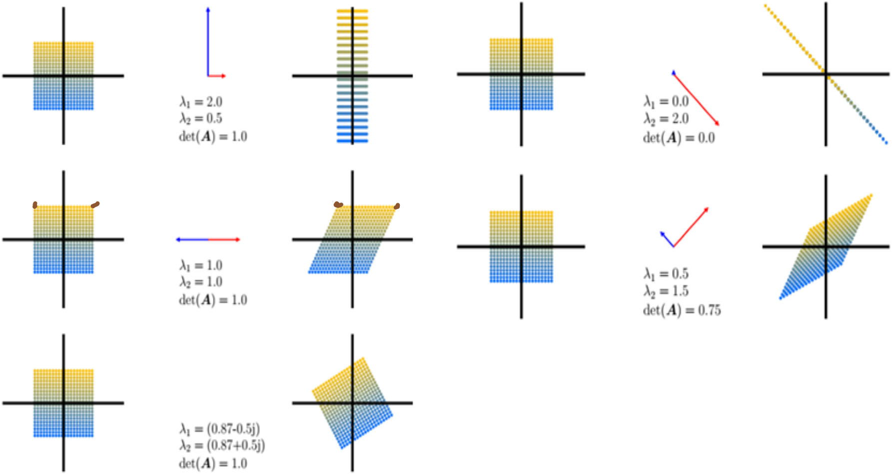
$\xrightarrow{\lambda = -4} \vec{x} = b\begin{pmatrix} 1 \\ 2 \end{pmatrix} \rightarrow m_g(-4) = 1$

$+ \quad 2 = n$

$B = \left\{ \begin{pmatrix} 3 \\ 1 \end{pmatrix}, \begin{pmatrix} 1 \\ 2 \end{pmatrix} \right\}$ define una base de $\mathbb{R}^2$

Interpretación gráfica (Mathematics for Machine Learning)

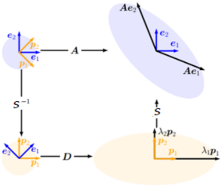
Sea $A \in \mathbb{K}^{n\times n}$: Área: $det(A) = \prod_{i=1}^{n} \lambda_i$ y Perímetro: $Tr(A) = \sum_{i=1}^{n} \lambda_i$



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

$\mathcal{B} = \{x_1, \ldots, x_n\}$

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:



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

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

donde D es una matriz *diagonal*.

$[A]_{\mathcal{B}} = D$

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

$(\lambda_1, x_1), \ldots, (\lambda_n, x_n) \rightarrow S = \begin{pmatrix} | & & | \\ x_1 & \cdots & x_n \\ | & & | \end{pmatrix}, \; D = \begin{pmatrix} \lambda_1 & & 0 \\ & \ddots & \\ 0 & & \lambda_n \end{pmatrix}$

$\downarrow S^{-1}$

$y = A \cdot x$ $\qquad x \downarrow S^{-1}$

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

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

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

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

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

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.