Análisis Matemático para Inteligencia Artificial

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

Especialización en Inteligencia Artificial

05/05/2023

Sea $\mathbb{V} - \mathbb{K}$ e.v., donde $\mathbb{K} = \mathbb{R}$ *o* $\mathbb{C}$, un producto interno sobre $\mathbb{V}$ es una función $\Phi : \mathbb{V} \times \mathbb{V} \to \mathbb{R}$ (*o* $\mathbb{C}$) que satisface:

$z = \underbrace{a}_{\mathbb{R}e} + i \cdot \underbrace{b}_{\mathbb{I}m}$

$\bar{z} = a + i(-b)$

$\nabla_i\, z \in \mathbb{R} \Rightarrow z = \bar{z}$

1. Para cada $\alpha \in \mathbb{R}$ (*o* $\mathbb{C}$), $u, v, w \in \mathbb{V}$.
   - $\Phi(u + v, w) = \Phi(u, w) + \Phi(v, w)$
   - $\Phi(\alpha \bullet u, v) = \alpha \bullet \Phi(u, v)$] lineal
2. $\Phi(u, v) = \overline{\Phi(v, u)}$ → – $\mathbb{R}$ → bilineal
3. $\Phi(v, v) \geq 0$, y $\Phi(v, v) = 0$ sii $v = 0_v$

Notación: $\Phi(u, v) = \langle u, v \rangle$

Definición: A un espacio vectorial real (complejo) provisto de un producto interno se lo llama espacio euclídeo (espacio unitario).

Obs: El p.i. es una generalización del producto escalar en $\mathbb{R}^n$ (o $\mathbb{C}^n$).

$\langle (1,2,3), (3,2,2) \rangle = 1 \cdot 3 + 2 \cdot 2 + 3 \cdot 2 = 3 + 4 + 6 = 13$

Sea $\mathcal{V}$ el espacio de las funciones continuas de valor real (o complejo) en el intervalo $-1 \leq x \leq 1$ (se nota $\mathcal{C}([-1,1])$) con p.i.

$$\langle f, g\rangle = \int_{-1}^{1} f(x)\overline{g(x)}dx$$

Verificar que cumple:

1. Para cada $\alpha \in \mathbb{R}$ (o $\mathbb{C}$), $f, g, h \in \mathcal{C}([-1,1])$.
   - $\Phi(f+g, h) = \Phi(f,h) + \Phi(g,h)$
   - $\Phi(\alpha \bullet f, g) = \alpha \bullet \Phi(f,g)$
2. $\Phi(f,g) = \overline{\Phi(g,f)}$
3. $\Phi(f,f) \geq 0$, y $\Phi(f,f) = 0$ sii $f(x) = 0, \forall x$

1) a) $<f+g, h> = \int_{-1}^{1} [f(x)+g(x)] \cdot \overline{h(x)}\, dx = \int_{-1}^{1} f(x)\cdot\overline{h(x)} + g(x)\overline{h(x)}\, dx = \int_{-1}^{1} f(x)\overline{h(x)}\, dx + \int_{-1}^{1} g(x)\overline{h(x)}\, dx = <f,h> + <g,h>$ ✓

2) $<f,g> = \int_{-1}^{1} f(x)\cdot\overline{g(x)}\, dx = \overline{\int_{-1}^{1} \overline{f(x)\cdot\overline{g(x)}}\, dx} = \overline{\int_{-1}^{1} \overline{f(x)}\cdot g(x)\, dx} = \overline{\int_{-1}^{1} g(x)\cdot\overline{f(x)}\, dx} = \overline{<g,f>}$ ✓

3) $<f,f> = \int_{-1}^{1} f(x)\cdot\overline{f(x)}\, dx \overset{③}{=} \int_{-1}^{1} \underbrace{\|f(x)\|^2}_{\geq 0\ \forall x} dx \geq 0$

$z = a+bi \Rightarrow z\cdot\overline{z} = (a+i(b))\cdot(a+i(-b)) = a^2+b^2+i\cancel{(-ab+ab)} = a^2+b^2 = \|z\|^2 \geq 0$ ③

② $\|z\|^2 = 0 \Leftrightarrow a,b=0 \Leftrightarrow z=0$

y $= 0$ $\Leftrightarrow$ ② $f(x) = 0$ $\forall x$

Definición de Norma

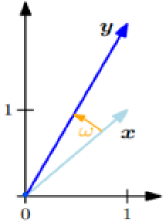
Sea $(\mathbb{V}, \langle ., . \rangle)$ un e.v. real (complejo) con p.i.. Sea $v \in \mathbb{V}$, se define la norma de v asociada a $\langle ., . \rangle$.

Notación: $||v|| = \langle v, v \rangle^{1/2}$

Es la generalización de la longitud de un vector en $\mathbb{R}^n$ (o $\mathbb{C}^n$).

Def: A partir de un p.i. se puede se puede definir el ángulo w entre dos vectores x, y

$$\cos(\omega) = \frac{\langle x, y \rangle}{||x|| \, ||y||}$$



$x-y$

$\|x-y\|^2 = \|x\|^2 + \|y\|^2 - 2\cdot\|x\|\|y\|\cdot\cos(w)$

$\langle x-y, x-y\rangle = \langle x,x\rangle + \langle y,y\rangle - 2\cdot\|x\|\|y\|\cos(w)$

$\langle x,x\rangle + \langle y,y\rangle - 2\langle x,y\rangle \Rightarrow \langle x,y\rangle = \|x\|\cdot\|y\|\cdot\cos(w)$

$\Leftrightarrow \cos(w) = \frac{\langle x,y\rangle}{\|x\|\|y\|}$

ley de cos

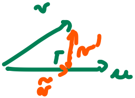
$a^2 = b^2 + c^2 - 2\cdot b\cdot c\cdot\cos(A)$

Propiedades de la Norma

1. $\forall v \in \mathbb{V},\ ||v|| \geq 0$, y $||v|| = 0$ sii $v = 0$. ← linealidad del p.i. (3)
2. Sean $\alpha \in \mathbb{R}(o\ \mathbb{C})$, $v \in \mathbb{V}$, $||\alpha \bullet v|| = |\alpha|\ ||v||$. ← idea $\langle \alpha v, \alpha v\rangle = \alpha\bar{\alpha}\langle v, v\rangle$; $\sqrt{\ }$; $|\alpha|\cdot||v||$; $\sqrt{\ }$ x def
3. Desigualdad de Cauchy Schwartz: si $u, v \in \mathbb{V}$ entonces

$$|\langle u, v\rangle| \leq ||u||\ ||v||$$

idea



$v = \tilde{v} + v'$

$||v||^2 = ||\tilde{v}||^2 + ||v'||^2$

$\langle u, \tilde{v} + v'\rangle =$

$\langle u, \tilde{v}\rangle + \cancel{\langle u, v'\rangle}$ $\quad u \perp v'$

$= ||u||\cdot||\tilde{v}|| \leq ||u||\cdot||v||$

pero $||\tilde{v}|| \leq ||v||$

$\cos(\omega) = \dfrac{\langle u, v\rangle}{||u||\,||v||}$

4. Desigualdad Triangular: si $u, v \in \mathbb{V}$ entonces

$$||u + v|| \leq ||u|| + ||v||$$

idea $||u+v|| = ||u|| + ||v||$ sólo si u y v no colineales

Ortogonalidad

$\langle u, v_1\rangle > 0$
$\langle u, v_2\rangle = 0$
$\langle u, v_3\rangle < 0$

Def: $(\mathbb{V}, \langle ., . \rangle)$ un $\mathbb{K}$-EV (con $\mathbb{K} = \mathbb{R}$ *o* $\mathbb{C}$) con p.i. dos vectores $u, v \in \mathbb{V}$ se dicen ortogonales si $\langle u, v\rangle = 0$.

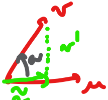
$\langle u, v\rangle = \|u\| \|v\| \cos(\omega)$

Teorema de Pitágoras: Si $u, v \in \mathbb{V}$ son ortogonales entonces $\|u + v\|^2 = \|u\|^2 + \|v\|^2$.

Def: $(\mathbb{V}, \langle ., . \rangle)$ un $\mathbb{K}$-EV (con $\mathbb{K} = \mathbb{R}$ *o* $\mathbb{C}$) con p.i.. Se dice que $\{v_1, ..., v_r\} \subset \mathbb{V}$ es un conjunto ortogonal si $\langle v_i, v_j\rangle = 0, \ \forall i \neq j$. Si $\|v_i\| = 1, \ \forall i$ se dice que es un conjunto ortonormal.

La proyección ortogonal del vector v sobre el vector u es otro vector que notamos como $P_u(v)$, y se define:

$\cancel{\|u\|}\,\|v\|.\cos(\omega)$



$|\tilde{v}| = \|v\|.\cos(\omega)$

$$\tilde{v} = P_u(v) = \underbrace{\frac{\langle u, v\rangle}{\cancel{\|u\|}}}_{\text{norma}^*} \underbrace{\frac{\overbrace{u}^{\text{dir+sent}}}{\|u\|}}_{\check{u}} = \frac{\langle u, v\rangle}{\langle u, u\rangle} u$$

si $\|u\| = 1$

$$P_u(v) = \underbrace{\langle u, v\rangle}_{\text{norma}^*} \cdot \overbrace{u}^{\text{dir+sent}}$$

$\leq \|v\|$

$\oplus$ ó "sent"

Proceso de Ortogonalización de Gram Schmidt

Def: Una base ortonormal (BON) de un E.V. es una base $B = \{v_1, ..., v_n\}$ que satisface:

$$\langle v_i, v_j \rangle = 0, \ \forall i \neq j$$
$$\langle v_i, v_i \rangle = 1, \ \forall i$$

Obs: Si sólo se cumple que $\langle v_i, v_j \rangle = 0, \ \forall i \neq j$ se dice que es una base ortogonal.

Para transformar una base en una base ortonormal usamos el proceso de Gram-Schmidt:

$$k_1 = v_1$$
$$k_2 = v_2 - Proy_{k_1}(v_2)$$
$$\vdots$$
$$k_n = v_n - \sum_{i=1}^{n-1} Proy_{k_i}(v_n)$$

$\check{k}_1 = \frac{k_1}{||k_1||}$

$w' = w - \tilde{w} = w - P_{k_1}(w)$

$k_3 \leftarrow w' = w - \tilde{w}_1 - \tilde{w}_2 = w - P_{k_1}(w) - P_{k_2}(w)$

Y así, $\tilde{B} = \{\check{k}_1, ..., \check{k}_n\}$ pidiendo que $||k_i|| = 1$ resulta una BON.

Sea $\mathbb{V}$ un EV de dimensión $n < \infty$ y $S \subset \mathbb{V}$ un SEV de dimensión $m \leq n$. El complemento ortogonal $(S^{\perp})$ es un SEV de de dimensión $n - m$ que satisface:

$$S \cup S^{\perp} = \mathbb{V} \quad \wedge \quad S \cap S^{\perp} = \emptyset$$

Ejemplo:

$S = \left\langle \overset{u}{\begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}}, \overset{v}{\begin{pmatrix} 1 \\ 2 \\ 0 \end{pmatrix}} \right\rangle$ con prod. escalar.

$\rightarrow$ busco w / $\langle u, w \rangle = 0$, $\langle w, v \rangle = 0$

opc 1) como EV $= \mathbb{R}^3$ puedo usar $u \times v$ $\Rightarrow w = \begin{vmatrix} \check{i} & \check{j} & \check{k} \\ 1 & 1 & 1 \\ 1 & 2 & 0 \end{vmatrix} = (-2, 1, 1)$

opc 2) busco un w^* potencial LI de $S \Leftrightarrow \notin S$

$$w = w^* - P_{k_1}(w^*) - P_{k_2}(w^*) \quad \text{con } k_1 = u, \quad k_2 = v - P_u(v)$$

Distancia

Sea $\mathbb{V}$-$\mathbb{K}, (\mathbb{R}$ *o* $\mathbb{C})$ EV con p.i. $\langle ., . \rangle$ se define la distancia $d : \mathbb{V} \times \mathbb{V} \to \mathsf{K}$ como $d(u, v) = ||u - v||.$ $= <u-v, u-v>^{1/2}$

Propiedades:

1. $d(u, v) \geq 0, \ \forall u, v \in \mathbb{V}$
2. $d(u, v) = 0 \Leftrightarrow u = v$
3. $d(u, v) = d(v, u), \ \forall u, v \in \mathbb{V}$
4. $d(u, v) \leq d(u, w) + d(w, v), \ \forall u, v, w \in \mathbb{V}$

Observación: Existen distancias que no están asociadas a ninguna norma.

Matrices definidas positivas

Una matriz $A \in \mathbb{R}^{n \times n}$ se dice definida positiva si es simétrica y vale que:

$$x^T A x > 0, \ \forall x \in \mathbb{R}^n - \{0\}$$

Si vale que $x^T A x \geq 0$ se la llama semi definida positiva.

Ejemplo: $A = \begin{pmatrix} 2 & 3 \\ 3 & 9 \end{pmatrix} \in \mathbb{R}^{2\times 2} \rightarrow A^T = \begin{pmatrix} 2 & 3 \\ 3 & 9 \end{pmatrix} \rightarrow$ simétrica ✓

$x = \begin{pmatrix} a \\ b \end{pmatrix} \in \mathbb{R}^2$ ✓

$$\underbrace{\underbrace{\underbrace{(a \ b)}_{1\times n} \underbrace{\begin{pmatrix} 2 & 3 \\ 3 & 9 \end{pmatrix}}_{n\times n}}_{1\times 1} \underbrace{\begin{pmatrix} a \\ b \end{pmatrix}}_{n\times 1}} = \underbrace{(2a+3b, \ 3a+9b)}_{1\times 2} \underbrace{\begin{pmatrix} a \\ b \end{pmatrix}}_{2\times 1} = 2a^2 + 3ab + 3ab + 9b^2 = 2a^2 + 6ab + 9b^2$$

$$= 2a^2 + 2(3b)\cdot a + (3b)^2 = a^2 + [a^2 + 2a3b + (3b)^2]$$

$$= \underbrace{a^2}_{\geq 0} + \underbrace{(a+3b)^2}_{\geq 0} \geq 0 \quad \text{y si } a \text{ ó } b \neq 0 \Rightarrow > 0$$

Teorema: Sea $\mathbb{V}$ un EV de dimensión finita, y B una base de $\mathbb{V}$, vale que es un p.i. sii existe una matriz definida positiva tal que:

$$\langle x, y \rangle = \tilde{x}^T A \tilde{y} \quad > 0 \quad \forall x, y \neq 0_v$$

donde $\tilde{x}, \ \tilde{y}$, son las representaciones de $x, \ y$ en la base B.

Transformaciones

Sea $T : \mathbb{V} \to \mathbb{W}$ una transformación, donde $\mathbb{V}, \ \mathbb{W}$ son dos conjuntos arbitrarios. Se dice que T es:

- Inyectiva: si $\forall x, \ y \in \mathbb{V} : T(x) = T(y) \to x = y$
- Suryectiva: si $T(\mathbb{V}) = \mathbb{W}$
- Biyectiva: si es inyectiva y suryectiva.

Transformaciones Lineales

Sean $\mathbb{V}, \ \mathbb{W}$ dos EV, $L : \mathbb{V} \to \mathbb{W}$ es una transformación lineal si:

$$L(\alpha x + \beta y) = \alpha L(x) + \beta L(y) \ \ \forall \alpha, \beta \in \mathbb{K}, \ \forall x, y \in \mathbb{V}$$

- Isomorfismo: $L : \mathbb{V} \to \mathbb{W}$ es lineal y biyectiva.
- Endomorfismo: si $L : \mathbb{V} \to \mathbb{V}$ es lineal.
- Automorfismo: $L : \mathbb{V} \to \mathbb{V}$ es lineal y biyectiva.

Teorema: Sea $\mathbb{V}$ y $\mathbb{W}$, dos espacios vectoriales de dimensión finita son un isomorfismo sii $dim(\mathbb{V}) = dim(\mathbb{W})$.

Teorema: Sea $\mathbb{V}$ un EV, $dim(\mathbb{V}) = n < \infty$ tiene un isomorfismo con $\mathbb{R}^n$. Si consideramos la base $B = \{v_1, ..., v_n\}$, todo $v \in \mathbb{V}$ puede escribirse como $v = \alpha_1 v_1 + ... + \alpha_n v_n$. Luego las coordenadas de v en la base B resulta:

$$\alpha = (\alpha_1, ..., \alpha_n)^T \in \mathbb{R}^n$$

$\mathbb{V} = \mathcal{P}_2 \longrightarrow \mathbb{R}^3$

$dim(\mathbb{V}) = 3$

$B_{\mathbb{V}} = \left\{ \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ x \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 0 \\ x^2 \end{pmatrix} \right\}$ $\rightarrow$ $v = 3 + 5x^2 - 10x^2 \in \mathcal{P}_2$

$[v]_B = (3, 5, -10) \in \mathbb{R}^3$

Núcleo e Imagen de una transformación

Sea $L : \mathbb{V} \to \mathbb{W}$, se define:

$$\underbrace{A}_{m\times n} \cdot \underbrace{\overset{\in \mathbb{R}^n}{v}}_{\substack{n\times 1 \\ in}} = \underbrace{\overset{\in \mathbb{R}^m}{b}}_{\substack{m\times 1 \\ out}}$$

- Núcleo (o Kernel) $Nu(L) = \{v \in \mathbb{V} : L(v) = 0_W\}$,
- Imagen $Im(L) = L(\mathbb{V}) = \{w \in \mathbb{W} : \exists v \in \mathbb{V},\ w = L(v)\}$

Teorema: **Toda transformación lineal se puede representar de forma matricial.** $L(v) = A \cdot v$

si $v_1, v_2 \,/\, Av_1 = Av_2 = 0_v \Rightarrow A(\alpha v_1 + \beta v_2) = 0_w$

- Espacio Nulo de A: es un subespacio de $\mathbb{R}^n$ formado por todas las soluciones del sistema lineal homogéneo $Av = 0$, $A \in \mathbb{R}^{m\times n}$.

$$N(A) = \{v \in \mathbb{R}^{n\times 1} : Av = 0\}$$

- Espacio columna de A: es el subespacio de $\mathbb{R}^m$ generado por los n vectores columna de A: $A = (A_1\ A_2 \dots A_n)$ $EC(A) = \langle A_1, A_2, \dots, A_n \rangle$

$$EC(A) = \{\alpha_1(a_{11}, ..., a_{m1})^T + ... + \alpha_m(a_{1n}, ..., a_{mn})^T, \alpha_i \in \mathbb{R}\}$$

- Espacio fila de A: e el subespacio de R^n generado por los m vectores fila de A: $EF(A) = \langle -A_1-^T, -A_2-^T, \dots, -A_m^T- \rangle$

$$EF(A) = \{\alpha_1(a_{11}, ..., a_{1n}) + ... + \alpha_m(a_{m1}, ..., a_{mn}), \alpha_i \in \mathbb{R}\}$$

Veamos un ejemplo...

$L:\mathbb{R}^2\to\mathbb{R}^2 \qquad L\begin{pmatrix}x\\y\end{pmatrix}=\begin{pmatrix}x+y\\x-y\end{pmatrix}$

- ¿Es TL? $L(\alpha v_1+\beta v_2) = L\left(\alpha\begin{pmatrix}x_1\\y_1\end{pmatrix}+\beta\begin{pmatrix}x_2\\y_2\end{pmatrix}\right) = L\begin{pmatrix}\alpha x_1+\beta x_2\\ \alpha y_1+\beta y_2\end{pmatrix} = \begin{pmatrix}\alpha x_1+\beta x_2+\alpha y_1+\alpha y_2\\ \alpha x_1+\beta x_2-\alpha y_1-\beta y_2\end{pmatrix}$

$= \begin{pmatrix}\alpha(x_1+y_1)+\beta(x_2+y_2)\\ \alpha(x_1-y_1)+\beta(x_2-y_2)\end{pmatrix} = \alpha\begin{pmatrix}x_1+y_1\\x_1-y_1\end{pmatrix}+\beta\begin{pmatrix}x_2+y_2\\x_2-y_2\end{pmatrix} = \alpha L(v_1)+\beta L(v_2)$ ✓

- $Nu(L)=\left\{v\in\mathbb{R}^2 \,/\, L(v)=\begin{pmatrix}0\\0\end{pmatrix}\right\}$ $\quad\begin{cases}x+y=0\\x-y=0\end{cases} \to x=y,\ 2x=0 \Rightarrow Nu(L)=\left\{\begin{pmatrix}0\\0\end{pmatrix}\right\}$

$\Rightarrow L$ es inyectiva

- $Im(L)$: $L\begin{pmatrix}1\\0\end{pmatrix}=\begin{pmatrix}1\\1\end{pmatrix}, \quad L\begin{pmatrix}0\\1\end{pmatrix}=\begin{pmatrix}1\\-1\end{pmatrix}$

$\underbrace{Im(L)}_{Im(L)} \underset{=}{=} \left\langle\begin{pmatrix}1\\1\end{pmatrix},\begin{pmatrix}1\\-1\end{pmatrix}\right\rangle = \underbrace{\mathbb{R}^2}_{V} \to L$ es suryectiva

$\Downarrow$

L es isomorfismo
automorfismo

Seguimos con el ejemplo...

$A = \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix}$ $\quad N(A) = \{v / A \cdot v = 0\} = \left\{\begin{pmatrix} 0 \\ 0 \end{pmatrix}\right\}$ $\quad \dim(N(A)) = 0$

$\begin{array}{cc} \uparrow & \uparrow \\ L(e_1) & L(e_2) \end{array}$

- $EC(A) = \left\langle \begin{pmatrix} 1 \\ 1 \end{pmatrix}, \begin{pmatrix} 1 \\ -1 \end{pmatrix} \right\rangle = \mathbb{R}^2$ $\quad \dim(EC(A)) = \dim(EF(A)) = 2$

- $EF(A) = \left\langle \begin{pmatrix} 1 \\ 1 \end{pmatrix}, \begin{pmatrix} 1 \\ -1 \end{pmatrix} \right\rangle = \mathbb{R}^2$

Conclusiones ...

Teorema: Sea $A \in \mathbb{R}^{m\times n}$, $dim(EC(A)) = dim(EF(A)) = r(A)$. Donde $r(A)$ se denomina rango de la matriz.

Definición: Se denomina nulidad de una matriz A a la dimensión de su espacio nulo $N(A)$, $n(A) = dim(N(A))$, siendo $N(A) = \{v \in \mathbb{R}^n, Av = 0\}$

Teorema de Rango-Nulidad: Para toda matriz $A \in \mathbb{R}^{m\times n}$ se verifica:

$$\underbrace{r(A)}_{\#\text{vectores mapeables}} + \underbrace{n(A)}_{\#\text{vectores ``perdidos''}} = n$$

$$\dim(V) = \dim(\mathrm{Im}(L)) + \dim(\mathrm{Nu}(L))$$

si L es lineal