Análisis Matemático para Inteligencia Artificial

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

Especialización en Inteligencia Artificial

10/03/2023

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:

1. Para cada $\alpha \in \mathbb{R}$ (*o* $\mathbb{C}$), $u, v, w \in \mathbb{V}$.
   - $\Phi(u + v, w) = \Phi(u, w) + \Phi(v, w)$
   - $\Phi(\alpha \bullet u, v) = \alpha \bullet \Phi(u, v)$

   en $\mathbb{V}$ → prod. en $\mathbb{R}$ (o $\mathbb{C}$)

   linealidad
2. $\Phi(u, v) = \Phi(v, u)$ → simetría
3. $\Phi(v, v) \geq 0$, y $\Phi(v, v) = 0$ sii $v = 0$

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 $\mathbb{C}^n$).

$\langle \underbrace{(1,2,3)}_{\vec{u}}, (2,3,5) \rangle = 1.2 + 2.3 + 3.5 = 23$

$\langle \vec{u}, \vec{v} \rangle = \sum u_i \cdot v_i$

También hay otros espacios con productos internos ...

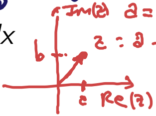
Sea $\mathcal{V}$ el espacio de las funciones continuas de valor real (o complejo) en el intervalo $-1 \leq x \leq 1$ (se nota $\mathcal{C}([-1,1])$) con p.i.

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

conjugado

$\boxed{z = a + bi}$

$a = Re(z)$ $\quad b = Im(z)$



$z = a + bi$

i es el nro imag.

$x^2 + 1 = 0$

$i^2 = -1$ $\quad x = i$

Verificar que cumple:

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

- $\bar{z}$ se lee z conjugado

$\bar{z} = a - bi$

prop $\bar{\bar{z}} = \overline{a - bi} = a + bi = z$

$z \cdot \bar{z} = (a+bi)(a-bi) = a^2 - \cancel{abi} + \cancel{abi} - b^2 i^2 = a^2 + b^2 = |z|^2$ (*)

$|z| = \sqrt{a^2 + b^2}$

$\overline{u \cdot v} = \bar{u} \cdot \bar{v}$

(1) $\langle f+g, h\rangle = \int_{-1}^{1} (f+g)(x)\overline{h(x)}dx = \int_{-1}^{1} [f(x)+g(x)] \cdot \overline{h(x)}dx$

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

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

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

(3) $\langle f, f\rangle = \int_{-1}^{1} f^2(x)\,dx \geq 0$ (*)

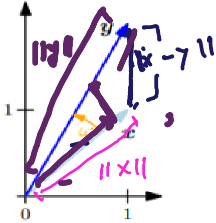
Sea $(\mathbb{V}, \langle .,. \rangle)$ un e.v. real (complejo) con p.i.. Sea $v \in \mathbb{V}$, se define la norma de v asociada a $\langle .,. \rangle$.

Notación: $||v|| = \langle v, v \rangle^{1/2}$ ≥ 0 salvo $v = 0$ $||v|| = 0$

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

$$\|x\|\,\|y\| \cos(\omega) = \frac{\langle x, y \rangle}{\|x\|\,\|y\|} \quad \neq 0 \qquad (*)$$



Ley de cosenos

dem $||x-y||^2 \overset{(2)\,?}{=} ||x||^2 + ||y||^2 - 2||x||\,||y|| \cos w \leftarrow (A)$

$(\langle x-y, x-y \rangle^{1/2})^2 \overset{(1)}{=} \langle x, x-y \rangle + \langle -y, x-y \rangle \overset{(1)}{=}$

$\langle x,x \rangle + \langle x, -y \rangle + \langle -y, x \rangle + \langle -y, -y \rangle = ||x||^2 + ||y||^2 - 2\langle x, y \rangle \leftarrow (B)$

(A) tiene que ser = (B) (*)

Ley de cosenos

$a^2 = b^2 + c^2 - 2\, b.c \cos \alpha$

α es áng. opuesto a a

Propiedades de la Norma

1. $\forall v \in \mathbb{V}, \ ||v|| \geq 0$, y $||v|| = 0$ sii $v = 0$.
2. Sean $\alpha \in \mathbb{R}(o\ \mathbb{C})$, $v \in \mathbb{V}$, $||\alpha \bullet v|| = |\alpha|\,||v||$.
3. Desigualdad de Cauchy Schwartz: si $u, v \in \mathbb{V}$ entonces

$$\boxed{|\underbrace{\langle u, v\rangle}_{\in \mathbb{R} \text{ ó } \mathbb{C}}| \leq ||u||\ ||v||}$$

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

Para la dem $||u+v||^2 = \langle u+v, u+v\rangle$

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

Truco para la demostración de Cauchy-Schwartz

Se define $\boxed{w = u - \frac{\langle u, v\rangle}{||v||^2} \cdot v}$ $v \neq 0$ Luego $||w||^2 = \left\langle u - \frac{\langle u,v\rangle}{||v||^2} \cdot v,\ u - \frac{\langle u,v\rangle}{||v||^2} v \right\rangle$

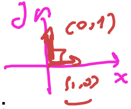
completar con las prop de linealidad...

$$\underbrace{||u||^2}_{\geq 0} - 2\frac{\langle u, v\rangle}{||v||^2}\langle u, v\rangle + \left(\frac{\langle u, v\rangle}{||v||^2}\right)^2 ||v||^2 \geq 0$$

Ortogonalidad

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 $\langle u, v \rangle = 0$.

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



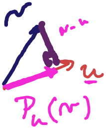
Def: $(\mathbb{V}, \langle ., . \rangle)$ un $\mathbb{K}$-EV (con $\mathbb{K} = \mathbb{R}$ *o* $\mathbb{C}$) con p.i.. Se dice que $\{v_1, ..., v_r\} \subset \mathbb{V}$ es un conjunto ortogonal si $\langle v_i, v_j \rangle = 0, \ \forall i \neq j$. Si $||v_i|| = 1, \ \forall i$ se dice que es un conjunto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

$||(1,0)|| = \sqrt{1^2 + 0^2} = 1$
$= \sqrt{\langle (1,0), (1,0) \rangle}$

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

$$P_u(v) = \frac{\langle u, v \rangle}{||u||} \frac{u}{||u||} = \boxed{\frac{\langle u, v \rangle}{\langle u, u \rangle}} u$$

$\in \mathbb{K}$



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:

v defino $\frac{v}{||v||} = u$

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

$||u|| = \left\| \frac{v}{||v||} \right\| = \frac{||v||}{||v||} = 1$

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:

$B = \left\{ \begin{pmatrix}1\\1\end{pmatrix}, \begin{pmatrix}1\\2\end{pmatrix} \right\}$ $V = \mathbb{R}^2$ p.i. canónico

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

G-S: $k_1 = \begin{pmatrix}1\\1\end{pmatrix}$

$k_2 = \begin{pmatrix}1\\2\end{pmatrix} - P_{k_1}\left(\begin{pmatrix}1\\2\end{pmatrix}\right) = \underbrace{\begin{pmatrix}1\\2\end{pmatrix}}_{v_2} - \frac{\left\langle \begin{pmatrix}1\\2\end{pmatrix}, \begin{pmatrix}1\\1\end{pmatrix} \right\rangle}{\underbrace{\left\| \begin{pmatrix}1\\1\end{pmatrix} \right\|^2}_{= \langle k_1, k_1 \rangle}} \begin{pmatrix}1\\1\end{pmatrix}$ $v_1 = k_1$

Y así, $\tilde{B} = \{k_1, ..., k_n\}$ pidiendo que $||k_i|| = 1$ resulta una BON.

$= \begin{pmatrix}1\\2\end{pmatrix} - \frac{3}{2}\begin{pmatrix}1\\1\end{pmatrix} = \begin{pmatrix}-1/2\\1/2\end{pmatrix}$

$\tilde{B} = \left\{ \begin{pmatrix}1\\1\end{pmatrix}, \begin{pmatrix}-1/2\\1/2\end{pmatrix} \right\}$ verif $\left\langle \begin{pmatrix}1\\1\end{pmatrix}, \begin{pmatrix}-1/2\\1/2\end{pmatrix} \right\rangle = -\frac{1}{2} + \frac{1}{2} = 0$ ✓

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:

$S = \text{gen}\left\{ \overset{u}{\begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}}, \overset{v}{\begin{pmatrix} 1 \\ 2 \\ 0 \end{pmatrix}} \right\}$ con p.i. canónico

$\mathbb{V} = \mathbb{R}^3$
$S = \mathbb{R}^2$
$S^{\perp} = \mathbb{R}$

para buscar $S^{\perp}$ hay q buscar un vector q sea ortogonal a ambos vectores de S

Op. 1 (en $\mathbb{R}^3$)

$$u \times v = \begin{vmatrix} \hat{i} & \hat{j} & \hat{k} \\ 1 & 1 & 1 \\ 1 & 2 & 0 \end{vmatrix} =$$

Op 2 ($\forall \mathbb{V}$) G. S. $\quad k_3 = v_3 - \left(P_{k_1}(v_3) + P_{k_2}(v_3)\right)$

Para eso $k_1 = u \quad k_2 = k_1 - P_{k_1}(v)$ invento v_3 cualq $(\neq u, \neq v)$

por ej. $\begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}$

Distancia

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

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.

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

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

$a_{ij} = a_{ji} \ \forall i,j \ i \neq j$

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

Ejemplo: $A = \begin{bmatrix} 2 & 3 \\ 3 & 9 \end{bmatrix}$ es simetrica $a_{12} = a_{21}$ ó lo q' es lo mismo $A = A^T$

$$\underbrace{(x, y)}_{1\times 2} \underbrace{\begin{bmatrix} 2 & 3 \\ 3 & 9 \end{bmatrix}}_{2\times 2} \underbrace{\begin{pmatrix} x \\ y \end{pmatrix}}_{2\times 1} = (2x + 3y, \ 3x + 9y) \begin{pmatrix} x \\ y \end{pmatrix} = 2x^2 + \overbrace{3xy + 3xy}^{2.3xy} + \underbrace{9y^2}_{(3y)^2} =$$

$$= x^2 + \underbrace{x^2 + 2x(3y) + (3y)^2}_{} = \underbrace{x^2}_{\geq 0} + \underbrace{(x + 3y)^2}_{>0} > 0$$

def. positiva

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:

$$\langle x, y \rangle = \tilde{x}^T A \tilde{y} \qquad x^T A x > 0$$

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

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.

Representaciones

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:

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

Núcleo e Imagen de una transformación

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

- 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.**

- 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:

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

- Espacio fila de A: e el subespacio de R^n generado por los m vectores fila de A:

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

Veamos un ejemplo...

Seguimos con el ejemplo...

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:

$$r(A) + n(A) = n$$