$V = (\mathbb{V}, +, \mathbb{K}, \cdot)$ $\leftarrow \Phi$

Análisis Matemático para Inteligencia Artificial

Martín Errázquin (merrazquin@fi.uba.ar)

Especialización en Inteligencia Artificial

Clase 2

$0 = 0 + i \cdot 0$

$z = a + ib \quad a, b \in \mathbb{R} \qquad i^2 = -1$

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

$z \cdot \bar{z} = (a+ib)(a+i(-b)) = a^2 - (ib)^2 = a^2 - i^2 b^2$
$= a^2 + b^2 > 0 \quad \forall z \neq 0$

$+ : \mathbb{V} \times \mathbb{V} \to \mathbb{V}$

$\bullet : \mathbb{K} \times \mathbb{V} \to \mathbb{V}$

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:

$z = a + ib \in \mathbb{C}$

$a, b \in \mathbb{R}$

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

   $\bar{z} = a + i(-b)$
2. $\Phi(u, v) = \overline{\Phi(v, u)}$ $\Leftrightarrow$ @ $\mathbb{R}$ $\quad \Phi(u,v) = \Phi(v,u)$
3. $\Phi(v, v) \geq 0$, y $\Phi(v, v) = 0$ sii $v = 0$ $\to$ si $v = 0_{\mathbb{V}} \Rightarrow \Phi(v,v) = 0$

   si $v \neq 0_{\mathbb{V}} \Rightarrow \Phi(v,v) > 0$

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

$\langle u, v \rangle$

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 $\mathbb{C}^n$).

$$\begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix} \bullet \begin{pmatrix} -1 \\ -2 \\ -3 \end{pmatrix} = 1 \cdot (-1) + 1 \cdot (-2) + 1 \cdot (-3) = -6 \to A = I, \quad B = E \quad (1\ 1\ 1) \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} -1 \\ -2 \\ -3 \end{pmatrix}$$

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. $[-1,1] \to \mathbb{C}$

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

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

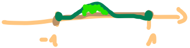
$= \langle f,h\rangle + \langle g,h\rangle$

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

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

3) $\langle f,f\rangle = \int_{-1}^{1} f(x)\overline{f(x)}\,dx = \int_{-1}^{1} \underbrace{\|f(x)\|^2}_{\geq 0\ \forall x} dx = \begin{cases} 0 & \text{si } f(x)\equiv 0 \quad (f(x)=0\ \forall x) \\ >0 & \text{si } f(x)\not\equiv 0 \end{cases}$

$z \in \mathbb{C}$ $\quad z\cdot\bar{z} = \|z\|^2 \begin{cases} >0 & \text{si } z\neq 0 \\ =0 & \text{si } z=0 \end{cases}$



$\langle v, v\rangle \in \mathbb{R}_0^+$ PI $\hookrightarrow$ Norma

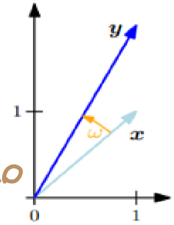
Sea $(\mathbb{V}, \langle ., .\rangle)$ un e.v. real (complejo) con p.i.. Sea $v \in \mathbb{V}$, se define la norma de v asociada a $\langle ., .\rangle$.

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

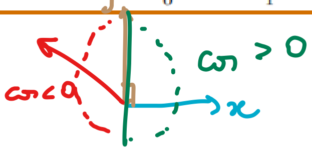
$$\cos(\omega) = \frac{\langle x, y\rangle}{||x||\ ||y||} \in [-1, +1]$$



$\cos = 0$

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

$$= \frac{\langle x, y\rangle}{\sqrt{\langle x, x\rangle \cdot \langle y, y\rangle}}$$



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

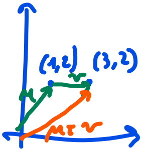
$$0 \leq |\langle u, v\rangle| \leq \underline{||u||\ ||v||}$$

$$\langle u, v\rangle = ||u|| \cdot ||v|| \underbrace{\cos(w)}_{[-1,1]}$$

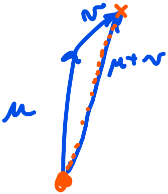
$$|\langle u, v\rangle| = ||u|| \cdot ||v|| \cdot \underbrace{|\cos(w)|}_{[0,1]}$$

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

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



$u = (1,2)$
$v = (2,0)$
$u + v = (3,2)$



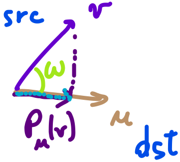
Def: $(\mathbb{V}, \langle ., . \rangle)$ un $\mathbb{K}$-EV (con $\mathbb{K} = \mathbb{R}$ *o* $\mathbb{C}$) con p.i. dos vectores $u, v \in \mathbb{V}$ se dicen ortogonales si $\langle u, v \rangle = 0$.

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

G ortog. $M = \begin{pmatrix} \ddots & 0 \\ 0 & \ddots \end{pmatrix}$ G orton. $M = I = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$

Def: $(\mathbb{V}, \langle ., . \rangle)$ un $\mathbb{K}$-EV (con $\mathbb{K} = \mathbb{R}$ *o* $\mathbb{C}$) con p.i.. Se dice que $G = \{v_1, ..., v_r\} \subset \mathbb{V}$ 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:



$$P_u(v) = \frac{\langle u, v \rangle}{||u||} \overbrace{\frac{u}{||u||}}^{\check{u}} = \frac{\langle u, v \rangle}{\langle u, u \rangle} u$$

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

Proceso de Ortogonalización de Gram Schmidt

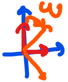
Def: Una base ortonormal (BON) de un E.V. es una base $B = \{v_1, ..., v_n\}$ que satisface:

B: • generador de $\mathbb{V}$ • LI
BOG: • ortogonal
BON: • ortonormal

$\mathbb{R}^2$: $\left\{ \begin{pmatrix} 1 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \end{pmatrix} \right\}$

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

con $\omega \in (0, 2\pi)$



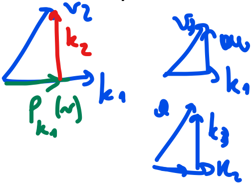
Obs: Si sólo se cumple que $\langle v_i, v_j \rangle = 0, \; \forall i \neq j$ se dice que es una base ortogonal.

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

$\{\} \xrightarrow{+k_1} \{k_1\} \xrightarrow{+k_2} \{k_1, k_2\} \xrightarrow{+k_3} \{k_1, k_2, k_3\} ...$

$+ \frac{k_1}{||k_1||}$

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



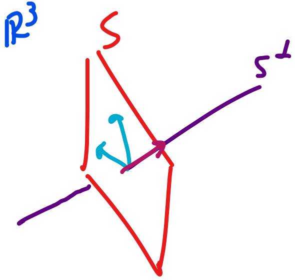
Y así, $\tilde{B} = \{k_1, ..., k_n\}$ pidiendo que $||k_i|| = 1$ resulta una BON.

Complemento Ortogonal

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:



BOG $B_V = \{v_1, \ldots, v_m, v_{m+1}, \ldots, v_n\}$

$B_S = \{v_1, \ldots, v_m\}$

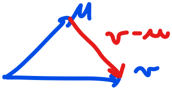
$B_{S^{\perp}} = \{v_{m+1}, \ldots, v_n\}$

$\forall\, u \in S,\ v \in S^{\perp}$

$u \perp v \iff \langle u, v \rangle = 0$

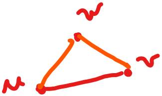
Distancia

PI → NORMA → DISTANCIA



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

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: $\to A^T = \begin{pmatrix} 2 & 3 \\ 3 & 9 \end{pmatrix} = A$

$A = \begin{pmatrix} 2 & 3 \\ 3 & 9 \end{pmatrix}$

$x = \begin{pmatrix} a \\ b \end{pmatrix}$

$(a \ b) \begin{pmatrix} 2 & 3 \\ 3 & 9 \end{pmatrix} \begin{pmatrix} a \\ b \end{pmatrix} = (a \ b) \begin{pmatrix} 2a+3b \\ 3a+9b \end{pmatrix} = 2a^2 + 3ab + 3ab + 9b^2 = 2a^2 + 6ab + 9b^2$

$= a^2 + (a^2 + 6ab + 9b^2) = \underbrace{a^2}_{\geq 0} + \underbrace{(a+3b)^2}_{\geq 0} \geq 0 \ \forall a, b$

$> 0 \ \forall x \neq \vec{0}$

$x \neq 0 \Rightarrow a \neq 0 \lor b \neq 0$

$\Rightarrow A$ es def. pos.

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: $A = I$

$P_2 \quad \langle x, y \rangle = x \cdot y \quad \tilde{x}^T A \tilde{y}$

$B = \{1, x, x^2\}$

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

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

$v \in$ $v = (v_1, \dots, v_n)$ $\|v\|^2 = v_1^2 + \dots + v_n^2$

$\frac{d\|v\|^2}{dv} = (2v_1, \dots, 2v_n) = 2 \cdot v$ $\frac{d\|v\|^2}{dv_1} = 2v_1, \dots, \frac{d\|v\|^2}{dv_i} = 2 \cdot v_i$

El mundo del Data Science y las competencias de Machine Learning cobraron mucha relevancia con la competencia de Netflix, que prometía **1M USD** a quien pudiera obtener una performance de un 10% por sobre su algoritmo de base para el problema de recomendación de películas.

En este contexto surgió una nueva variante de modelo en base a factorización de matrices.

$x \in \mathbb{R}^{1 \times n}$ $y \in \mathbb{R}^{n \times 1}$ $(x \cdot y)'_x = y^T$ $(x \cdot y)'_y = x^T$

¿Lo interesante? Con las dos clases de AM ya nos alcanza para entenderlo!

Antes de ir: ¿Por qué funciona la inversa? (1/2)

Demostremos que: $\exists A^{-1} \Leftrightarrow det(A) \neq 0 \Leftrightarrow \nexists \tilde{x} \neq \vec{0} \,/\, A \cdot \tilde{x} = \vec{0}$

Si $A \in \mathbb{R}^{n \times n}$ **es una matriz definida positiva, entonces es invertible:**
Dem.: (por método del absurdo)

Supongamos que el determinante de A es cero, y llegaremos a una solución absurda.

Si $det(A) = 0$ y queremos resolver un sistema $Ax = 0$ significa que además de la solución trivial hay infinitas soluciones, es decir $rango(A) < n$, entonces

$$\exists \tilde{x} \neq 0 : A\tilde{x} = 0 \Rightarrow \tilde{x}^T A \tilde{x} = \tilde{x}^T 0 = 0$$

y entonces no se cumple que $\tilde{x}^T A \tilde{x} > 0$.

A def pos $\Leftrightarrow x^T A x > 0 \quad \forall x \neq \vec{0}$

si $det(A) = 0$

$\Rightarrow \exists \tilde{x} \neq \vec{0} \,/\, A\tilde{x} = \vec{0}$

$\tilde{x}^T A \tilde{x} = \tilde{x}^T \cdot \vec{0} = \vec{0}$

¿Por qué funciona la inversa? (2/2)

$A = x_1 \cdot x_1^T + \dots + x_n \cdot x_n^T + \lambda I_k$

Demostremos que: $\lambda \in \mathbb{R}^+$

$A = \underbrace{\underbrace{x}_{k\times 1} \cdot \underbrace{x^T}_{1\times k}}_{k\times k} + \lambda \underbrace{I_k}_{} $ **es definida positiva**, es decir, $y^T A y > 0, \ \forall y \in \mathbb{R}^k - \{0\}$

$+ \begin{pmatrix} \lambda & & 0 \\ & \ddots & \\ 0 & & \lambda \end{pmatrix}$

Dem.: Sea $y \in \mathbb{R}^k - \{0\}$

$y^T x \qquad x^T y \qquad \lambda \cdot y^T \cdot \cancel{I} \cdot y$

$$y^T(x \cdot x^T + \lambda I_k)y = \overbrace{y^T x}\overbrace{x^T y} + y^T \lambda I_k y$$

(recordemos que el p.i. es $< u, v >= u^T v$), prop. asociativa,

$$< y, x >< x, y > +\lambda < y, y >= \underbrace{\underbrace{< x, y >^2}_{\geq 0} + \underbrace{\lambda}_{>0} \underbrace{||y||^2}_{>0}}_{>0} > 0.$$

$\therefore x \cdot x^T + \lambda I_k$ es definida positiva.

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.