

$$V = (V, +, \cdot, \cdot) \quad \langle \cdot \rangle$$

$\underbrace{\quad}_8 \subseteq V \quad \underbrace{\quad}_\Phi$

# Análisis Matemático para Inteligencia Artificial

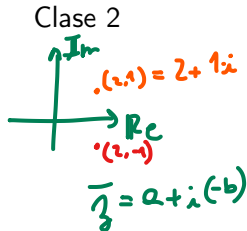
Martín Errázquin (merrazquin@fi.uba.ar)

Especialización en Inteligencia Artificial

$$i \doteq \sqrt{-1}$$

$$z = a + ib \quad a, b \in \mathbb{R}$$

$$= \rho e^{i\varphi} = \rho e^{i\psi}$$



$$\|z\|^2 = a^2 + b^2 \geq 0 \quad \forall z \in \mathbb{C}$$

$$\|z\| = 0 \Leftrightarrow z = 0 + i \cdot 0$$

$$\bar{\bar{z}} = z$$

# Espacios con Producto Interno: Definición

$$x = z = z + 0 \cdot i \quad \forall z \in \mathbb{R} \quad \bar{x} = x$$

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

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

$$\rightarrow \Phi(u + v, w) = \Phi(u, w) + \Phi(v, w)$$

$$\rightarrow \Phi(\alpha \bullet u, v) = \alpha \bullet \Phi(u, v)$$

lineal

$$\Phi(\alpha u + \beta v, w) =$$

$$\Phi(\alpha u, w) + \Phi(\beta v, w) =$$

$$\alpha \Phi(u, w) + \beta \Phi(v, w)$$

- 2  $\Phi(u, v) = \overline{\Phi(v, u)}$  @  $\mathbb{R} \quad \Phi(u, v) = \Phi(v, u)$

- 3  $\Phi(v, v) \geq 0$ , y  $\Phi(v, v) = 0$  si  $v = 0$

$$\hookrightarrow \text{si } v = 0_v, \Phi(v, v) = 0$$

$$\text{si } v \neq 0_v, \Phi(v, v) > 0$$

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  $\mathbb{C}^n$ ).

$$u, v \in \mathbb{R}^n \quad u \cdot v = \sum_{i=1}^n u_i \cdot v_i = u^T \cdot v = v^T \cdot u \quad (1 \ 2) \begin{pmatrix} 2 \\ 1 \end{pmatrix} = 9$$

$\mathbb{C}^n$

## 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.  $f: [-1, 1] \rightarrow \mathbb{C}$  continuas

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

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

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

$$3) \langle f, f \rangle = \int_{-1}^1 \underbrace{f(x) \overline{f(x)}}_{|f(x)|^2} dx = \int_{-1}^1 |f(x)|^2 dx \begin{cases} \rightarrow 0 & \text{si } f(x) \equiv 0 \\ \rightarrow > 0 & \text{si } f(x) \neq 0 \end{cases}$$

# Definición de Norma

EV +  $\Phi \rightsquigarrow$  NORMA

$$v = (1,1)$$

$$\|v\|_2 = \sqrt{2}$$

$$\|v\|_1 = 2$$

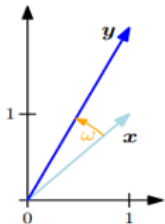
Sea  $(\mathbb{V}, \langle \cdot, \cdot \rangle)$  un e.v. real (complejo) con p.i.. Sea  $v \in \mathbb{V}$ , se define la **norma de  $v$**  asociada a  $\langle \cdot, \cdot \rangle$ .

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

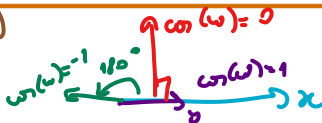
Es la generalización de la longitud de un vector en  $\mathbb{R}^n$  (o  $\mathbb{C}^n$ ).

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

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



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



# Propiedades de la Norma

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

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

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

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

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



# Ortogonalidad

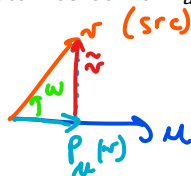
**Def:**  $(\mathbb{V}, \langle \cdot, \cdot \rangle)$  un  $\mathbb{K}$ -EV (con  $\mathbb{K} = \mathbb{R}$  o  $\mathbb{C}$ ) con p.i. dos vectores  $u, v \in \mathbb{V}$  se dicen **ortogonales** si  $\langle u, v \rangle = 0$ .

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

mag. div. sent.  
 $\forall v \in \mathbb{V} \quad v = \|v\| \cdot \tilde{v}$

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



$\tilde{v} = v - P_u(v) \perp u$

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

$\sim \|P_u(v)\| = \frac{\|u\| \cdot \|v\| \cdot \cos(\theta)}{\|u\|}$

# Proceso de Ortogonalización de Gram Schmidt

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

$$\begin{aligned}\langle v_i, v_j \rangle &= 0, \quad \forall i \neq j \\ \langle v_i, v_i \rangle &= 1, \quad \forall i\end{aligned}$$

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

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

$\{v_1, v_2, v_3\}$   
 $\{v_1, k_2, k_3\}$

$$\begin{aligned}k_1 &= v_1 \\ k_2 &= v_2 - \text{Proy}_{k_1}(v_2)\end{aligned}$$

$\vdots$

$$k_n = v_n - \sum_{i=1}^{n-1} \text{Proy}_{k_i}(v_n)$$

$$k_3 = v_3 - \text{Proy}_{k_1}(v_3) - \text{Proy}_{k_2}(v_3)$$

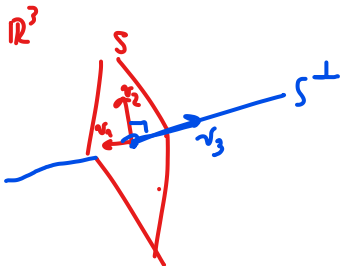
Y así,  $\tilde{B} = \{k_1, \dots, k_n\}$  pidiendo que  $\|k_i\| = 1$  resulta una BON.

# 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 dimensión  $n - m$  que satisface:

$$S \cup S^\perp = \mathbb{V} \quad \wedge \quad S \cap S^\perp = \{0\}$$

Ejemplo:



$$\text{p.e. } \mathcal{B}_{\mathbb{V}} = \{v_1, \dots, v_m, v_{m+1}, \dots, v_n\}$$

$$\mathcal{B}_S = \{v_1, \dots, v_m\}$$

$$\mathcal{B}_{S^\perp} = \{v_{m+1}, \dots, v_n\}$$

$$\forall u \in \mathcal{B}, v \in \mathcal{B}^\perp$$

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



# Distancia

$$EV + \mathbb{E} \rightarrow NORMA \sim \mathbb{D} \quad d(u, v) = \|u - v\| = \sqrt{\langle u - v, u - v \rangle}$$

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

# 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:

$A = \begin{pmatrix} 2 & 3 \\ 3 & 9 \end{pmatrix} \sim A^T = \begin{pmatrix} 2 & 3 \\ 3 & 9 \end{pmatrix} = A \quad \checkmark$

$x = \begin{pmatrix} a \\ b \end{pmatrix} \Rightarrow x^T A x = (a \ b) \begin{pmatrix} 2 & 3 \\ 3 & 9 \end{pmatrix} \begin{pmatrix} a \\ b \end{pmatrix} = (a \ b) \begin{pmatrix} 2a+3b \\ 3a+9b \end{pmatrix} = 2a^2 + 3ab + 3ab + 9b^2$

$= 2a^2 + 6ab + 9b^2 = a^2 + (a^2 + 6ab + 9b^2) = a^2 + (a+3b)^2$

$\geq 0$  (si  $a=0, b \neq 0 \Rightarrow \geq 0$ )

$\geq 0$  (si  $a \neq 0, b \neq 0 \Rightarrow \geq 0$ )

$\geq 0$  (si  $a \neq 0, b = 0 \Rightarrow \geq 0$ )

$\geq 0$  (si  $a = 0, b = 0 \Rightarrow \geq 0$ )

$A \rightarrow \text{def.}$

**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:

$B = \{x^1, x^2\}$   
 $y = x^1 - 3x^2$   
 $\tilde{y} = (-3, 0, 1)$

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

$$\exists \Leftrightarrow \exists (A, \Lambda)$$

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

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.

¿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:

$$\Leftrightarrow x^T A x > 0 \quad \forall x \neq 0_V$$

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  $\text{rango}(A) < n$ , 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  $\tilde{x}^T A\tilde{x} > 0$ .

$$p \Rightarrow q \quad \text{no INV} \Rightarrow \text{no def. pos.}$$

$$\begin{array}{c} \Downarrow \\ q \Rightarrow p \end{array}$$

$$\text{def. pos} \Rightarrow \text{INV}$$

## ¿Por qué funciona la inversa? (2/2)

Demostremos que:

$$A = x_1 \cdot x_1^T + x_2 \cdot x_2^T + \dots + x_n \cdot x_n^T + \lambda I_k$$

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

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

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

*Handwritten notes:* "en el" above the first term, "Id" below the second term.

(recordemos que el p.i. es  $\langle u, v \rangle = u^T v$ ), prop. asociativa,

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

*Handwritten notes:* "y^T I\_k y = \lambda \cdot y^T y = \lambda \cdot y^T y" in the top right; "y^T x x^T y" above the first term; "Id" below the second term; "y^T x x^T y" above the first term; "y^T x x^T y" above the first term; "y^T x x^T y" above the first term.

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

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.