Análisis Matemático para Inteligencia Artificial

Martín Errázquin (merrazquin@fi.uba.ar)

Especialización en Inteligencia Artificial

Clase 1

Administrativo

TODO el material excepto las grabaciones de las clases está disponible en el aula virtual del campus!

Filosofía del curso:

- Aprendizaje jerarquizado: ignorar info de más > quedarse sin aprender
- Aplicaciones siempre que se pueda
- No sobrepresionar, pero premiar dedicación

Forma de evaluación:

1. Autoevaluaciones individuales:
   - Uno por clase
   - Todos deben aprobarse (60+%)
   - Infinitos intentos, nota es el máximo
2. Trabajo Práctico (grupal):
   - máximo 5 integrantes, se recomienda 4-5
   - No obligatorio pero recomendable (consigna ya en el campus)
   - Se entrega código **documentado**

aev.1: 9, 4, 3, 2, 1 $\rightsquigarrow$ 9

⋮

aev.8: 4, 3, 5 $\rightsquigarrow$ 5

5

Nota = $0.6 \cdot \min\{\text{Cuest}_i\} + 0.5 \cdot \text{TP}$

Cierre de Entregas: W8 - SÁBADO - 12 pm GMT-3

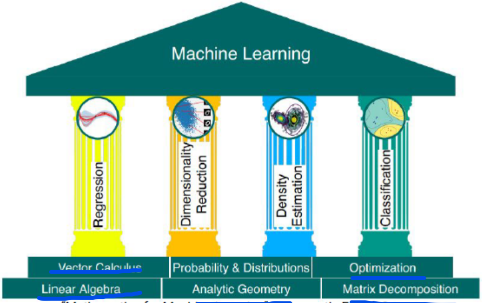
mediodía

Presentación de la Materia

¿Por qué estudiar Análisis Matemático?

A medida que Machine Learning se vuelve más común, y los paquetes de software se vuelven más simples de usar, uno se abstrae cada vez más de los detalles técnicos que hay detrás. **Modelo de caja negra.**

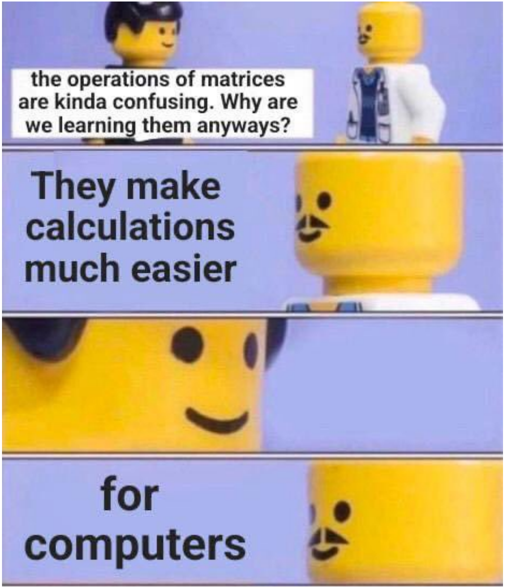
Esto trae el **peligro** de desconocer las decisiones de diseño y las limitaciones de cada algoritmo.



"Mathematics for Machine Learning", Deisenroth, Faisal, Ong

Bibliografía Recomendada: **Mathematics for Machine Learning**
Marc Peter Deisenroth, A. Aldo Faisal, and Cheng Soon Ong.
Published by Cambridge University Press (2020).
Está disponible gratis en `http://mml-book.github.io/`

Motivación: performance



Motivación: performance y legibilidad

Quiero hacer $y = A \cdot x + b$:

```
def matvec_naive(A,x,b):
  y = np.empty_like(b)
  for i in range(A.shape[0]):
    aux = b[i]
    for j in range(A.shape[1]):
        aux += A[i][j] * x[j]
    y[i] = aux
  return y
```

```
def matvec_np(A,x,b):
    return A @ x + b
```

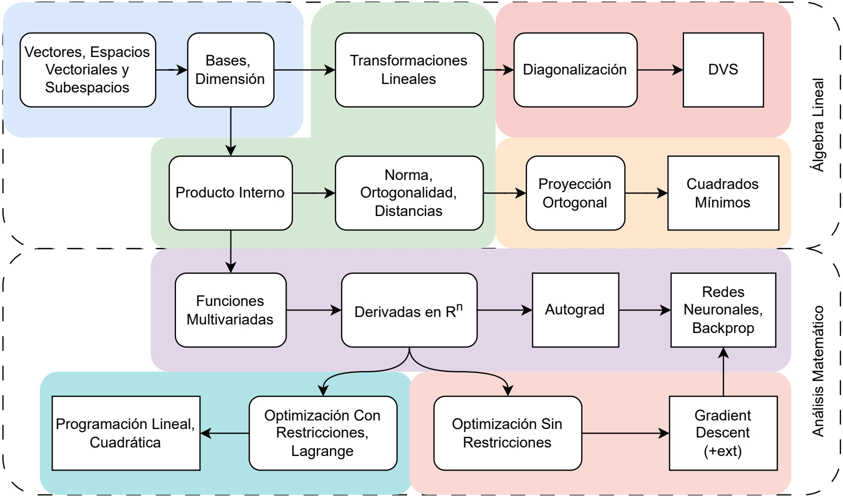
- ¿Cuál creen que funciona más rápido?
- ¿Cuál creen que es más fácil de leer (y mantener)?

Early colab: Operaciones vectorizadas y GPUs!

¿Qué buscamos?

- Objetivo primario 1: proveer el sustento teórico matemático requerido para poder comprender la formulación de los modelos clásicos.
- Objetivo primario 2: entrenar lectura e interpretación de lenguaje matemático formal y avanzado.
- Objetivo secundario: cuando sea posible, mostrar aplicaciones directas de la teoría vista.
- Bonus: concientizar sobre importancia y ventajas de optimización en tiempo y memoria, y cómo la matemática es la vía para la misma.

Mapa del temario



 2^2^3 ↗ 2^(2^3)=2^8=256 ↘ (2^2)^3 = 4^3 = 64

Diremos que $\mathcal{V} = (\mathbb{V}, +, \mathbb{K}, \bullet)$ es un espacio vectorial si $\mathbb{K}$ y $\mathbb{V}$ son conjuntos no vacíos y la operación $+ : \mathbb{V} \times \mathbb{V} \to \mathbb{V}$, y la acción $\bullet : \mathbb{K} \times \mathbb{V} \to \mathbb{V}$ cumplen:

1. $+$ es asociativa $\forall u, v, w \in \mathbb{V} \quad u + (v + w) = (u + v) + w$
2. $+$ tiene elemento neutro $\exists 0_{\mathbb{V}} \in \mathbb{V} : u + 0_{\mathbb{V}} = 0_{\mathbb{V}} + u = u \quad \forall u \in \mathbb{V}$
3. $+$ tiene elemento inverso "opuesto" $\forall v \in \mathbb{V} \, \exists \tilde{v} \in \mathbb{V} / v + \tilde{v} = \tilde{v} + v = 0_{\mathbb{V}}$
4. $+$ es conmutativa $\forall u, v \in \mathbb{V} \quad u + v = v + u$
5. $\alpha \bullet (v \overset{\mathbb{V}}{+} w) = \alpha \bullet v + \alpha \bullet w, \; \forall \alpha \in \mathbb{K}, \forall v, w \in \mathbb{V}$ distr. de $\bullet$ wrt $+_{\mathbb{V}}$
6. $(\alpha \underset{\mathbb{K}}{+} \beta) \bullet v = \alpha \bullet v + \beta \bullet v, \; \forall \alpha, \beta \in \mathbb{K}, \forall v \in \mathbb{V}$ distr. de $\bullet$ wrt $+_{\mathbb{K}}$
7. $\bullet$ tiene elemento neutro: $1 \bullet v = v, \forall v \in \mathbb{V}$
8. $\bullet$ es asociativa:

$$\alpha \bullet (\beta \bullet v) = (\underset{\cdot_{\mathbb{K}}}{\alpha\beta}) \bullet v, \; \forall \alpha, \beta \in \mathbb{K}, \forall v \in \mathbb{V}$$

Subespacios Vectoriales: definición

Sea $\mathcal{V} = (\mathbb{V}, +, \mathbb{K}, \bullet)$ un espacio vectorial, un subconjunto $S \subseteq \mathbb{V}$, $S \neq \emptyset$ se dice que es un subespacio de $\mathbb{V}$ si la suma y el producto por escalares de $\mathbb{V}$ son una operación y una acción en S que lo convierten en un $\mathbb{K}$-espacio vectorial. $\mathcal{S} = (S, +, \mathbb{K}, \bullet)$ $S \subseteq \mathbb{V}$, $S \neq \emptyset$

Condiciones necesarias y suficientes para caracterizar subespacios

S es un subespacio en un $\mathbb{K}$-espacio vectorial sii:

1. $S \neq \emptyset$ $(0 \in S)$ $\rightsquigarrow$ $0_V = 0_S \in S$
2. $v, w \in S \to v + w \in S$ $\to$ $+$ cerrada en S
3. $\alpha \in \mathbb{K}, v \in S \to \alpha \bullet v \in S$ $\to$ $\bullet$ " "

Subespacios triviales

- $\{0\} \subseteq \mathbb{V}$
- $\mathbb{V} \subseteq \mathbb{V}$

Definición: Sea $\mathcal{V}$ un espacio vectorial, y $G = \{v_1, ..., v_r\} \subseteq \mathbb{V}$. Una combinación lineal de G es un elemento $v \in \mathbb{V}$ tal que $v = \sum_{i=1}^{r} \alpha_i \bullet v_i$, donde $\alpha_i \in \mathbb{K}$, para $i = 1, ..., r$.

$G = \left\{ \begin{pmatrix} 1 \\ 0 \\ 1 \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \\ 1 \end{pmatrix} \right\} \subseteq \mathbb{R}^3$

$\begin{pmatrix} 1 \\ 0 \\ 1 \end{pmatrix}$ es CL de G? $\Leftrightarrow \exists \alpha_1, \alpha_2 \in \mathbb{R} \,/\, \begin{pmatrix} 1 \\ 0 \\ 1 \end{pmatrix} = \alpha_1 \cdot \begin{pmatrix} 1 \\ 0 \\ 1 \end{pmatrix} + \alpha_2 \begin{pmatrix} 0 \\ 1 \\ 1 \end{pmatrix}$

sí, $\alpha_1 = 1$, $\alpha_2 = 0$

Definición: Sea $\mathcal{V}$ un espacio vectorial, y $G \subseteq \mathbb{V}$. Se dice que G es un sistema de generadores de $\mathcal{V}$ si todo elemento de $\mathbb{V}$ es una combinación lineal de G.

Notación: $\langle G \rangle = \mathbb{V}$.

$\langle G \rangle = \{ v \in \mathbb{V} \,/\, v$ es CL de $G \}$

$G \subseteq \langle G \rangle \subseteq \mathbb{V}$

Sea $G = \left\{ \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}, \begin{pmatrix} 2 \\ 1 \\ 0 \end{pmatrix}, \begin{pmatrix} 4 \\ 3 \\ 2 \end{pmatrix} \right\}$ dado un vector cualquiera $v \in \mathbb{R}^3$,

¿podemos escribirlo como combinación lineal de los vectores de G? $\langle G \rangle \overset{?}{=} \mathbb{R}^3$

$\forall \begin{pmatrix} x \\ y \\ z \end{pmatrix} \exists \alpha_1, \alpha_2, \alpha_3 \in \mathbb{R} / \alpha_1 \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix} + \alpha_2 \begin{pmatrix} 2 \\ 1 \\ 0 \end{pmatrix} + \alpha_3 \begin{pmatrix} 4 \\ 3 \\ 2 \end{pmatrix} = \begin{pmatrix} x \\ y \\ z \end{pmatrix}$

$$\begin{cases} \alpha_1 + 2\alpha_2 + 4\alpha_3 = x & (1) \\ \alpha_1 + \alpha_2 + 3\alpha_3 = y & (2) \\ \alpha_1 + 2\alpha_3 = z & (3) \end{cases}$$

de (3) $\alpha_1 = z - 2\alpha_3$

en (2) $z - 2\alpha_3 + \alpha_2 + 3\alpha_3 = y$

$\alpha_2 = y - z - \alpha_3$

en (1):

$z - 2\alpha_3 + 2(y - z - \alpha_3) + 4\alpha_3 =$

$z \underbrace{- 2\alpha_3} + 2y - 2z \underbrace{- 2\alpha_3} + \underbrace{4\alpha_3} =$

$\boxed{\underbrace{2y - z} = x}$ no vale $\forall x, y, z$

$\Downarrow$

$\langle G \rangle \neq \mathbb{R}^3$

Independencia Lineal

$S' = \left\{ \begin{pmatrix} 1 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \end{pmatrix}, \begin{pmatrix} 2 \\ 3 \end{pmatrix} \right\}$

Dentro de los conjuntos generadores, nos interesan aquellos que son mínimos (menor cantidad de elementos).

Sea $S \subseteq \mathcal{V}$ un subespacio vectorial, y sea:

- $\{v_1, ..., v_n\} \subseteq \mathbb{V}$. Entonces $\langle v_1, ..., v_n \rangle \subseteq S$ sii $v_i \in S$, $\forall 1 \leq i \leq n$.
- $\{v_1, ..., v_n, v_{n+1}\} \subseteq \mathbb{V}$. Entonces $\langle v_1, ..., v_n, v_{n+1} \rangle = \langle v_1, ..., v_n \rangle$ sii $v_{n+1} \in \langle v_1, ..., v_n, \rangle$ $\Leftrightarrow$ v_{n+1} es CL de $\{v_1, ..., v_n\}$

Definición: Sea $\mathcal{V}$ un espacio vectorial, y sea $\{v_\alpha\}_{\alpha \in I}$ una familia de vectores en $\mathbb{V}$; se dice que $\{v_\alpha\}_{\alpha \in I}$ es linealmente independiente (l.i.) sii

$$\sum_{\alpha \in I} k_\alpha \bullet v_\alpha = 0 \rightarrow k_\alpha = 0, \ \forall \alpha \in I$$

$\alpha_1 \cdot v_1 + ... + \alpha_n \cdot v_n = 1 \cdot v_{n+1}$

$G = \{v_1, ..., v_{n+1}\}$ $\quad \alpha_1 \cdot v_1 + ... + \alpha_n \cdot v_n + (-1) \cdot v_{n+1} = 0_V$

$\Updownarrow$

$\exists \alpha_1, ..., \alpha_n \neq \vec{0}$

Observar:

- $\{0\}$ es linealmente dependiente (l.d.)
- si $v \neq 0$, $\{v\}$ es l.i.
- si $v_1 \propto v_2$ (colineales), $\{v_1, v_2\}$ es l.d.
- si v_1, v_2 no nulos, ni proporcionales, $\{v_1, v_2\}$ es l.i.

Definición: Sea $\mathcal{V}$ un espacio vectorial, un conjunto $\{v_\alpha\}_{\alpha \in I}$ se llama base de $\mathcal{V}$ si $\{v_\alpha\}_{\alpha \in I}$ es un conjunto $\underbrace{\text{linealmente independiente}}_{\text{no sobra nada}}$ de $\mathbb{V}$ que satisface $\underbrace{\langle v_\alpha \rangle_{\alpha \in I}}_{\text{no falta nada}} = \mathbb{V}$.

Definición: Sean $\mathcal{V}$ un espacio vectorial, $B = \{v_1, ..., v_n\}$ una base de $\mathcal{V}$. Diremos que n es la dimensión de $\mathcal{V}$, donde $n < \infty$.
Comentario: Tener en cuenta que existen espacios vectoriales con dimensión infinita.

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

$B' = \left\{ \begin{pmatrix} 1 \\ 1 \end{pmatrix}, \begin{pmatrix} 1 \\ -1 \end{pmatrix} \right\}$ BOG

$B = \left\{ \begin{pmatrix} 1 \\ 1 \end{pmatrix}, \begin{pmatrix} 1 \\ 2 \end{pmatrix} \right\}$

$B'' = \left\{ \begin{pmatrix} 1/\sqrt{2} \\ 1/\sqrt{2} \end{pmatrix}, \begin{pmatrix} 1/\sqrt{2} \\ -1/\sqrt{2} \end{pmatrix} \right\}$ BON

$[u]_B = \begin{pmatrix} \alpha_1 \\ \alpha_2 \end{pmatrix}$ $\quad \forall u \in \mathbb{R}^2 \ \exists! \, \alpha_1, \alpha_2 \ / \ \alpha_1 \cdot v_1 + \alpha_2 \cdot v_2 = u$ $\quad B = \{v_1, v_2\}$

Matriz de Cambio de Base

Sea $B = \{v_1, ..., v_n\}$ una base de $\mathbb{R}^n$ y sea la matriz de vectores de B:

$S \cdot \alpha = x$

$\begin{pmatrix} | \\ v_1 \\ | \end{pmatrix} \cdot \alpha_1 + \dots + \alpha_n \cdot \begin{pmatrix} | \\ v_n \\ | \end{pmatrix} = x$

$$S = \begin{pmatrix} | & | & \dots & | \\ v_1 & v_2 & \dots & v_n \\ | & | & \dots & | \end{pmatrix} \in \mathbb{R}^{n \times n}$$

Sea $\alpha = (\alpha_1, \dots, \alpha_n) \in \mathbb{R}^n$. Resolver $S \cdot \alpha = x$ para un cierto $x \in \mathbb{R}^n$ significa hallar la (única) CL de vectores de B que resulta en x. Pero, como S admite inversa, eso es simplemente calcular $\alpha = S^{-1} \cdot x$.

Tenemos entonces que para una base B y S la matriz que tiene en sus columnas a los vectores de la base,

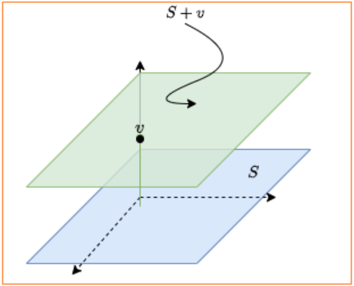
- S es la matriz de cambio de base $B \to E$.
- S^{-1} es la matriz de cambio de base $E \to B$.

Notación: a α se lo suele denotar $[x]_B$, el *vector de coordenadas de x en la base B*.

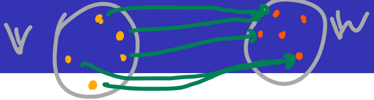
Variedad lineal

Sea $\mathcal{V}$ un espacio vectorial, M es una variedad lineal $M \subseteq \mathbb{V}$ es un conjunto de la forma

$M = \{s + v, \ donde \ s \in S\}$, siendo S subespacio de $\mathcal{V}$, y $v \in \mathbb{V}$.



Transformaciones



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

$T(x) \neq T(y)$

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

Importante: Para toda transformación biyectiva existe una (única) transformación **inversa**.

En particular, nos interesan (por ahora) las **transformaciones lineales**. T es una TL cuando $\mathbb{V}$ y $\mathbb{W}$ son EVs y además se verifica que

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

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

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

$B_V = \{v_1, \ldots, v_n\}$

$B_W = \{w_1, \ldots, w_n\}$

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:

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

Importante: esto justifica que siempre usemos $\mathbb{R}^n$.

Ejemplo: $\mathcal{P}_2[x]$ y $\mathbb{R}^3$

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

$B_{\mathbb{R}^3} = \left\{ \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix} \right\}$

$v = 3x + 5x^2 \rightarrow [v]_B = (0, 3, 5)$

$v' = \begin{pmatrix} 0 \\ 3 \\ 5 \end{pmatrix} \longleftarrow B_{\mathbb{R}^3}$

Espacios nulo, columna y fila

Teorema: Toda transformación lineal $T : \mathbb{R}^n \to \mathbb{R}^m$ se puede representar de forma matricial como $T(x) = Ax$ con $A \in \mathbb{R}^{m\times n}$.

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

$$A = \begin{pmatrix} | & & | \\ A_1 & \cdots & A_n \\ | & & | \end{pmatrix} = \begin{pmatrix} — A^{(1)} — \\ \vdots \\ — A^{(m)} — \end{pmatrix}$$

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

- Espacio columna de A: es el subespacio de $\mathbb{R}^m$ generado por los n vectores columna de A: $\left\langle \left\{ \begin{pmatrix} | \\ A_1 \\ | \end{pmatrix}, \dots, \begin{pmatrix} | \\ A_n \\ | \end{pmatrix} \right\} \right\rangle \subseteq \mathbb{R}^m$

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

- Espacio fila de A: es el subespacio de $\mathbb{R}^n$ generado por los m vectores fila de A:

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

$$= \left\langle \left\{ A^{(1)}, \dots, A^{(m)} \right\} \right\rangle \subseteq \mathbb{R}^n$$

Mini ejemplo

$$L\begin{pmatrix} x \\ y \end{pmatrix} = \underbrace{\begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix}}_{A} \begin{pmatrix} x \\ y \end{pmatrix}$$

NA: $A v = \begin{pmatrix} 0 \\ 0 \end{pmatrix} \rightsquigarrow \begin{cases} 1x + 1y = 0 \\ 1x - 1y = 0 \rightsquigarrow x = y \end{cases}$ $x = y = 0$ $\Rightarrow N(A) = \left\langle \begin{pmatrix} 0 \\ 0 \end{pmatrix} \right\rangle$

$\dim(N(A)) = 0$

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

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

Relevancia

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

Observación: Toda matriz tiene la misma cantidad de filas y columnas LI.

Definición: Se denomina **nulidad** de la matriz A a la dimensión de su espacio nulo, $n(A) = dim(N(A))$.

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

$0 \leq r(A) \leq \min(m, n)$

$0 \leq n(A) \leq n$

$$\underbrace{r(A)}_{\text{\#dim aprovechadas}} + \underbrace{n(A)}_{\text{\#dim ``perdidos''}} = \underbrace{n}_{\text{\#dim entrada}}$$

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.