

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

mín max {Nota}
sem. intentos

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**

mono-notebook
↗ repo forkable
mediodía

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

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

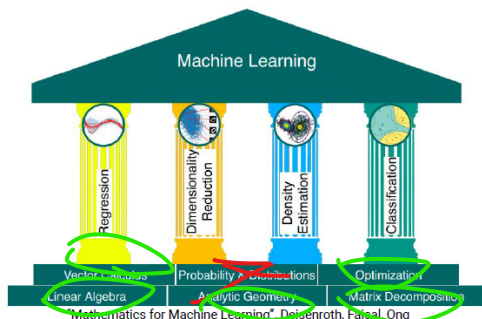
Presentación de la Materia

¿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.



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 (I): performance y legibilidad

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

versión escalera

```
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
```

versión vectorizada

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

$$y_i = \left(\sum_{j=1}^n A_{ij} \cdot x_j \right) + b_i$$

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

Early colab: Operaciones vectorizadas y GPUs!

Motivación (II): Más tipos de error que debuggear

- Error de programación: error de sintaxis
- Error de programación: el código está mal pensado

Aparecen nuevos!

- Error de matemática: el modelo está mal planteado
- Error de planteo: no están dadas las condiciones para la aplicación del modelo seleccionado
- Error de traducción: el código no refleja el modelo planteado

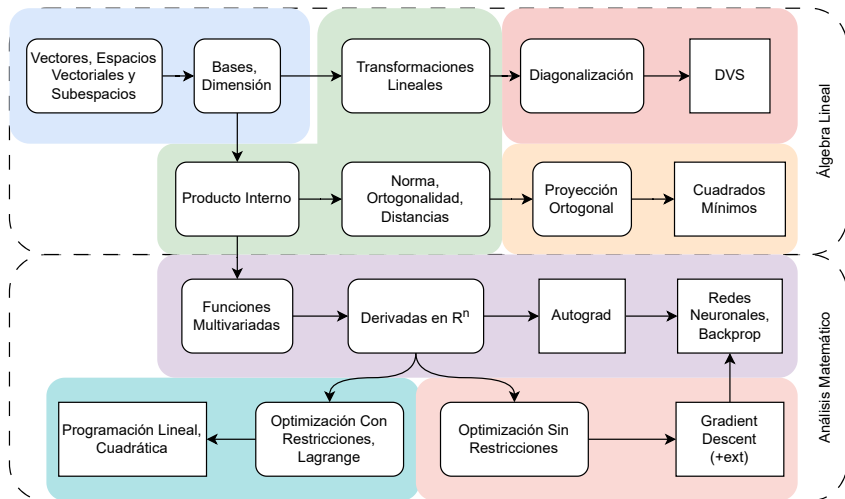
Motivación (III): entender los modelos

- Regresión logística asume superposición entre clases
- K-Means asume clusters esféricos
- Árboles de decisión tienen fronteras de decisión en forma de hiperplanos
- ¿Por qué las Redes Neuronales se entrenan más rápido usando GPUs?
- ¿Por qué en las redes neuronales importa la escala y en los árboles de decisión no?
- En kNN ¿Es lo mismo maximizar producto interno que minimizar distancia euclídea?

¿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



Clase 1: Espacios Vectoriales

$$\begin{aligned} 2^1 3^1 2 &\rightarrow (2^1 3)^1 2 = 6^1 2 = 6^1 \\ &\rightarrow 2^1 (3^1 2) = 2^1 6 = 2^1 6 \\ &= 512 \end{aligned}$$

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 $+: \mathbb{V} \times \mathbb{V} \rightarrow \mathbb{V}$, y la acción

$\bullet: \mathbb{K} \times \mathbb{V} \rightarrow \mathbb{V}$ cumplen:

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

$$\underbrace{\alpha \bullet (\beta \bullet v)}_{\mathbb{K} \times \mathbb{V}} = \underbrace{(\alpha \beta) \bullet v}_{\mathbb{K} \times \mathbb{K}}$$

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.

$$S = (S, +, \mathbb{K}, \bullet)$$

$$\mathbb{R}^2 \subset \mathbb{R}^3$$

Condiciones necesarias y suficientes para caracterizar subespacios

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

$$(x, y, 0) \in \mathbb{R}^3$$

① $S \neq \emptyset$ ($0 \in S$) $\rightarrow 0_{\mathcal{V}} \in S$

② $v, w \in S \rightarrow v + w \in S$ \rightarrow suma cerrada en S

③ $\alpha \in \mathbb{K}, v \in S \rightarrow \alpha \bullet v \in S$ \rightarrow prod. + escalar " " "

Subespacios triviales

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

Representación de subespacios

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

$$v = 3 \cdot \begin{pmatrix} 1 \\ 2 \\ 1 \end{pmatrix} + 2 \cdot \begin{pmatrix} 0 \\ -1 \\ 1 \end{pmatrix} = \begin{pmatrix} 3 \\ -2 \\ 5 \end{pmatrix}$$

$$v' = 0 \cdot \begin{pmatrix} 1 \\ 2 \\ 1 \end{pmatrix} + 1 \cdot \begin{pmatrix} 0 \\ -1 \\ 1 \end{pmatrix} = \begin{pmatrix} 0 \\ -1 \\ 1 \end{pmatrix}$$

$$G = \left\{ \begin{pmatrix} 1 \\ 2 \\ 1 \end{pmatrix}, \begin{pmatrix} 0 \\ -1 \\ 1 \end{pmatrix} \right\} \subseteq \mathbb{R}^3 \quad \langle G \rangle = \left\{ \alpha \cdot \begin{pmatrix} 1 \\ 2 \\ 1 \end{pmatrix} + \beta \cdot \begin{pmatrix} 0 \\ -1 \\ 1 \end{pmatrix} \mid \alpha, \beta \in \mathbb{R} \right\} \subseteq \mathbb{R}^3$$

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

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

$$\langle G \rangle = \{ v \in \mathcal{V} \mid v \text{ es CL de } G \}$$

\uparrow
conj. de vectores $\subseteq \mathcal{V}$

$$\forall G \subseteq \mathcal{V}$$

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

Ejemplo

$$.2 + .1$$

! el 3º es CL del 1º y 2º

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 ?

$$\textcircled{1} \begin{pmatrix} x \\ y \\ z \end{pmatrix} = a \cdot \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix} + b \cdot \begin{pmatrix} 2 \\ 1 \\ 0 \end{pmatrix} + c \cdot \begin{pmatrix} 4 \\ 3 \\ 2 \end{pmatrix}$$

$$\exists a, b, c \in \mathbb{R} / \forall x, y, z \in \mathbb{R} \quad \textcircled{2}$$

$$\begin{cases} a + 2b + 4c = x & \textcircled{1} \\ a + b + 3c = y & \textcircled{2} \\ a + 2c = z & \textcircled{3} \end{cases}$$

$$\text{de } \textcircled{3} \quad a = z - 2c$$

$$\text{en } \textcircled{2} \quad b = y - z - c$$

$$\text{en } \textcircled{1} \quad (z - 2c) + 2(y - z - c) + 4c = x$$

$$[2y - z = x]$$

IMPOSIBLE
CONDICIONES
 x, y, z

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

Independencia Lineal

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:

- $\{v_1, \dots, v_n\} \subseteq \mathbb{V}$. Entonces $\langle v_1, \dots, v_n \rangle \subseteq S$ sii $v_i \in S, \forall 1 \leq i \leq n$.
 - $\{v_1, \dots, v_n, v_{n+1}\} \subseteq \mathbb{V}$. Entonces $\langle v_1, \dots, v_n, v_{n+1} \rangle = \langle v_1, \dots, v_n \rangle$ sii $v_{n+1} \in \langle v_1, \dots, v_n \rangle$
- $\langle \text{red}, \text{yellow} \rangle = \langle \text{red}, \text{yellow}, \text{red} \rangle$

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

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

1. $v_{n+1} = d_1 \cdot v_1 + \dots + d_n \cdot v_n$

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.

$G = \{v_1, \dots, v_n, v_{n+1}\}$

$d_i \in \mathbb{K}$

$d_1 \cdot v_1 + \dots + d_n \cdot v_n + (-1) \cdot v_{n+1} = 0$
 \uparrow
 $\exists d_1, \dots, d_n, d_{n+1} \neq 0$

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 linealmente independiente de \mathcal{V} que satisface $\langle v_\alpha \rangle_{\alpha \in I} = \mathcal{V}$.

cubre todo \Leftrightarrow no falta nada

no sobre nada

Definición: Sean \mathcal{V} un espacio vectorial, $B = \{v_1, \dots, v_n\}$ 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

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

$$E = \left\{ \begin{pmatrix} 1 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \end{pmatrix} \right\}$$

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

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

BOG

BON

Matriz de Cambio de Base

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

$$S \cdot \alpha = x$$
$$\alpha_1 \begin{pmatrix} | \\ v_1 \\ | \end{pmatrix} + \dots + \alpha_n \begin{pmatrix} | \\ v_n \\ | \end{pmatrix} = \begin{pmatrix} | \\ x \\ | \end{pmatrix}$$
$$S = \begin{pmatrix} | & | & \dots & | \\ v_1 & v_2 & \dots & v_n \\ | & | & \dots & | \end{pmatrix} \in \mathbb{R}^{n,n}$$

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,

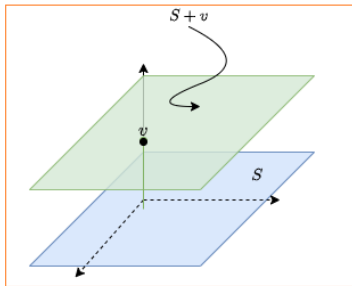
- S es la matriz de cambio de base $B \rightarrow E$.
- S^{-1} es la matriz de cambio de base $E \rightarrow B$.

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

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



Transformaciones



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

- **Inyectiva:** si $\forall x, y \in \mathbb{V} : T(x) = T(y) \rightarrow x = y$ $\forall x \neq y \Rightarrow T(x) \neq T(y)$
- **Surjectiva:** si $T(\mathbb{V}) = \underline{\mathbb{W}}$
- **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

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

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

Representaciones

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

TL biyectiva

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

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

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

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

$$B_{\mathcal{P}_2} = \{1, x, x^2\} \Rightarrow \dim(\mathcal{P}_2) = 3$$

$\downarrow \quad \downarrow \quad \downarrow$

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

$$v = 5 - 3x + 2x^2$$
$$\alpha = (2, -3, 5) \in \mathbb{R}^3$$

Espacios nulo, columna y fila

Teorema: Toda transformación lineal $T : \mathbb{R}^n \rightarrow \mathbb{R}^m$ 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 $Av = 0$.

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

“conjunto de preimágenes del 0”

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

$$\left\langle \begin{pmatrix} a_{11} \\ \vdots \\ a_{m1} \end{pmatrix}, \dots, \begin{pmatrix} a_{1n} \\ \vdots \\ a_{mn} \end{pmatrix} \right\rangle \subseteq \mathbb{R}^m$$

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

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

$$\left\langle \begin{pmatrix} a_{11} & \dots & a_{1n} \end{pmatrix}, \dots, \begin{pmatrix} a_{m1} & \dots & a_{mn} \end{pmatrix} \right\rangle \subseteq \mathbb{R}^n$$

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

Mini ejemplo

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

$$NA: A \cdot v = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$$

$$N(A) = \left\{ \begin{pmatrix} 0 \\ 0 \end{pmatrix} \right\} = \left\langle \begin{pmatrix} 0 \\ 0 \end{pmatrix} \right\rangle \rightarrow \dim(N(A)) = 0$$

$$\begin{cases} x+y=0 & 2x=0 \Leftrightarrow x=0 \\ x-y=0 \rightarrow x=y & y=0 \end{cases}$$

$$EC: A = \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix}$$

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

$$\dim(EC(A)) = 2$$

$$EF: A = \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix}$$

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

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:

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

$$n(A) \leq n$$

$$\underbrace{r(A)}_{\substack{\# \dim \\ \text{"aprovechados"}}} + \underbrace{n(A)}_{\substack{\# \dim \\ \text{"perdidos"}}} = \underbrace{n}_{\substack{\# \dim \\ \text{entradas}}}$$

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.

A handwritten table with three rows and two columns. The first column contains student identifiers $w1$, $w2$, and $w3$. The second column contains their scores: $10, 1, 3$ for $w1$, $8, 7, 10$ for $w2$, and $1, 1, 1, 9$ for $w3$. A vertical line separates these from the final scores: 10 , 10 , and 9 . An orange arrow points from the word 'MAX' to the score 9 for $w3$. Another orange arrow points down from the right side of the table to the word 'MIN'.

$w1$	$10, 1, 3$	10
$w2$	$8, 7, 10$	10
$w3$	$1, 1, 1, 9$	9

MAX \rightarrow 9

\downarrow MIN