→ Por qué el ML funciona?

ML → Alg. estadístico

ML → Datos

ML → Alguien lo programó

ML → ~~Datos~~ → Big Data.

ML → Matemáticas

→ Partes del ML

Observación + Análisis

training

Datos ↓

→ EDA

DATOS → Modelo

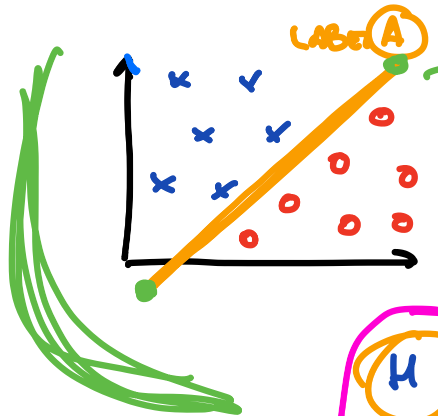
DATA → MATHEMATICAL MODEL → OUTPUT HYPOTHESIS

Hypothesis class.

LOSS FUNCTION

OPTIMIZATION METHOD

gradient descent



Recta

$$\boxed{y = mx + n}$$

$$H = \{ \underbrace{mx + n}_{\infty \quad \infty} \}$$

$|H| =$

$y = 3x + 2$ // $y = \underset{m}{-251.2}\,x + \underset{n}{4.2}$

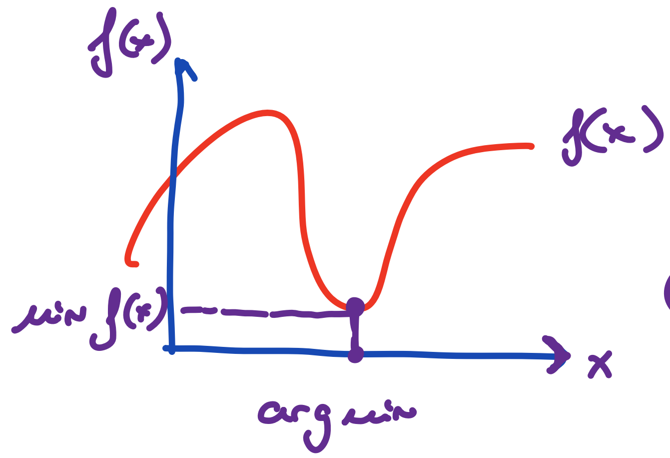
LEAST SQUARES (MÍNIMOS CUADRADOS)

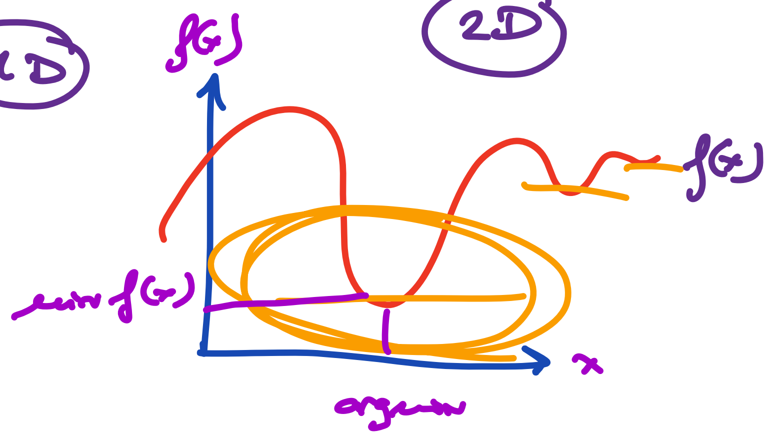
$$\boxed{\underset{h}{\arg\min} \ \frac{1}{n} \sum_{i=1}^{N} (y_i - h(x_i))^2}$$

m n

2D

1D





Mínimo global.

→ **Best classifier**

$$\underbrace{\arg\min_h}\ \frac{1}{|D|}\sum_{i=1}^{D} \underbrace{1_{\{y_i \neq h(x_i)\}}}$$

examen

1 → B (D) ✗

2 → C (C) ✓

3 → D (A) ✗

loss function

$S = (x_i, y_i)$

Pregunta 1, B

Pregunta 2, C

Pregunta 3, D

□ → coche $\quad 1\{y_i \neq h(x_i)\}$

□ → gato $\quad 1\{y_i \neq h(x_i)\}$

□ → casa. $\quad 1\{y_i \neq h(x_i)\}$

y_i

$$\sum_{i=1}^{D} 1_{\{\underbrace{y_i}\neq h(x_i)\}}$$

modelo gpt2 $(x_i) \rightarrow \hat{y}_i$ Prediction

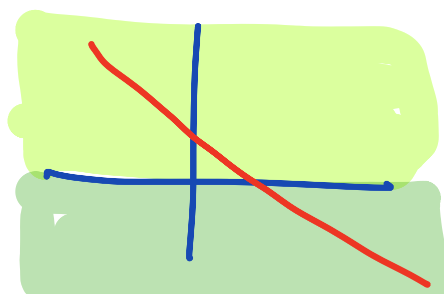
$y_i \neq$ salida del modelo → +1

→ y_i

$$\arg\min_h \frac{1}{n}\sum_{i=1}^{n} \left(y_i - h(x_i)\right)^2 \qquad \Big| \qquad \arg\min_h \frac{1}{n}\sum_{i=1}^{n} \left|y_i - h(x_i)\right|$$

- positivo
- incrementar resultado

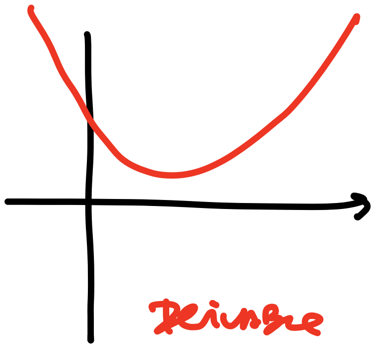
$$\arg\min_h \frac{1}{|D|}\sum_{i=1}^{D} 1_{\{y_i \neq h(x_i)\}}$$



(+)

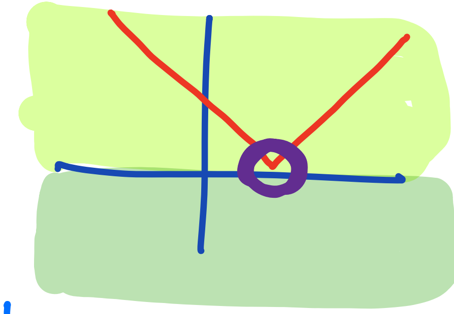
(−)

$(y_i - h(x_i))^2$ $\quad\quad$ $|y_i - h(x_i)|$



$\frac{df(x)}{dx} = 0$

Derivable en todo su dominio



$\frac{df(x)}{dx} = 0$

No es derivable

$\text{argmin}_h \frac{1}{|D|} \sum_{i=1}^{D} 1_{\{y_i \neq h(x_i)\}}$ → 0-1 loss function

Best classifier

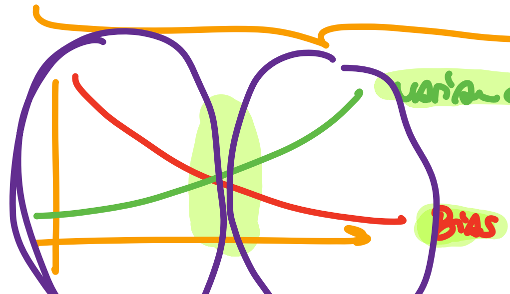
0-1 loss function → surrogate loss function → classifier calibrated

→ Generalización de los modelos de m.

Overfitting

underfitting

$\underbrace{E = \text{var} + \text{bias}}$ $+ e^2$



$H = \{ \ldots \}$

surrogate. Loss Functions

→ **VECTOR ?**

Física

→ Flecha
- Dirección
- módulo
- Sentido

Vector

Física

- Flecha / array
  - ↳ Dirección.
  - ↳ módulo.
  - ↳ Sentido.

Magnitud

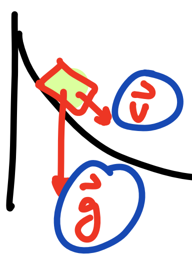
Matemáticas

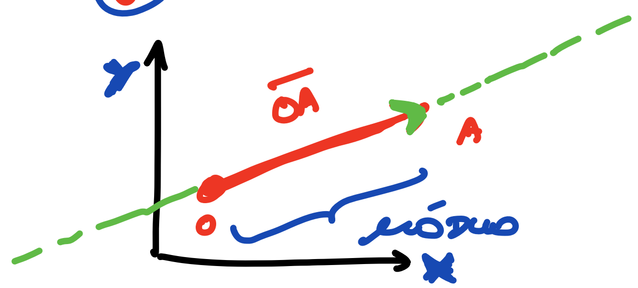
?

Software engineer

Lista de elementos.

[– , – , – , ..., —]

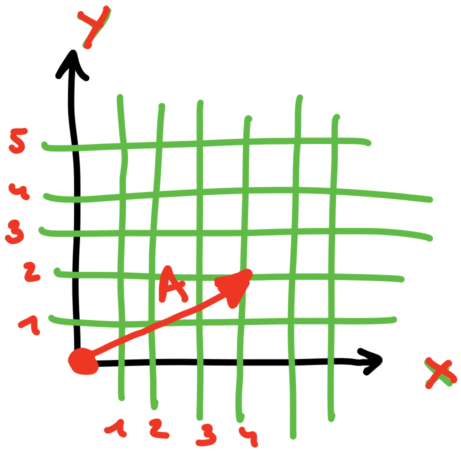




Interpretación geométrica de los vectores (Física)

$OA \in \mathbb{R}^2$

TENSOR → MATRIZ → VECTOR



$$\vec{A} = \begin{bmatrix} 4 \\ 2 \end{bmatrix}$$

$$\underbrace{[\,-,\ -,\ -,\ -,\ \ldots,\ -\,]}_{N}$$

["Hola", "que", "tal", "cómo", "estás", "?"] = B

1 2 3 4 5 6

$\boxed{B \in \mathbb{R}^6}$

→ (1, 0)

→ GloVe Word2Vec

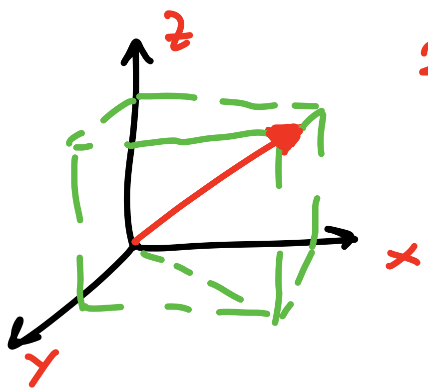
→ Transformer (seq2seq)

$\boxed{\text{"rey" − "mujer" = "reina"}}$

$$\rightarrow \underbrace{[\,-,\ -,\ -,\ -,\ -,\ -,\ -,\ \ldots,\ -\,]}_{N}$$

Performance modelo

Interpretabilidad

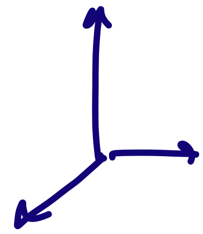


3D / 2D

Dimensionality Reduction

1

2 → 3D



→ PCA

$$\underbrace{[\,200,\ 300,\ 400,\ \ldots,\ 700,\ \ldots\ 900\,]}_{n\,=\,250}$$

↳ Componentes seleccionados

Sesión 2

DATA → Modelo matemático → Output hypothesis

Hypothesis class → Modelo matemático

Loss function → Modelo matemático

optimización → Loss function

/ grad. descent

{ } → Hypothesis class

H → $|H| = \infty$ → x

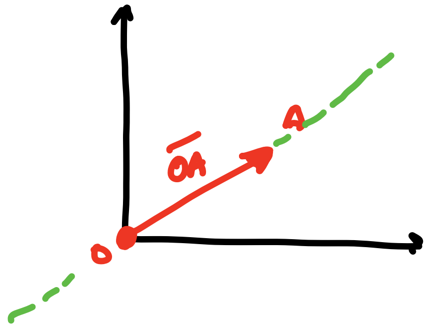
H → $|H| \neq \infty$ → x

→ **Vectores**

→ Física (geométrica)

→ Matemático ②

→ Software engineer



[_ , _ , _ , _ , ... , _]

- Vector es una tupla de elementos reales.

$\vec{v} = [a_1, a_2, a_3, \dots, a_n]$; $\vec{v} \in \mathbb{R}^n$

$\vec{v} = [\underbrace{1, 2, 3, 4}_{n=4}]$; $\vec{v} \in \mathbb{R}^4$

→ ["Hola", "que", "tal", ...]

vector numérico → [200, 2, 3, ...]

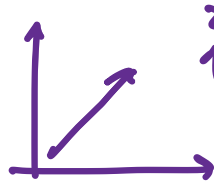
LIBRO → $[\_, \_, \_, \_, \dots, n]$; $\vec{v} \in \mathbb{R}^n$

120000

Memoria CPU

ML

Interpretabilidad.



2D

1200

Dimensionality Reduction

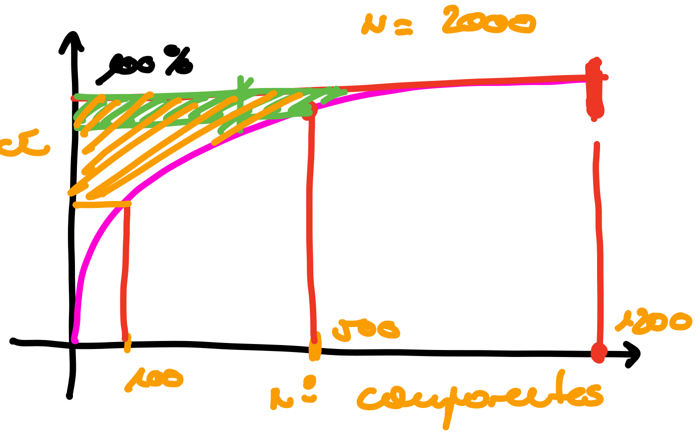
PCA

768

[1, 2, 30, 100, 200, 550, ..., 1560] $\to \mathbb{R}^N \to \mathbb{R}^{2000}$

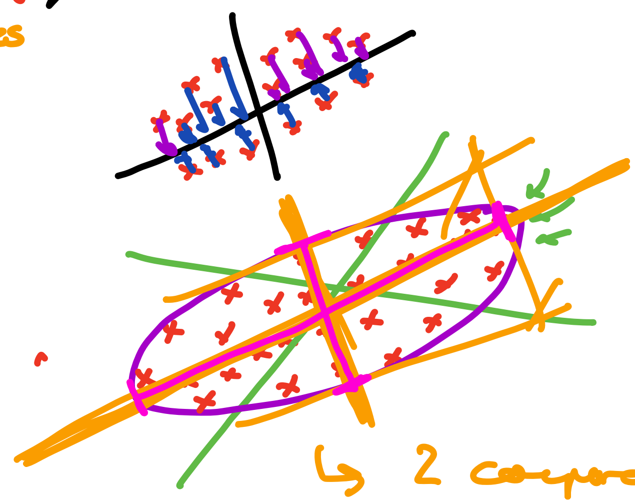
$N = 2000$

% de representación de los datos originales



PCA → maximiza la varianza de las componentes

vector $\mathbb{R}^{120000}$



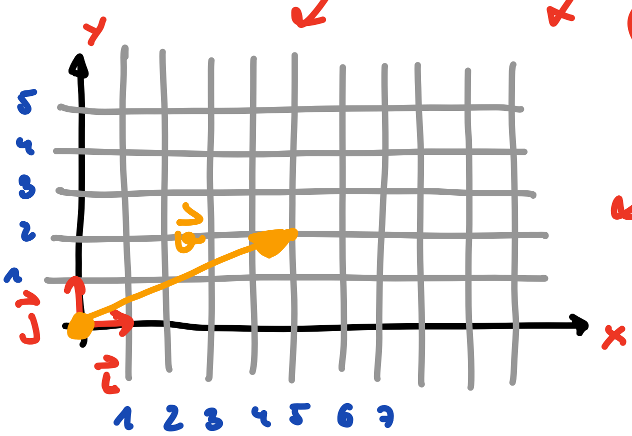
↳ 2 componentes

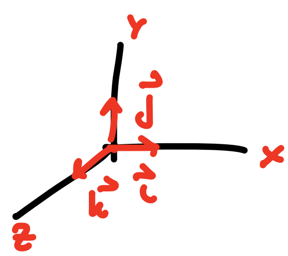
$w \in \mathbb{R}^{2000} \to \mathbb{R}^2$

RANGO

2D

↳ $\mathbb{R}^N$





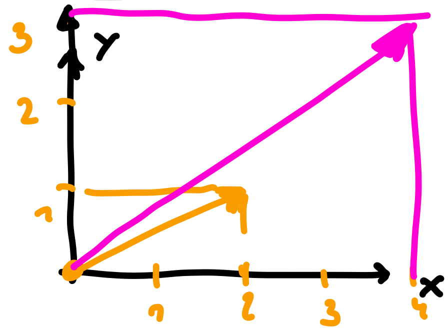
$\vec{u} = [5, 2]$

eje x

eje y

$\vec{u} = (5\vec{i}, 2\vec{j})$

- **Suma de un vector con un escalar**



$$\vec{v} = [2, 1], \quad a = 2$$

$$\vec{r} = (a + v_1, a + v_2) =$$

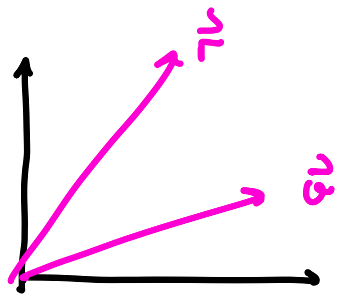
$$= (2 + 2, 2 + 1) = (4, 3)$$

$$\boxed{\vec{r} = (4, 3)}$$

- **Producto escalar / interno**

$$\vec{r} = (r_1, r_2, r_3, \dots r_n)$$

$$\vec{v} = (v_1, v_2, v_3, \dots v_n)$$



$$\boxed{\vec{v} \cdot \vec{r} = \sum_{i=1}^{n} v_i \cdot r_i} \rightarrow \vec{v} \cdot \vec{r} = v_1 \cdot r_1 + v_2 \cdot r_2 + v_3 \cdot r_3 + \dots + v_n \cdot r_n$$

PRODUCTO ESCALAR

$$\vec{r} = [2, 3, 1]$$

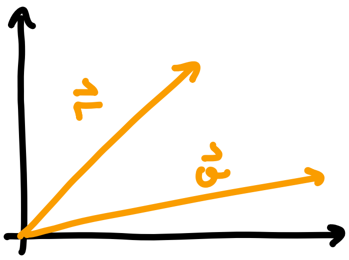
$$\vec{v} = [1, 4, 5]$$

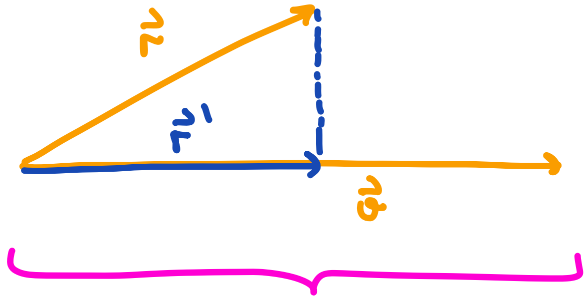
$$\vec{r} \cdot \vec{v} = \sum_{i=1}^{n} r_i \cdot v_i =$$

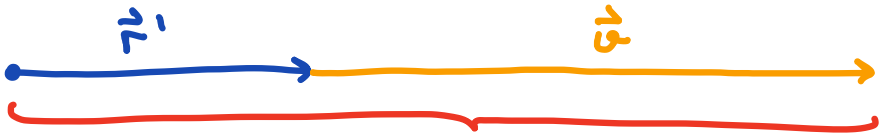
$$= 2 \cdot 1 + 3 \cdot 4 + 1 \cdot 5 =$$

$$= 2 + 12 + 5 = \boxed{19}$$

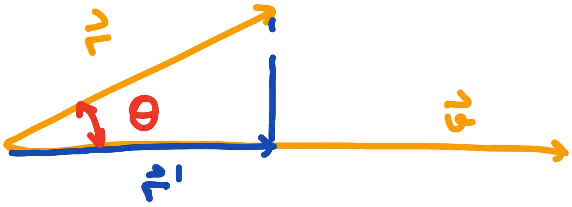
ESCALAR





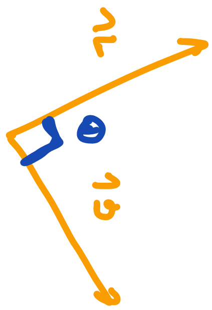


$= \boxed{19}$



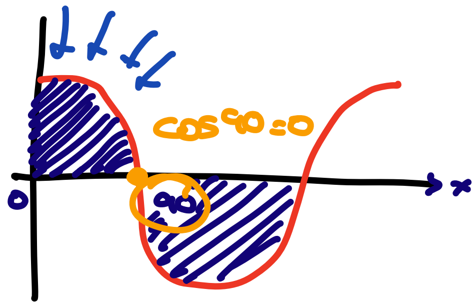
$$\boxed{\vec r \cdot \vec v = |r||v| \cos\theta}$$

Casos especiales



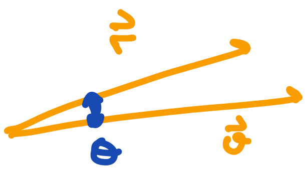
Caso 1

$$\boxed{\vec r \cdot \vec v = 0}$$



Caso 2

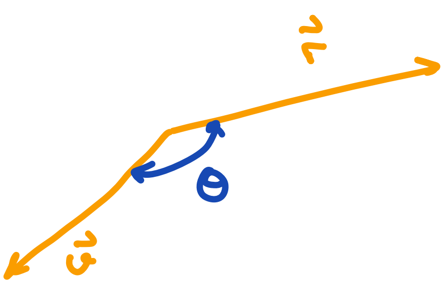
$\theta < 90°$



$$\boxed{\vec r \cdot \vec v > 0}$$

caso 3

$\theta > 90°$

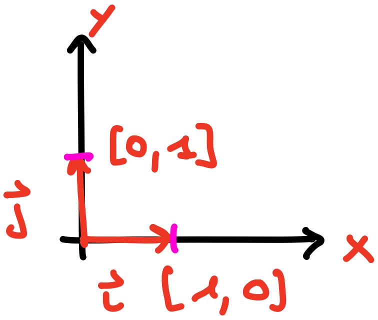


$$\boxed{\vec r \cdot \vec v < 0}$$

$\boxed{\vec{r}\cdot\vec{s} = 0}$ → vectores perpendiculares → **ORTOGONALES**

$|\vec{r}| = 1$

$|\vec{s}| = 1$ → **ORTONORMALES**



$$\begin{cases} \vec{i} = [1,0] \rightarrow \sqrt{1^2+0^2} = \boxed{1} & |\vec{i}| = 1 \\ \vec{j} = [0,1] \rightarrow \sqrt{0^2+1^2} = \boxed{1} & |\vec{j}| = 1 \end{cases}$$

$$\underbrace{\vec{i}\cdot\vec{j} = 1\cdot 0 + 0\cdot 1 = \boxed{0}}_{\text{ORTONORMALES}}$$

→ **Cosine Similarity**

$$\vec{r}\cdot\vec{s} = |\vec{r}||\vec{s}|\cos\theta$$

$$\boxed{\cos\theta = \frac{\vec{r}\cdot\vec{s}}{|\vec{r}||\vec{s}|}}$$

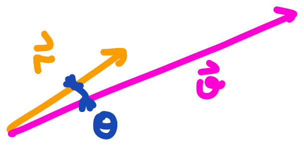
Cosine similarity

$$\boxed{\text{SIMILITUD} = \frac{\vec{r}\cdot\vec{s}}{|\vec{r}||\vec{s}|}}$$

prod. escalar (es un número)

prod del módulo de los vectores (número)

CASO 1



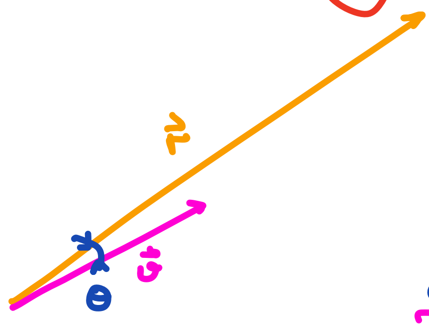
$|\vec{r}| = 2$

$|\vec{v}| = 4$

$\theta = 20°$

0.8

CASO 2



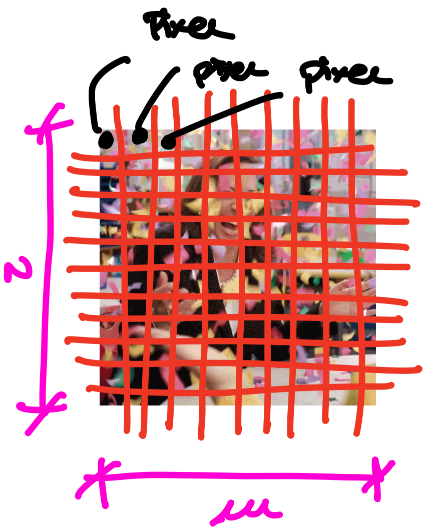
$|\vec{r}| = 20$

$|\vec{v}| = 3$

$\theta = 20°$

0.8

$$\boxed{\text{Similitud} = \frac{\vec{r}\cdot\vec{v}}{\underbrace{|\vec{r}||\vec{v}|}_{\text{normalizar}}}}$$



→ image (256 × 256) →

$$\begin{bmatrix} 256, & 128, & 20, & \dots, & 30 \\ 124, & 72, & 84, & \dots, & 20 \\ \vdots & \vdots & \vdots & \dots & \vdots \\ \vdots & \vdots & \vdots & \dots & \vdots \end{bmatrix}_{n \times m}$$

$$\to M_{256 \times 256}$$

$$\begin{bmatrix} 256, & 128, & 20, & \dots, & 30 \\ 124, & 72, & 84, & \dots, & 20 \\ \vdots & \vdots & \vdots & \dots & \vdots \\ \vdots & \vdots & \vdots & \dots & \vdots \end{bmatrix} \to [256, 128, 20, \dots, 30, 124, 72, 84, \dots, 20] \dots$$

$$\begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 9 \end{bmatrix}_{3\times 3} \to \underbrace{[\underbrace{1, 2, 3}_{1\ \text{fila}}, \underbrace{4, 5, 6}_{2\ \text{fila}}, \underbrace{7, 8, 9}_{3\ \text{fila}}]}_{\vec{v}}$$

$3 \times 3 = 9$

Tamaño

$$\begin{bmatrix} 9 & 8 & 7 \\ 6 & 5 & 4 \\ 3 & 2 & 1 \end{bmatrix} \longrightarrow [9, 8, 7, 6, 5, 4, 3, 2, 1]_{3\times 3 = 9}$$

TAMAÑO

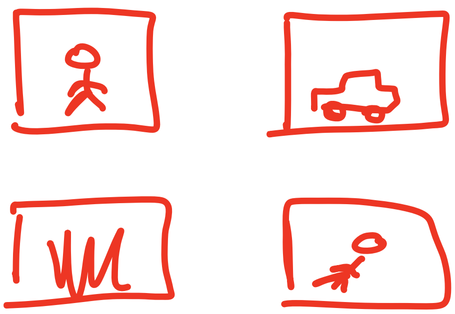
$$\text{SIMILITUD} = \frac{\vec{r}\cdot\vec{v}}{|\vec{r}||\vec{v}|}$$

Son parecidas o no.

Similitud.

contenido

color



0 → 256

256, 20, 50
--

256, 20, 50

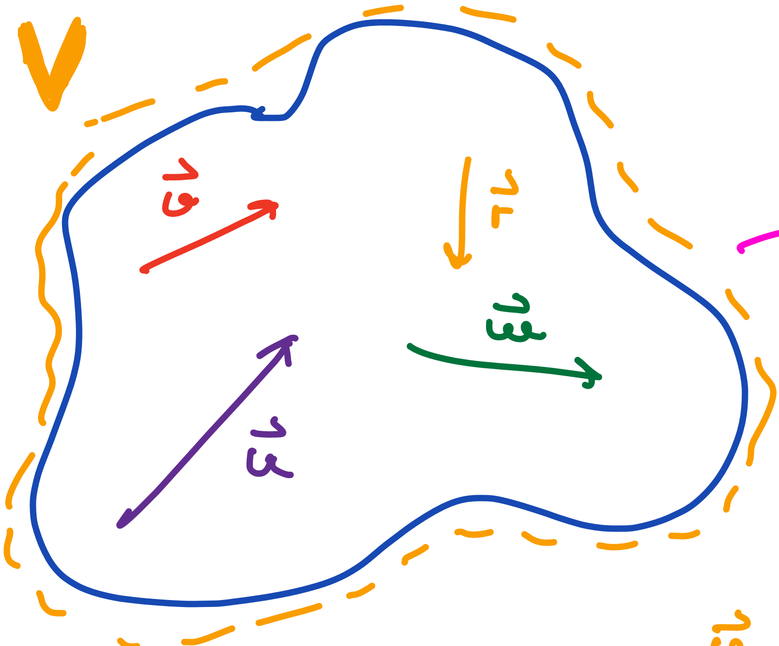
$$\begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 9 \end{bmatrix}$$

1 2 3 4 5 6 7 8 9

→ [1, 4, 7, 2, 5, 8, 3, 6, 9].

ESPACIOS VECTORIALES

Espacio vectorial → Dominio conjunto de vectores
Espacio vectorial → 3 dimensiones



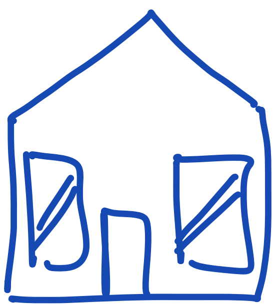
→ Tienen las mismas propiedades

↓

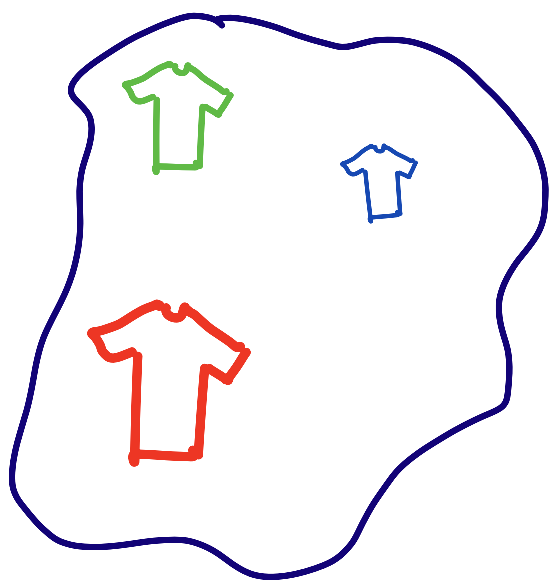
Operaciones con ellas

$\vec{v}$ { – Cosine similarity.

↳ $\underbrace{[\quad]}_{N}$, $\underbrace{[\quad]}_{N}$



- Precio
- TALLA
- COLOR
- Tipo
- Tipo Tela
- H/M



[25 €, 40, amarillo, camiseta, algodón, H]

[72 €, 43, Azul, camiseta, algodón, M]

...

- Talla
- precio
- Tipo
- Tipo tela
- H/M

[25 €, 40, amarillo, camiseta, algodón, H]

1 720 23 1

[72 €, 43, Azul, camiseta, algodón, M]

3 720 23 2

[25, 40, 1, 720, 23, 1]

[72, 43, 3, 720, 23, 2]

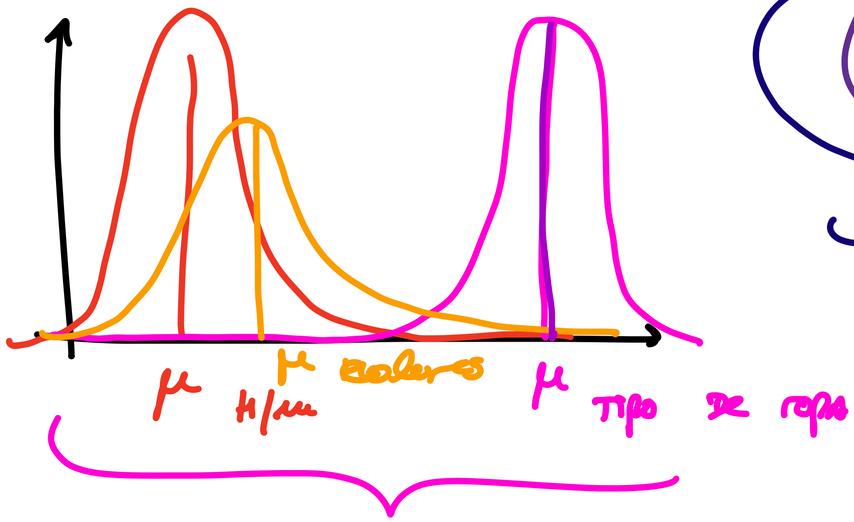
ejemplo.

(100 – 1500)

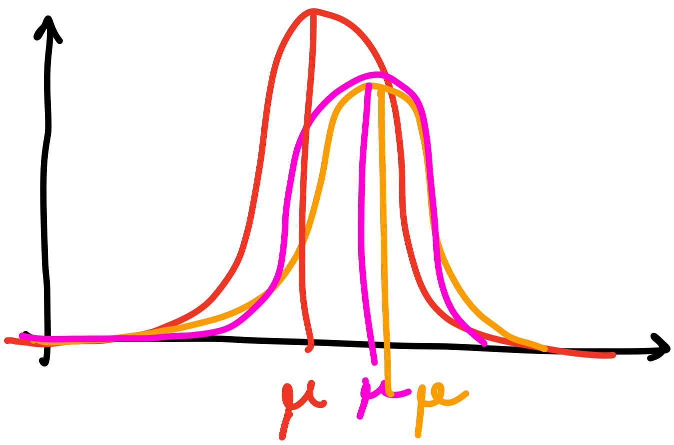
SIMILITUD

(-2 → 2)

→ NORMALIZACIÓN



$$\underbrace{\left(\frac{x_i - \mu}{\sigma}\right)}_{\approx 0} \rightarrow \text{PCA}$$



$$\vec{v}_1 = [1, 2]$$

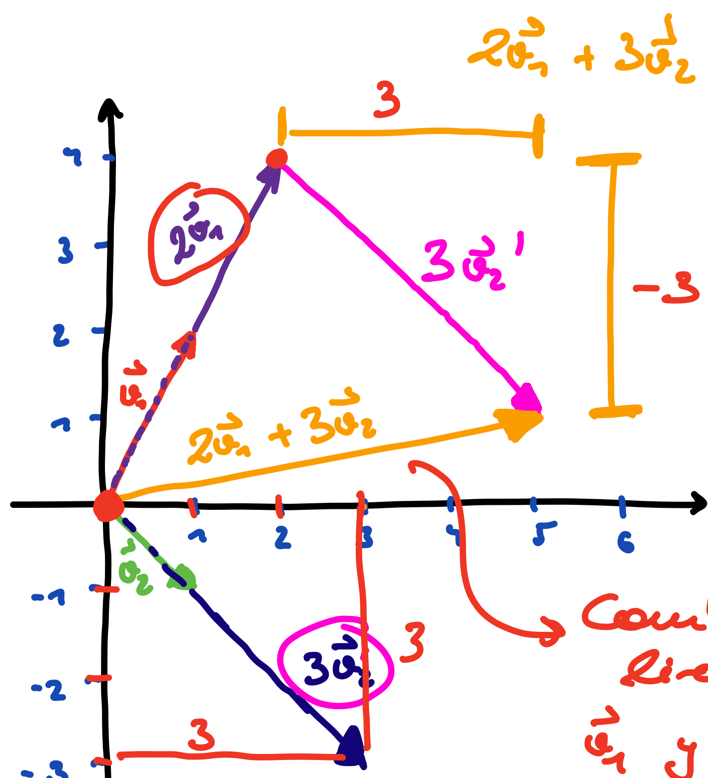
$$\vec{v}_2 = [1, -1]$$

$$\vec{z} = a_1 \cdot \vec{v}_1 + a_2 \cdot \vec{v}_2$$

$$\vec{z} = 2\vec{v}_1 + 3\vec{v}_2$$

$2\vec{v}_1 \rightarrow \vec{v}_1 = [1, 2]$

$2 \cdot \vec{v}_1 = [2, 4]$



Combinación lineal de $\vec{v}_1$ y $\vec{v}_2$

$3\vec{v}_2 \rightarrow$

$3 \cdot [1, -1] = [3, -3]$

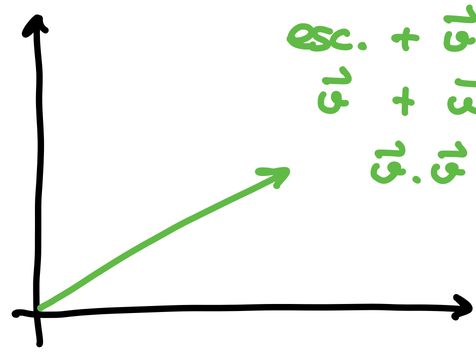
$2\vec{v}_1 + 3\vec{v}_2 =$

$= [2, 4] + [3, -3] =$

$= \boxed{[5, 1]}$

Sesion 3

Vectores
- Física
- matemát.
- Soft. eng.



esc. + $\vec{v}$

$\vec{v} + \vec{w}$

$\vec{v} \cdot \vec{v}$ (prod. esc.)

$\vec{v} = [-, -, -, \dots, -]$ $\quad \vec{v} \in \mathbb{R}^N$

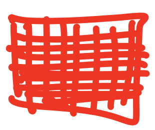
$< 90°$

$> 90°$

$$\boxed{\text{similitud} = \frac{\vec{r} \cdot \vec{v}}{|\vec{r}|\,|\vec{v}|}}$$

Cosine Similarity

→ Imagenes → [—, —, —, ...]



→ Texto → [—, —, —, ..., —]

Bag of words

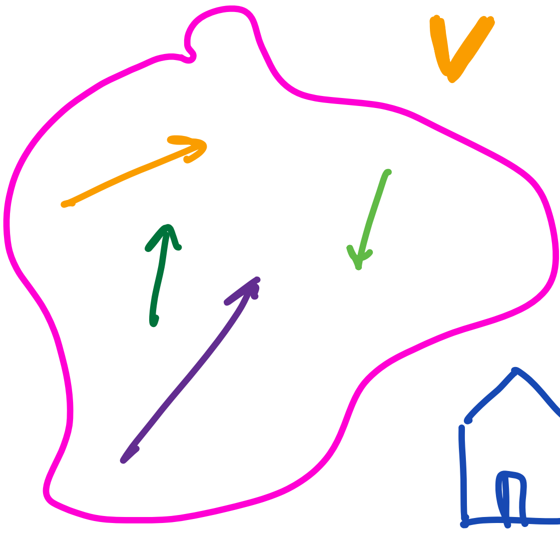
perro

[illegible]

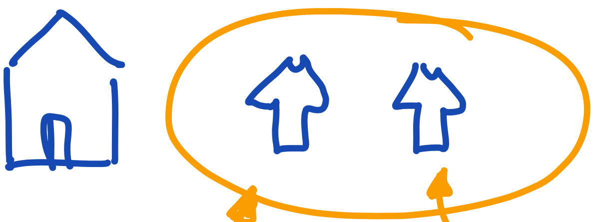
...

$\left\{ \begin{array}{l} [1, 0, 1, 1] \\ [0, 1, 1, 1] \end{array} \right.$

→ Espacio Vectorial

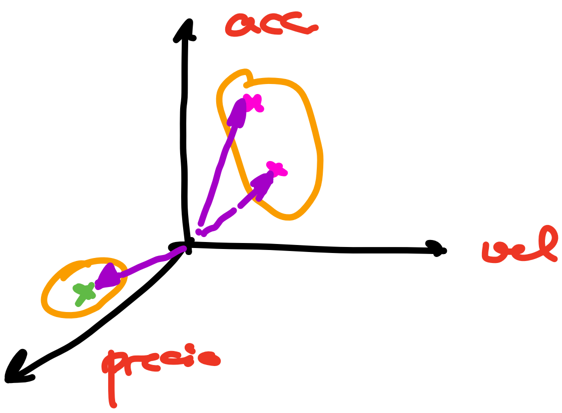


Han sido creados usando las mismas propiedades



vector 1 vector 2

cosine similarity



MATRICES

→ Conjunto de vectores

→ Conjunto de valores que son accesibles ($M_{2\times 2}$)

→ " " " + operaciones.

$$
A = \begin{array}{c} \overbrace{\hphantom{1 \quad 2 \quad 3}}^{N} \\ \begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 9 \end{bmatrix} \\ m \times n \end{array} \left.\vphantom{\begin{bmatrix} 1 \\ 4 \\ 7 \end{bmatrix}}\right\} m
$$

m → filas

N → columnas.

vector $\rightarrow$ $\underbrace{[3, 5, 7, 8]}_{4}$; $\vec{v} \in \mathbb{R}^4$

matriz $\rightarrow$ $\begin{bmatrix} 3 & 6 \\ 4 & 7 \\ 5 & 8 \end{bmatrix}_{3\times2}$; $A \in \mathbb{R}^{3\times2}$

$\cancel{d}\,(=1) \times m \times N$ $\mathbb{R}^{d\times m\times N}$ $\longrightarrow$ $A_{(m\times N)} \in \mathbb{R}^{(\cancel{m}\,(=1) \times N)}$ $\longrightarrow$ $\vec{v}_N \in \mathbb{R}^{(\cancel{m}\,(=1) \times N)}$

TENSOR

MATRIZ $\rightarrow$ VECTOR

$$\begin{bmatrix} 3 & 6 \\ \cancel{4} & \cancel{7} \\ \cancel{5} & \cancel{8} \end{bmatrix}_{\cancel{3}\times 2 \;\to\; 1\times 2} \Big\}\, m \longrightarrow \underbrace{[3, 6]}_{N}$$

$\underbrace{\quad\quad}_{N}$

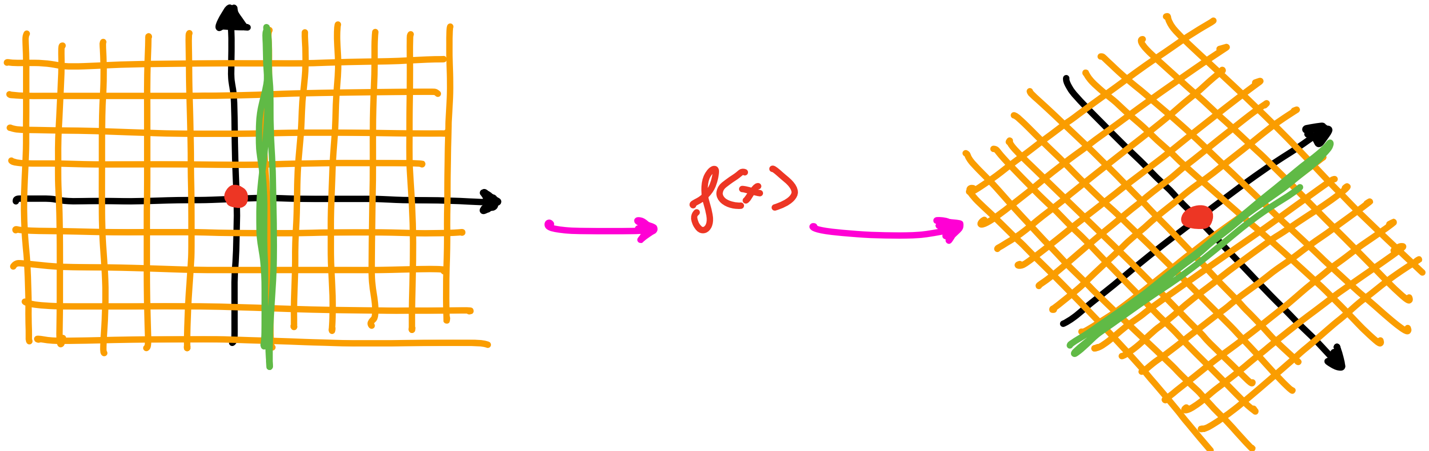
→ TRANSFORMACIÓN LINEAL

- Operación que transforma un vector/entidad en otro en el mismo vect.
- TRANSF. → modificar algo con algunas reglas
  LINEAL → no potencias.
- ¿¿¿¿¿¿¿¿¿

TRANS. LINEAL → $f(x)$

$2 \rightarrow \boxed{f(x)} \rightarrow 8$

25 $\quad\quad$ 25

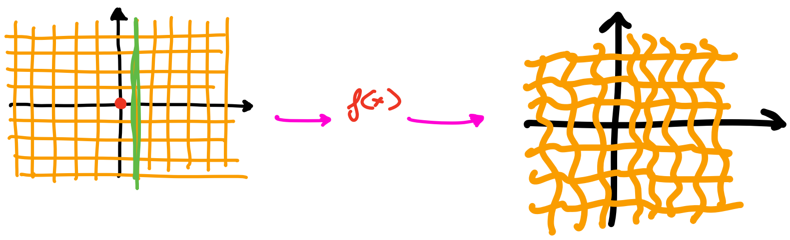


Transf. Lineal

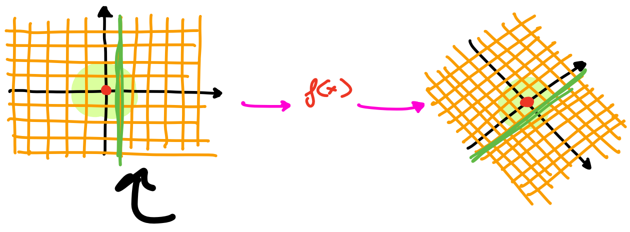
Mantiene la proporción

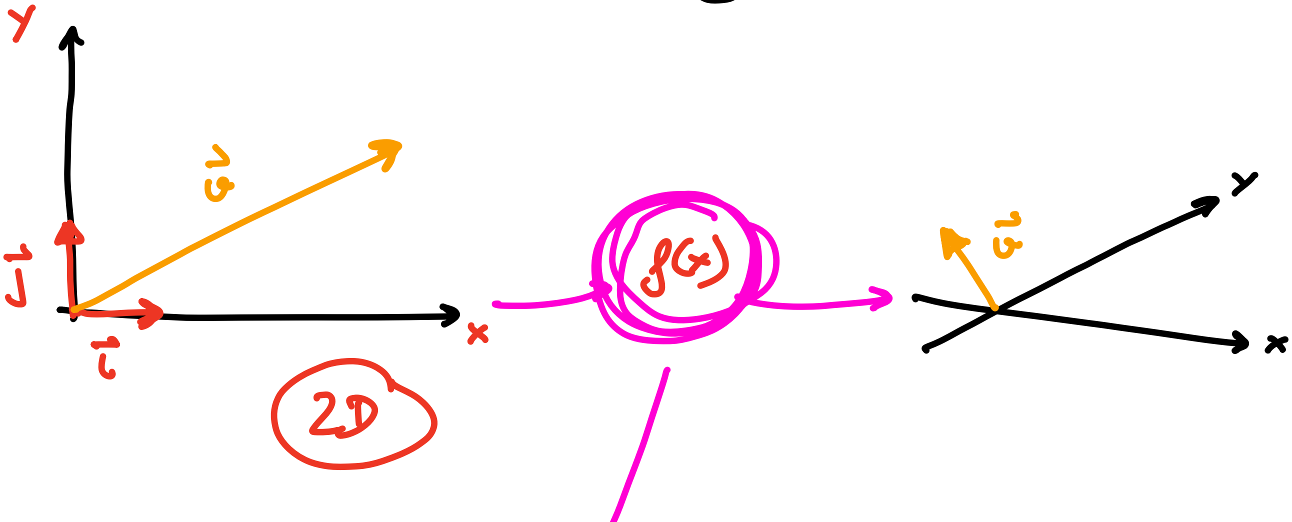
No hay distorsión

Las líneas siguen siendo rectas.



El centro permanece en la misma posición.

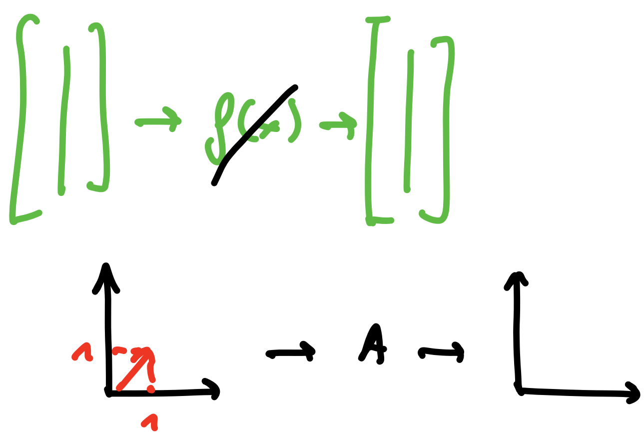




Donde acaban las componentes del vector en el eje $\vec{i}$

Donde acaban las comp. del vector en el eje $\vec{j}$ dep. de la transf.

$$f(x) = A = \begin{array}{cc} \begin{bmatrix} 1 & 3 \\ -2 & 0 \end{bmatrix} \\ \vec{i} \quad \vec{j} \end{array}$$



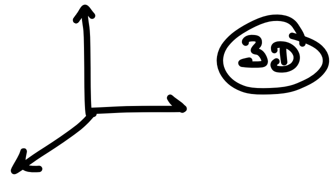
$$\begin{pmatrix} 1 \\ 1 \end{pmatrix} \rightarrow [1, 1]$$

$$\begin{pmatrix} 1 \\ 1 \end{pmatrix} \rightarrow \underbrace{\begin{bmatrix} 1 & 3 \\ -2 & 0 \end{bmatrix}}_{\text{eje } x \quad \text{eje } y} \rightarrow ¿?$$

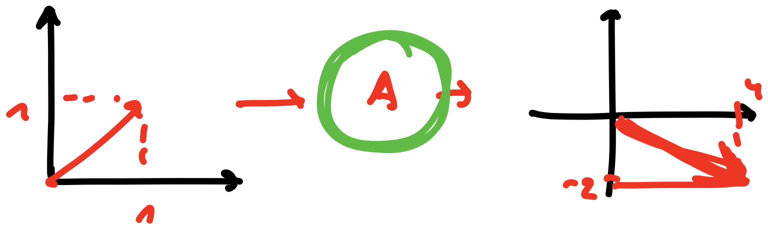
$f(x)$

$$\begin{bmatrix} 1x + 3y \\ -2x + 0y \end{bmatrix} \rightarrow \begin{bmatrix} 1 \cdot 1 + 3 \cdot 1 \\ -2 \cdot 1 + 0 \cdot 1 \end{bmatrix} = \begin{bmatrix} 1 + 3 \\ -2 + 0 \end{bmatrix} = \begin{bmatrix} 4 \\ -2 \end{bmatrix}$$

$$\begin{pmatrix}1\\1\end{pmatrix} \rightarrow A \rightarrow \begin{pmatrix}4\\-2\end{pmatrix}$$



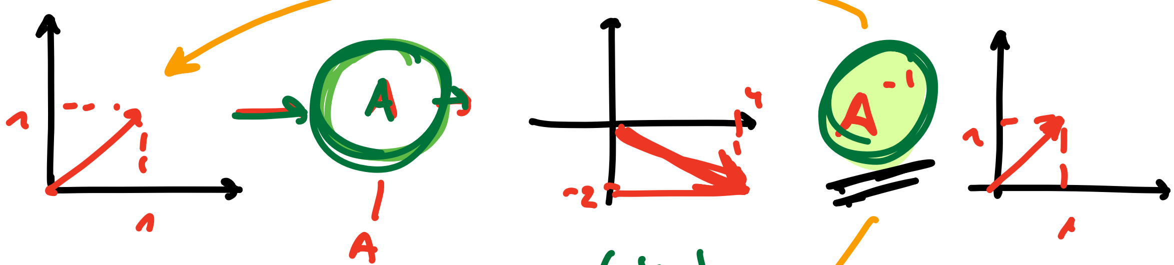
[–, –, –, ..., –]



$$\boxed{Ax = B}$$

$$\begin{pmatrix}1\\1\end{pmatrix} \qquad \begin{pmatrix}4\\-2\end{pmatrix}$$

→ Inversa



$$\begin{pmatrix}1\\1\end{pmatrix} = x \qquad B = \begin{pmatrix}4\\-2\end{pmatrix} \qquad x = \begin{pmatrix}1\\1\end{pmatrix}$$

$$Ax = B$$

$$\underbrace{\cancel{A^{-1} \cdot A}}_{=1} x = A^{-1} \cdot B$$

$$\overset{=1}{\cancel{I}} x = A^{-1} \cdot B$$

$$\boxed{x = A^{-1} \cdot B}$$

$$A = \begin{bmatrix}1 & 2\\ 3 & 0\end{bmatrix} ; \quad A^{-1} = \begin{bmatrix}3 & 1\\ 0 & 2\end{bmatrix}$$

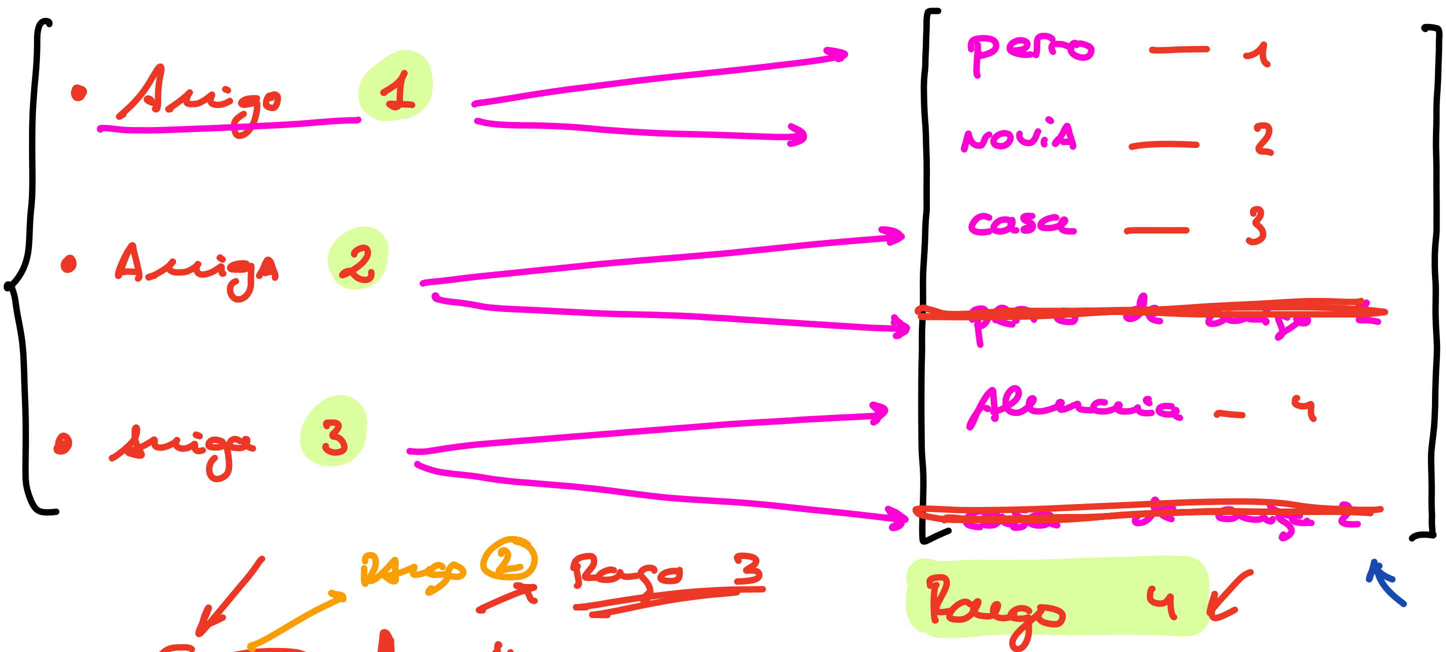
$$x = \begin{pmatrix}1\\1\end{pmatrix}$$

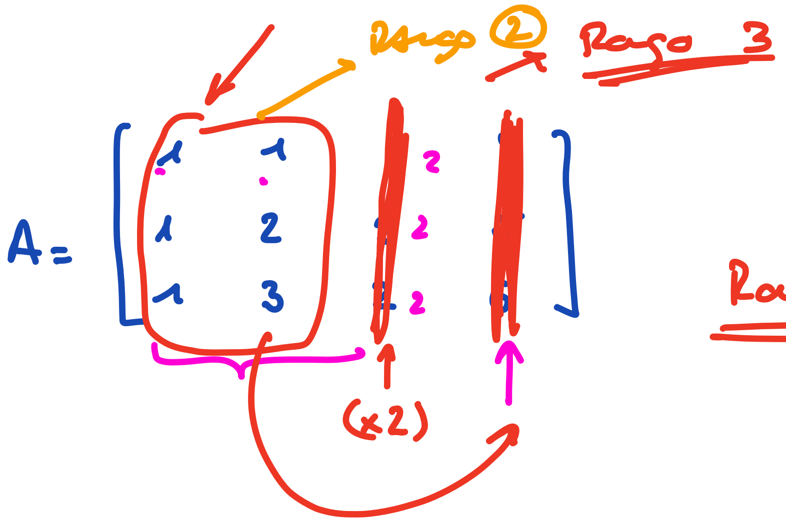
$$B = \begin{pmatrix}4\\-2\end{pmatrix}$$

$$1x + 2y = ¿a$$
$$3x + 0y = ¿b$$

→ Rango

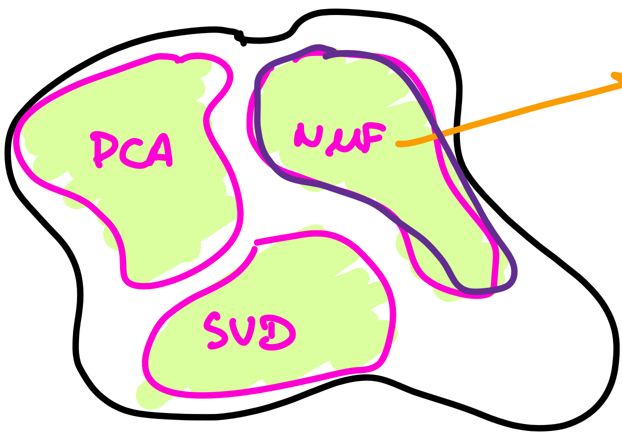
↳ es la cantidad de información única que contiene la entidad.





Rango = 2

→ Dictionary Learning



non-negative matrix factorization

$$(A - DR) = 0$$

DATOS

$\approx \hat{A}$

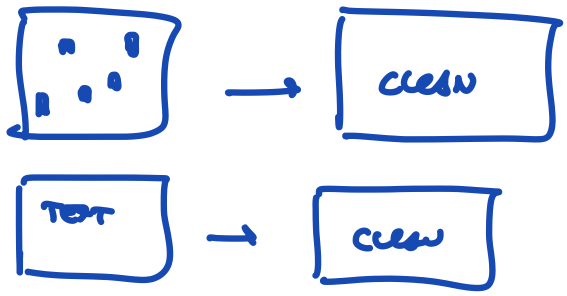
$(A - DR) = 0$ Loss Function Dict. learning.

Reducir los datos restando el prod. del $D \times R$.

D y R

$(D \times R) \approx A \rightarrow (A - D.R \approx 0)$

→ Recont. imagen



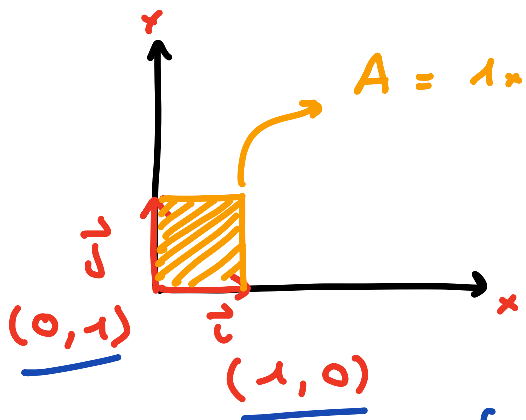
A 250000×1200000 → • D • R → $\approx A$

→ D, R

↳ Rango sea lo menor posible.

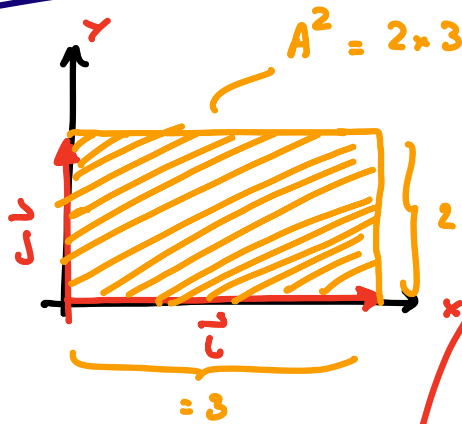
→ **Determinante** → Representa a cada matriz.

↳ Una medida de la distorsión del vector tras la transf. lineal.



$A = 1 \times 1 = \boxed{1}$

$\rightarrow \begin{bmatrix} 3 & 0 \\ 0 & 2 \end{bmatrix} \rightarrow$



$A^2 = 2 \times 3 = 6$

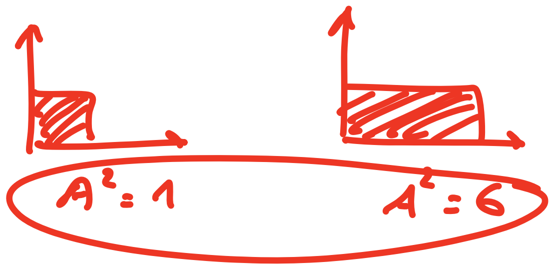
$\vec{i} = (1, 0) \rightarrow \begin{bmatrix} 3 & 0 \\ 0 & 2 \end{bmatrix} = \begin{pmatrix} 3x + 0y \\ 0x + 2y \end{pmatrix} \rightarrow \begin{pmatrix} 3.1 + 0 \\ 0.1 + 2.0 \end{pmatrix} = \begin{pmatrix} 3 \\ 0 \end{pmatrix}$

$$\vec{j} = (0,1) \rightarrow \begin{bmatrix} 3 & 0 \\ 0 & 2 \end{bmatrix} = \begin{pmatrix} 3x + 0y \\ 0x + 2y \end{pmatrix} \rightarrow \begin{pmatrix} 3.0 + 0.1 \\ 0.0 + 2.1 \end{pmatrix} = \begin{pmatrix} 0 \\ 2 \end{pmatrix}$$

$$\det(A) = \det\left(\begin{bmatrix} 3 & 0 \\ 0 & 2 \end{bmatrix}\right) = 3 \times 2 - 0 \times 0 = \boxed{6}$$

$$\boxed{\det(A) = 6}$$

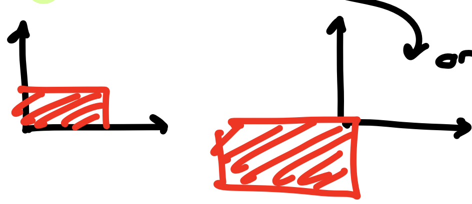
$$\boxed{\det(A) = 6}$$

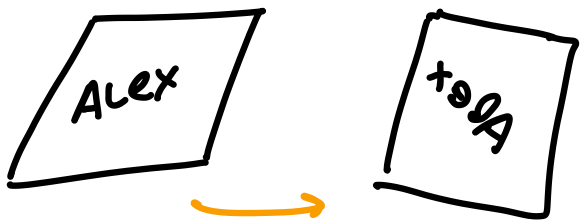


$$\boxed{\det(A) = -6}$$

factor

orientación





CAMBIO DE ORIENTACIÓN

Sesion 4

$$A = \begin{bmatrix} | & | & | & | \end{bmatrix} \qquad A \in$$

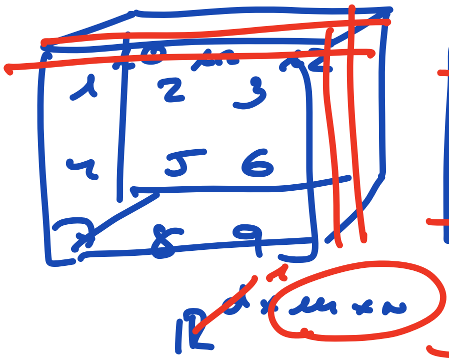
$$\cancel{d}^{=1} \times m \times N \longrightarrow A_{(m \times N)} \in \mathbb{R}^{(\cancel{m}^{=1} \times N)} \longrightarrow \vec{v}_N \in \mathbb{R}^{(\cancel{m}^{=1} \times N)}$$

$$\mathbb{R}^{d \times m \times N}$$

TENSOR

MATRIZ

VECTOR



$\mathbb{R}^{\cancel{1} \times m \times n}$

$$\begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 9 \end{bmatrix}$$

$m \times n = 3 \times 3 = 1 \times 3$

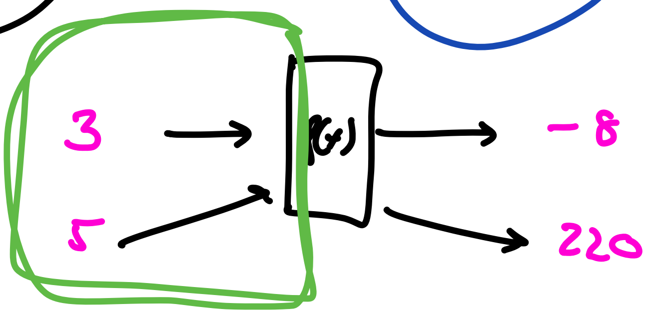
$$\overbrace{\begin{bmatrix} 4 & 5 & 6 \end{bmatrix}}^{n}$$

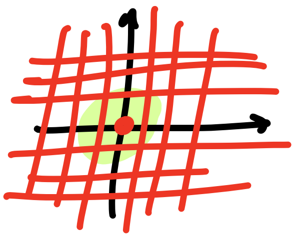
$\mathbb{R}^{(\cancel{1} \times 3)}$

TRANS. Lineal $\longrightarrow$ $f(x)$

$3 \longrightarrow f(x) \longrightarrow -8$

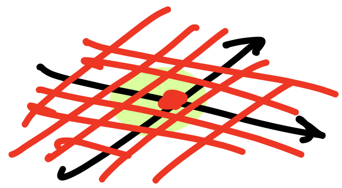
$5 \longrightarrow f(x) \longrightarrow 220$





1º No hay distorsión

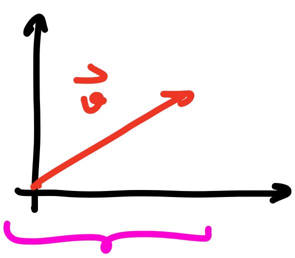
Centro fijo.



$$f(x) = \begin{bmatrix} 3 & 1 \\ 0 & 2 \end{bmatrix}$$

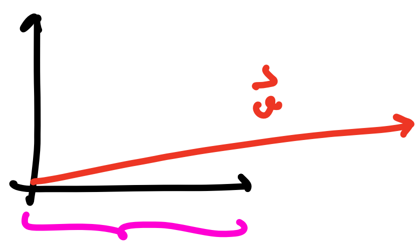
$\parallel$ A

Matriz de transf.



$\vec{v}$

$\underbrace{}_{X}$ $\longrightarrow$ $\underbrace{A}_{\text{Matriz de transf.}}$ $\longrightarrow$ $\underbrace{}_{B}$



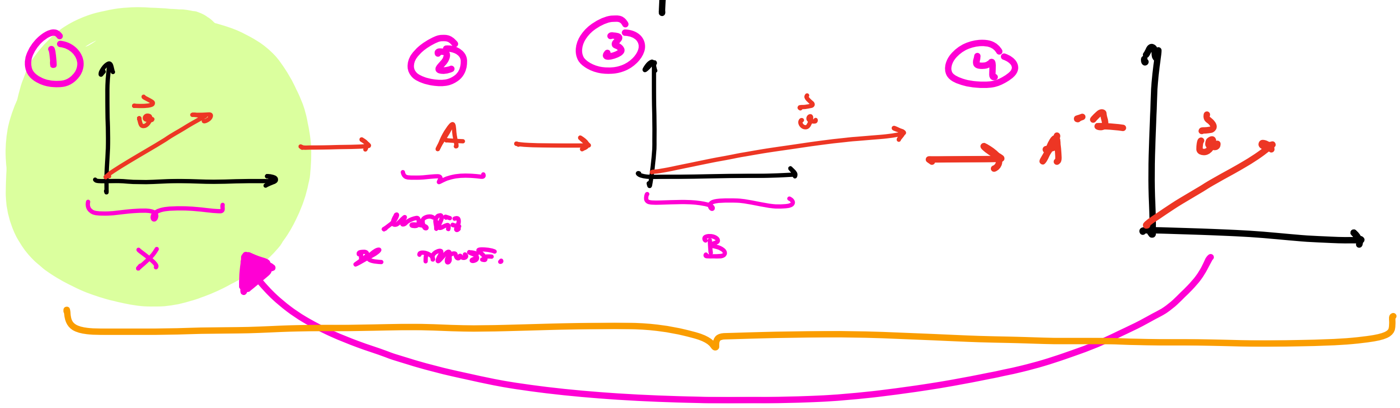
$\vec{v}$

$$\boxed{Ax = B}$$

Trans. Lineal.

input salida

→ matriz inversa A^{-1}



$$Ax = B$$

$$\underbrace{A^{-1} \cdot A}_{Id} x = A^{-1} \cdot B$$

$$\boxed{x = A^{-1} \cdot B}$$

entrada de la transf.

$\boxed{Ax = B}$

$\boxed{x = A^{-1} \cdot B}$

→ matrices cuadradas con $\det(A) \neq 0$

S. DET.

$\det(A) \neq 0 \rightarrow A^{-1} \rightarrow x = A^{-1} \cdot B \rightarrow$ única solución

$\det(A) = 0$

→ Regla gauss → inf. soluciones (S.C.I) ↓ simplificar

→ Regla gauss → no tiene solución (S.I)

Sistemas de ecuaciones lineales.

→ Regresión lineal → Mínimos cuadrados (least squares)

→ grad. Descent.

Datos → modelo mat. → output hypothesis

$H = \{ \}$ → modelo mat.

Loss function → modelo mat.

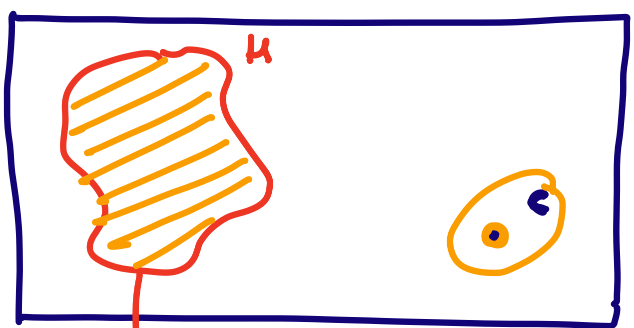
optimización → Loss function

↳ número de elementos en el conjunto

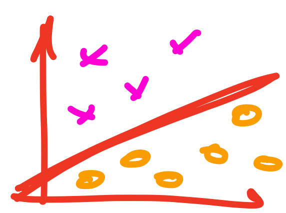
$|H| = \infty$ → Deep learning

↳ PAC learnable

↳ $|H| = \infty$; nuestro modelo puede ser aprendido.



Universal function space



↳ elegimos el resolver un prob. de ML

- SVM
- Modelo lineal
- LSTM

→ Si c, que es el "mejor modelo" no está incluido dentro de nuestra H, no lo podemos aprender.

→ 0-1 loss function

↳ no es derivable

↳ no convexa

} optimización

0-1 loss function → loss convexas funct. / derivables

surrogate loss functions

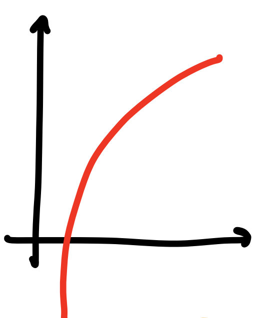
↳ classifier calibrated

→ **Least Squares**

$$\boxed{h_\theta = \arg\min \frac{1}{N}\sum_{i=0}^{N}(y_i - h(x_i))^2}$$

- Hinge loss (svm)
- Cross - entropy (—)

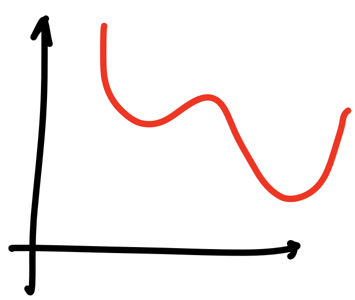
- Cauchy loss → no conv.



Fun ①

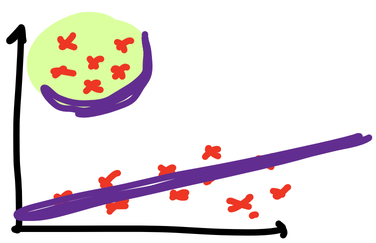
↙

<u>clase</u>



Fun ②

↙



$f(x) = e^x$

→ Convex : Cuando la Hessian Matrix de la función $f(x)$ es positive semidef.

Jaccobian Matrix — 1° derivada parcial.

Hessian Matrix — 2° derivada parcial de la función

→ Los autovalores de la Hessian Matrix sean > 0.

$\underbrace{\qquad\qquad\qquad\qquad\qquad\qquad}_{\text{Definición de convexo.}}$

$$f(x, y) = e^{\frac{x}{2}} \sin(y)$$

1° derivada

$$\frac{\partial f(x,y)}{\partial x} = \frac{1}{2} e^{\frac{x}{2}} \cdot \sin(y)$$

$$\frac{\partial f(x,y)}{\partial y} = e^{\frac{x}{2}} \cdot \cos(y)$$

$$f(x) = x^2$$

$$\frac{df(x)}{dx} = 2x \quad \text{1° derivada}$$

$$\frac{d^2 f(x)}{dx^2} = 2 \quad \text{2° derivada}$$

$$\nabla f(x,y) = \begin{bmatrix} \frac{\partial f(x,y)}{\partial x} \\ \frac{\partial f(x,y)}{\partial y} \end{bmatrix} = \boxed{\begin{bmatrix} \frac{1}{2} e^{\frac{x}{2}} \cdot \sin(y) \\ e^{x/2} \cdot \cos(y) \end{bmatrix}}$$

nabla

x
y
x
y

Jaccobian Matrix

2° derivada

$$H = \underbrace{\begin{bmatrix} \frac{\partial^2 f}{\partial x^2} & \frac{\partial^2 f}{\partial x \partial y} \\ \frac{\partial^2 f}{\partial y \partial x} & \frac{\partial^2 f}{\partial y^2} \end{bmatrix}}_{\text{matrix Hessiana}} = \boxed{\begin{bmatrix} \frac{1}{4} e^{\frac{x}{2}} \sin(y) & \frac{1}{2} e^{\frac{x}{2}} \cdot \cos(y) \\ \frac{1}{2} e^{\frac{x}{2}} \cdot \cos(y) & -e^{\frac{x}{2}} \sin(y) \end{bmatrix}}$$

$$\frac{\partial f(x,y)}{\partial x} = \frac{1}{2} e^{\frac{x}{2}} \cdot \sin(y)$$

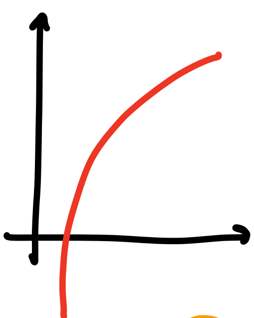
$$\frac{\partial f(x,y)}{\partial y} = e^{\frac{x}{2}} \cos(y)$$

Autovalores de la matrix hessiana.

$$\lambda \geq 0$$

$$e^x = e^x \cdot 1$$

$$e^{x/2} = e^{\frac{x}{2}} \cdot \frac{1}{2}$$



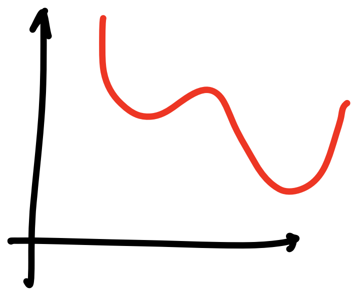
Fun ①

CLASE

$$f(x) = \ln(x)$$

$$\frac{df(x)}{dx} = \frac{1}{x}$$

$$\frac{d^2 f(x)}{dx^2} = -\frac{1}{x^2}$$

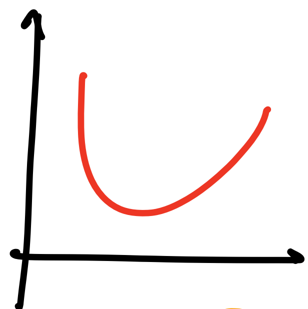


Fun ②

$$f(x) = \underbrace{x^3 + x^2}$$

$$\frac{df(x)}{dx} = 3x^2 + 2x$$

$$\frac{d^2 f(x)}{dx^2} = 6x + 2$$



Fun ③

$$(x-y)^2$$

$$\begin{cases} \frac{\partial f(x,y)}{\partial x} = 2(x-y) \\ \frac{\partial f(x,y)}{\partial y} = -2(x-y) \end{cases}$$

$$2(x-y) \rightarrow$$

$$2x - \cancel{2y} = 0 \rightarrow 2$$

$$A = H = \begin{bmatrix} 2 & -2 \\ -2 & 2 \end{bmatrix}$$

$$\geq 0 \rightarrow \text{Convex}$$

$\det(A - \lambda I) \rightarrow 2 \ \lambda$

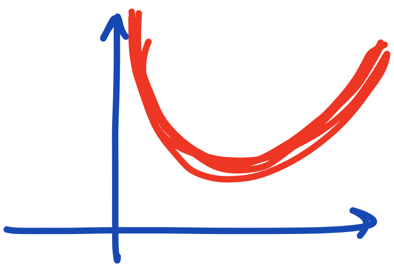
True

Salida del modelo

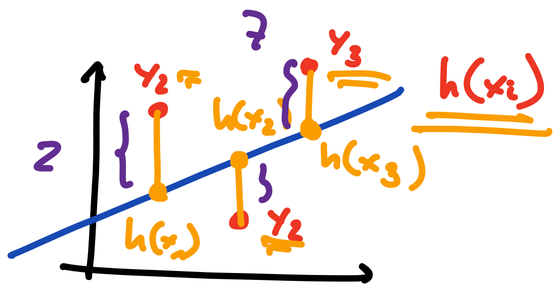
$$\boxed{w_{LS} = \arg\min_{w} \sum_{i=1}^{N} (y_i - h(x_i))^2}$$

Least squares

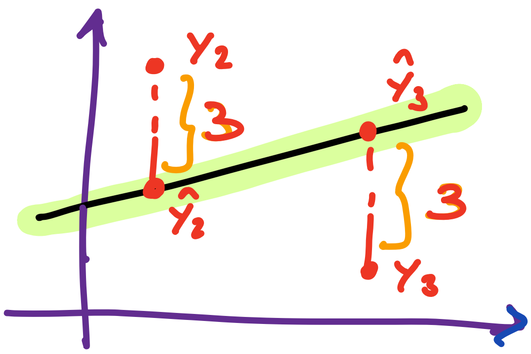
mínimos cuadrados



$\rightarrow$ convex



$\rightarrow$ estandarizar

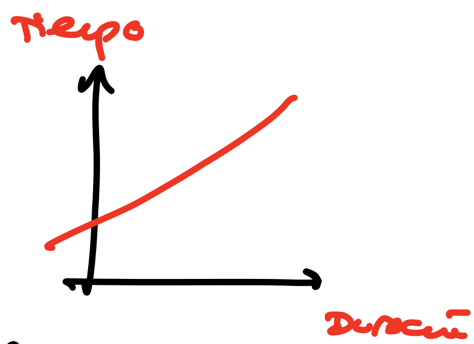


1 variable

$2 + 3 + 7 \rightarrow \bar{x}$

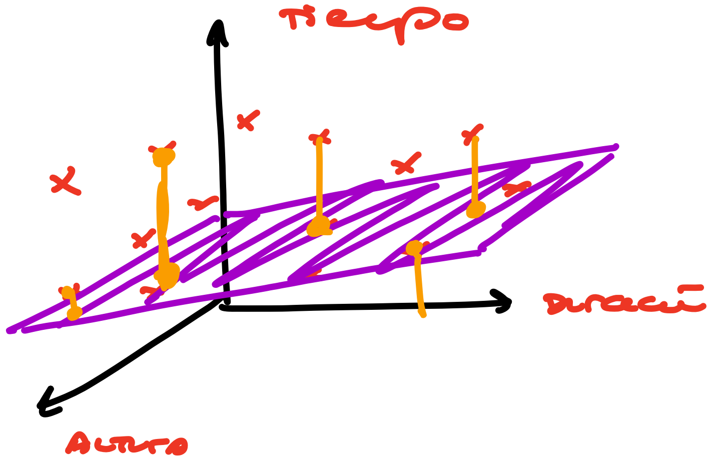
$$\sum_{i=1}^{N} (y_i - \hat{y}_i)^2 \Rightarrow$$

$$\underbrace{y_i^2 - 2 y_i \cdot \hat{y}_i + \hat{y}_i^2}_{0}$$



$$\underbrace{\text{Tiempo}_{\text{espera}} = w_0 + \text{duración} \cdot w_1}$$

$$\text{Tiempo}_{\text{espera}} = w_0 + \text{duración} \cdot w_1 + \text{altura} \cdot w_2$$



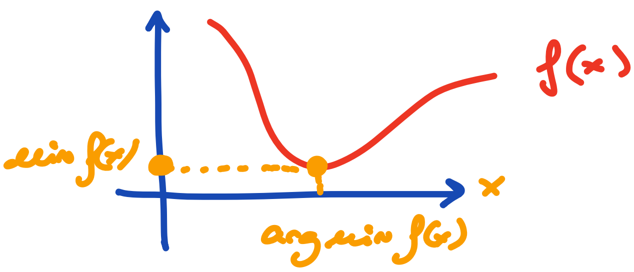
→ sis. ecuaciones

$$\boxed{w' = (x^T \cdot x)^{-1} \cdot x^T \cdot y}$$

↳ x → lineal
  ↳ polinomial

↳ y → $\begin{bmatrix} \\ \\ \end{bmatrix}$

$$\boxed{w_{LS} = \arg\min_{w} \sum_{i=1}^{N} (y_i - h(x_i))^2}$$



argmin f(x) → valores que hacen mínima a una func.

→ mínimo de la func.

$$\boxed{\frac{df(x)}{dx} = 0}$$ mínimo

→ mínimo $f(x,y)$

$$\boxed{\nabla f(x,y) = \left[\frac{\partial f(x,y)}{\partial x} , \frac{\partial f(x,y)}{\partial y} \right]}$$

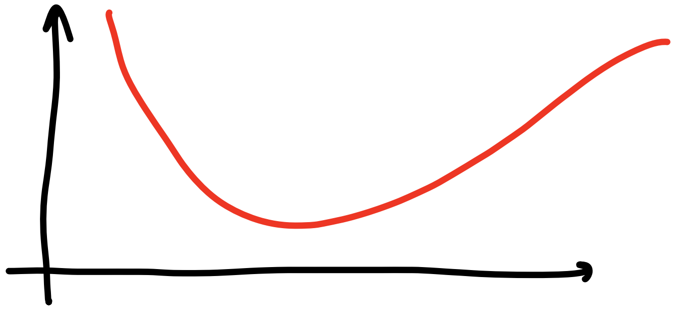
Derivada respecto a x

derivada respecto a y

$$\boxed{\nabla f(x,y) = 0}$$

$$\underbrace{\nabla f(x,y)}_{\text{2 variables}} = \left[\frac{\partial f(x,y)}{\partial x} , \cancel{\frac{\partial f(x,y)}{\partial y}} \right]$$

↳ 1 variable → ↳ $\left[\frac{\partial f(x,\cancel{y})}{\partial x} \right] \rightarrow \left[\frac{df(x)}{dx} \right]$

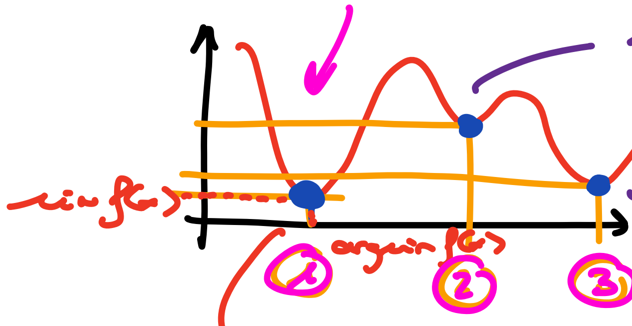


CONVEX

↳ Hessiana

λ ≥ 0

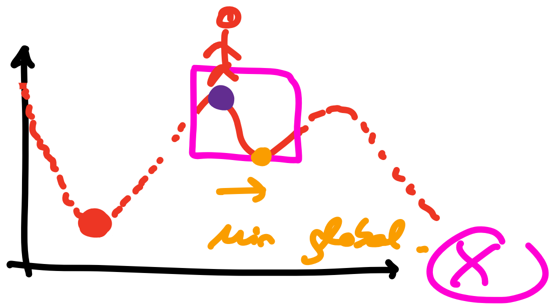
mínimo global



min local

min local

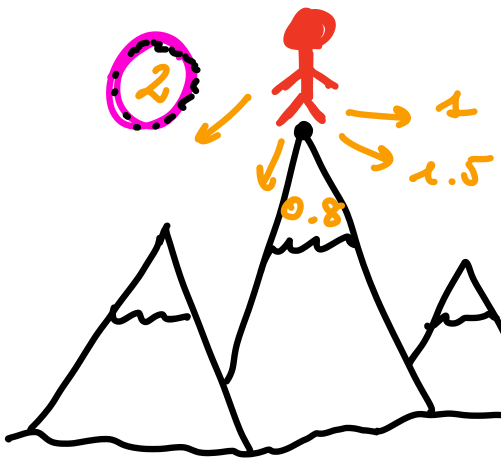
global min.



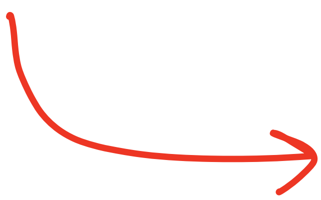
→ **Convexa** → tener un solo mínimo que coincide con el mínimo global.

→ Gradient Descent

STEP 1



Base

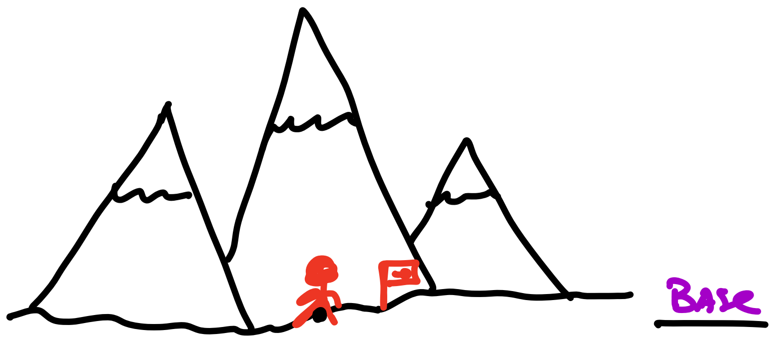


step 2

Base

step 3

step 4



gradient DESCENT

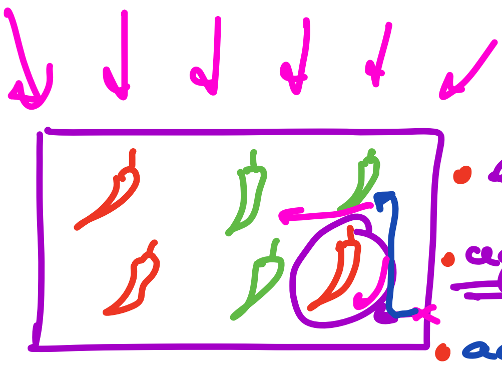
↳ un método iterativo de optimización.

gradient descent → gradient

descent → Busca el valor mínimo

ascent → Busca la máxima de una prob.

Reinforcement learning.



- env. → habitación.
- agent
- actions

} rewards.

-20

+20

+20

-20

⋮

Gradient Ascent.

- <u>gradiente</u>

$$f(x,y) = x^2y + xy^2$$

$$\nabla f(x,y) = \left[\frac{\partial f(x,y)}{\partial x}, \frac{\partial f(x,y)}{\partial y}\right] =$$

$$= \underbrace{\left[2xy + y^2, \; x^2 + 2xy\right]}_{\text{Derivada 1ª de } f(x,y)}$$

$\nabla f(x,y) \rightarrow$ <u>gradiente</u>

<u>gradiente nos indica la dirección de mayor cambio.</u>

Descent

Ascent

$$\nabla f(x,y,z,w) = \left[\frac{\partial f}{\partial x}, \frac{\partial f}{\partial y}, \frac{\partial f}{\partial z}, \frac{\partial f}{\partial w}\right]$$

gradient descent } → Iterative

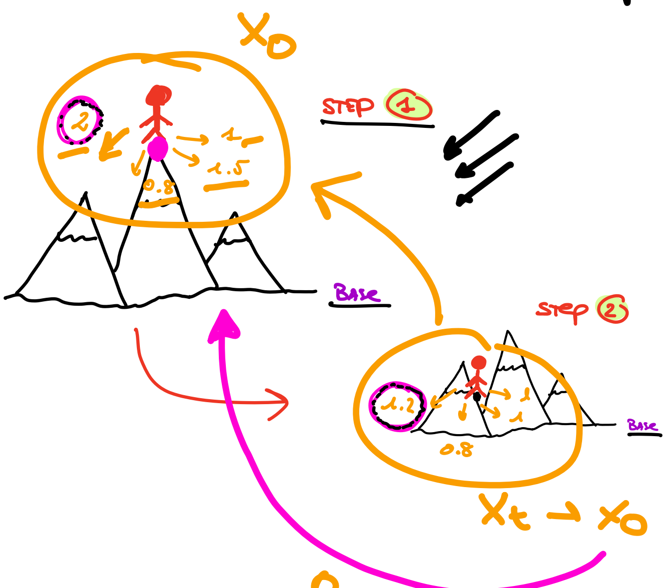
$\underbrace{\text{Iterative}}$

↓

Updating rule → $\boxed{\underbrace{x_t = x_{t-1}}_{x_t \to x_{t-1}} - \underbrace{\gamma \nabla f(x)}_{\text{dirección de mayor cambio}}}$

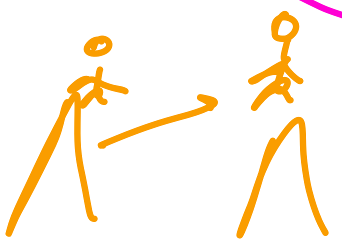
learning rate

descent.



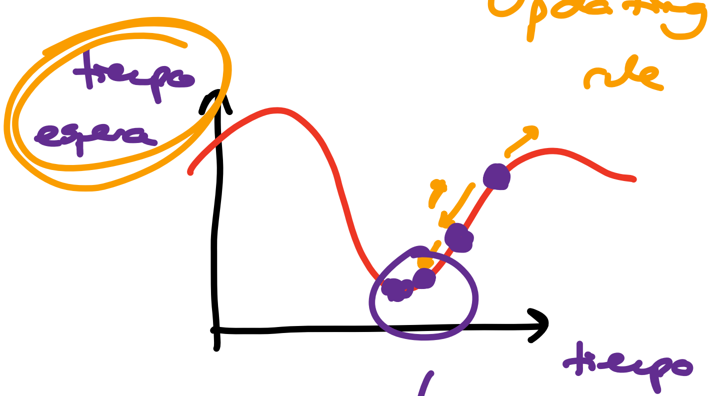
step 1

$x_{t-1} = x_0 =$



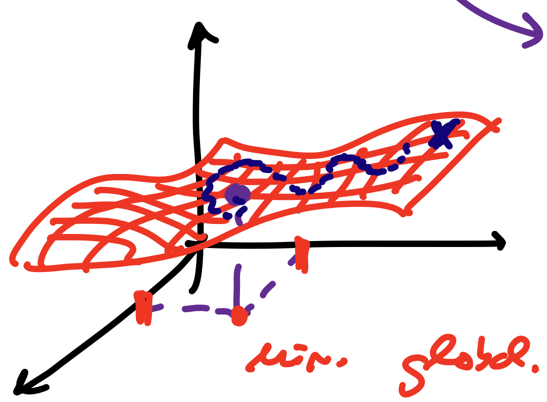
$\boxed{x_t = x_{t-1} - \underbrace{\gamma \nabla f(x)}_{\text{Dirección de mayor cambio}}}$

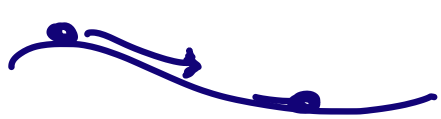
Updating rule del grad. Descent.



tiempo espera $= w_0 + w_1 \cdot$ tiempo

grad. descent = encontrar el mínimo.





grad. descent

momentum