PECE Programa de Educação Continuada
Escola Politécnica da USP

INTELIGÊNCIA ARTIFICIAL
Análise Estatística de Dados

AULA 01

Larissa Driemeier
driemeie@usp.br

PROGRAMA DO CURSO

Aula	Data	Conteúdo da Aula
01	03/03	Introdução ao Curso. Vetores, matrizes. Operações básicas, normas, determinante, inversa, traço, dependência linear. Problema de autovalor e autovetor. Decomposição de valor singular.
02	17/03	Variáveis independentes e não independentes. Estatística Descritiva e Indutiva. Definições de medidas de dispersão e tendência central.
03	10/03	Distribuições de probabilidade. Curva Normal.
04	24/03	Probabilidade condicional. Regra da cadeia da probabilidade condicional. Teorema de Bayes.
05	31/03	Estatística Bayesiana.
06	07/04	Modelo de Markov
07	14/04	Modelo de Markov oculto.
08	28/04	Intervalo de confiança e teste de hipótese
09	05/05	Trabalho de conclusão do curso de Estatística

AVALIAÇÃO

média final

Soma das notas dos Trabalhos de cada aula

$$M = \sum_i T_i \leq 10$$

Critério de Aprovação: $M \geq 7{,}0$

Recuperação: Para alunos cuja média final estiver contida no intervalo $4{,}0 \leq M < 7{,}0$ será oferecida uma prova de recuperação, que **substituirá a nota total, e vale 10,0.**

REFERÊNCIAS BIBLIOGRÁFICAS

Bussab, W.; Morettin, P.L. *Estatística Básica*, Saraiva, 6ª Edição, 2009 .

Hastie, T; Tibshirani, R.; Friedman, J. *The Elements of Statistical Learning*, 2nd edition, Springer-Verlag, 2009.

REVISÃO DE ÁLGEBRA LINEAR

Notação básica
Operações e propriedades
Cálculo matricial

MOTIVAÇÃO

A álgebra linear fornece uma maneira compacta de representar e operar conjuntos de equações lineares!

Exemplo:

$$2m + 3b = 8$$

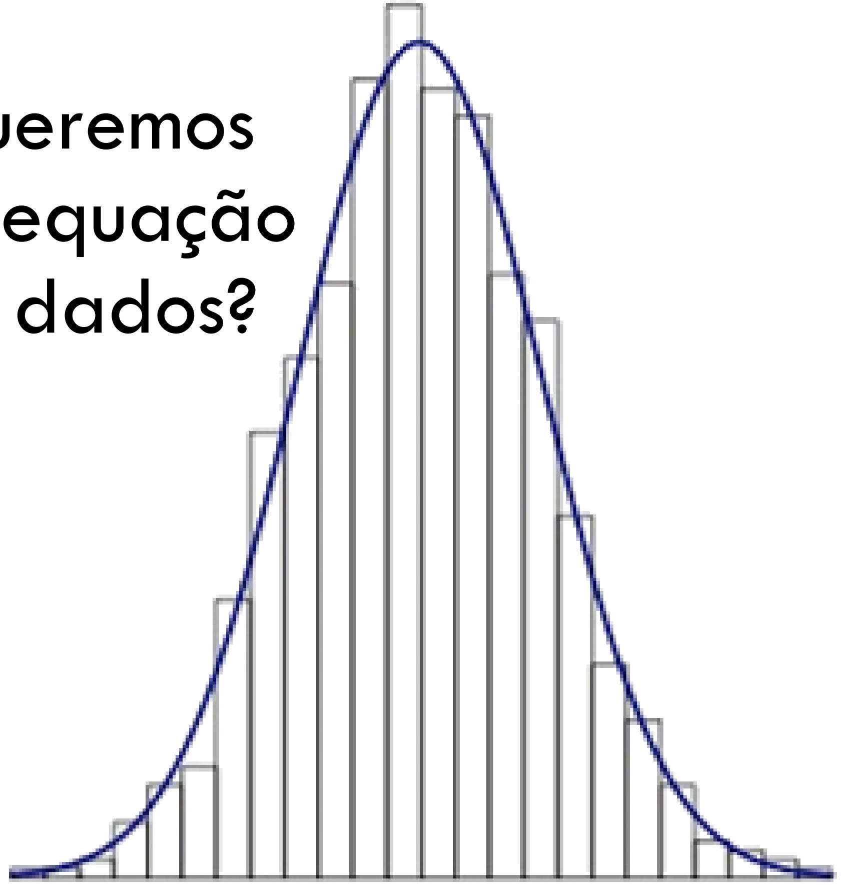
$$10m + 1b = 13$$

$$\begin{pmatrix} 2 & 3 \\ 10 & 1 \end{pmatrix} \begin{bmatrix} m \\ b \end{bmatrix} = \begin{pmatrix} 8 \\ 13 \end{pmatrix}$$

$$Ax = b$$

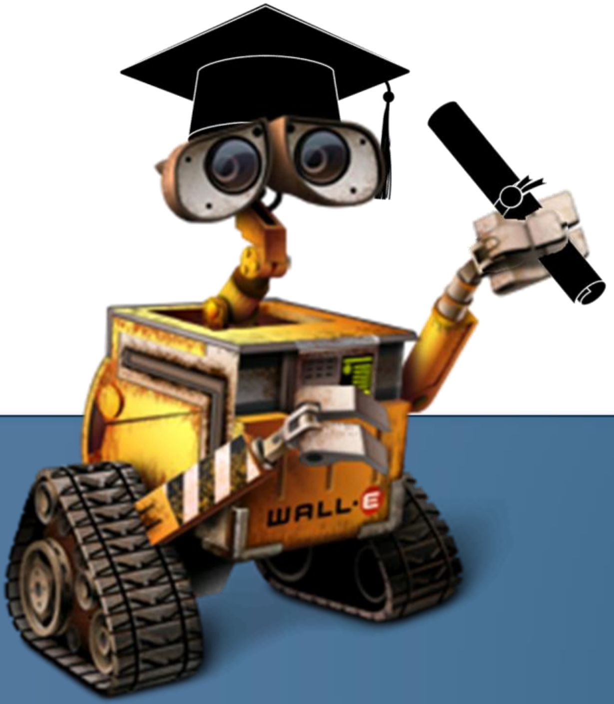
MAS...

E quando queremos ajustar uma equação para alguns dados?



$$f(x) = \frac{1}{\sqrt{2\pi}\sigma_x} e^{-\frac{1}{2}\left(\frac{x-\mu_x}{\sigma_x}\right)^2} \qquad -\infty \leq x \leq \infty$$

PECE Programa de Educação Continuada



VETORES

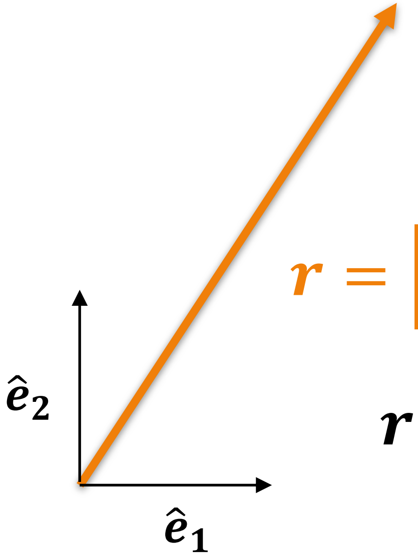
Vetores são um elemento fundamental da álgebra linear.

O QUE É UM VETOR?

Generalizando, para nosso universo de IA, vetor é uma lista de atributos de um objeto!

450 m^2
5 quartos
6 banheiro
850000 reais

$$\begin{bmatrix} 450 \\ 5 \\ 6 \\ 850 \end{bmatrix}$$



$$\boldsymbol{r} = \begin{bmatrix} r_1 \\ r_2 \end{bmatrix}$$

$$\boldsymbol{r} = r_1\hat{\boldsymbol{e}}_1 + r_2\hat{\boldsymbol{e}}_2$$

$\boldsymbol{x} \in \mathbb{R}^n$

Vetor coluna, n linhas e 1 coluna

$$\boldsymbol{x} = \begin{bmatrix} x_1 \\ x_2 \\ \dots \\ x_n \end{bmatrix}$$

x_i é o i-ésimo elemento do vetor x

Para expressar o vetor linha,

$$\boldsymbol{x}^T = [x_1 \quad x_2 \quad \dots \quad x_n]$$

calcula-se o transposto de $\boldsymbol{x}$, isto é, um vetor de 1 linha e n colunas.

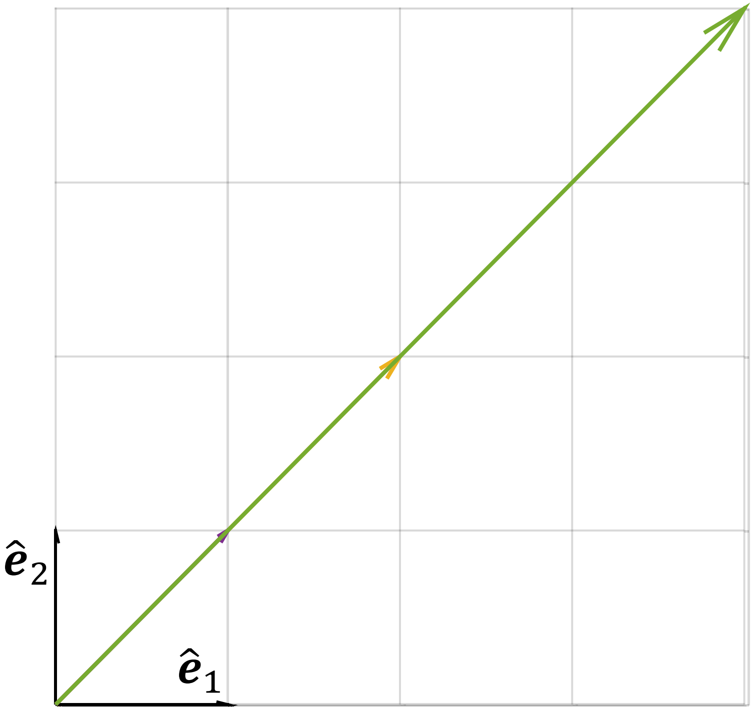
OPERAÇÕES BÁSICAS COM VETORES

$$\frac{\boldsymbol{r}}{2} = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$$

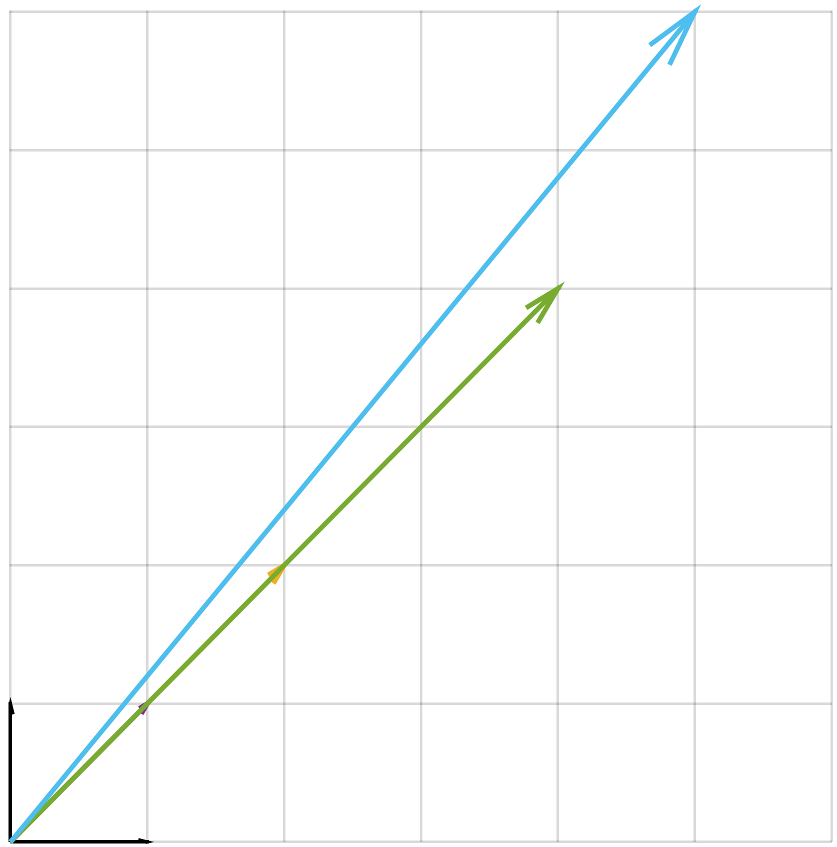
$$\boldsymbol{r} = \begin{bmatrix} 2 \\ 2 \end{bmatrix}$$

#quartos
#banheiros

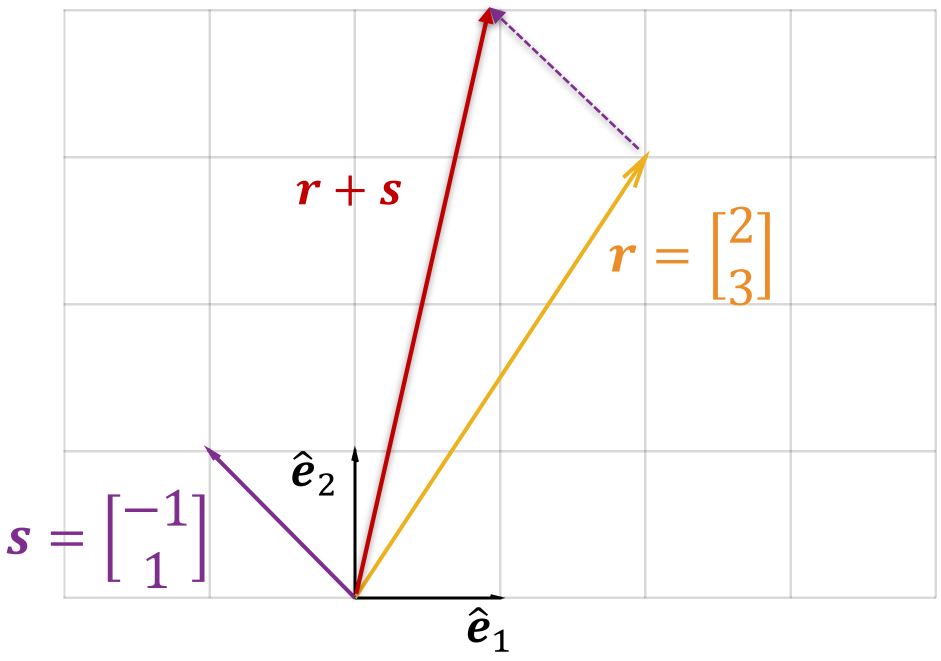
$$2\boldsymbol{r} = \begin{bmatrix} 4 \\ 4 \end{bmatrix}$$



$$\boldsymbol{r}' = \boldsymbol{r} + \begin{bmatrix} 3 \\ 4 \end{bmatrix} = \begin{bmatrix} 5 \\ 6 \end{bmatrix}$$



OPERAÇÕES BÁSICAS COM VETORES



$$\boldsymbol{s}+\boldsymbol{r}=\boldsymbol{r}+\boldsymbol{s}=\begin{bmatrix}1\\4\end{bmatrix}$$

$$(\boldsymbol{s}+\boldsymbol{r})+\boldsymbol{t}=\boldsymbol{s}+(\boldsymbol{r}+\boldsymbol{t})$$

$$\boldsymbol{r}+(-\boldsymbol{r})=\begin{bmatrix}0\\0\end{bmatrix}$$

PRODUTO INTERNO ENTRE VETORES

O produto escalar é a principal ferramenta para calcular projeções de vetores, decomposições de vetores e determinar ortogonalidade.

$\boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}^n$ calcula-se $\boldsymbol{x}^T\boldsymbol{y} \in \mathbb{R}$

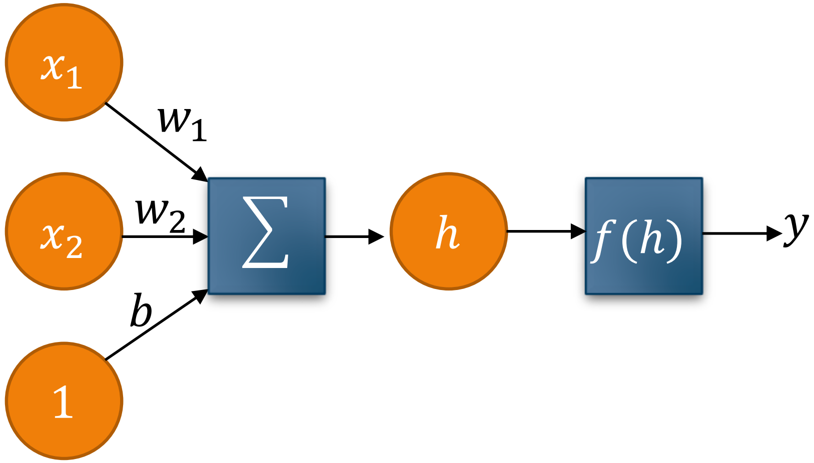
$$\boldsymbol{v} = \boldsymbol{x}^T\boldsymbol{y} = \begin{bmatrix} x_1 & x_2 & \cdots & x_n \end{bmatrix} \begin{bmatrix} y_1 \\ y_2 \\ \cdots \\ y_n \end{bmatrix} = x_1y_1 + x_2y_2 + \cdots + x_ny_n = \sum_{i=1}^{n} x_iy_i$$

POR EXEMPLO...

Ache o produto interno entre os seguintes vetores,

$$\boldsymbol{r} = \begin{bmatrix} -5 \\ 3 \\ 2 \\ 8 \end{bmatrix} \quad \boldsymbol{s} = \begin{bmatrix} 1 \\ 2 \\ -1 \\ 0 \end{bmatrix}$$

PORQUE PRODUTO INTERNO É IMPORTANTE...



$$y = f(h)$$

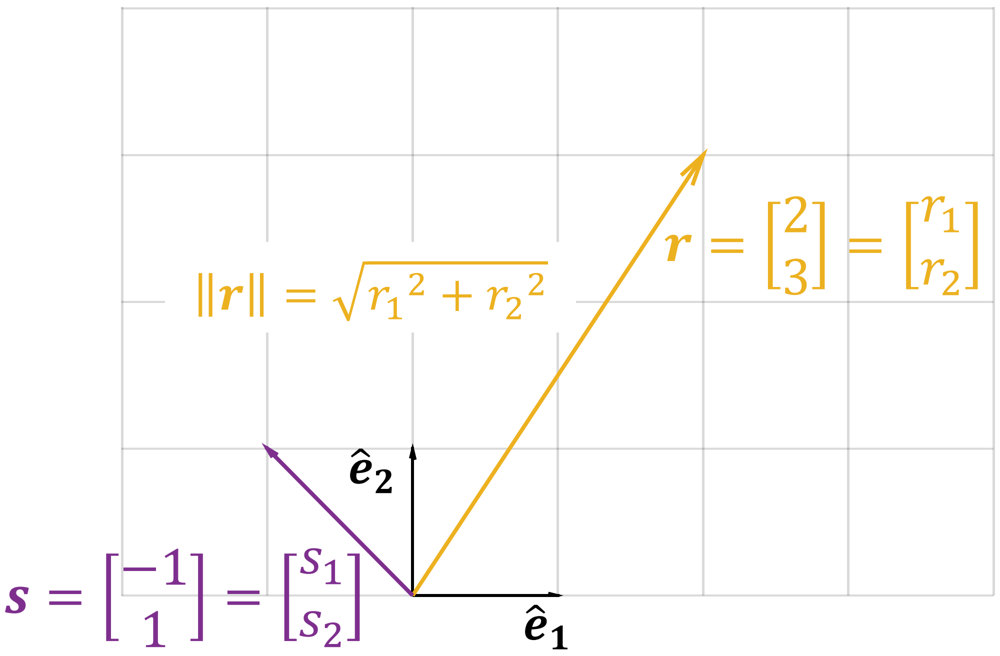
$$y = f(w_1x_1 + w_2x_2 + b)$$

$$y = f\left(\sum_i w_ix_i + b\right)$$

$$h = \begin{bmatrix} x_1 & x_2 \end{bmatrix} \begin{bmatrix} w_1 \\ w_2 \end{bmatrix}$$

PECE Programa de Educação Continuada

CARACTERÍSTICAS DO PRODUTO INTERNO



$$\boldsymbol{r} \cdot \boldsymbol{s} = \boldsymbol{s} \cdot \boldsymbol{r} = -1x2 + 1x3 = 1$$

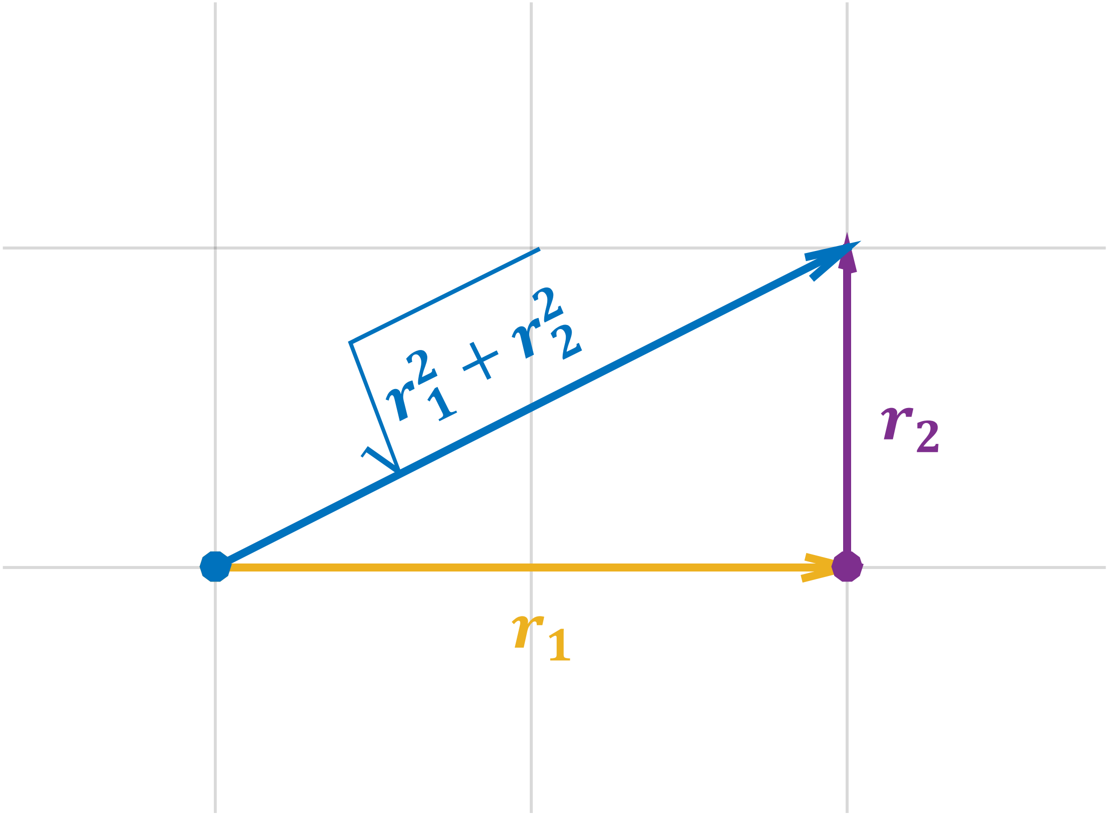
$$\boldsymbol{r} \cdot (\boldsymbol{s} + \boldsymbol{t}) = \boldsymbol{r} \cdot \boldsymbol{s} + \boldsymbol{r} \cdot \boldsymbol{t}$$

$$\boldsymbol{r} \cdot (a\boldsymbol{s}) = a(\boldsymbol{r} \cdot \boldsymbol{s})$$

$$\boldsymbol{r} = r_1\hat{\boldsymbol{e}}_1 + r_2\hat{\boldsymbol{e}}_2 = 2\hat{\boldsymbol{e}}_1 + 3\hat{\boldsymbol{e}}_2$$

$$\boldsymbol{r} \cdot \boldsymbol{r} = r_1^2 + r_2^2 = \left(\sqrt{r_1^2 + r_2^2}\right)^2 = \|\boldsymbol{r}\|^2$$

EXTRAPOLANDO...

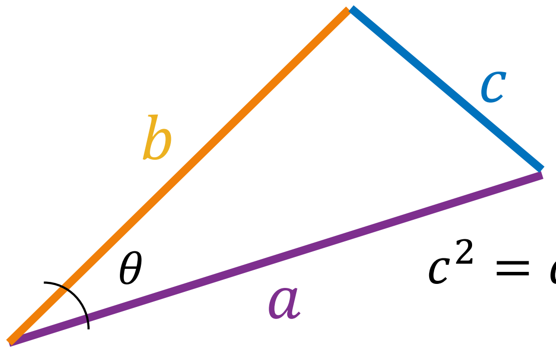


Podemos extender para qualquer dimensão: o tamanho de um vetor é a raíz quadrada da soma dos quadrados de seus componentes. Por exemplo,

$$\boldsymbol{s} = \begin{bmatrix} 1 \\ 3 \\ 4 \\ 2 \end{bmatrix}$$

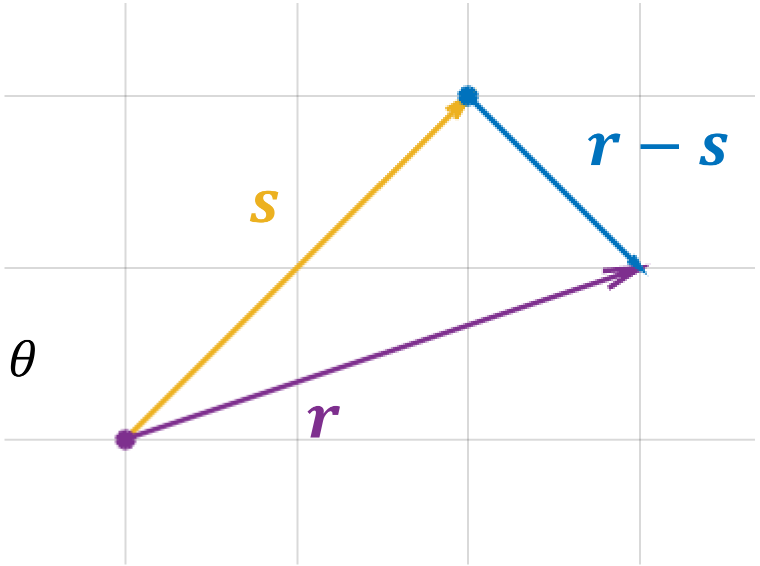
$$|\boldsymbol{s}| = \sqrt{30}$$

PRODUTO INTERNO E O ÂNGULO ENTRE VETORES



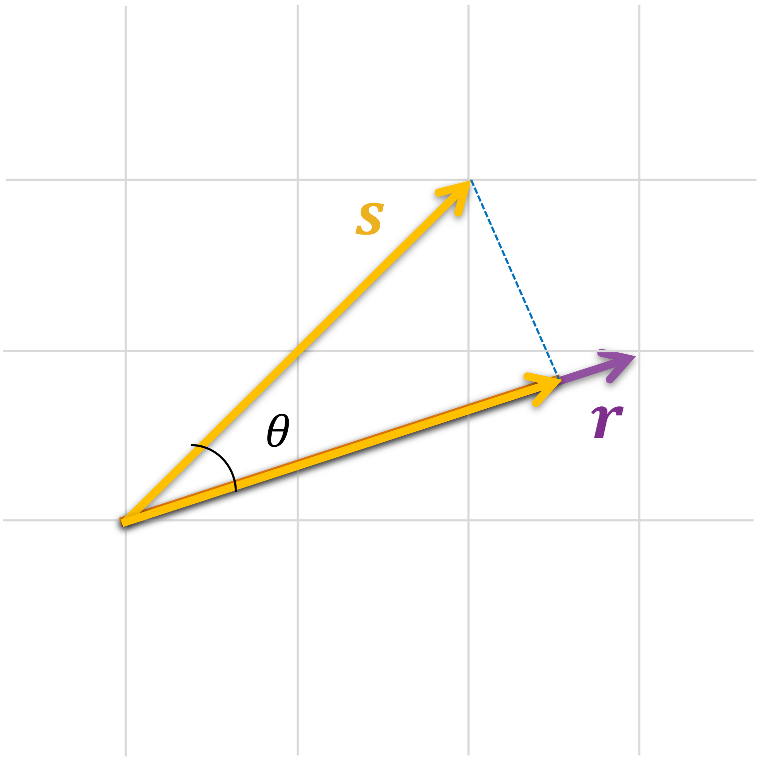
$$c^2 = a^2 + b^2 - 2ab\cos\theta$$

$$\|\boldsymbol{r} - \boldsymbol{s}\|^2 = \|\boldsymbol{r}\|^2 + \|\boldsymbol{s}\|^2 - 2\|\boldsymbol{r}\|\|\boldsymbol{s}\|\cos\theta$$



PROJEÇÃO ESCALAR E VETORIAL

$$\|\boldsymbol{r}-\boldsymbol{s}\|^2 = \|\boldsymbol{r}\|^2 + \|\boldsymbol{s}\|^2 - 2\|\boldsymbol{r}\|\|\boldsymbol{s}\|\cos\theta$$



Projeção escalar é o *tamanho* do vetor laranja.

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

Projeção vetorial é um escalonamento do vetor $\boldsymbol{r}$

$$\frac{\boldsymbol{r}\cdot\boldsymbol{s}}{\|\boldsymbol{r}\|}\frac{\boldsymbol{r}}{\|\boldsymbol{r}\|} = \frac{\boldsymbol{r}\cdot\boldsymbol{s}}{\boldsymbol{r}\cdot\boldsymbol{r}}\boldsymbol{r}$$

VETORES ORTOGONAIS E ORTONORMAIS

Dois vetores são ortogonais, $\cos\theta = 90^o$, se seu produto interno é nulo,

$$\frac{\boldsymbol{r}\cdot\boldsymbol{s}}{\|\boldsymbol{r}\|} = \|\boldsymbol{s}\|\cos\theta \rightarrow \boldsymbol{r}\cdot\boldsymbol{s} = 0$$

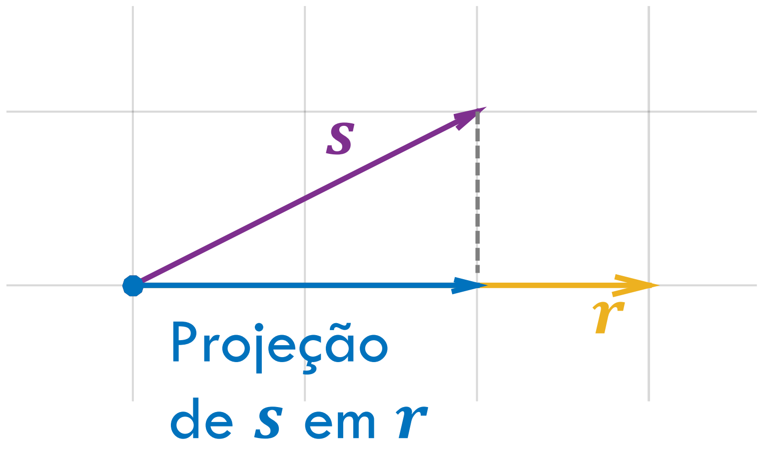
Um vetor é normalizado se sua dimensão é unitária, ié,

$$|\boldsymbol{r}| = 1$$

Podemos normalizar um vetor...

$$\boldsymbol{r}' = \frac{\boldsymbol{r}}{\|\boldsymbol{r}\|}$$

EXEMPLO



A projeção pode ser feita em qualquer dimensão. Por exemplo,

$$\boldsymbol{r} = \begin{bmatrix} 3 \\ -4 \\ 0 \end{bmatrix} \quad \boldsymbol{s} = \begin{bmatrix} 10 \\ 5 \\ -6 \end{bmatrix}$$

Qual a projeção escalar e vetorial de s em r?

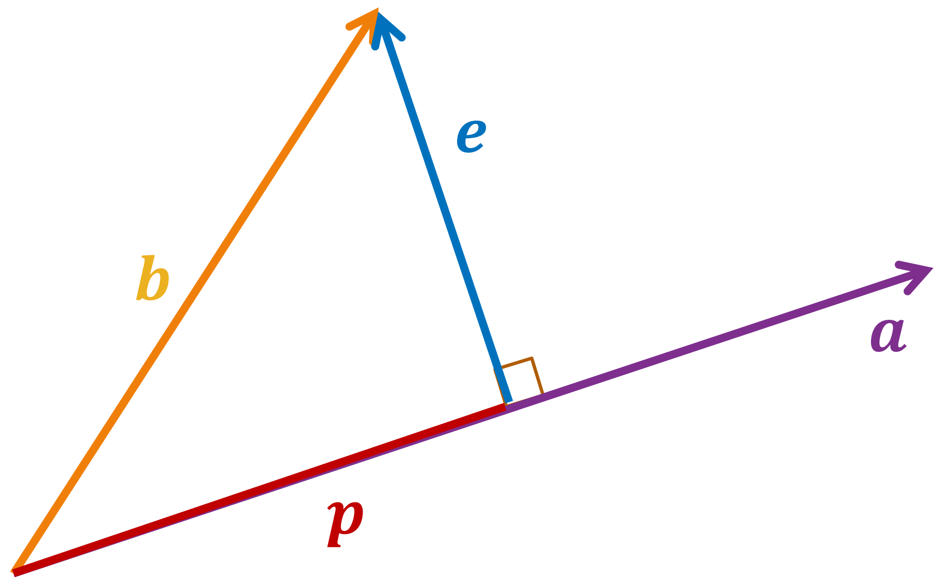
EXEMPLO

Com respeito a um dado sistema de coordenadas, um balão viaja com velocidade dada por $\begin{bmatrix} 1 \\ 2 \end{bmatrix}$, e o vento sopra na direção $\begin{bmatrix} 1 \\ 1 \end{bmatrix}$. Pergunta-se:

- Qual a intensidade da velocidade do balão na direção do vento?
- Qual a velocidade do balão na direção do vento?

EXEMPLO

$\boldsymbol{e}$ é perpendicular a $\boldsymbol{p}$ e conecta $\boldsymbol{a}$ e $\boldsymbol{b}$. Em uma perspectiva, a projeção tenta minimizar esse vetor $\boldsymbol{e}$ que pode ser visto como um ***erro*** ou uma ***margem***.



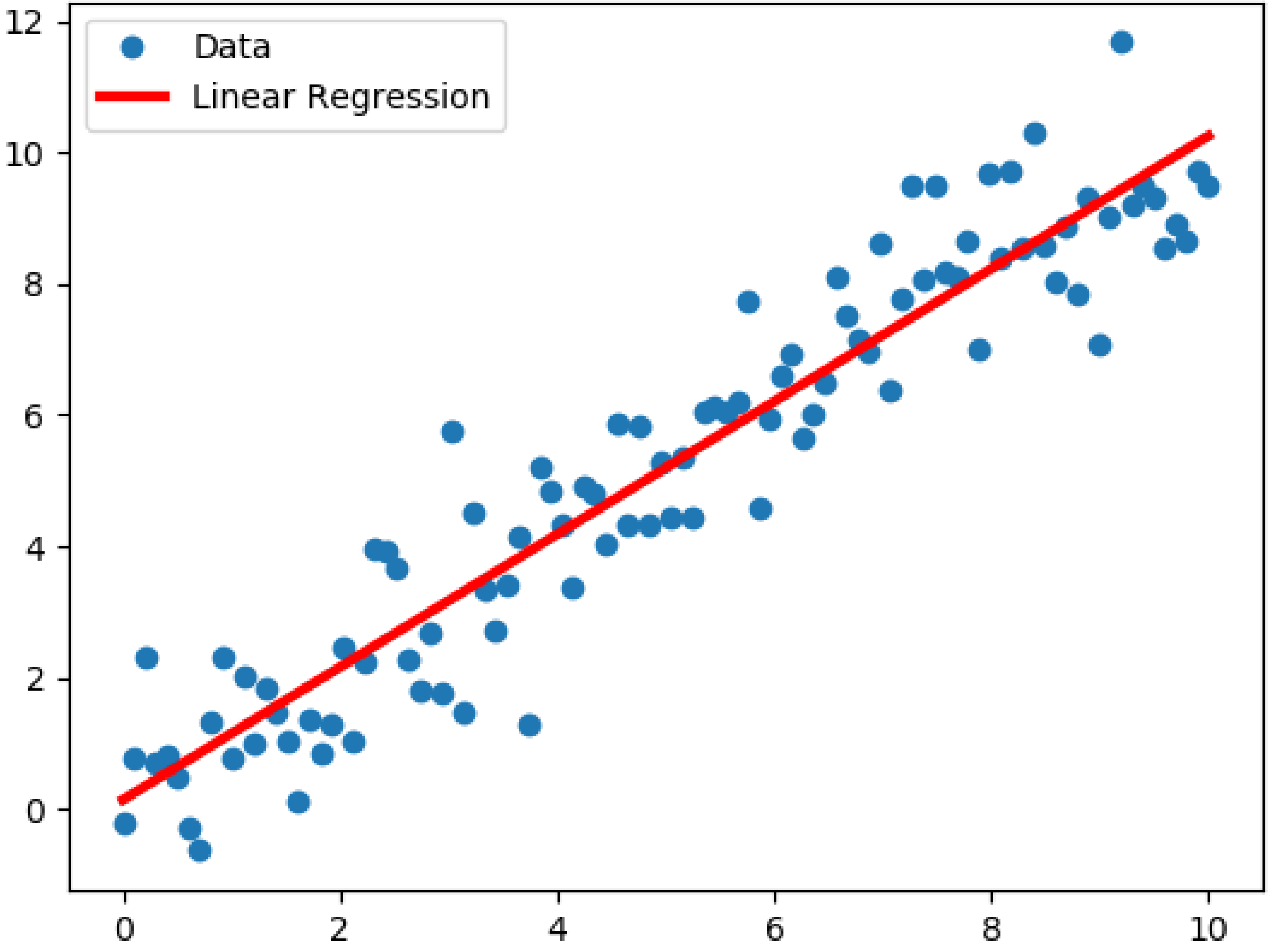
Componente ao longo de $\boldsymbol{a}$

$$\boldsymbol{p} = \frac{\boldsymbol{a} \cdot \boldsymbol{b}}{\boldsymbol{a} \cdot \boldsymbol{a}} = \frac{\boldsymbol{a}^T \boldsymbol{b}}{\boldsymbol{a}^T \boldsymbol{a}} \rightarrow \boldsymbol{a}^T \boldsymbol{b} = \boldsymbol{p} \cdot \|\boldsymbol{a}\|$$

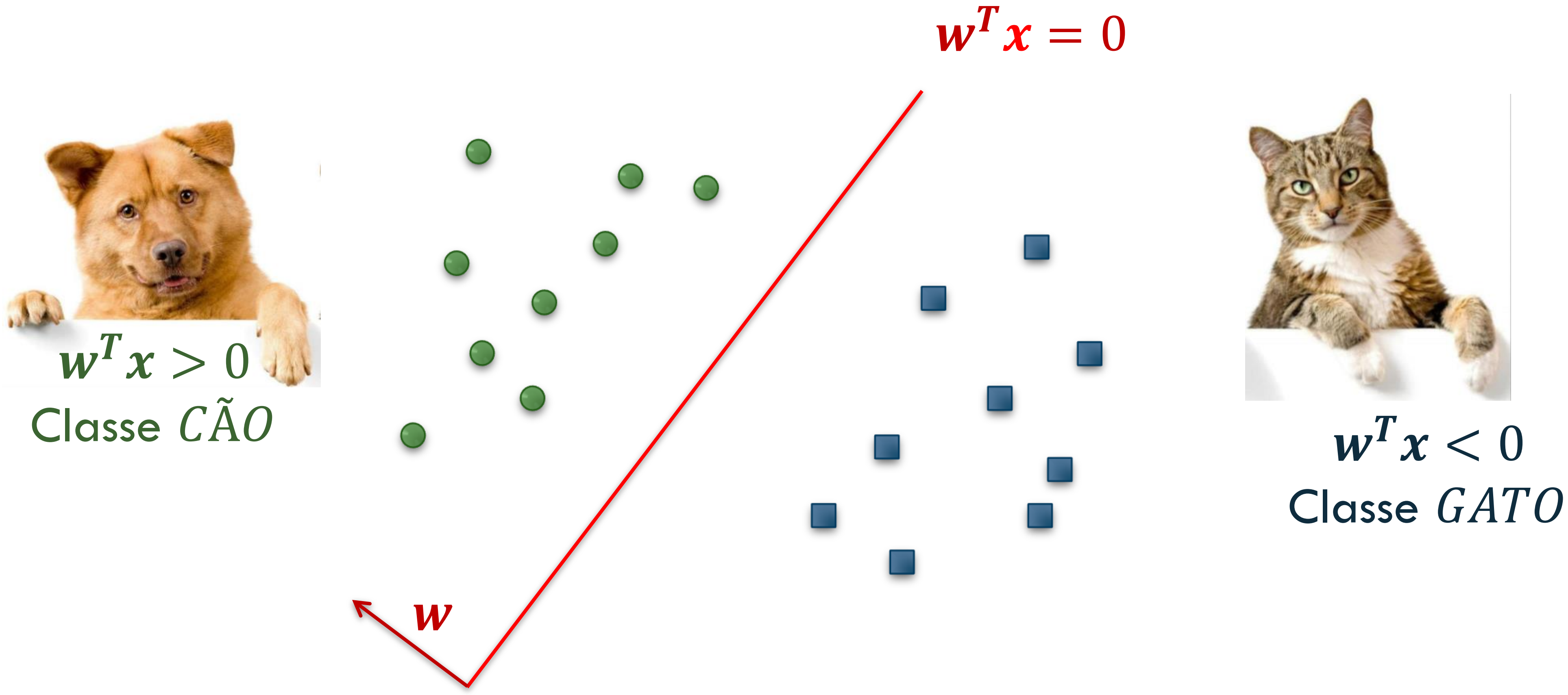
Normaliza o vetor $\boldsymbol{a}$

$$\boxed{\boldsymbol{e} \cdot \boldsymbol{a} = \boldsymbol{e}^T \boldsymbol{a} = 0}$$

APLICAÇÃO



APLICAÇÃO: CLASSIFICAÇÃO

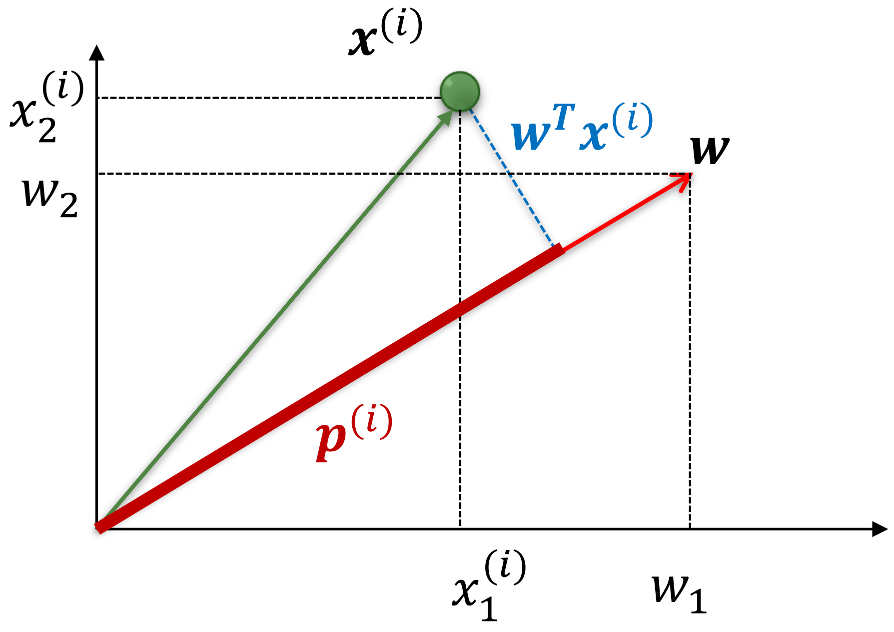


SVM

Escrevendo de forma simplificada

$$\min_{\boldsymbol{w}} \frac{1}{2}\sum_{j=1}^{n} w_j^2$$

$$\boldsymbol{w}^T\boldsymbol{x} \geq 1 \quad \text{se } y^{(i)} = 1$$
$$\boldsymbol{w}^T\boldsymbol{x} \leq -1 \quad \text{se } y^{(i)} = 0$$



$$\boldsymbol{w}^T\boldsymbol{x} = p^{(i)}\|\boldsymbol{w}\|$$

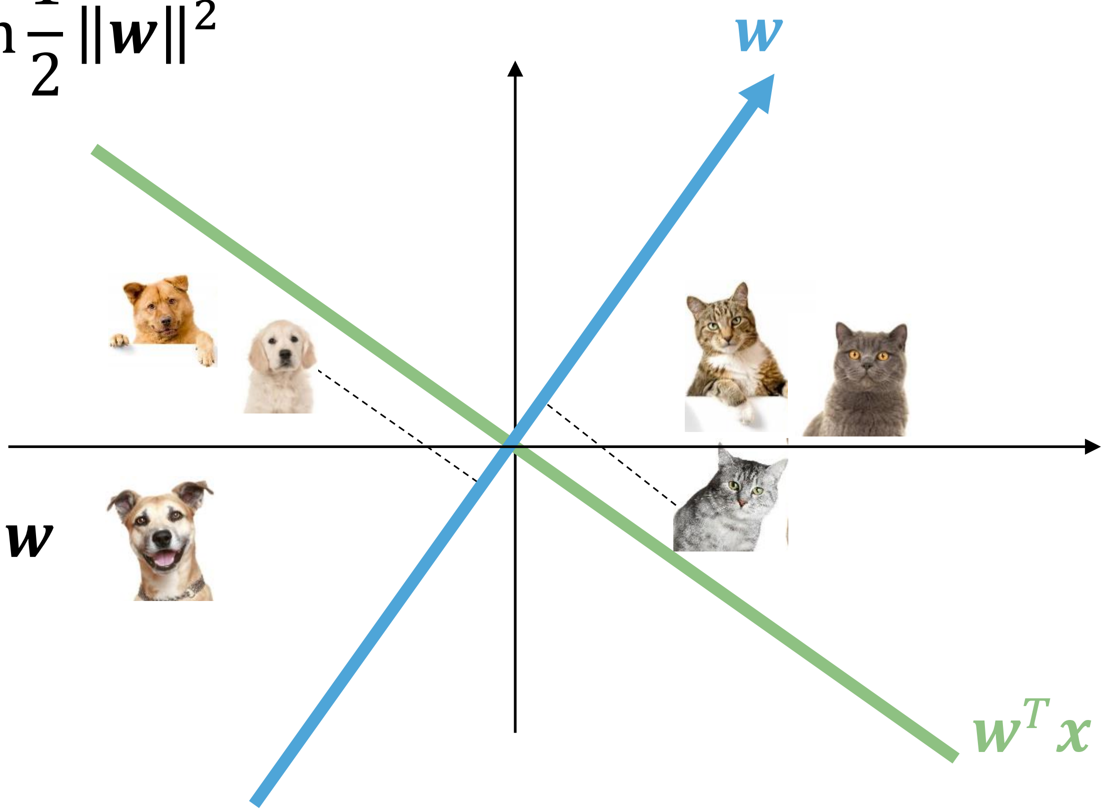
REESCREVENDO SVM

$$\min_{\boldsymbol{w}} \frac{1}{2}\sum_{j=1}^{n} w_j^2 = \min_{\boldsymbol{w}} \frac{1}{2}\|\boldsymbol{w}\|^2$$

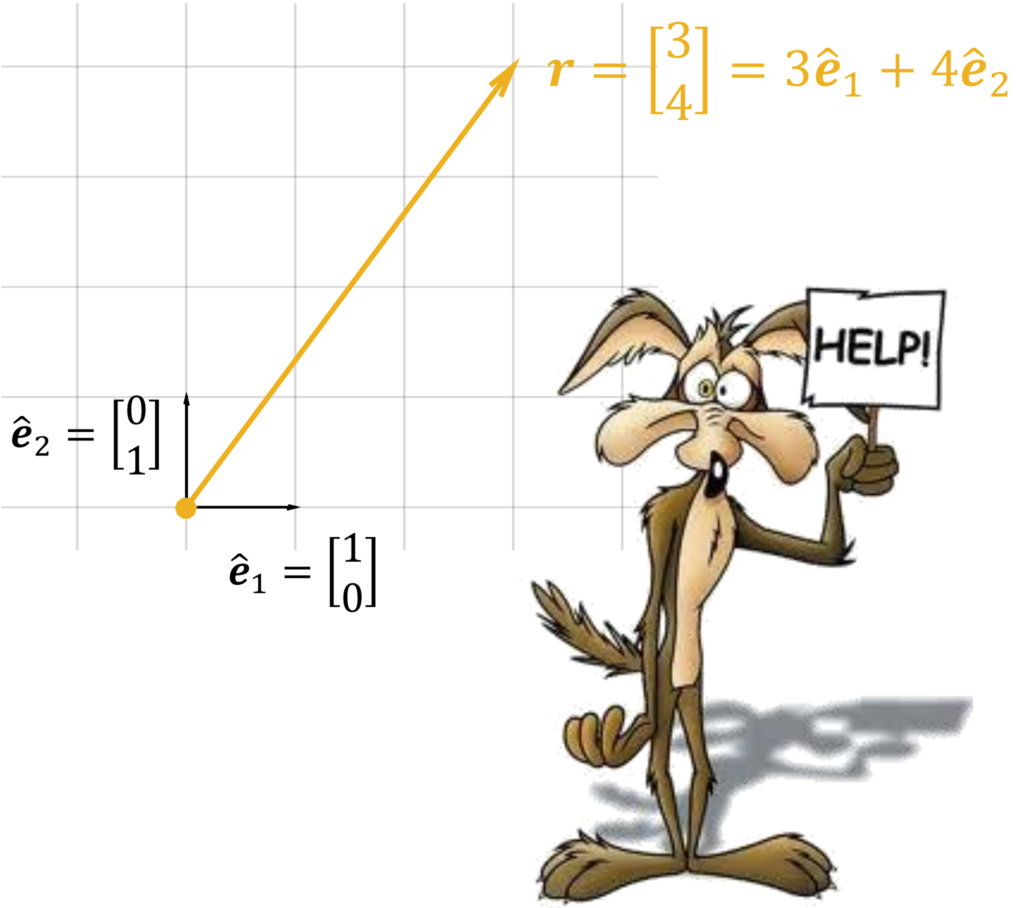
$p^{(i)} \cdot |\boldsymbol{w}| \geq 1 \quad$ se $y^{(i)} = 1$

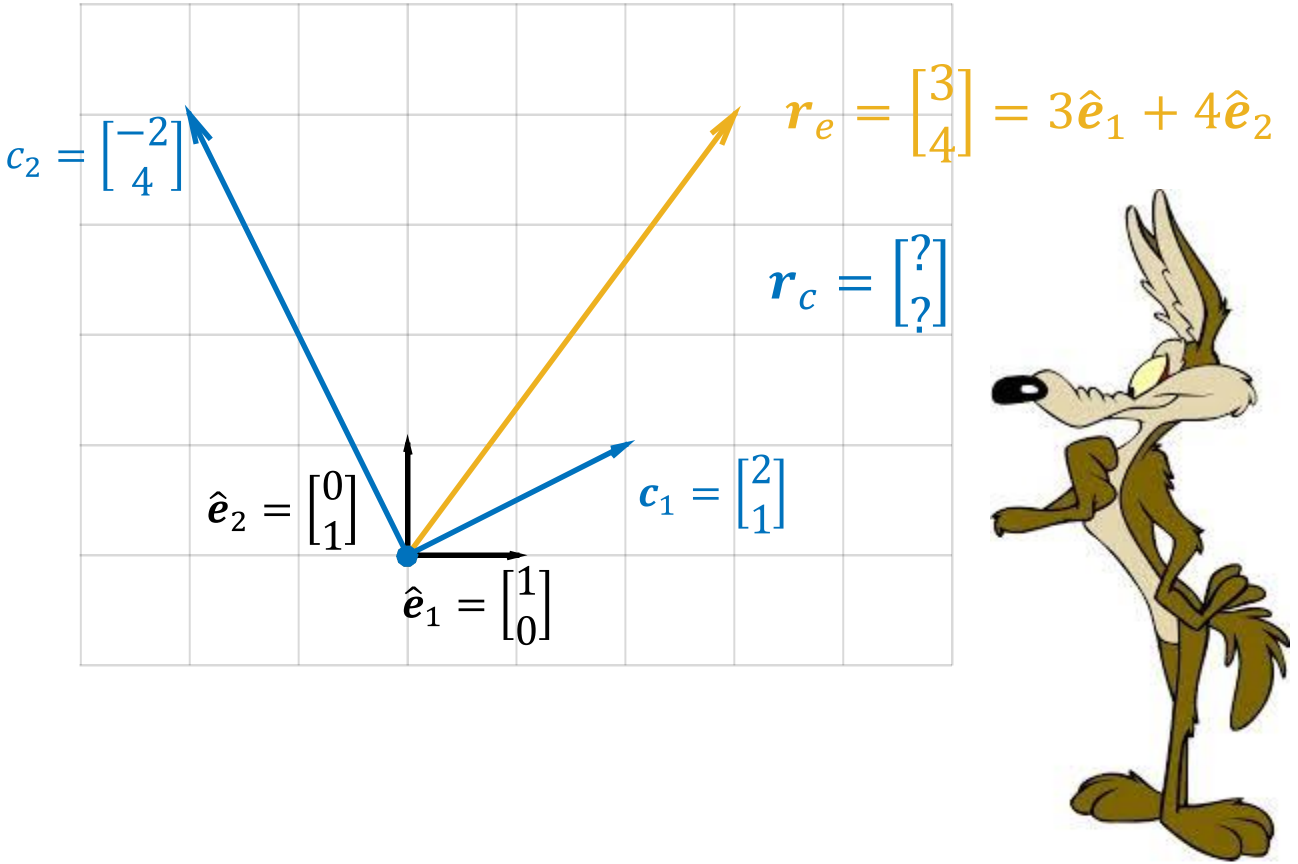
$p^{(i)} \cdot |\boldsymbol{w}| \leq -1 \quad$ se $y^{(i)} = 0$

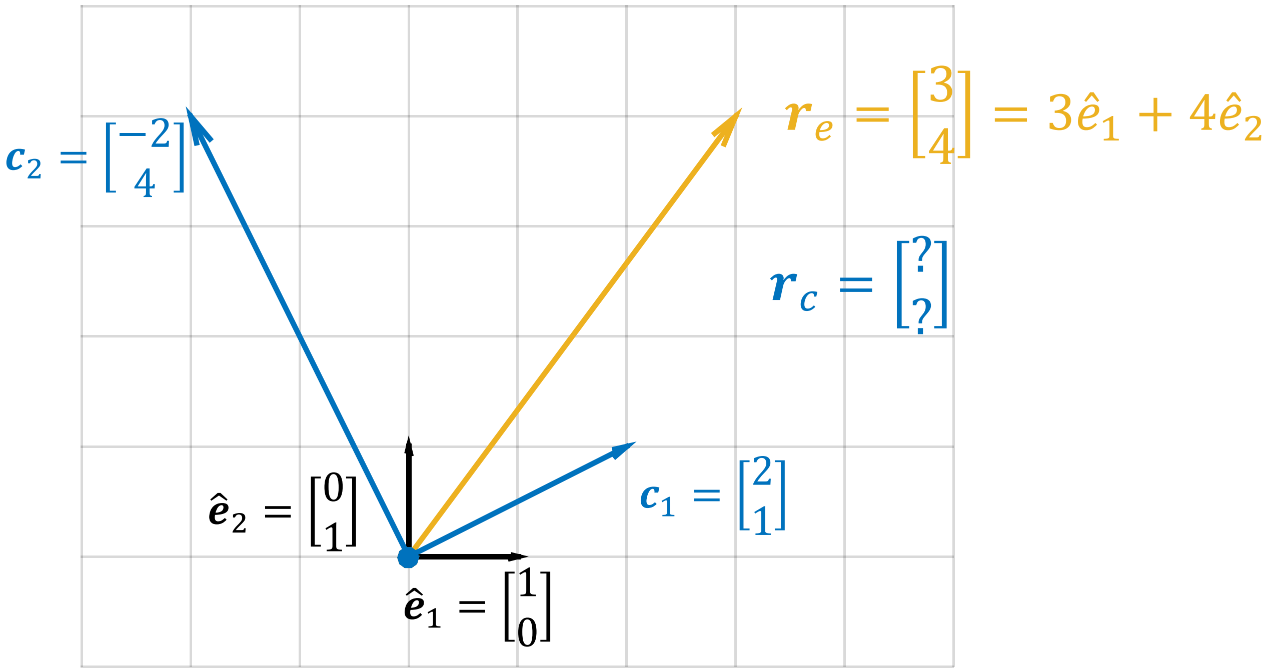
onde $p^{(i)}$ é a projeção de $\boldsymbol{x}^{(i)}$ sobre o vetor $\boldsymbol{w}$



TROCA DE BASE







$$\frac{\boldsymbol{r}_e \cdot \boldsymbol{c}_1}{|\boldsymbol{c}_1|^2} = 2$$

$$\boldsymbol{c}_1 \cdot \boldsymbol{c}_2 = \begin{bmatrix} 2 \\ 1 \end{bmatrix} \cdot \begin{bmatrix} -2 \\ 4 \end{bmatrix} = -4 + 4 = 0$$

Portanto, os vetores da nova base são ortogonais.

$$\frac{\boldsymbol{r}_e \cdot \boldsymbol{c}_2}{|\boldsymbol{c}_2|^2} = \frac{1}{2}$$

$$\boldsymbol{r}_c = \begin{bmatrix} 2 \\ 1/2 \end{bmatrix}$$

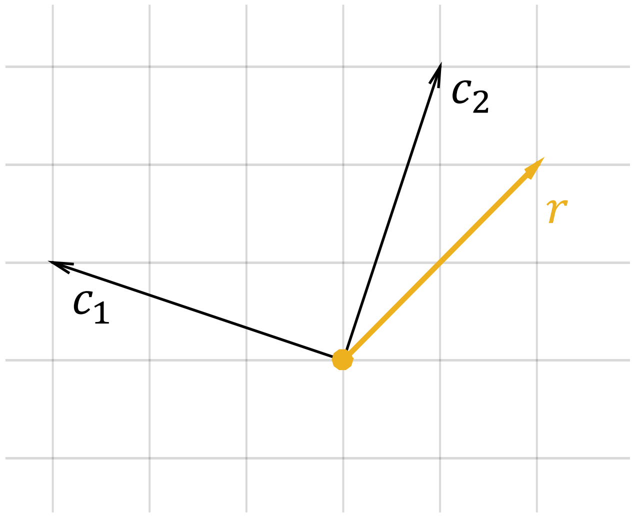
$$\boldsymbol{r}_e = 2\begin{bmatrix} 2 \\ 1 \end{bmatrix} + \frac{1}{2}\begin{bmatrix} -2 \\ 4 \end{bmatrix} = \begin{bmatrix} 3 \\ 4 \end{bmatrix}$$

EXEMPLO

Dados os seguintes vetores escritos em uma base comum

$$\boldsymbol{r} = \begin{bmatrix} 2 \\ 2 \end{bmatrix}, \boldsymbol{c}_1 = \begin{bmatrix} -3 \\ 1 \end{bmatrix}, \ \boldsymbol{c}_2 = \begin{bmatrix} 1 \\ 3 \end{bmatrix}$$

Como é o vetor $\boldsymbol{r}_c$, ié, o vetor $\boldsymbol{r}$ escrito na base definida por $\boldsymbol{c}_1$, $\boldsymbol{c}_2$? Sabe-se que $\boldsymbol{c}_1$, $\boldsymbol{c}_2$ formam uma base ortogonal.



EXTRAPOLANDO PARA OUTRAS DIMENSÕES

Dados os seguintes vetores escritos em uma base comum

$$\boldsymbol{r} = \begin{bmatrix}1\\1\\1\end{bmatrix}, \boldsymbol{c}_1 = \begin{bmatrix}2\\1\\0\end{bmatrix}, \quad \boldsymbol{c}_2 = \begin{bmatrix}1\\-2\\-1\end{bmatrix}, \quad \boldsymbol{c}_3 = \begin{bmatrix}-1\\2\\-5\end{bmatrix}$$

Como é o vetor $\boldsymbol{r}_c$, ié, o vetor r escrito na base definida por $\boldsymbol{c}_1$, $\boldsymbol{c}_2$, $\boldsymbol{c}_3$? Sabe-se que $\boldsymbol{c}_1$, $\boldsymbol{c}_2$, $\boldsymbol{c}_3$ formam uma base ortogonal.

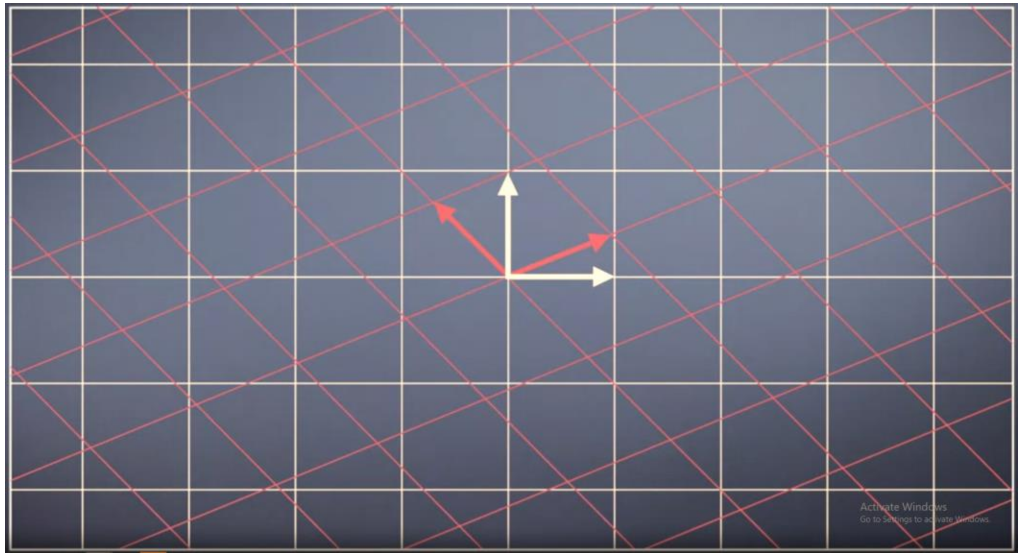
- ○ $\boldsymbol{r}_c = \begin{bmatrix}3/5\\-1/3\\-2/15\end{bmatrix}$
- ○ $\boldsymbol{r}_c = \begin{bmatrix}3\\-1\\-2\end{bmatrix}$
- ○ $\boldsymbol{r}_c = \begin{bmatrix}-3/5\\-1/3\\2/15\end{bmatrix}$
- ○ $\boldsymbol{r}_c = \begin{bmatrix}-3/5\\-1/3\\-2/15\end{bmatrix}$

BASE DE UM CONJUNTO DE VETORES

Diz-se que n vetores formam uma base para um conjunto de vetores de dimensão n se:

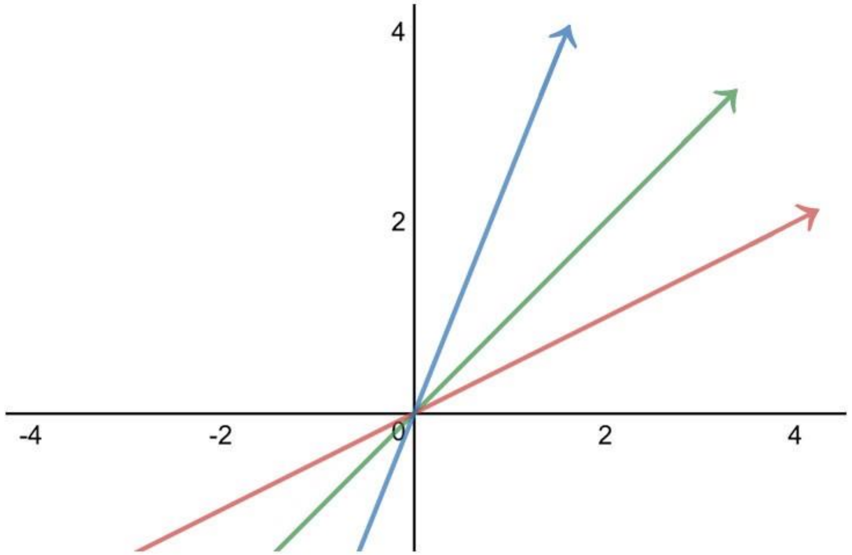
- ❑Não são uma combinação linear entre eles (são linearmente independentes, voltaremos nesse assunto)
- ❑Abrangem todo o espaço n-dimensional.

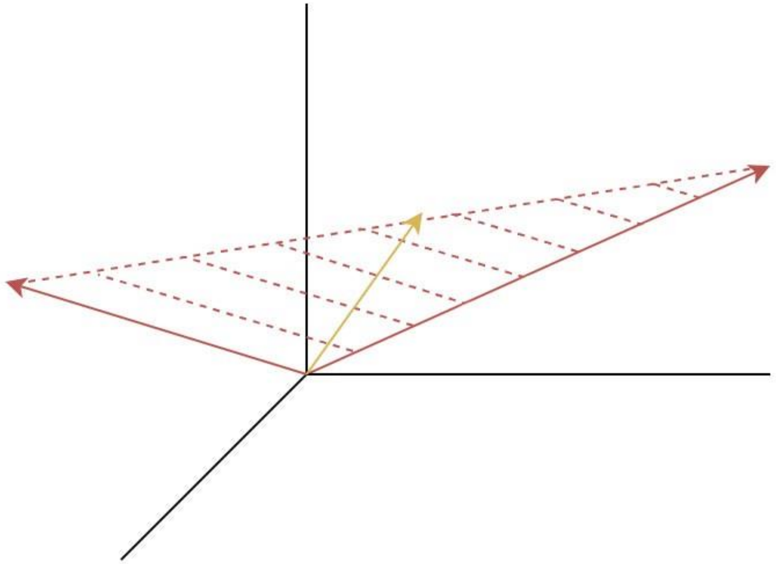
qualquer mapeamento que fazemos de um conjunto de vetores de base, de um sistema de coordenadas para outro, mantém o espaço vetorial sendo uma grade regularmente espaçada



ESPAÇO COMPOSTO DE VETORES LI

Um espaço n-dimensional não pode ter mais que n vetores linearmente independentes.





PRODUTO EXTERNO ENTRE VETORES

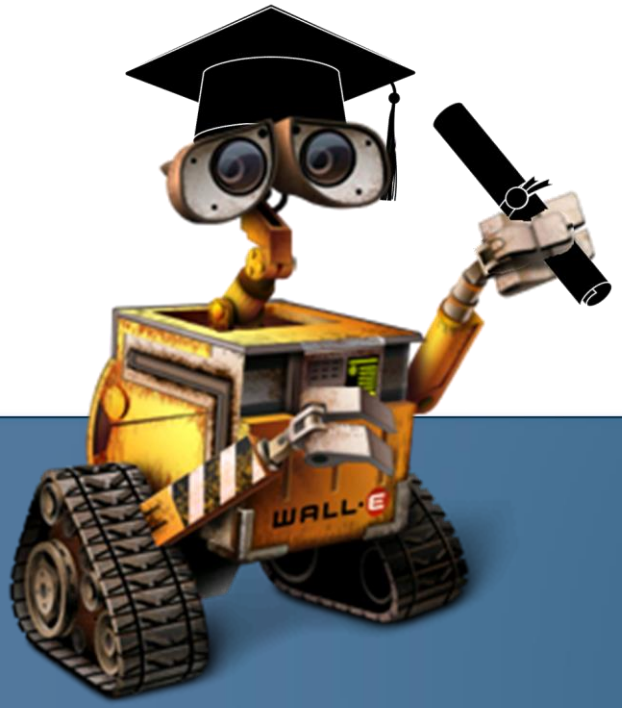
$\boldsymbol{x} \in \mathbb{R}^m, \boldsymbol{y} \in \mathbb{R}^n$ calcula-se $\boldsymbol{x}\boldsymbol{y}^T \in \mathbb{R}^{m \times n}$

$$\boldsymbol{A} = \boldsymbol{x}^T\boldsymbol{y} = \begin{bmatrix} x_1 \\ x_2 \\ \dots \\ x_m \end{bmatrix} \begin{bmatrix} y_1 & y_2 & \dots & y_n \end{bmatrix} = \begin{bmatrix} x_1y_1 & x_1y_2 & \dots & x_1y_n \\ x_2y_1 & x_2y_2 & \dots & x_2y_n \\ \dots & \dots & \dots & \dots \\ x_my_1 & x_my_2 & \dots & x_my_1 \end{bmatrix}$$

EXEMPLO

$\mathbf{1} \in \mathbb{R}^n$ é um vetor linha n-dimensional, cujas entradas são todas iguais a 1. Além disso, considere a matriz $A \in \mathbb{R}^{nxm}$ cujas colunas são todas iguais a algum vetor $\boldsymbol{x} \in \mathbb{R}^n$. Usando produto externo, podemos representar $\boldsymbol{A}$ compactamente como,

$$\boldsymbol{A} = \begin{bmatrix} | & | & | & | \\ \boldsymbol{x} & \boldsymbol{x} & \cdots & \boldsymbol{x} \\ | & | & | & | \end{bmatrix} = \begin{bmatrix} x_1 & x_1 & \cdots & x_1 \\ x_2 & x_2 & \cdots & x_2 \\ \vdots & \vdots & \ddots & \vdots \\ x_m & x_m & \cdots & x_m \end{bmatrix} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_m \end{bmatrix} \begin{bmatrix} 1 & 1 & \cdots & 1 \end{bmatrix} = \boldsymbol{x}\mathbf{1}^T$$



MATRIZES

Está na hora de falar de matrizes. Matrizes são usadas para definir transformações lineares.

MATRIZES

Seja $A \in \mathbb{R}^{mxn}$ uma matriz com m linhas e n colunas

$$\boldsymbol{A} = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{bmatrix}$$

a_{ij} é o elemento da *i-ésima* linha da *j-ésima* coluna de $\boldsymbol{A}$

A *i-ésima* coluna, $\boldsymbol{a}_i = a_{:,i}$

$$\boldsymbol{A} = \begin{bmatrix} | & | & & | \\ \boldsymbol{a}_1 & \boldsymbol{a}_2 & \cdots & \boldsymbol{a}_n \\ | & | & & | \end{bmatrix}$$

A *j-ésima* linha, $\boldsymbol{a}_j' = a_{j,:}$

$$\boldsymbol{A} = \begin{bmatrix} - & \boldsymbol{a}_1' & - \\ - & \boldsymbol{a}_2' & - \\ - & \vdots & - \\ - & \boldsymbol{a}_m' & - \end{bmatrix}$$

MATRIZ TRANSPOSTA

A transposição de uma matriz resulta na troca das linhas pelas colunas.

Dada uma matriz $\boldsymbol{A} \in \mathbb{R}^{mxn}$, sua transposição, escrita $\boldsymbol{A}^T \in \mathbb{R}^{nxm}$, é a matriz $n \times m$ cujas entradas são dadas por

$$(\boldsymbol{A}^T)_{ij} = a_{ji}$$

$$(\boldsymbol{A}^T)^T = \boldsymbol{A}$$
$$(\boldsymbol{AB})^T = \boldsymbol{B}^T\boldsymbol{A}^T$$
$$(\boldsymbol{A} + \boldsymbol{B})^T = \boldsymbol{A}^T + \boldsymbol{B}^T$$

Quando $\boldsymbol{A}^T = \boldsymbol{A}$ **a matriz é dita simétrica.**

TRAÇO DE UMA MATRIZ

O traço de uma matriz é a soma de sua diagonal principal,

$$\operatorname{tr} \boldsymbol{A} = \sum_{i=1}^{n} a_{ii}$$

$\boldsymbol{A} \in \mathbb{R}^{\boldsymbol{nxn}}, \qquad \operatorname{tr} \boldsymbol{A} = \operatorname{tr} \boldsymbol{A}^T$

$\boldsymbol{A}, \boldsymbol{B} \in \mathbb{R}^{\boldsymbol{nxn}}, \qquad \operatorname{tr}(\boldsymbol{A} + \boldsymbol{B}) = \operatorname{tr} \boldsymbol{A} + \operatorname{tr} \boldsymbol{B}$

$\boldsymbol{A} \in \mathbb{R}^{\boldsymbol{nxn}}, t \in \mathbb{R} \quad \operatorname{tr}(t\boldsymbol{A}) = t \operatorname{tr} \boldsymbol{A}$

$\boldsymbol{A}, \boldsymbol{B}$ tal que AB é quadrada $\operatorname{tr}(\boldsymbol{AB}) = \operatorname{tr} \boldsymbol{BA}$

$\boldsymbol{A}, \boldsymbol{B}, \boldsymbol{C}$ tal que ABC é quadrada $\operatorname{tr}(\boldsymbol{ABC}) = \operatorname{tr} \boldsymbol{BCA} = \operatorname{tr} \boldsymbol{CAB}$

MATRIZ SIMÉTRICA E ANTI-SIMÉTRICA, $A \in \mathbb{R}^{mxn}$

Simétrica

$$\boldsymbol{A} = \boldsymbol{A}^T$$

Anti-simétrica

$$\boldsymbol{A} = -\boldsymbol{A}^T$$

Toda matriz pode ser escrita como uma soma de uma parcela simétrica e outra anti-simétrica,

$$\boldsymbol{A} = \frac{1}{2}(\boldsymbol{A} + \boldsymbol{A}^T) + \frac{1}{2}(\boldsymbol{A} - \boldsymbol{A}^T)$$

PRODUTO ENTRE MATRIZ E VETOR

$\boldsymbol{A} \in \mathbb{R}^{mxn}$ e $\boldsymbol{v}, \boldsymbol{x} \in \mathbb{R}^{nx1}$

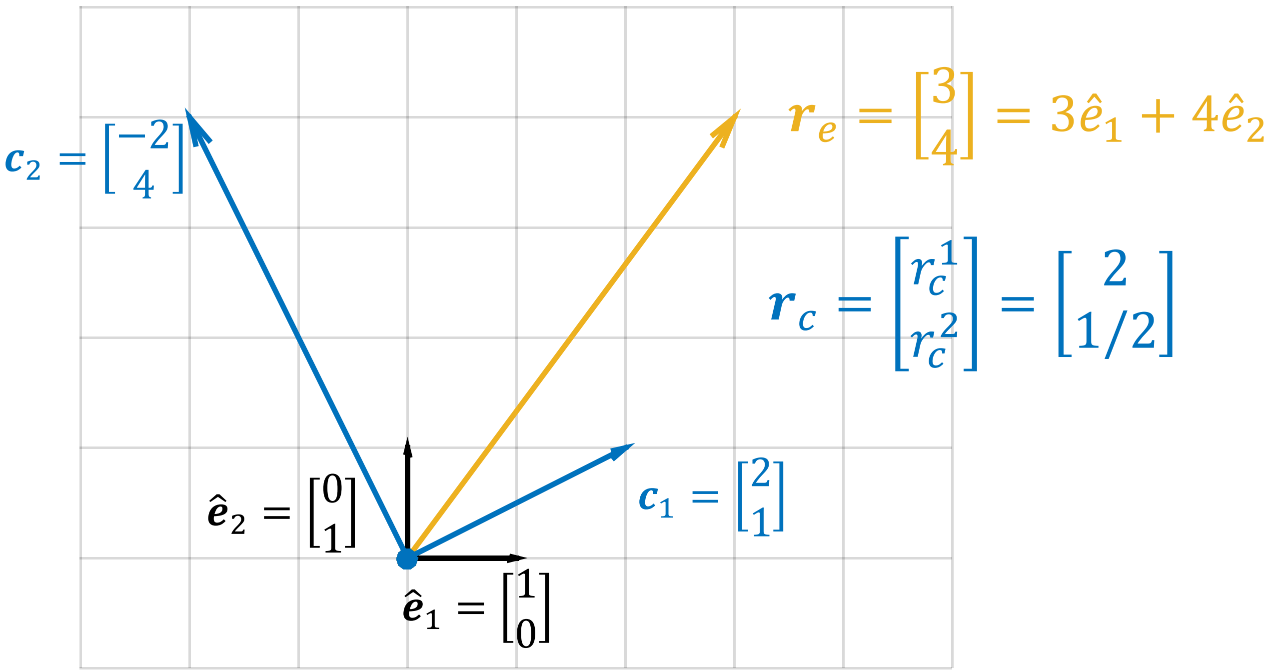
$$\boldsymbol{v} = \boldsymbol{A}\boldsymbol{x} = \begin{bmatrix} - & \boldsymbol{a}_1' & - \\ - & \boldsymbol{a}_2' & - \\ - & \vdots & - \\ - & \boldsymbol{a}_m' & - \end{bmatrix} \boldsymbol{x} = \begin{bmatrix} \boldsymbol{a}_1'\boldsymbol{x} \\ \boldsymbol{a}_2'\boldsymbol{x} \\ \vdots \\ \boldsymbol{a}_m'\boldsymbol{x} \end{bmatrix}$$

a *i-ésima* entrada de $\boldsymbol{v}$ é igual ao produto interno da *i-ésima* linha de $\boldsymbol{A}$ por $\boldsymbol{x}$, $v_i = \boldsymbol{a}_i'\boldsymbol{x}$

$$\boldsymbol{v} = \boldsymbol{A}\boldsymbol{x} = \begin{bmatrix} | & | & & | \\ \boldsymbol{a}_1 & \boldsymbol{a}_2 & \cdots & \boldsymbol{a}_n \\ | & | & & | \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} = \begin{bmatrix} \boldsymbol{a}_1 \end{bmatrix} x_1 + \begin{bmatrix} \boldsymbol{a}_2 \end{bmatrix} x_2 + \cdots + \begin{bmatrix} \boldsymbol{a}_n \end{bmatrix} x_n$$

$\boldsymbol{v}$ é uma combinação linear das colunas de $\boldsymbol{A}$, onde os coeficientes da combinação linear são dados pelas entradas de $\boldsymbol{x}$.

VOLTEMOS AO NOSSO EXEMPLO DE MUDANÇA DE BASE



$$\boldsymbol{r}_e = \begin{bmatrix} 3 \\ 4 \end{bmatrix} = r_c^1 \begin{bmatrix} 2 \\ 1 \end{bmatrix} + r_c^2 \begin{bmatrix} -2 \\ 4 \end{bmatrix}$$

$$\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} 3 \\ 4 \end{bmatrix} = \begin{bmatrix} 2 & -2 \\ 1 & 4 \end{bmatrix} \begin{bmatrix} r_c^1 \\ r_c^2 \end{bmatrix}$$

VOLTEMOS AOS VETORES LI...

Matematicamente, para que um conjunto de vetores $\boldsymbol{v}_i$ seja linearmente independente, em um espaço n-dimensional, a expressão

$$c_1\boldsymbol{v}_1 + c_2\boldsymbol{v}_2 + \ldots + c_n\boldsymbol{v}_n = 0$$

deve ser possível apenas se todos os fatores lineares c_i forem 0.

Em resumo, qualquer vetor não pode ser expresso como uma combinação linear de outros.

Matricialmente...

$$\boldsymbol{Ac} = \begin{bmatrix} | & | & & | \\ \boldsymbol{v}_1 & \boldsymbol{v}_2 & \cdots & \boldsymbol{v}_n \\ | & | & & | \end{bmatrix} \begin{bmatrix} c_1 \\ c_2 \\ \vdots \\ c_n \end{bmatrix} = \begin{bmatrix} \boldsymbol{a}_1 \end{bmatrix} c_1 + \begin{bmatrix} \boldsymbol{a}_2 \end{bmatrix} c_2 + \cdots + \begin{bmatrix} \boldsymbol{a}_n \end{bmatrix} c_n = \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix}$$

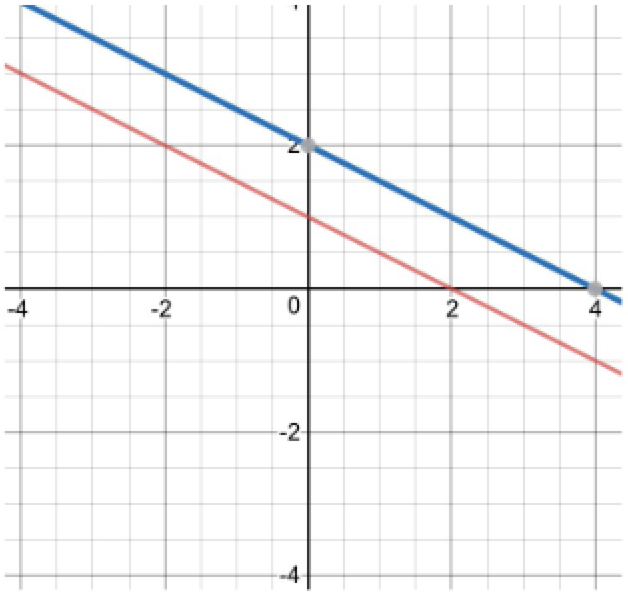
se e somente se $\boldsymbol{c}^T = [0 \quad 0 \quad \ldots \quad 0]$.

RANK

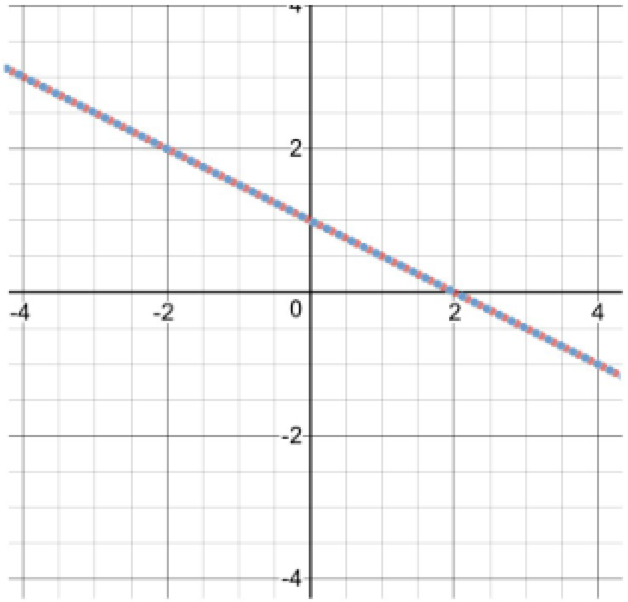
O ***rank*** mede a independência linear de uma matriz. O rank de uma matriz é igual ao número de linhas ou colunas indepententes.

$$\begin{pmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \end{pmatrix} \quad \begin{pmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 9 \end{pmatrix} \quad \begin{pmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 8 \end{pmatrix} \quad \begin{pmatrix} 1 \\ 3 \end{pmatrix} \begin{pmatrix} 5 & 6 \end{pmatrix} = \begin{pmatrix} 5 & 6 \\ 15 & 18 \end{pmatrix}$$

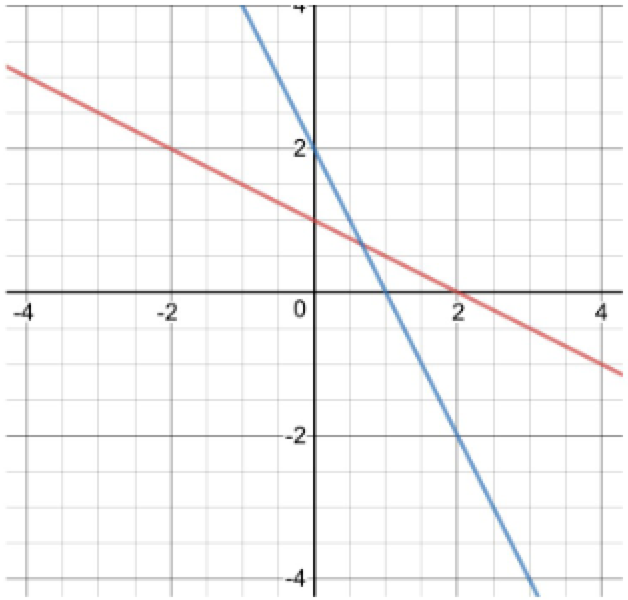
EXEMPLO



$$x + 2y = 2$$
$$2x + 4y = 8$$
$$\begin{bmatrix} 1 & 2 \\ 2 & 4 \end{bmatrix}\begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} 2 \\ 8 \end{bmatrix}$$



$$x + 2y = 2$$
$$2x + 4y = 4$$
$$\begin{bmatrix} 1 & 2 \\ 2 & 4 \end{bmatrix}\begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} 2 \\ 4 \end{bmatrix}$$



$$x + 2y = 2$$
$$2x + y = 2$$
$$\begin{bmatrix} 1 & 2 \\ 2 & 1 \end{bmatrix}\begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} 2 \\ 2 \end{bmatrix}$$

MATRIZ DIAGONAL

Uma matriz diagonal é uma matriz em que todos os termos fora da diagonal são 0.
Denomina-se, normalmente,

$$\boldsymbol{D} = \mathrm{diag}\,(d_1, d_2, \dots, d_n)$$

Com

$$D_{ij} = \begin{cases} d_i & i = j \\ 0 & i \neq j \end{cases}$$

A matriz identidade é um caso particular da matriz diagonal

$$\boldsymbol{I} = \mathrm{diag}\,(1,1, \dots, 1)$$

UMA MATRIZ ESPECIAL...

Primeiro, vamos pensar
em uma matriz que
não muda nada.

$$\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \begin{bmatrix} m \\ b \end{bmatrix} = \begin{pmatrix} m \\ b \end{pmatrix}$$

MATRIZ IDENTIDADE, $\boldsymbol{I} \in \mathbb{R}^{n \times n}$

É uma matriz quadrada de dimensões $n \times n$, definida como valor 1 nos termos da diagonal principal e 0 nos demais,

$$I_{ij} = \begin{cases} 1 & i = j \\ 0 & i \neq j \end{cases}$$

Para $\boldsymbol{A} \in \mathbb{R}^{m \times n}$,

$$\boldsymbol{AI} = \boldsymbol{A} = \boldsymbol{IA}$$

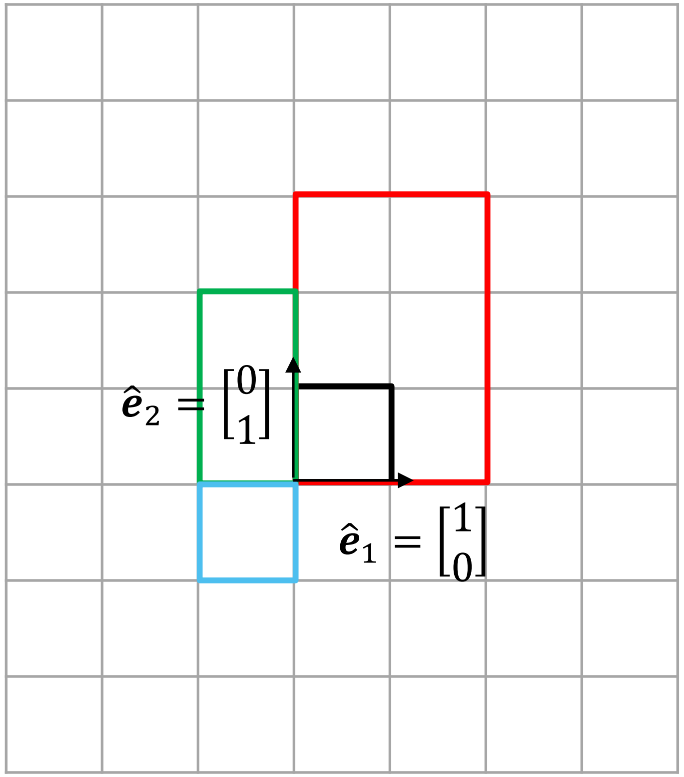
MODIFICANDO A IDENTIDADE...

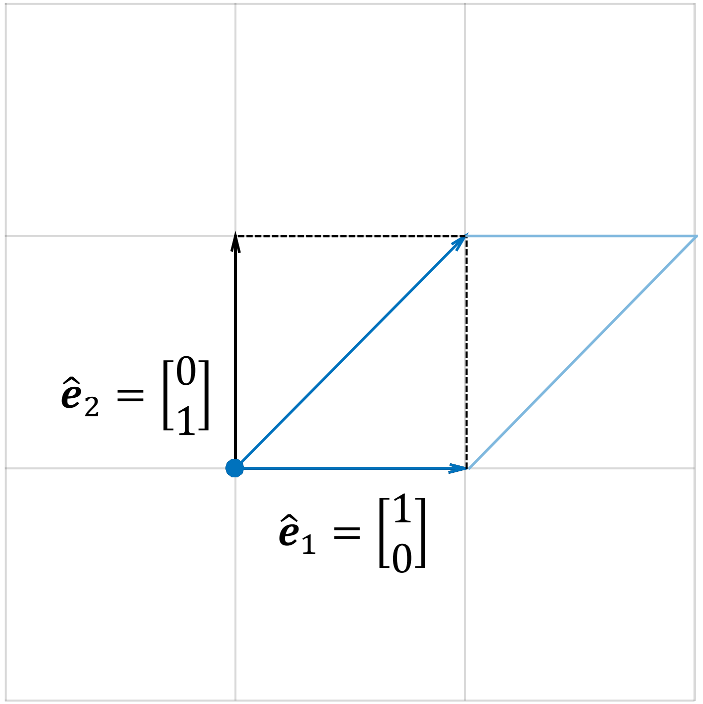
$$\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$$

$$\begin{pmatrix} 2 & 0 \\ 0 & 3 \end{pmatrix}$$

$$\begin{pmatrix} -1 & 0 \\ 0 & 2 \end{pmatrix}$$

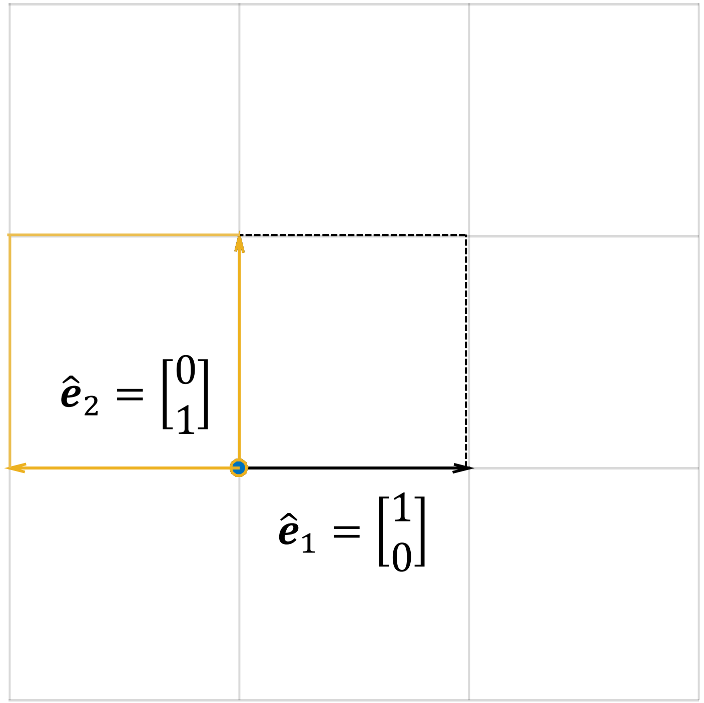
$$\begin{pmatrix} -1 & 0 \\ 0 & -1 \end{pmatrix}$$





Distorção

$$\begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}$$



Rotação

$$\begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}$$

Rotação: $\begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix}$

Qual a diferença entre essa rotação e $\begin{pmatrix} -1 & 0 \\ 0 & 1 \end{pmatrix}$?

MATRIZES ORTOGONAIS

Uma matriz quadrada $\boldsymbol{U} \in \mathbb{R}^{n \times n}$ é **ortogonal** se todas as suas colunas forem ortogonais entre si e normalizadas (as colunas são chamadas de ortonormais). Segue-se imediatamente da definição de ortogonalidade e normalidade que,

$$\boldsymbol{U}^T \boldsymbol{U} = I = \boldsymbol{U}\boldsymbol{U}^T$$

Exemplo clássico é a matriz de rotação... $\begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix}$

PRODUTO ENTRE MATRIZES

$$\boldsymbol{A} \in \mathbb{R}^{mxn}, \boldsymbol{B} \in \mathbb{R}^{nxp} \Rightarrow \boldsymbol{C} \in \mathbb{R}^{mxp}$$

$$\boldsymbol{C} = \boldsymbol{AB} = \begin{bmatrix} - & \boldsymbol{a}_1' & - \\ - & \boldsymbol{a}_2' & - \\ - & \vdots & - \\ - & \boldsymbol{a}_m' & - \end{bmatrix} \begin{bmatrix} | & | & & | \\ \boldsymbol{b}_1 & \boldsymbol{b}_2 & \cdots & \boldsymbol{b}_p \\ | & | & & | \end{bmatrix} = \begin{bmatrix} \boldsymbol{a}_1'\boldsymbol{b}_1 & \boldsymbol{a}_1'\boldsymbol{b}_2 & \cdots & \boldsymbol{a}_1'\boldsymbol{b}_p \\ \boldsymbol{a}_2'\boldsymbol{b}_1 & \boldsymbol{a}_2'\boldsymbol{b}_2 & \cdots & \boldsymbol{a}_2'\boldsymbol{b}_p \\ \vdots & \vdots & \ddots & \vdots \\ \boldsymbol{a}_m'\boldsymbol{b}_1 & \boldsymbol{a}_m'\boldsymbol{b}_2 & \cdots & \boldsymbol{a}_m'\boldsymbol{b}_p \end{bmatrix}$$

Multiplicação de matrizes vista como produto interno

$$\begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix} \begin{bmatrix} 10 & 20 \\ 30 & 40 \end{bmatrix} = \begin{bmatrix} 70 & 100 \\ 150 & 220 \end{bmatrix}$$

VÁRIAS VISÕES DE $\boldsymbol{C = AB}$

$$\boldsymbol{C = AB} = \begin{bmatrix} | & | & & | \\ \boldsymbol{a}_1 & \boldsymbol{a}_2 & \cdots & \boldsymbol{a}_n \\ | & | & & | \end{bmatrix} \begin{bmatrix} - & \boldsymbol{b}_1' & - \\ - & \boldsymbol{b}_2' & - \\ - & \vdots & - \\ - & \boldsymbol{b}_n' & - \end{bmatrix} = \sum_{i=1}^{n} \boldsymbol{a}_i \boldsymbol{b}_i'$$

Multiplicação de matrizes vista como produto externo

$$\boldsymbol{C = AB = A} \begin{bmatrix} | & | & & | \\ \boldsymbol{b}_1 & \boldsymbol{b}_2 & \cdots & \boldsymbol{b}_p \\ | & | & & | \end{bmatrix} = \begin{bmatrix} | & | & & | \\ \boldsymbol{Ab}_1 & \boldsymbol{Ab}_2 & \cdots & \boldsymbol{Ab}_p \\ | & | & & | \end{bmatrix},$$

$$\boldsymbol{c}_i = \boldsymbol{Ab}_i$$

Colunas de $\boldsymbol{C}$ vistas como produtos entre a matriz A e os vetores formados pelas colunas de $\boldsymbol{B}$

$$\boldsymbol{C = AB} = \begin{bmatrix} - & \boldsymbol{a}_1' & - \\ - & \boldsymbol{a}_2' & - \\ - & \vdots & - \\ - & \boldsymbol{a}_m' & - \end{bmatrix} \boldsymbol{B} = \begin{bmatrix} - & \boldsymbol{a}_1'\boldsymbol{B} & - \\ - & \boldsymbol{a}_2'\boldsymbol{B} & - \\ - & \vdots & - \\ - & \boldsymbol{a}_m'\boldsymbol{B} & - \end{bmatrix},$$

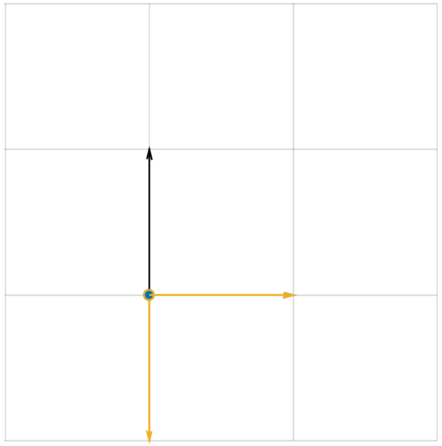
$$\boldsymbol{c}_i' = \boldsymbol{a}_i'\boldsymbol{B}$$

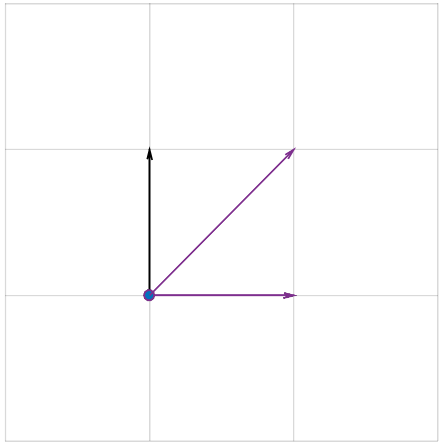
EXEMPLO

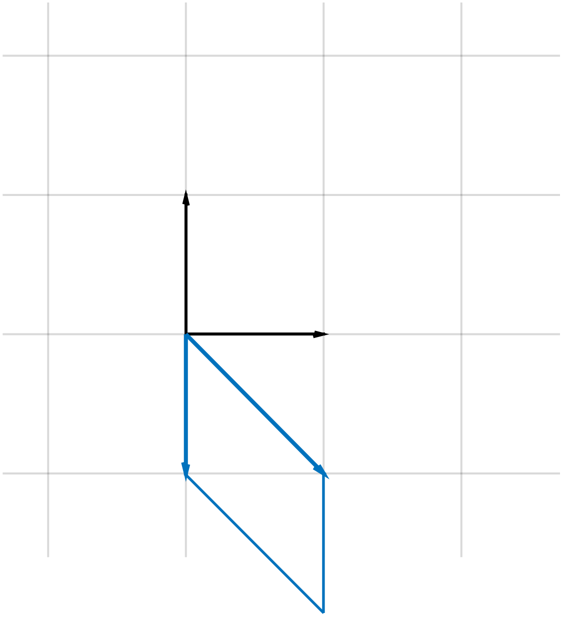
Dados,

$$\boldsymbol{T}_1 = \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix}, \qquad \boldsymbol{T}_2 = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}$$

$$\boldsymbol{T}_1\boldsymbol{T}_2 = \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix}\begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix} = \begin{bmatrix} 0 & 1 \\ -1 & -1 \end{bmatrix}$$

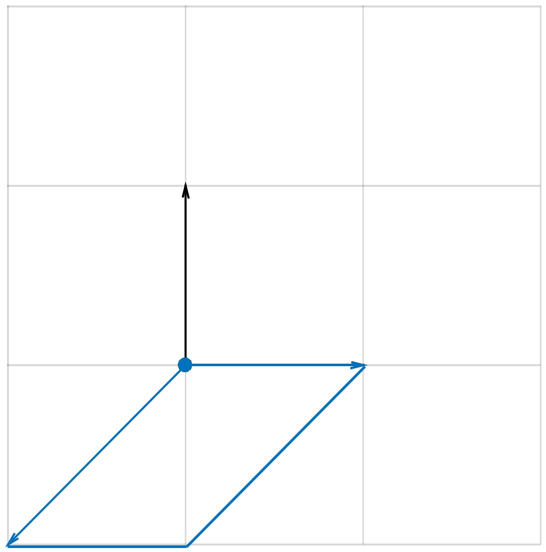






CUIDADO

$$\boldsymbol{T}_2\boldsymbol{T}_1 = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}\begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix} = \begin{bmatrix} -1 & 1 \\ -1 & 0 \end{bmatrix} \neq \boldsymbol{T}_1\boldsymbol{T}_2$$



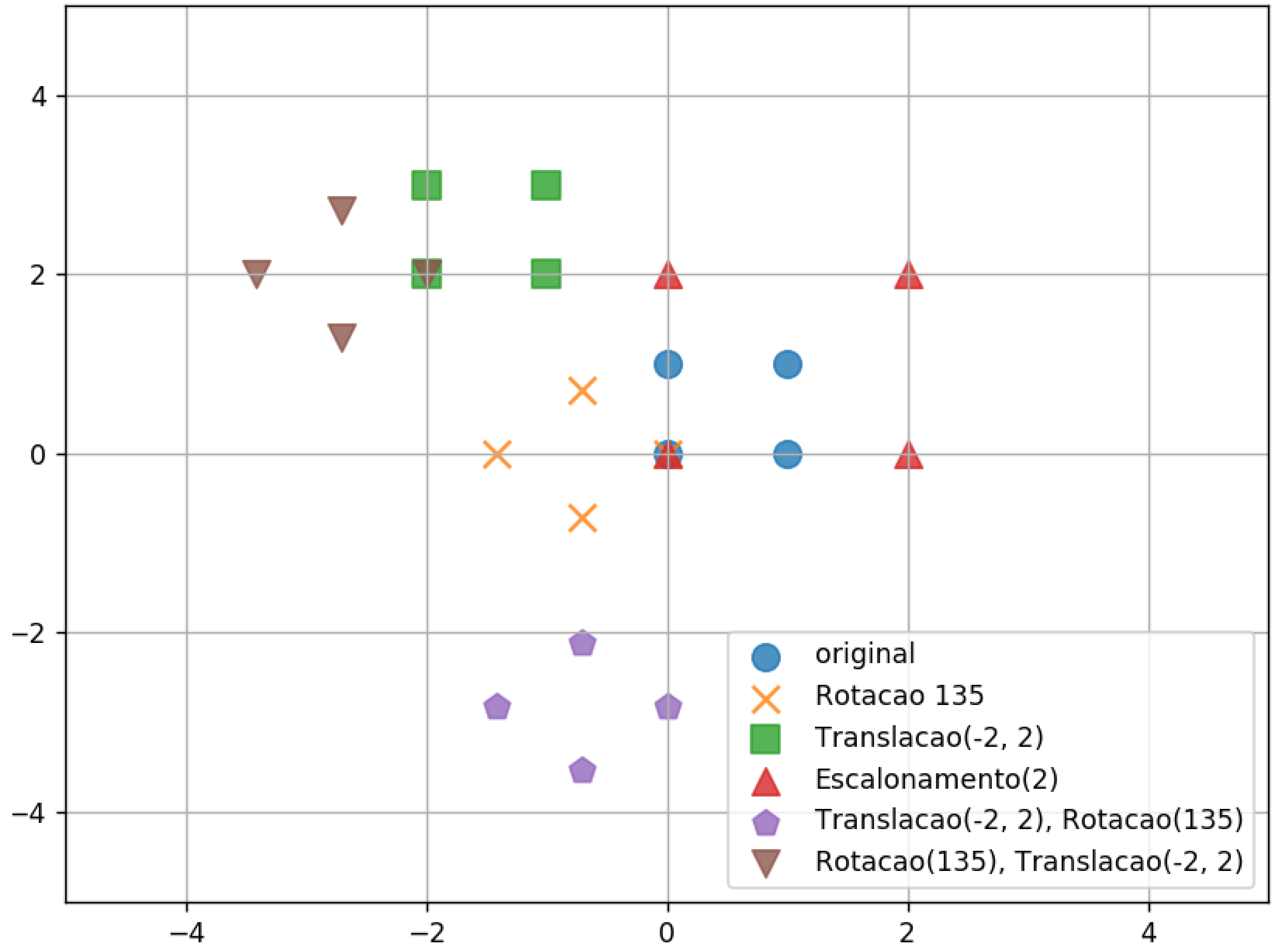
PROPRIEDADES DA MULTIPLICAÇÃO DE MATRIZES

Multiplicação de matrizes é distributiva e associativa,

$$(\boldsymbol{AB})\boldsymbol{C} = \boldsymbol{A}(\boldsymbol{BC})$$
$$\boldsymbol{A}(\boldsymbol{B} + \boldsymbol{C}) = \boldsymbol{AB} + \boldsymbol{AC}$$

mas somente em casos especiais é comutativa. Em geral,

$$\boldsymbol{AB} \neq \boldsymbol{BA}$$



PECE Programa de Educação Continuada

TRANSFORMAÇÃO GEOMÉTRICA DA IMAGEM

Envie cada pixel p da posição original (x, y) para a correspondente posição $(x', y') = T(x, y)$ na imagem transformada

$$x' = x + t_x$$
$$y' = y + t_y$$
$$\begin{bmatrix} 1 & 0 & t_x \\ 0 & 1 & t_y \\ 0 & 0 & 1 \end{bmatrix}$$

translação

$$x' = x * cos\Theta - y * sin\Theta$$
$$y' = x * cos\Theta + y * sin\Theta$$
$$\begin{bmatrix} cos\Theta & sin\Theta & 0 \\ -sin\Theta & cos\Theta & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

rotação

$$x' = c_x * x$$
$$y' = c_y * y$$
$$\begin{bmatrix} c_x & 0 & 0 \\ 0 & c_y & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

escalonamento

$$x' = x + s_v * y$$
$$y' = x * s_h + y$$
$$\begin{bmatrix} 1 & s_h & 0 \\ s_v & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

distorção

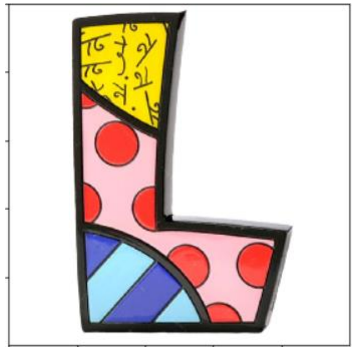
POR EXEMPLO...

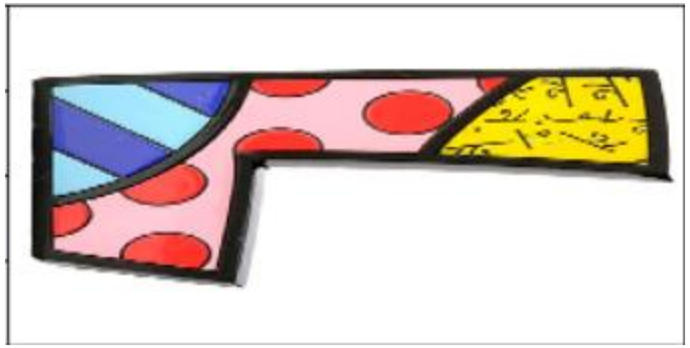
Sabendo que a letra **L** da figura ao lado sofreu seguinte transformação,

$$\begin{bmatrix} x' \\ y' \\ w' \end{bmatrix} = \boldsymbol{T} \begin{bmatrix} x \\ y \\ w \end{bmatrix}$$ onde $\boldsymbol{T}$ é composta das matrizes de escalonamento e rotação, respectivamente,

$$\boldsymbol{C} = \begin{pmatrix} 2 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}, \boldsymbol{R} = \begin{pmatrix} 0 & 1 & 0 \\ -1 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

Defina a matriz $\boldsymbol{T}$ para as seguintes transformações,





MATRIZ IDENTIDADE E MATRIZ INVERSA

$$\begin{pmatrix} 2 & 3 \\ 10 & 1 \end{pmatrix} \begin{bmatrix} m \\ b \end{bmatrix} = \begin{pmatrix} 8 \\ 13 \end{pmatrix}$$

$$\boldsymbol{Ar} = \boldsymbol{s}$$

$$\boldsymbol{A}^{-1}\boldsymbol{A} = \boldsymbol{I}$$

$\boldsymbol{A}^{-1}$ é a matriz inversa de $\boldsymbol{A}$

$$\boldsymbol{A}^{-1}\boldsymbol{Ar} = \boldsymbol{A}^{-1}\boldsymbol{s}$$

$$\boxed{\boldsymbol{r} = \boldsymbol{A}^{-1}\boldsymbol{s}}$$

O PROBLEMA É ACHAR A INVERSA...

$$\begin{pmatrix} 1 & 1 & 3 \\ 1 & 2 & 4 \\ 1 & 1 & 2 \end{pmatrix} \begin{pmatrix} b_{11} & b_{12} & b_{13} \\ b_{21} & b_{22} & b_{23} \\ b_{31} & b_{32} & a_{33} \end{pmatrix} = I = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

Vamos por etapas...

$$\begin{pmatrix} 1 & 1 & 3 \\ 1 & 2 & 4 \\ 1 & 1 & 2 \end{pmatrix} \begin{pmatrix} b_{11} \\ b_{21} \\ b_{31} \end{pmatrix} = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}$$

METODOLOGIA TRDICIONAL DE SUBSTITUIÇÃO REVERSA

$$\begin{pmatrix} 1 & 1 & 3 \\ 1 & 2 & 4 \\ 0 & 0 & 1 \end{pmatrix}\begin{pmatrix} b_{11} \\ b_{21} \\ b_{31} \end{pmatrix} = \begin{pmatrix} 1 \\ -1 \\ 1 \end{pmatrix}$$

$$\begin{pmatrix} 1 & 1 & 3 \\ 0 & 1 & 1 \\ 0 & 0 & 1 \end{pmatrix}\begin{pmatrix} b_{11} \\ b_{21} \\ b_{31} \end{pmatrix} = \begin{pmatrix} 1 \\ -1 \\ 1 \end{pmatrix}$$

$$\begin{pmatrix} 1 & 1 & 3 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}\begin{pmatrix} b_{11} \\ b_{21} \\ b_{31} \end{pmatrix} = \begin{pmatrix} 1 \\ -2 \\ 1 \end{pmatrix}$$

$$\begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}\begin{pmatrix} b_{11} \\ b_{21} \\ b_{31} \end{pmatrix} = \begin{pmatrix} 0 \\ -2 \\ 1 \end{pmatrix}$$

$$\begin{pmatrix} 1 & 1 & 3 \\ 1 & 2 & 4 \\ 1 & 1 & 2 \end{pmatrix}\begin{pmatrix} b_{11} \\ b_{21} \\ b_{31} \end{pmatrix} = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}$$

Eliminação

Substituição reversa

$$\begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}\begin{pmatrix} b_{11} \\ b_{21} \\ b_{31} \end{pmatrix} = \begin{pmatrix} 0 \\ -2 \\ 1 \end{pmatrix}$$

RESOLVENDO DE UMA VEZ SÓ...

$$\begin{pmatrix} 1 & 1 & 3 \\ 1 & 2 & 4 \\ 1 & 1 & 2 \end{pmatrix} \begin{pmatrix} b_{11} & b_{12} & b_{13} \\ b_{21} & b_{22} & b_{23} \\ b_{31} & b_{32} & a_{33} \end{pmatrix} = I = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

$$\begin{pmatrix} 1 & 1 & 3 \\ 0 & 1 & 1 \\ 0 & 0 & 1 \end{pmatrix} \qquad \begin{pmatrix} 1 & 0 & 0 \\ -1 & 1 & 0 \\ 1 & 0 & -1 \end{pmatrix}$$

$$\begin{pmatrix} 1 & 1 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \qquad \begin{pmatrix} -2 & 0 & 3 \\ -2 & 1 & 1 \\ 1 & 0 & -1 \end{pmatrix}$$

$$\begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \qquad \boxed{\begin{pmatrix} 0 & -1 & 2 \\ -2 & 1 & 1 \\ 1 & 0 & -1 \end{pmatrix}}$$

DESSA FORMA, PODEMOS RESOLVER O PROBLEMA DIRETAMENTE

$$\begin{pmatrix} 1 & 1 & 3 \\ 1 & 2 & 4 \\ 1 & 1 & 2 \end{pmatrix}\begin{pmatrix} a \\ b \\ c \end{pmatrix} = \begin{pmatrix} 15 \\ 21 \\ 13 \end{pmatrix}$$

$$\begin{pmatrix} 1 & 1 & 3 \\ 0 & 1 & 1 \\ 0 & 0 & 1 \end{pmatrix}\begin{pmatrix} a \\ b \\ c \end{pmatrix} = \begin{pmatrix} 15 \\ 6 \\ 2 \end{pmatrix}$$

$$\begin{pmatrix} 1 & 1 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}\begin{pmatrix} a \\ b \\ c \end{pmatrix} = \begin{pmatrix} 9 \\ 4 \\ 2 \end{pmatrix}$$

Eliminação

Substituição reversa

$$\begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}\begin{pmatrix} a \\ b \\ c \end{pmatrix} = \begin{pmatrix} 5 \\ 4 \\ 2 \end{pmatrix}$$

EXEMPLO

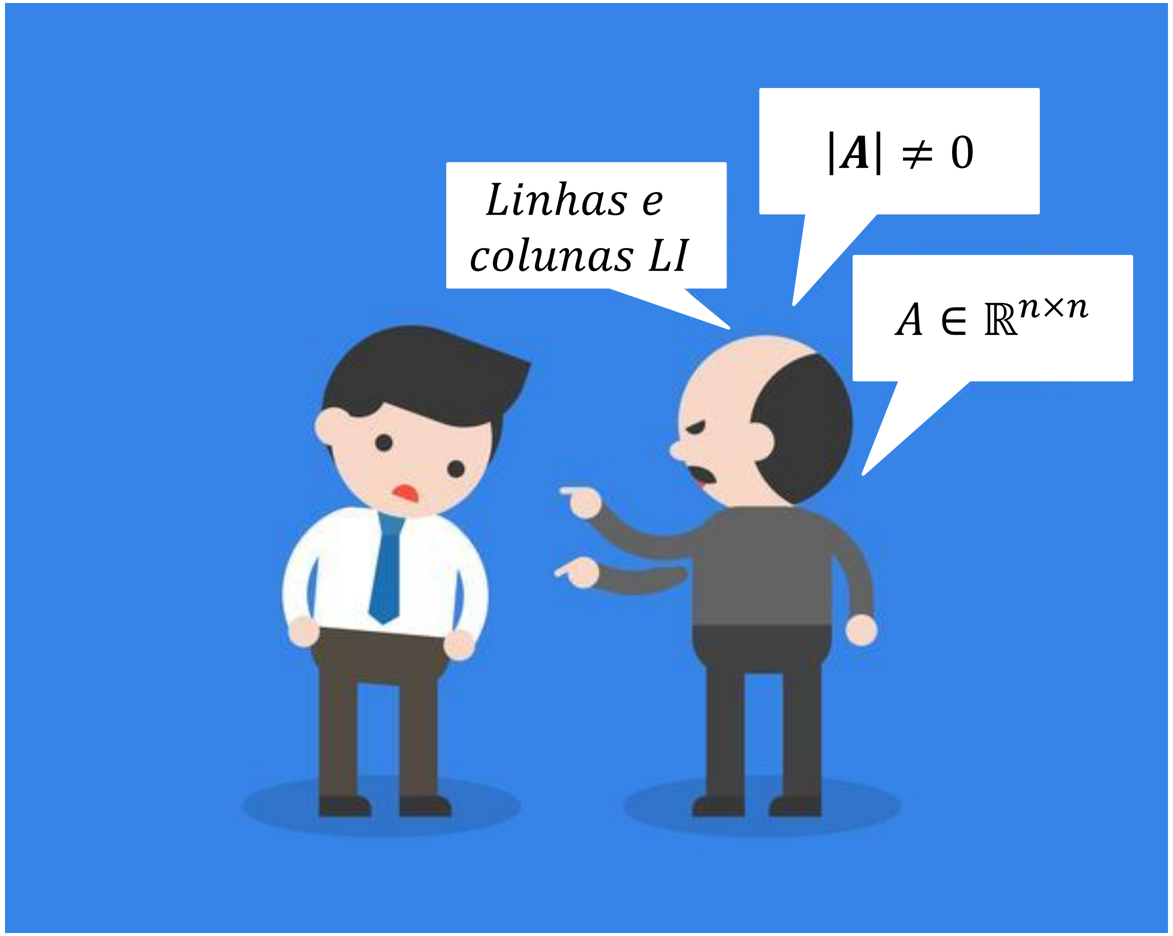
Use eliminação e substituição reversa para resolver o sistema:

$$\begin{pmatrix} 1 & 1 & 2 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} a \\ b \\ c \end{pmatrix} = \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}$$

PODE-SE INVERTER QUALQUER MATRIZ?

Não!

- A matriz deve ser quadrada e não singular!
- Matriz é singular quando tem linhas e/ou colunas linearmente dependentes.
- Por consequência, seu determinante não pode ser nulo.



DETERMINANTE DE $\boldsymbol{A} \in \mathbb{R}^{n \times n}$

O determinante de uma matriz $\boldsymbol{A}$ quadrada, $|\boldsymbol{A}|$ ou $\det(\boldsymbol{A})$, é um número, ié, $\mathbb{R}^{n \times n} \rightarrow \mathbb{R}$

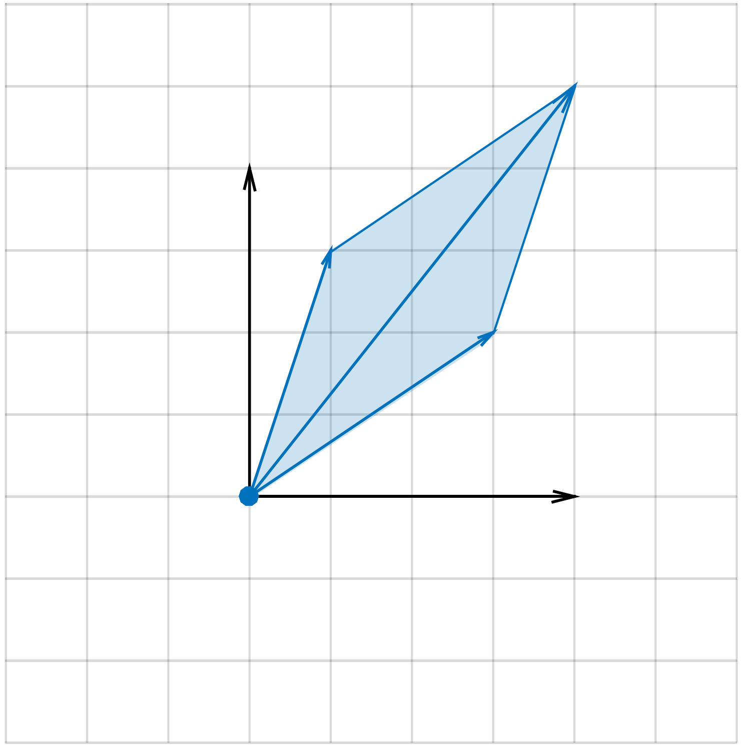
Começaremos com uma interpretação geométrica do determinante...

$$A = \begin{bmatrix} | & | & & | \\ a_1 & a_2 & \cdots & a_n \\ | & | & & | \end{bmatrix}$$

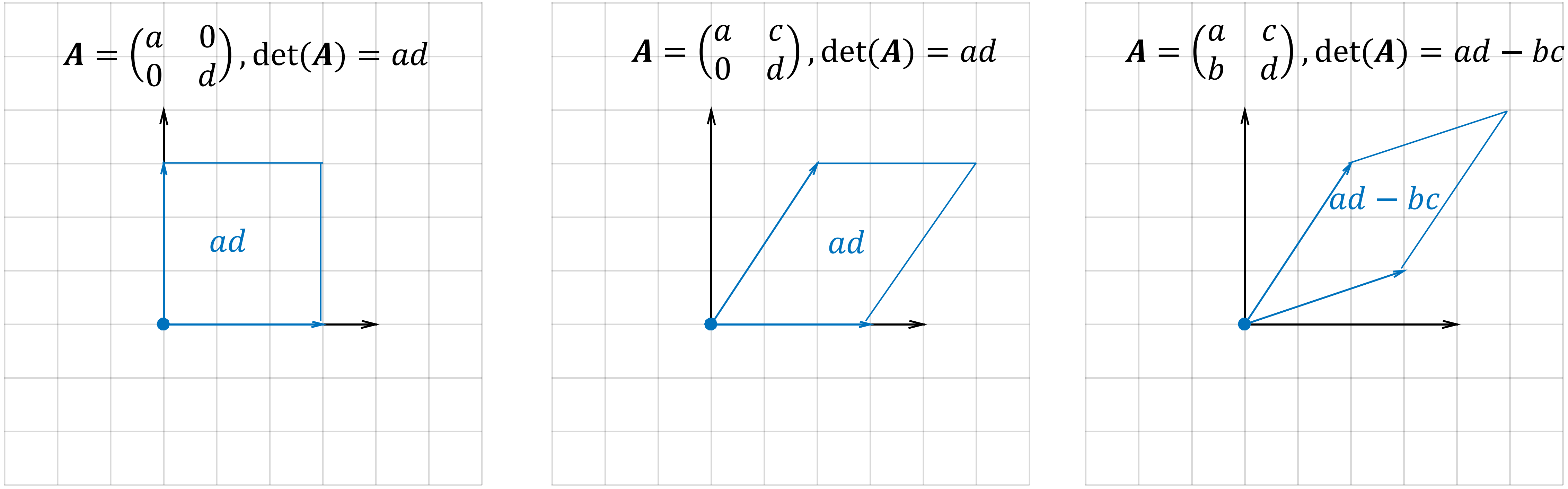
Considere o conjunto de pontos $S \in \mathbb{R}^n$ formado tomando todas as combinações lineares possíveis dos vetores de linha $\boldsymbol{a}_1 \ldots \boldsymbol{a}_n \in \mathbb{R}^n$ de A, onde os coeficientes da combinação linear estão todos entre 0 e 1.

$$\boldsymbol{A} = \begin{bmatrix} 1 & 3 \\ 3 & 2 \end{bmatrix}$$

$$\boldsymbol{a}_1 = \begin{bmatrix} 1 \\ 3 \end{bmatrix} \qquad \boldsymbol{a}_2 = \begin{bmatrix} 3 \\ 2 \end{bmatrix}$$

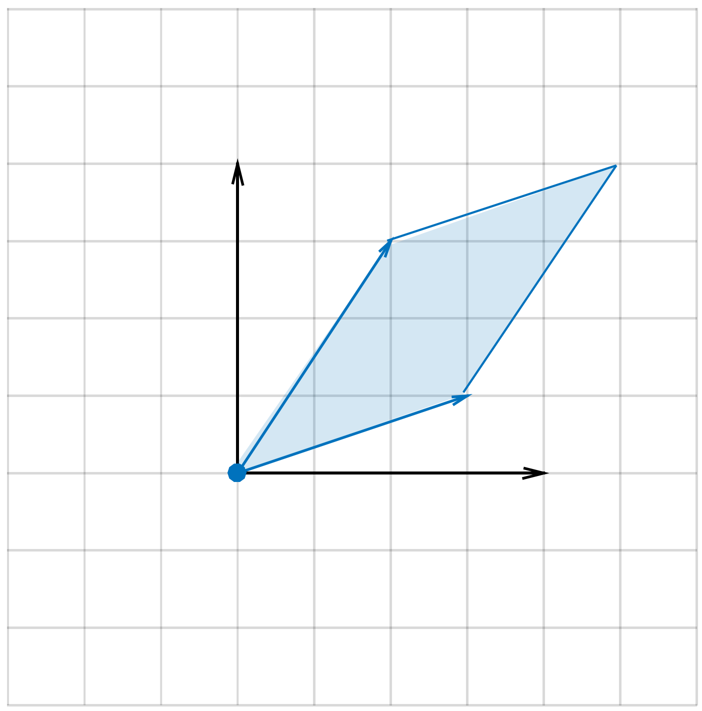


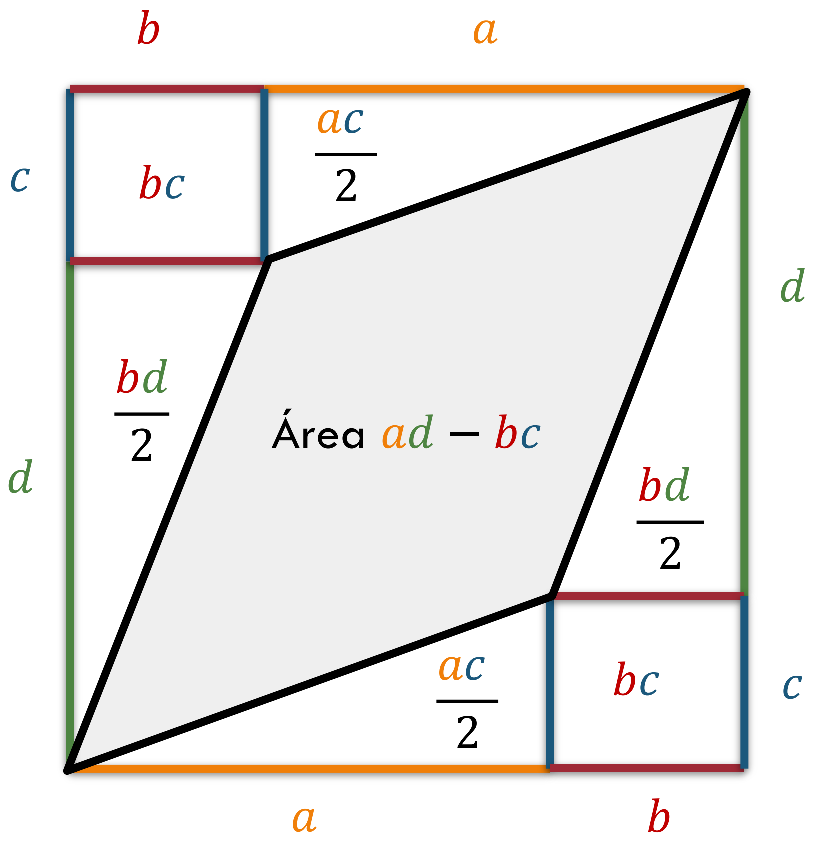
DETERMINANTE DE UMA MATRIZ



GENERALIZANDO O CASO 2D

$$\boldsymbol{A} = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$$





PORTANTO

Para $\boldsymbol{A} \in \mathbb{R}^{n \times n}$,

$$|\boldsymbol{A}| = \sum_{i=1}^{n} (-1)^{i+j} a_{ij} \left|\boldsymbol{A}_{\backslash \mathrm{i} \backslash \mathrm{j}}\right| \qquad \text{Para qualquer } j \in 1, \dots, n$$

$$= \sum_{j=1}^{n} (-1)^{i+j} a_{ij} \left|\boldsymbol{A}_{\backslash \mathrm{i} \backslash \mathrm{j}}\right| \qquad \text{Para qualquer } i \in 1, \dots, n$$

$\boldsymbol{A}_{\backslash \mathrm{i} \backslash \mathrm{j}} \in \mathbb{R}^{(n-1)\times(n-1)}$ é a matriz que resulta da exclusão da i-ésima linha e j-ésima coluna de $\boldsymbol{A}$.

MATRIZ ADJUNTA

$$\mathrm{adj}(\boldsymbol{A}) \in \mathbb{R}^{n\times n}, \qquad (\mathrm{adj}(\boldsymbol{A}))_{ij} = (-1)^{i+j}\left|\boldsymbol{A}_{\backslash \mathrm{i} \backslash \mathrm{j}}\right|$$

$$\boldsymbol{A}^{-1} = \frac{1}{|\boldsymbol{A}|}\mathrm{adj}(\boldsymbol{A})$$

ARRISCAMOS EXPLICITAR O DETERMINANTE DE ALGUMAS MATRIZES...

$$
\begin{aligned}
\left|\begin{bmatrix} a_{11} \end{bmatrix}\right| &= a_{11} \\
\left|\begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix}\right| &= a_{11}a_{22} - a_{12}a_{21} \\
\left|\begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix}\right| &= \begin{array}{l} a_{11}a_{22}a_{33} + a_{12}a_{23}a_{31} + a_{13}a_{21}a_{32} \\ \quad -a_{11}a_{23}a_{32} - a_{12}a_{21}a_{33} - a_{13}a_{22}a_{31} \end{array}
\end{aligned}
$$

PROPRIEDADES...

$\boldsymbol{A} \in \mathbb{R}^{n \times n}, |\boldsymbol{A}| = |\boldsymbol{A}^T|$

$\boldsymbol{A}, \boldsymbol{B} \in \mathbb{R}^{n \times n}, |\boldsymbol{AB}| = |\boldsymbol{A}|\ |\boldsymbol{B}|$
$|\boldsymbol{A}^{-1}| = 1/|\boldsymbol{A}|$

$$\begin{bmatrix} - & \boldsymbol{a}_2' & - \\ - & \boldsymbol{a}_1' & - \\ - & \vdots & - \\ - & \boldsymbol{a}_m' & - \end{bmatrix} = -|\boldsymbol{A}|$$

$$\begin{bmatrix} - & t\boldsymbol{a}_1' & - \\ - & \boldsymbol{a}_2' & - \\ - & \vdots & - \\ - & \boldsymbol{a}_m' & - \end{bmatrix} = t|\boldsymbol{A}|$$

IMPORTANTE PROPRIEDADE DO DETERMINANTE

O determinante é um *escalonamento* da inversa. Por exemplo, a inversa de uma matriz 2x2,

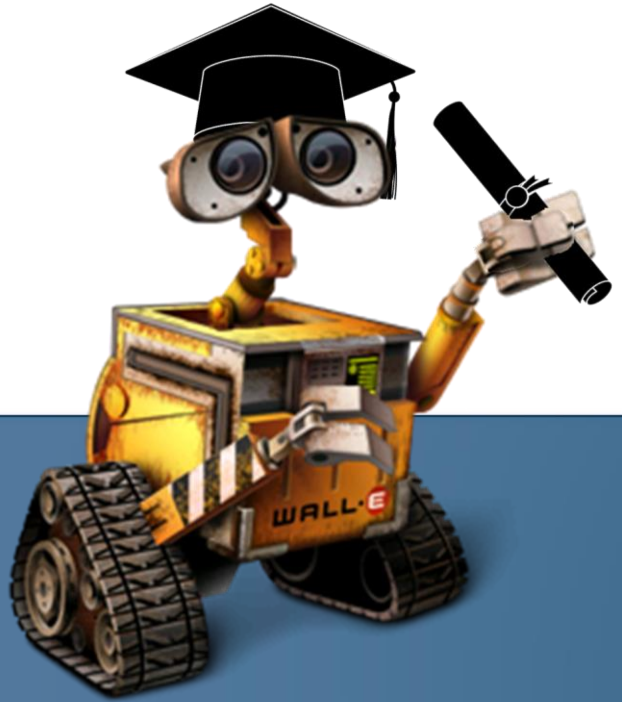
$$\boldsymbol{A} = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$$

é dada por,

$$\boldsymbol{A}^{-1} = \frac{1}{\det \boldsymbol{A}} \begin{bmatrix} d & -b \\ -c & a \end{bmatrix}$$

Portanto, uma matriz cujo determinante é nulo, não é formada somente de vetores LI. Em outras palavras, a matriz é singular.

PECE Programa de Educação Continuada

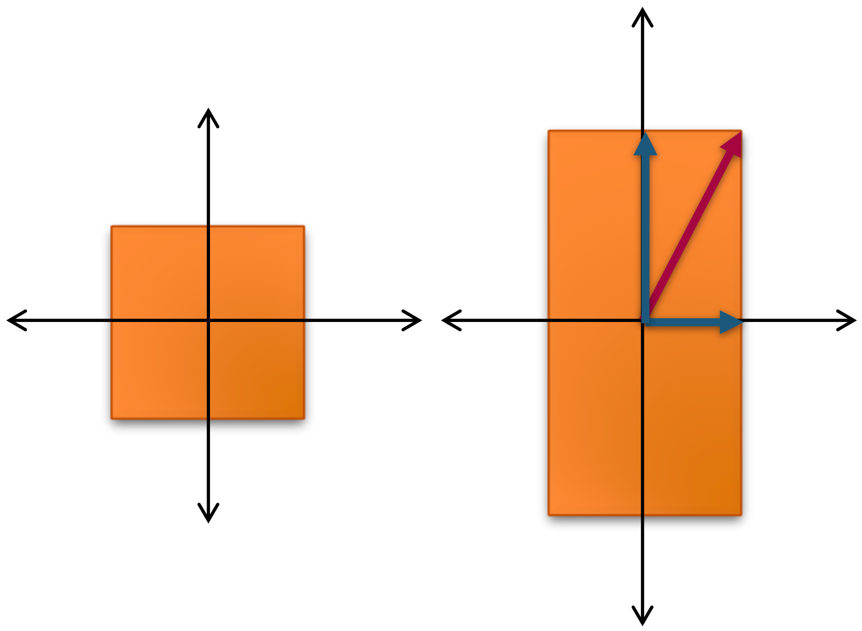


AUTOVALORES E AUTOVETORES

A palavra "eigen" é talvez mais útil se for traduzida do alemão como “característica”.

Então, quando falamos de um *eigenproblem (problema característico)*, estamos falando em encontrar as propriedades características de algo. Mas característica de quê?

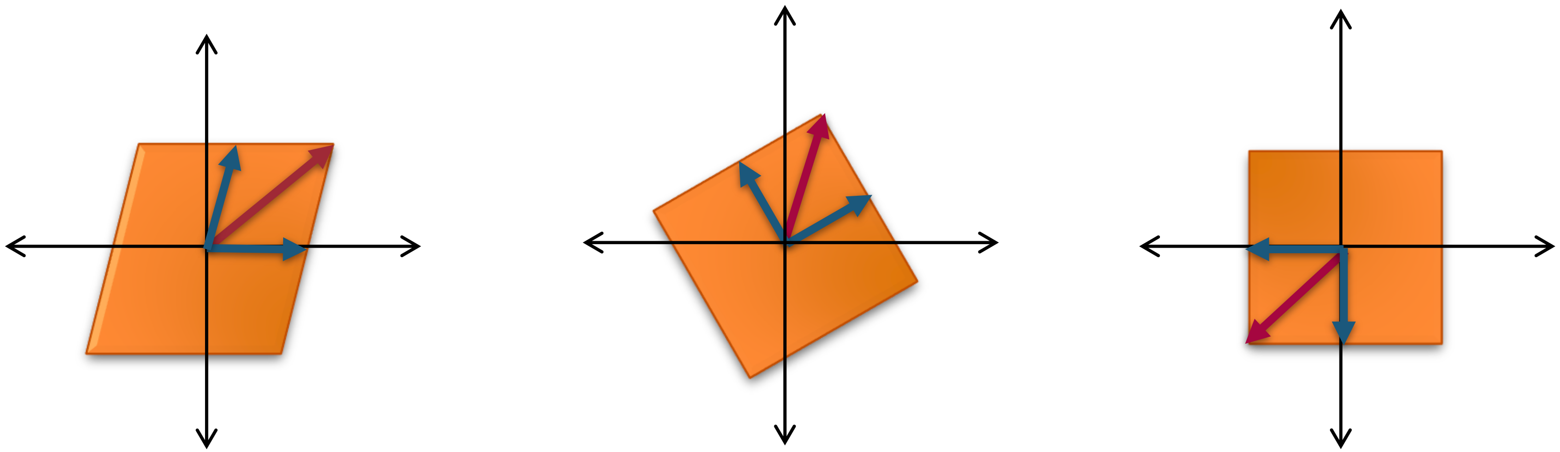
AUTOVALORES E AUTOVETORES

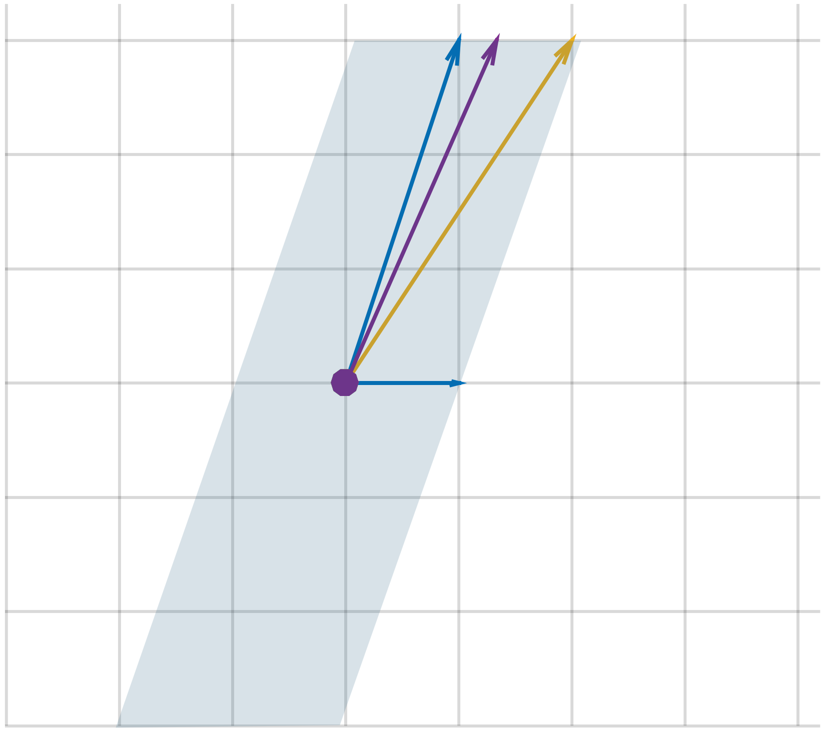


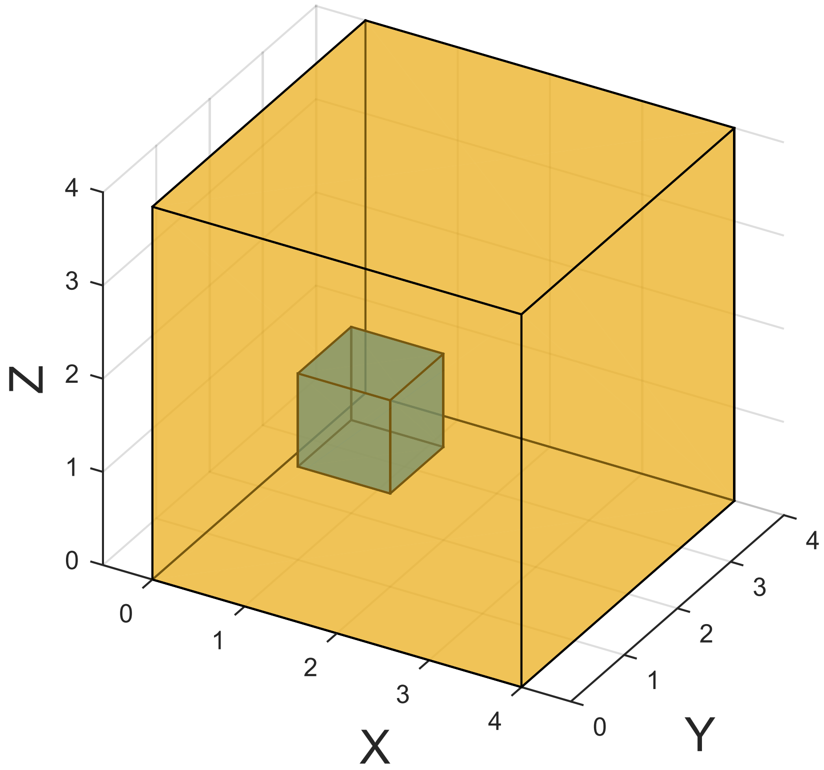
Além dos **vetores horizontal e vertical**, a direção de qualquer outro vetor foi alterado pelo escalonamento vertical que fizemos.

Portanto, os **vetores horizontal e vertical** são especiais, são característicos dessa transformação específica, e é por isso que nos referimos a eles como **autovetores**.

QUANTOS AUTOVETORES???






Z
0
1
2
3
4
X
0
1
2
3
4
Y
0
1
2
3
4

GENERALIZANDO...

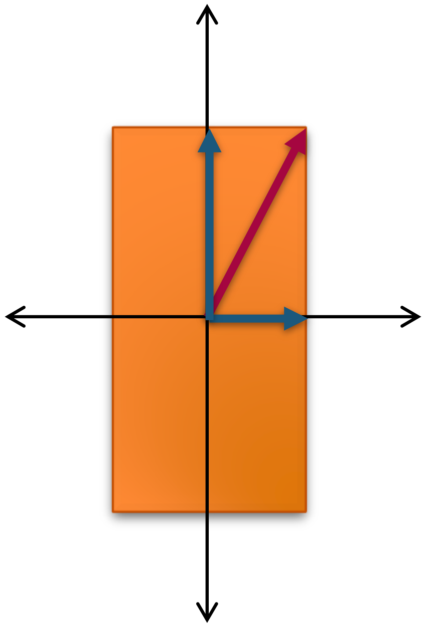
$$\boldsymbol{Av} = \lambda\boldsymbol{v}, \quad \boldsymbol{v} \neq 0,$$

$$(\lambda\boldsymbol{I} - \boldsymbol{A})\boldsymbol{v} = 0, \quad \boldsymbol{v} \neq 0,$$

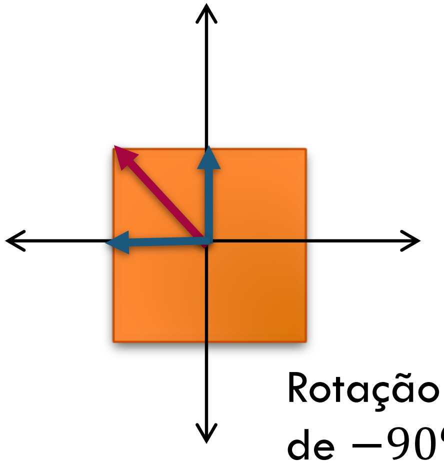
$$|(\lambda\boldsymbol{I} - \boldsymbol{A})| = 0$$

EXEMPLOS

$$|(\lambda \boldsymbol{I} - \boldsymbol{A})| = 0$$



$$\boldsymbol{A} = \begin{pmatrix} 1 & 0 \\ 0 & 2 \end{pmatrix}$$



Rotação de -90^o

$$\boldsymbol{A} = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}$$

Caso da matriz $\boldsymbol{A}$ 2×2

$$\lambda^2 - \mathrm{tr}(\boldsymbol{A})\,\lambda + \det \mathbf{A} = 0$$

EXEMPLO

Para matriz

$$\boldsymbol{A} = \begin{pmatrix} 5 & 4 \\ -4 & -3 \end{pmatrix}$$

Qual o *polinômio característico* e sua solução?

SAIBA QUE...

O traço de um $\boldsymbol{A}$ é igual à soma de seus autovalores,

$$\operatorname{tr}\boldsymbol{A} = \sum_{i=1}^{n} \lambda_i$$

O determinante de $\boldsymbol{A}$ é igual ao produto de seus autovalores,

$$\det\boldsymbol{A} = \prod_{i=1}^{n} \lambda_i$$

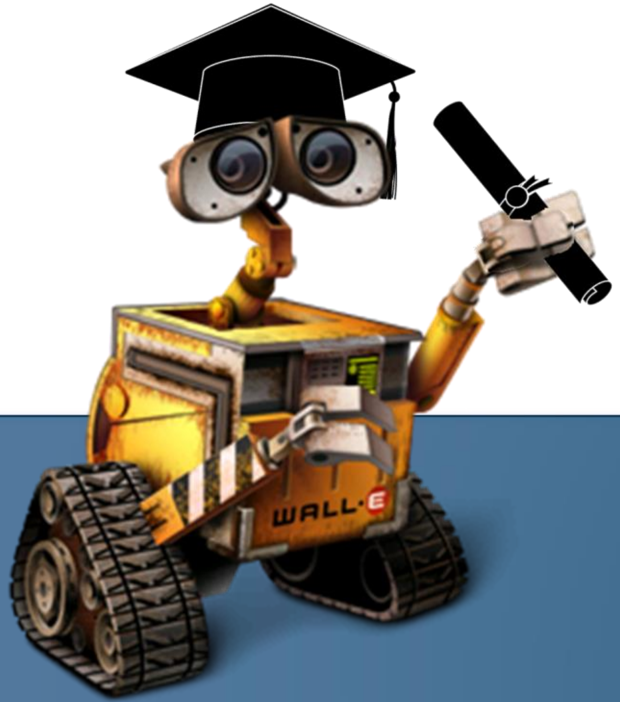
O rank de $\boldsymbol{A}$ é igual ao número de autovalores diferentes de zero de $\boldsymbol{A}$.

Os autovalores de uma matriz diagonal $\boldsymbol{D} = diag\,(d_1, d_2, \dots, d_n)$ são as entradas diagonais $d_1, d_2, \dots, d_n$.

MATRIZES SIMÉTRICAS

Duas propriedades importantes surgem quando olhamos para os autovalores e autovetores de uma matriz simétrica $\boldsymbol{A}$,

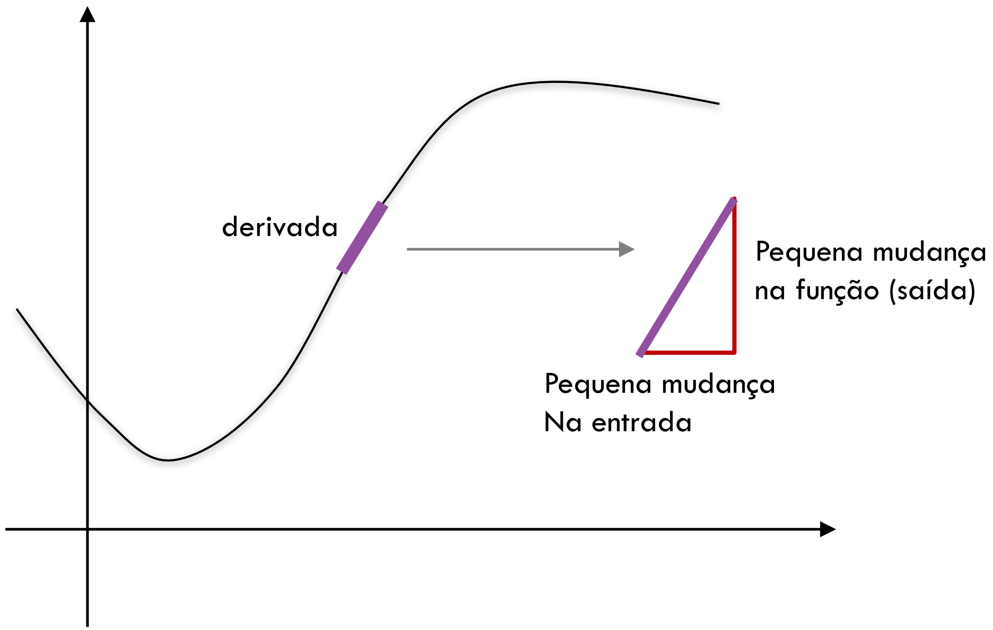
- todos os autovalores de $\boldsymbol{A}$ são reais;
- os autovetores de $\boldsymbol{A}$ são ortonormais, ou seja, a matriz composta dos autovetores é uma matriz ortogonal.



GRADIENTE

Para otimizar uma solução ou minimizar um erro, precisamos saber como diferenciar funções objetivo e erros , que são expressas com matrizes e vetores.

INTUIÇÃO INICIAL DE GRADIENTE



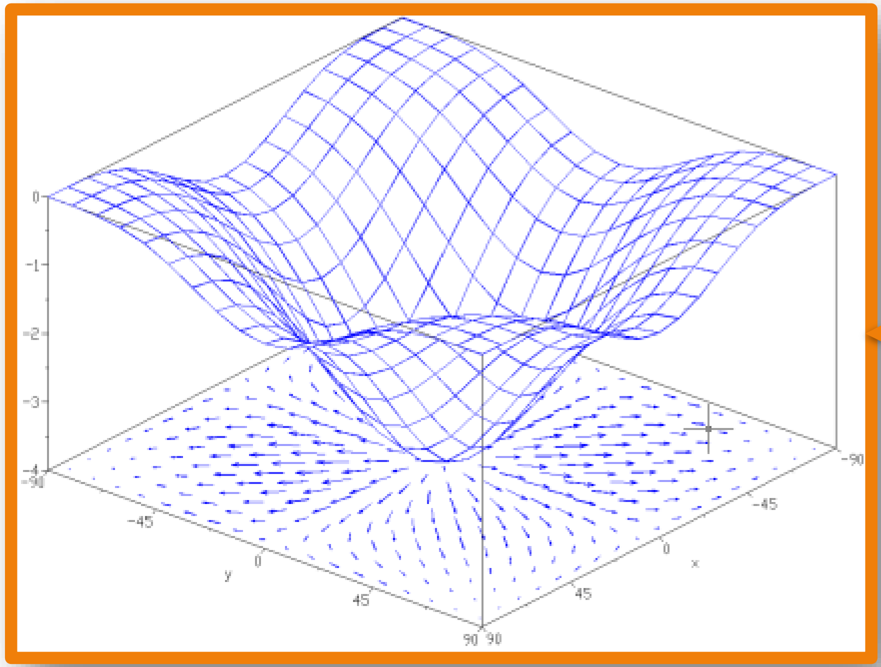
A derivada de uma função é basicamente o quanto a função muda se você alterar a entrada em uma quantia muito pequena. No espaço unidimensional, o gradiente é simplesmente a inclinação da função naquele ponto.

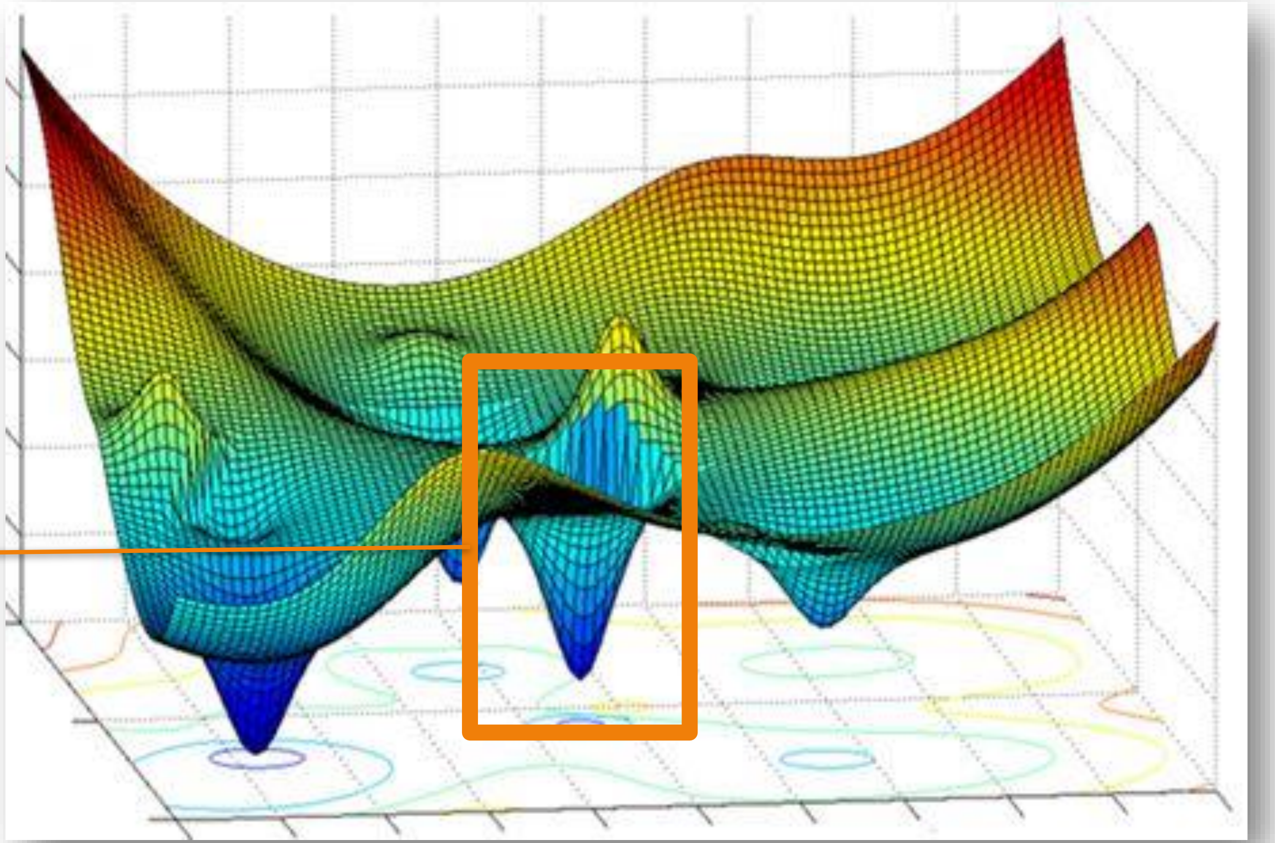
O gradiente da função $f(x_1, x_2, \dots)$ é simplesmente sua derivada, ié, a taxa de mudança, da função para cada direção x_i .

PECE Programa de Educação Continuada

GRADIENTE

Muitas funções têm múltiplas entradas, e cada entrada pode ser considerada uma direção diferente no espaço.





GRADIENTE

Suponha que $f\colon \mathbb{R}^{mxn} \to R$. Então o gradiente de f com respeito a $\boldsymbol{A} \in \mathbb{R}^{mxn}$, é uma matriz $n \times n$ de derivadas parciais,

$$\nabla_{\boldsymbol{A}} f(\boldsymbol{A}) \in \mathbb{R}^{mxn} = \begin{bmatrix} \frac{\partial f(\boldsymbol{A})}{\partial a_{11}} & \frac{\partial f(\boldsymbol{A})}{\partial a_{12}} & \cdots & \frac{\partial f(\boldsymbol{A})}{\partial a_{1n}} \\ \frac{\partial f(\boldsymbol{A})}{\partial a_{21}} & \frac{\partial f(\boldsymbol{A})}{\partial a_{22}} & \cdots & \frac{\partial f(\boldsymbol{A})}{\partial a_{2n}} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial f(\boldsymbol{A})}{\partial a_{m1}} & \frac{\partial f(\boldsymbol{A})}{\partial a_{m2}} & \cdots & \frac{\partial f(\boldsymbol{A})}{\partial a_{mn}} \end{bmatrix}$$

$$\left(\nabla_{\boldsymbol{A}} f(\boldsymbol{A})\right)_{ij} = \frac{\partial f(\boldsymbol{A})}{\partial a_{ij}}$$

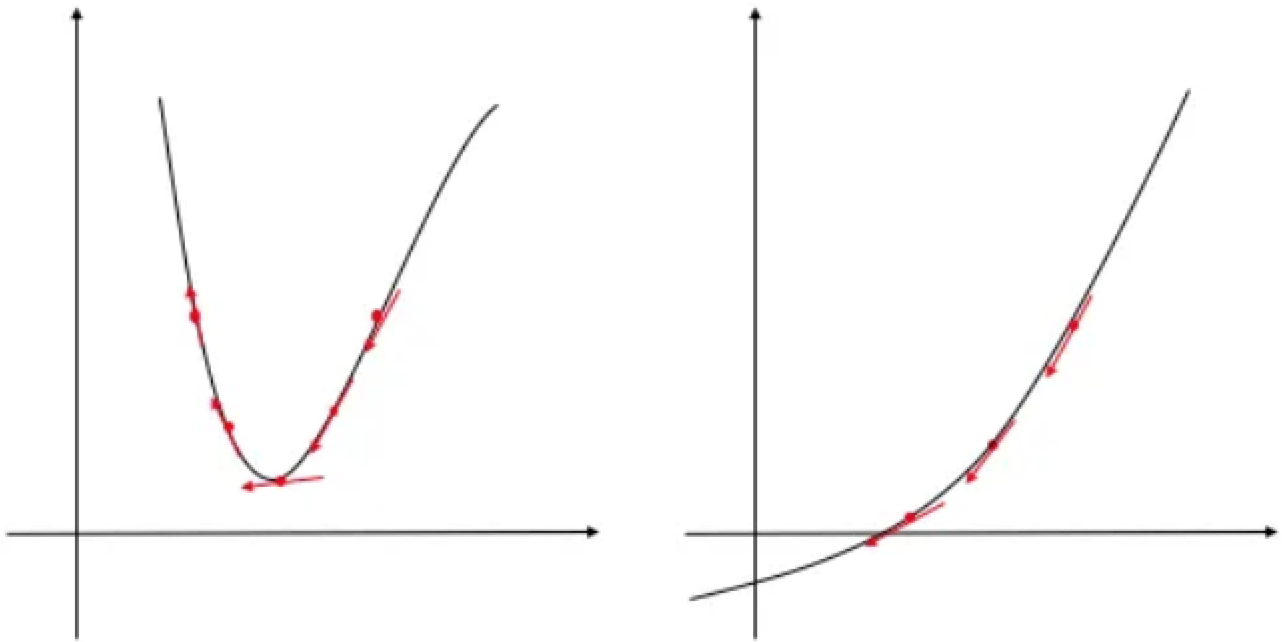
$\nabla_{\boldsymbol{A}} f(\boldsymbol{A})$ tem sempre a dimensão de $\boldsymbol{A}$. Supondo um vetor, $\boldsymbol{x} \in \mathbb{R}^n$, o gradiente é uma matriz $n \times 1$ de derivadas parciais,

$$\nabla_{\boldsymbol{x}} f(\boldsymbol{x}) = \begin{bmatrix} \dfrac{\partial f(\boldsymbol{x})}{\partial x_1} \\ \dfrac{\partial f(\boldsymbol{x})}{\partial x_2} \\ \vdots \\ \dfrac{\partial f(\boldsymbol{x})}{\partial x_m} \end{bmatrix}$$

Gradiente de uma função **é definido apenas se a função tem valor real escalar**. Não podemos, por exemplo, pegar o gradiente da função $\boldsymbol{A}\boldsymbol{x}$, $\boldsymbol{A} \in \mathbb{R}^{m \times n}$ e $\boldsymbol{x} \in \mathbb{R}^n$ em relação a $\boldsymbol{x}$.

DERIVADAS DE SEGUNDA ORDEM

E as derivadas de segunda ordem, ié, a derivada da derivada? Em outras palavras, é a taxa de variação da inclinação.



HESSIANA

A **Hessiana** é uma **matriz** que organiza todas as derivadas parciais de segunda ordem de uma função. Suponha que $f: \mathbb{R}^n \to R$. Então a matriz hessiana em relação a x, escrita $\nabla_{\boldsymbol{x}}^2 f(\boldsymbol{x}) = \boldsymbol{H}$, é uma matriz $n \times n$ simétrica de derivadas parciais,

Cada linha representa a alteração do gradiente em uma determinada direção.

$$\nabla_{\boldsymbol{x}}^2 f(\boldsymbol{x}) \in \mathbb{R}^{nxn} = \begin{bmatrix} \frac{\partial^2 f(\boldsymbol{x})}{\partial x_1^2} & \frac{\partial^2 f(\boldsymbol{x})}{\partial x_1 \partial x_2} & \cdots & \frac{\partial^2 f(\boldsymbol{x})}{\partial x_1 \partial x_n} \\ \frac{\partial^2 f(\boldsymbol{x})}{\partial x_2 \partial x_1} & \frac{\partial^2 f(\boldsymbol{x})}{\partial x_2^2} & \cdots & \frac{\partial^2 f(\boldsymbol{x})}{\partial x_2 \partial x_n} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial^2 f(\boldsymbol{x})}{\partial x_n \partial x_1} & \frac{\partial^2 f(\boldsymbol{x})}{\partial x_n \partial x_2} & \cdots & \frac{\partial^2 f(\boldsymbol{x})}{\partial x_n^2} \end{bmatrix}$$

$$\left(\nabla_{\boldsymbol{x}}^2 f(\boldsymbol{x})\right)_{ij} = \frac{\partial^2 f(\boldsymbol{x})}{\partial x_i \partial x_j}$$

MÉTODO DO GRADIENTE DESCENDENTE

Hipótese:

$$h_w = \boldsymbol{w}^T \boldsymbol{x} = w_0 x_0 + w_1 x_1 + w_2 x_2 + \cdots + w_n x_n$$

Parâmetros:

$$\boldsymbol{w} = [w_0 \;\; w_1 \;\; w_2 \;\; \ldots \;\; w_n]$$

m: número de dados
n: número de features

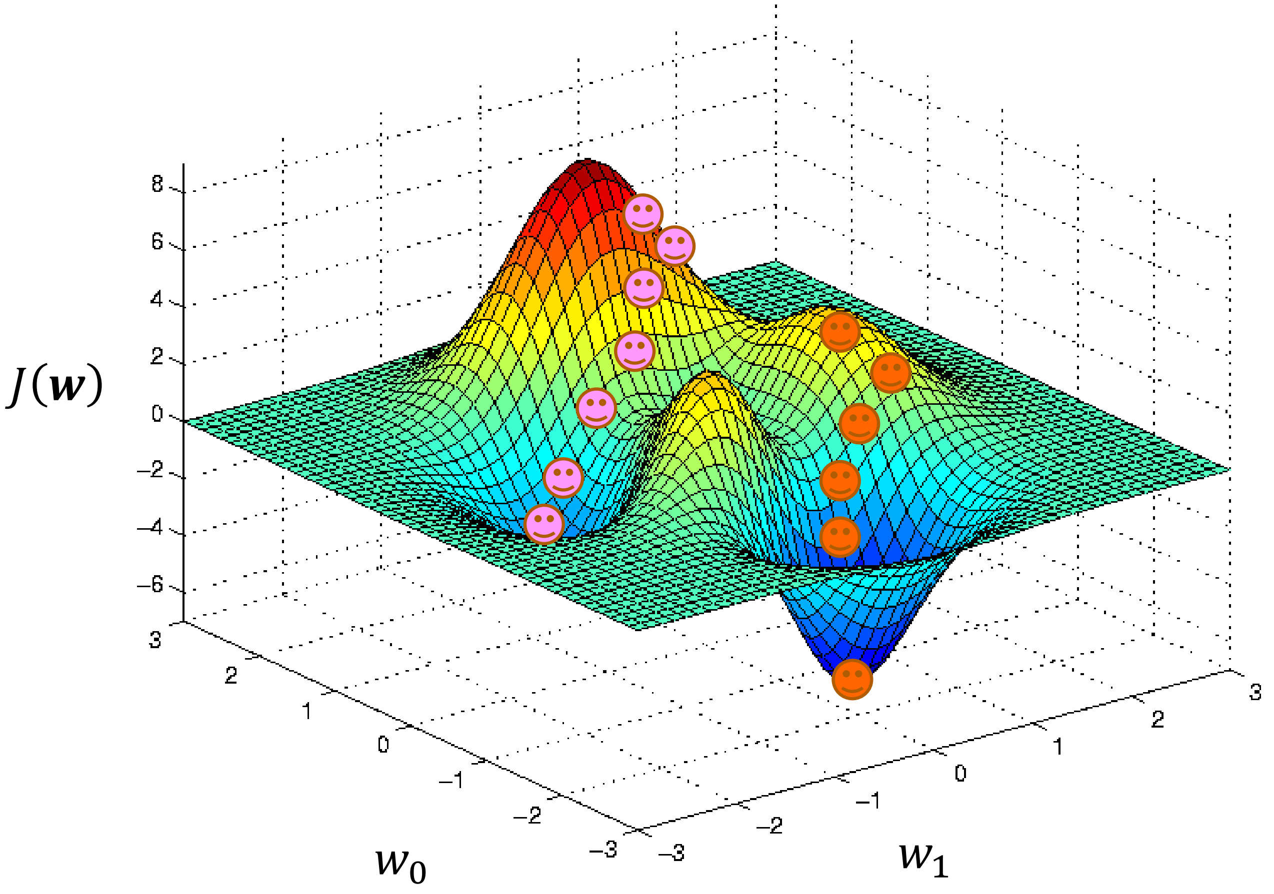
Função custo:

$$J(\boldsymbol{w}) = \frac{1}{2m} \sum_{i=1}^{m} \left(h_w\left(x^{(i)}\right) - y^{(i)}\right)^2$$

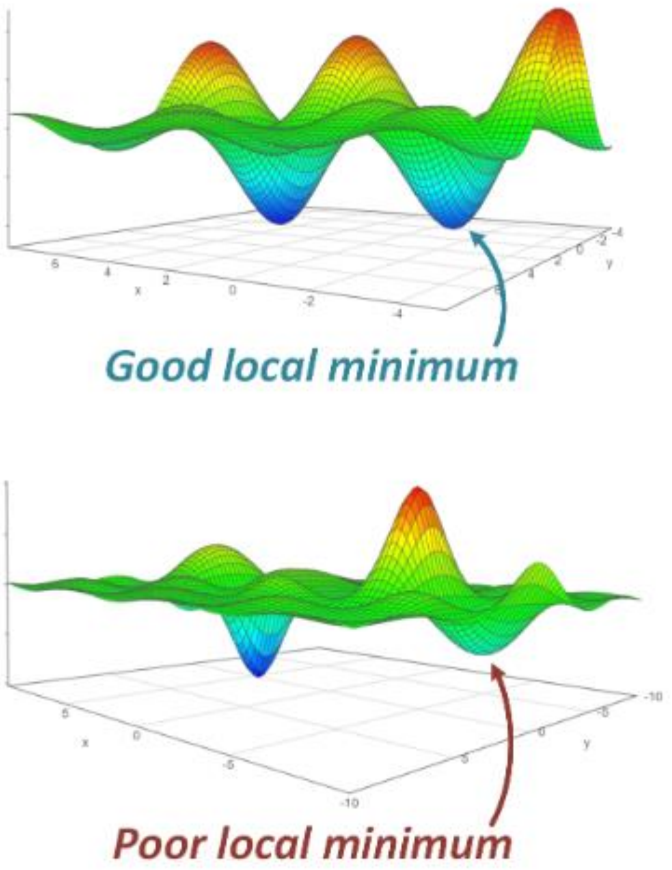
Objetivo:

$$\min_{\boldsymbol{w}} J(\boldsymbol{w})$$

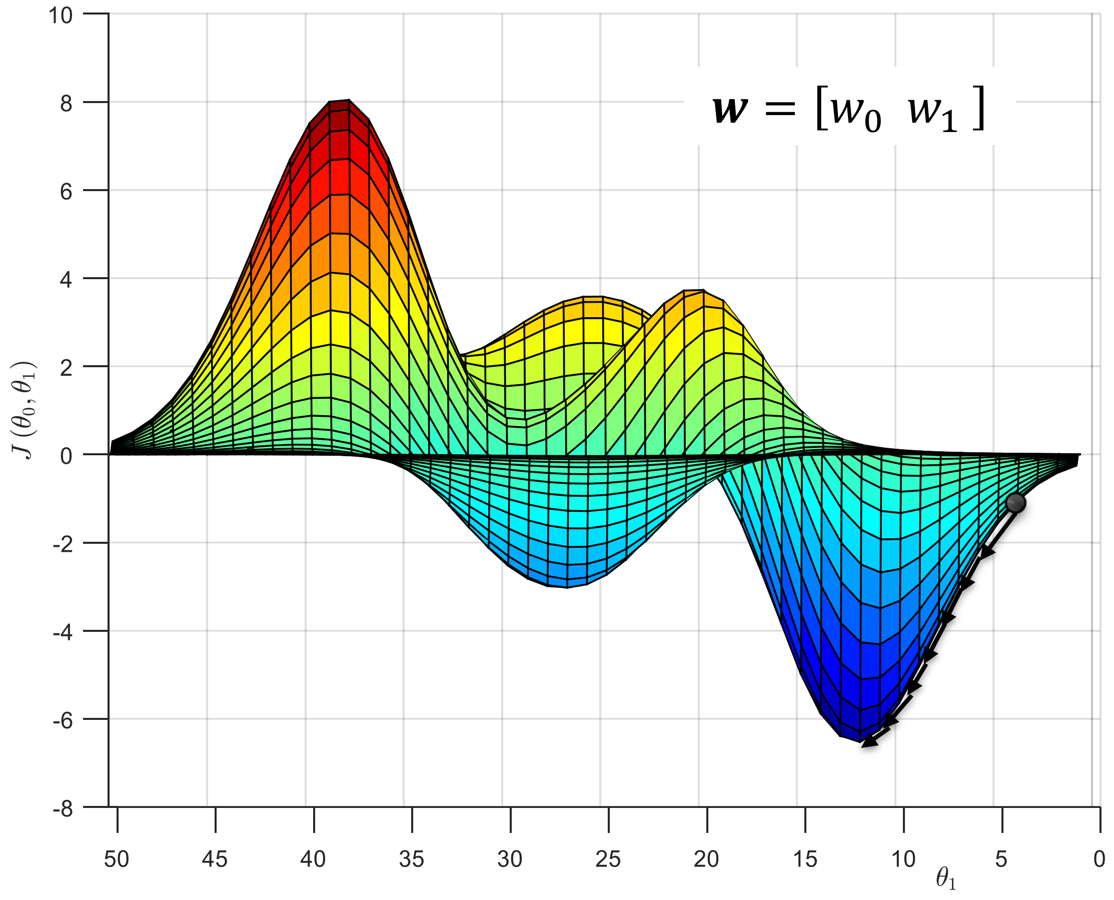
INTUIÇÃO SOBRE GRADIENTE DESCENDENTE



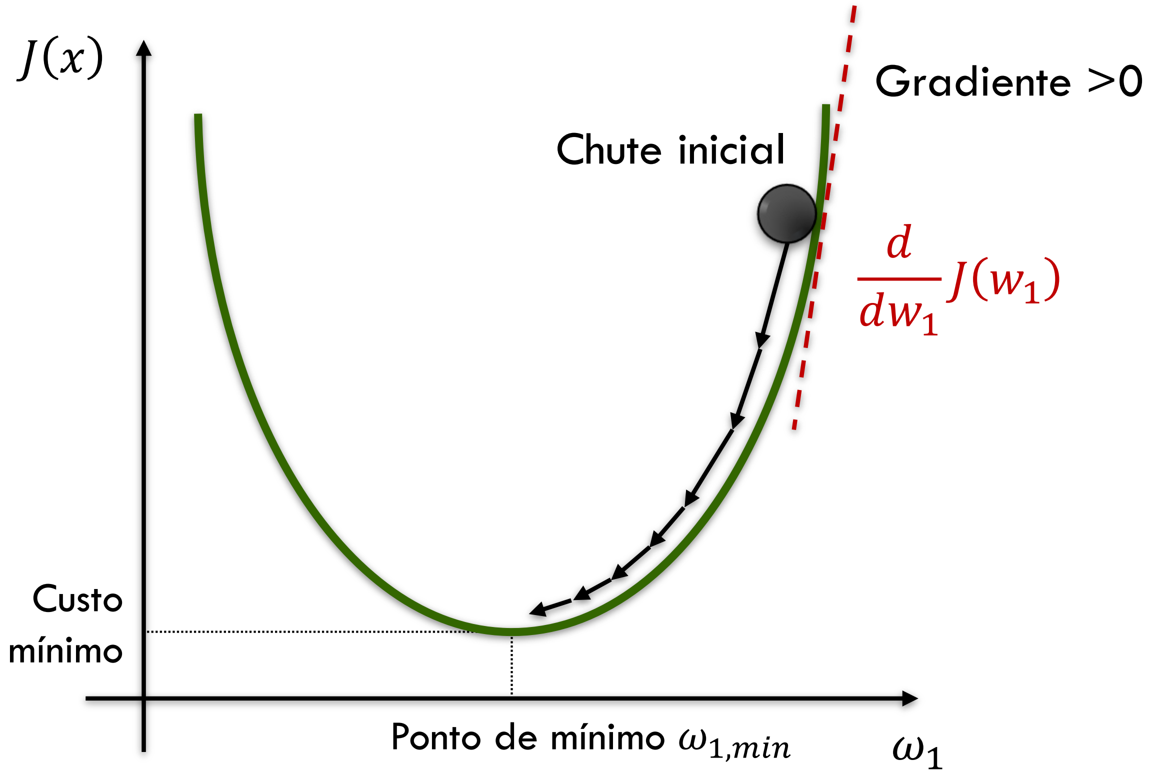
https://towardsdatascience.com/coding-deep-learning-for-beginners-linear-regression-gradient-descent-fcd5e0fc077d



https://www.offconvex.org/2018/11/07/optimization-beyond-landscape/



Supondo $w_0 = 0$ para ilustrar de forma clara o algoritmo



Enquanto $\left\|\nabla J\left(\mathbf{w}^{(0)}\right)\right\| > \varepsilon$:

$$w_k^{(i+1)} := w_k^{(i)} - \alpha \nabla J\left(\mathbf{w}^{(i)}\right), \qquad k = 0,1$$

$$i \mathrel{+}= 1$$

ERRO MÍNIMO QUADRADO

Suponha que recebamos as matrizes $\boldsymbol{A} \in \mathbb{R}^{m \times n}$ e um vetor $\boldsymbol{b} \in \mathbb{R}^{m}$, tal que $\boldsymbol{b} \notin \mathcal{R}(\boldsymbol{A})$. Nesta situação, não conseguiremos encontrar um vetor $\boldsymbol{x} \in \mathbb{R}^{n}$, tal que $\boldsymbol{Ax} = \boldsymbol{b}$, então, em vez disso, queremos encontrar um vetor $\boldsymbol{x}$ tal que $\boldsymbol{Ax}$ seja o mais próximo possível $\boldsymbol{b}$, medido pelo quadrado da norma euclidiana,

$$||\boldsymbol{Ax} - \boldsymbol{b}||^2$$

FUNÇÕES QUADRÁTICAS E LINEARES

$$\nabla_x \boldsymbol{b}^T \boldsymbol{x} = \boldsymbol{b}$$

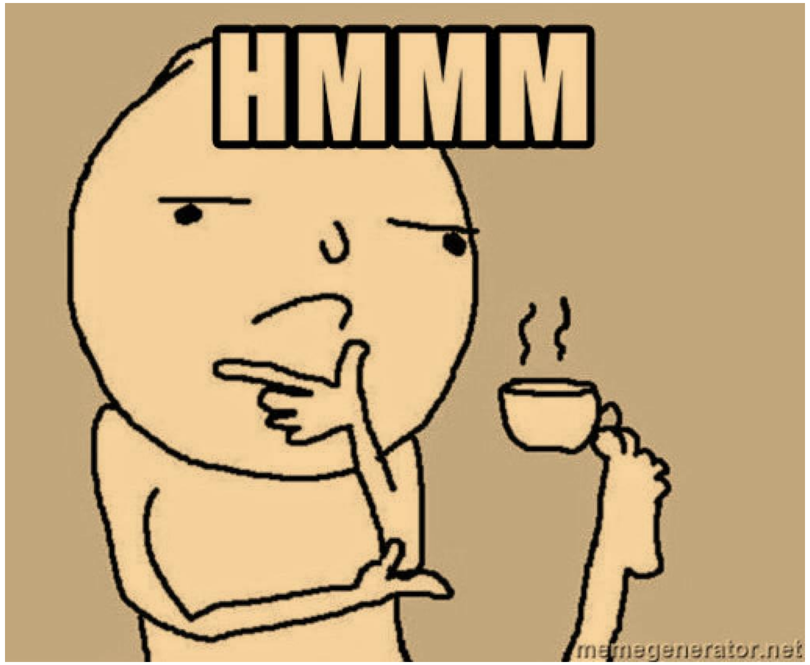
$$\nabla_x \boldsymbol{x}^T \boldsymbol{A}\boldsymbol{x} = \boldsymbol{2}\boldsymbol{A}\boldsymbol{x} \text{ (Se } \boldsymbol{A} \text{ é simétrica)}$$

$$\nabla_x^2 \boldsymbol{x}^T \boldsymbol{A}\boldsymbol{x} = \boldsymbol{2}\boldsymbol{A} \text{ (Se } \boldsymbol{A} \text{ é simétrica)}$$

$$\nabla_x \boldsymbol{2}\boldsymbol{b}^T \boldsymbol{A}\boldsymbol{x}$$

$$
\begin{aligned}
|\boldsymbol{Ax}-\boldsymbol{b}|^2 &= (\boldsymbol{Ax}-\boldsymbol{b})^T(\boldsymbol{Ax}-\boldsymbol{b}) \\
&= \boldsymbol{x}^T\boldsymbol{A}^T\boldsymbol{Ax} - \boldsymbol{2b}^T\boldsymbol{Ax} + \boldsymbol{b}^T\boldsymbol{b}
\end{aligned}
$$

$$
\begin{aligned}
\nabla_x(\boldsymbol{x}^T\boldsymbol{A}^T\boldsymbol{Ax} - \boldsymbol{2b}^T\boldsymbol{Ax} + \boldsymbol{b}^T\boldsymbol{b}) &= \nabla_x\boldsymbol{x}^T\boldsymbol{A}^T\boldsymbol{Ax} - \nabla_x\boldsymbol{2b}^T\boldsymbol{Ax} + \nabla_x\boldsymbol{b}^T\boldsymbol{b} \\
&= \nabla_x\boldsymbol{x}^T\boldsymbol{A}^T\boldsymbol{Ax} - \nabla_x\boldsymbol{2b}^T\boldsymbol{Ax} + \nabla_x\boldsymbol{b}^T\boldsymbol{b} \\
&= 2\boldsymbol{A}^T\boldsymbol{Ax} - \boldsymbol{2A}^T\boldsymbol{b}
\end{aligned}
$$



$$
\begin{aligned}
2\boldsymbol{A}^T\boldsymbol{Ax} - \boldsymbol{2A}^T\boldsymbol{b} &= \boldsymbol{0} \\
\boldsymbol{x} &= (\boldsymbol{A}^T\boldsymbol{A})^{-1}\boldsymbol{A}^T\boldsymbol{b}
\end{aligned}
$$

EQUAÇÃO NORMAL

A equação normal é uma solução de forma fechada para descobrir o valor dos parâmetros $\boldsymbol{w}$ que minimizam uma determinada função de custo. É chamado de *solução de forma fechada* no sentido de fornecer o resultado diretamente através da equação.

$$\boldsymbol{x} = (\boldsymbol{A}^T\boldsymbol{A})^{-1}\boldsymbol{A}^T\boldsymbol{b}$$

POR EXEMPLO...

Considerando a tabela abaixo, encontre os parâmetros ótimos para a previsão do preço de um imóvel. Primeiramente, considere apenas um parâmetro – dimensão – depois, aumente e inclua os demais. Compare os resultados.

$$y = w_0 x_0 + w_1 x_1, \qquad x_0 = 1$$

$x = (A^T A)^{-1} A^T b$???

Casa	Dimensão (em m^2)	#quartos	#andares	Idade (em anos)	Preço (em milhares de reais)
01	195	5	1	45	460
02	131	3	2	40	232
03	142	3	2	30	315
04	80	2	1	36	178

PRIMEIRA LIÇÃO

Ache o vetor $\boldsymbol{w}$ para a *regressão linear* do conjunto de dados da tabela ao lado. Na equação,

$$y = w_0x_0 + w_1x_1 + w_2x_2$$

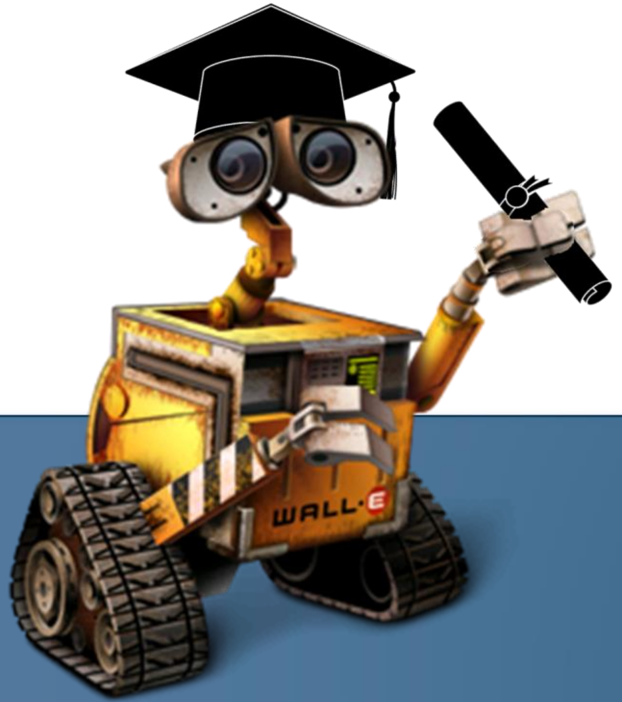
y é a taxa de gordura, $\boldsymbol{w}$ são os pesos e $\boldsymbol{x}$ os dados.

Índice	Peso (Kg)	Idade (anos)	Taxa de gordura no sangue
1	84	46	354
2	73	20	190
3	65	52	405
4	70	30	263
5	76	57	451
6	69	25	302
7	63	28	288
8	72	36	385
9	79	57	402
10	75	44	365
11	27	24	209
12	89	31	290
13	65	52	346
14	57	23	254
15	59	60	395
16	69	48	434
17	60	34	220
18	79	51	374
19	75	50	308
20	82	34	220
21	59	46	311
22	67	23	181
23	85	37	274
24	55	40	303
25	63	30	244

Fonte: https://people.sc.fsu.edu/~jburkardt/datasets/regression/x09.txt

PECE Programa de Educação Continuada

PECE Programa de Educação Continuada



SVD

Em ML, alguns dos conceitos mais importantes de álgebra linear são a decomposição em valores singulares (SVD) e a análise de componentes principais (PCA).

EXEMPLO

Com todos os dados brutos coletados, como podemos descobrir estruturas? Por exemplo, com as taxas de juros Selic dos últimos 20 anos, podemos entender sua composição para identificar *tendências*?

Mês/Ano	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013	2014	2015	2016	2017	2018	2019	2020
Janeiro (%)	1,46	1,27	1,53	1,97	1,27	1,38	1,43	1,08	0,93	1,05	0,66	0,86	0,89	0,60	0,85	0,94	1,06	1,09	0,58	0,54	0,38
Fevereiro (%)	1,45	1,02	1,25	1,83	1,08	1,22	1,15	0,87	0,80	0,86	0,59	0,84	0,75	0,49	0,79	0,82	1,00	0,87	0,47	0,49	
Março (%)	1,45	1,26	1,37	1,78	1,38	1,53	1,42	1,05	0,84	0,97	0,76	0,92	0,82	0,55	0,77	1,04	1,16	1,05	0,53	0,47	
Abri (%)	1,30	1,19	1,48	1,87	1,18	1,41	1,08	0,94	0,90	0,84	0,67	0,84	0,71	0,61	0,82	0,95	1,06	0,79	0,52	0,52	
Maio (%)	1,49	1,34	1,41	1,97	1,23	1,50	1,28	1,03	0,88	0,77	0,75	0,99	0,74	0,60	0,87	0,99	1,11	0,93	0,52	0,54	
Junho (%)	1,39	1,27	1,33	1,86	1,23	1,59	1,18	0,91	0,96	0,76	0,79	0,96	0,64	0,61	0,82	1,07	1,16	0,81	0,52	0,47	
Julho (%)	1,31	1,50	1,54	2,08	1,29	1,51	1,17	0,97	1,07	0,79	0,86	0,97	0,68	0,72	0,95	1,18	1,11	0,80	0,54	0,57	
Agosto (%)	1,41	1,60	1,44	1,77	1,29	1,66	1,26	0,99	1,02	0,69	0,89	1,07	0,69	0,71	0,87	1,11	1,22	0,80	0,57	0,50	
Setembro (%)	1,22	1,32	1,38	1,68	1,25	1,50	1,06	0,80	1,10	0,69	0,85	0,94	0,54	0,71	0,91	1,11	1,11	0,64	0,47	0,46	
Outubro (%)	1,29	1,53	1,65	1,64	1,21	1,41	1,09	0,93	1,18	0,69	0,81	0,88	0,61	0,81	0,95	1,11	1,05	0,64	0,54	0,48	
Novembro (%)	1,22	1,39	1,54	1,34	1,25	1,38	1,02	0,84	1,02	0,66	0,81	0,86	0,55	0,72	0,84	1,06	1,04	0,57	0,49	0,38	
Dezembro (%)	1,20	1,39	1,74	1,37	1,48	1,47	0,99	0,84	1,12	0,73	0,93	0,91	0,55	0,79	0,96	1,16	1,12	0,54	0,49	0,37	

VETORES E VALORES SINGULARES

Dada uma matriz $\boldsymbol{A} \in \mathbb{R}^{m \times n}$, as matrizes à esquerda e à direita de $\boldsymbol{A}$ são, respectivamente,

$$\underset{m \times m}{\boldsymbol{A}\boldsymbol{A}^T} \text{ e } \underset{n \times n}{\boldsymbol{A}^T\boldsymbol{A}}.$$

$\boldsymbol{A}\boldsymbol{A}^T$ e $\boldsymbol{A}^T\boldsymbol{A}$ são simétricas, reais e quadradas. Portanto, autovalores $\boldsymbol{A}\boldsymbol{A}^T$ ou $\boldsymbol{A}^T\boldsymbol{A}$ são reais não negativos, iguais ou distintos. Ambas as matrizes têm o mesmo rank r de $\boldsymbol{A}$.

VALORES SINGULARES DE UMA MATRIZ

Chamamos de **valores singulares** da matriz $\boldsymbol{A}$, $\lambda_1, \lambda_2, \ldots, \lambda_r$, as r raízes quadradas dos autovalores não nulos $\boldsymbol{A}\boldsymbol{A}^T$ ou $\boldsymbol{A}^T\boldsymbol{A}$;

$$\sigma_i = \sqrt{\lambda_i}$$

Os ***vetores singulares*** $\boldsymbol{u}_i$ à esquerda da matriz $\boldsymbol{A}$ são os autovetores de $\boldsymbol{A}\boldsymbol{A}^T, m \times m$.

Os ***vetores singulares*** $\boldsymbol{v}_i$ à direita da matriz $\boldsymbol{A}$ são os autovetores de $\boldsymbol{A}^T\boldsymbol{A}, n \times n$.

MATRIZES $\boldsymbol{U}$ E $\boldsymbol{V}$

Os autovetores de $\boldsymbol{A}\boldsymbol{A}^T$ formam as colunas da matriz $\boldsymbol{U}$ e os autovetores de $\boldsymbol{A}^T\boldsymbol{A}$ formam as colunas da matriz $\boldsymbol{V}$. Como são simétricas, os autovetores podem ser definidos ortonormais, ié, perpendiculares entre si e de comprimento unitário, de forma que:

$$\boldsymbol{U}^T\boldsymbol{U} = \boldsymbol{I} \qquad \boldsymbol{V}^T\boldsymbol{V} = \boldsymbol{I}$$

POR EXEMPLO...

https://www.d.umn.edu/~mhampton/m4326svd_example.pdf

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

$$\boldsymbol{A}^T\boldsymbol{A} = \begin{pmatrix} 13 & 12 & 2 \\ 12 & 13 & -2 \\ 2 & -2 & 8 \end{pmatrix}$$

$$|\boldsymbol{A}^T\boldsymbol{A} - \lambda\boldsymbol{I}| = 0$$

$$\lambda^3 - 34\lambda^2 + 225\lambda = \lambda(9-\lambda)(25-\lambda) = 0$$

$$\lambda_1 = 25, \qquad \lambda_2 = 9, \qquad \lambda_3 = 0$$

$$\sigma_1 = \sqrt{\lambda_1} = \sqrt{25} = 5$$

$$\sigma_2 = \sqrt{\lambda_2} = \sqrt{9} = 3$$

$$\sigma_3 = \sqrt{\lambda_3} = 0$$

$$\boldsymbol{A}\boldsymbol{A}^T = \begin{pmatrix} 17 & 8 \\ 8 & 17 \end{pmatrix}$$

$$|\boldsymbol{A}\boldsymbol{A}^T - \lambda\boldsymbol{I}| = 0$$

$$\lambda^2 - 34\lambda + 225 = (25-\lambda)(9-\lambda)$$

$$\lambda_1 = 25, \qquad \lambda_2 = 9$$

AUTOVETORES...

$$(\boldsymbol{A}\boldsymbol{A}^T - \lambda_i \boldsymbol{I})\boldsymbol{u}_i = \boldsymbol{0}$$

Por exemplo, para $\lambda = 25$,

$$\boldsymbol{A}\boldsymbol{A}^T - 25\boldsymbol{I} = \begin{pmatrix} 17-25 & 8 \\ 8 & 17-25 \end{pmatrix} = \begin{pmatrix} -8 & 8 \\ 8 & -8 \end{pmatrix}$$

$$\begin{pmatrix} -8 & 8 \\ 8 & -8 \end{pmatrix}\begin{pmatrix} u_{11} \\ u_{12} \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix} \Rightarrow \boldsymbol{u}_1 = \begin{pmatrix} u_{11} \\ u_{12} \end{pmatrix} = \begin{pmatrix} 1/\sqrt{2} \\ 1/\sqrt{2} \end{pmatrix}$$

Analogamente: $\boldsymbol{u}_2 = \begin{pmatrix} 1/\sqrt{2} \\ -1/\sqrt{2} \end{pmatrix}$

$$(\boldsymbol{A}^T\boldsymbol{A} - \lambda_i \boldsymbol{I})\boldsymbol{v}_i = \boldsymbol{0}$$

Por exemplo, para $\lambda = 25$,

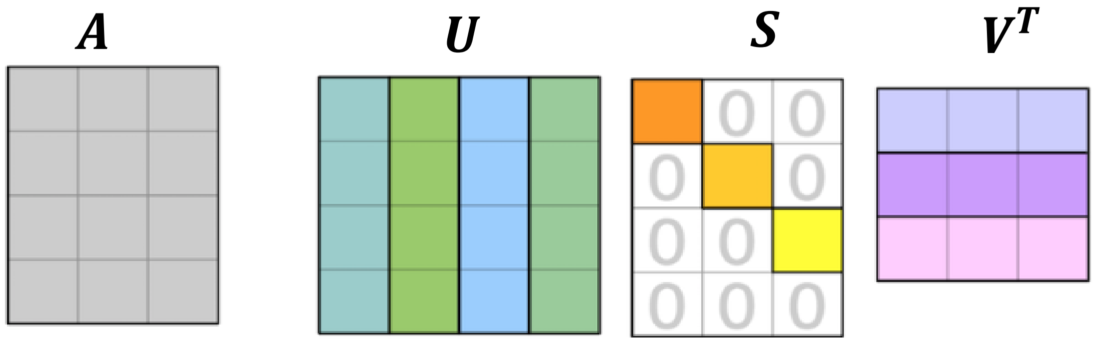
$$\boldsymbol{A}^T\boldsymbol{A} - 25\boldsymbol{I} = \begin{pmatrix} 13-25 & 12 & 2 \\ 12 & 13-25 & -2 \\ 2 & -2 & 8-25 \end{pmatrix} = \begin{pmatrix} -12 & 12 & 2 \\ 12 & -12 & -2 \\ 2 & -2 & -17 \end{pmatrix}$$

$$\begin{pmatrix} -12 & 12 & 2 \\ 12 & -12 & -2 \\ 2 & -2 & -17 \end{pmatrix}\begin{pmatrix} v_{11} \\ v_{12} \\ v_{13} \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix} \Rightarrow \boldsymbol{v}_1 = \begin{pmatrix} v_{11} \\ v_{12} \\ v_{13} \end{pmatrix} = \begin{pmatrix} 1/\sqrt{2} \\ 1/\sqrt{2} \\ 0 \end{pmatrix}$$

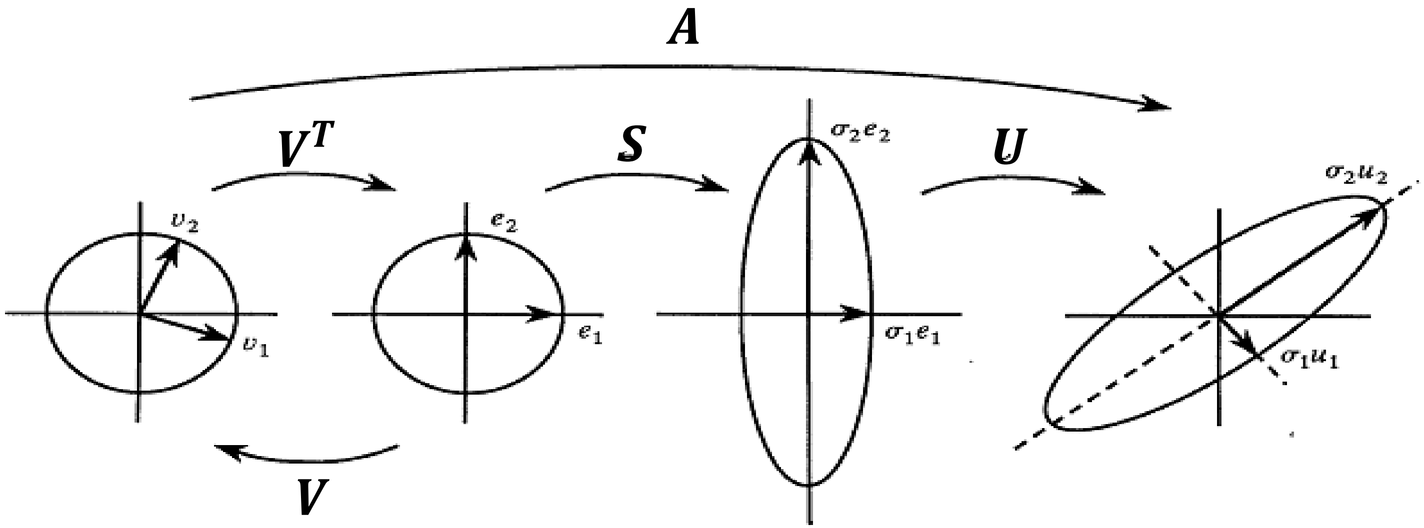
Analogamente: $\boldsymbol{v}_2 = \begin{pmatrix} 1/\sqrt{18} \\ -1/\sqrt{18} \\ 4/\sqrt{18} \end{pmatrix}, \boldsymbol{v}_3 = \begin{pmatrix} 2/3 \\ -2/3 \\ -1/3 \end{pmatrix}$

INTUIÇÃO DE SVD

$$A_{m\times n} = U_{m\times m} S_{m\times n} V^T_{n\times n}$$



$A_{m\times n}$ é uma matriz de valores reais
$U_{m\times m}$ é uma matriz ortogonal
$S_{m\times n}$ é uma matriz diagonal
$V^T_{n\times n}$ é uma matriz ortogonal



DECOMPOSIÇÃO EM VALORES SINGULARES

O **SVD (Singular Value Decomposition)** é um método para a fatorar matrizes:

$$\boldsymbol{A} = \boldsymbol{U}\boldsymbol{S}\boldsymbol{V}^T = \boldsymbol{U}\begin{bmatrix} \boldsymbol{D}_{rxr} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{0} \end{bmatrix}_{m \times n} \boldsymbol{V}^T$$

onde $\boldsymbol{U}$ e $\boldsymbol{V}$ são as matrizes ortogonais que definimos e $\boldsymbol{S}$ é diagonal.

Os valores da matriz diagonal $\boldsymbol{S}$ são os valores singulares que acabamos de aprender, e por isso a decomposição recebe este nome.

O número de valores singulares diferentes de zero é igual ao rank da matriz $\boldsymbol{A}$.

MATRIZ DIAGONAL S

Se $\boldsymbol{A}$ tem r valores singulares não nulos, $\mathbf{S}$ é uma matriz da forma:

$$\mathbf{S} = \begin{pmatrix} \sigma_1 & 0 & 0 & 0 & 0 & 0 \\ 0 & \sigma_2 & 0 & 0 & 0 & 0 \\ 0 & 0 & \ddots & 0 & 0 & 0 \\ 0 & 0 & 0 & \sigma_r & 0 & 0 \\ 0 & 0 & 0 & 0 & \ddots & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix}$$

Os elementos diagonais da matriz $\boldsymbol{S}$, ié, os valores singulares σ_i, em SVD são computados em ordem decrescente $\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_r$

MATRIZES ORTOGONAIS $\boldsymbol{U}$ E $\boldsymbol{V}$

$$\boldsymbol{V} = \begin{pmatrix} | & | & & | \\ \boldsymbol{v}_1 & \boldsymbol{v}_2 & \dots & \boldsymbol{v}_n \\ | & | & & | \end{pmatrix}_{n \times n} \qquad \boldsymbol{U} = \begin{pmatrix} | & | & & | \\ \boldsymbol{u}_1 & \boldsymbol{u}_2 & \dots & \boldsymbol{u}_m \\ | & | & & | \end{pmatrix}_{m \times m}$$

onde $\boldsymbol{v}_i$ são os autovetores de $\boldsymbol{A}^T\boldsymbol{A}$ e $\boldsymbol{u}_i$ são os autovetores de $\boldsymbol{A}\boldsymbol{A}^T$

$$\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_m$$

$$\begin{pmatrix} | & | & & | \\ \boldsymbol{u}_1 & \boldsymbol{u}_2 & \dots & \boldsymbol{u}_m \\ | & | & & | \end{pmatrix}$$

VOLTEMOS AO NOSSO EXEMPLO...

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

A matriz $\boldsymbol{A}$ é 2×3, portanto, $\boldsymbol{V}$ tem dimensão 2×3, $\boldsymbol{U}$ tem dimensão 2×2 e $\boldsymbol{S}$ tem dimensão 2×3

$$\boldsymbol{V}^T = \begin{pmatrix} 1/\sqrt{2} & 1/\sqrt{2} & 0 \\ 1/\sqrt{18} & -1/\sqrt{18} & 4/\sqrt{18} \\ 2/3 & -2/3 & -1/3 \end{pmatrix}$$

$$\boldsymbol{U} = \begin{pmatrix} 1/\sqrt{2} & 1/\sqrt{2} \\ 1/\sqrt{2} & -1/\sqrt{2} \end{pmatrix}$$

$$\boldsymbol{S} = \begin{pmatrix} 5 & 0 & 0 \\ 0 & 3 & 0 \end{pmatrix}$$

INSIGHT $\boldsymbol{AV} = \boldsymbol{US}$

$$\begin{array}{cccc} \boldsymbol{A}_{m\times n} & \boldsymbol{V}_{n\times r} & \boldsymbol{U}_{m\times r} & \boldsymbol{S}_{r\times r} \end{array}$$

$$\begin{pmatrix} \boldsymbol{a}_{11} & \boldsymbol{a}_{12} & \cdots & \boldsymbol{a}_{1n} \\ \vdots & \vdots & \ddots & \vdots \\ \boldsymbol{a}_{m1} & \boldsymbol{a}_{m2} & \cdots & \boldsymbol{a}_{mn} \end{pmatrix} \begin{pmatrix} | & | & & | \\ \boldsymbol{v}_1 & \boldsymbol{v}_2 & \dots & \boldsymbol{v}_r \\ | & | & & | \end{pmatrix} = \begin{pmatrix} | & | & & | \\ \boldsymbol{u}_1 & \boldsymbol{u}_2 & \dots & \boldsymbol{u}_r \\ | & | & & | \end{pmatrix} \begin{pmatrix} \sigma_1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \sigma_r \end{pmatrix}$$

$$A \begin{pmatrix} | \\ \boldsymbol{v}_1 \\ | \end{pmatrix} = \begin{pmatrix} | & | & & | \\ \boldsymbol{u}_1 & \boldsymbol{u}_2 & \dots & \boldsymbol{u}_r \\ | & | & & | \end{pmatrix} \begin{pmatrix} \sigma_1 \\ \vdots \\ 0 \end{pmatrix} = \begin{pmatrix} | \\ \boldsymbol{u}_1 \\ | \end{pmatrix} \sigma_1$$

$$\boldsymbol{Av}_1 = \sigma_1 \boldsymbol{u}_1$$

$$\boxed{\boldsymbol{Av}_i = \sigma_i \boldsymbol{u}_i}$$

$$\begin{pmatrix} | & & | \\ \boldsymbol{u}_1 & \cdots & \boldsymbol{u}_r \\ | & & | \end{pmatrix} \begin{pmatrix} \sigma_1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \sigma_r \end{pmatrix} = \begin{pmatrix} | & & | \\ \sigma_1 \boldsymbol{u}_1 & \cdots & \sigma_r \boldsymbol{u}_r \\ | & & | \end{pmatrix}$$

$$\begin{pmatrix} | & & | \\ \boldsymbol{u}_1 & \cdots & \boldsymbol{u}_r \\ | & & | \end{pmatrix} \begin{pmatrix} - & \boldsymbol{v}_1 & - \\ & \vdots & \\ - & \boldsymbol{v}_r & - \end{pmatrix} = \begin{pmatrix} | \\ \boldsymbol{u}_1 \\ | \end{pmatrix} \begin{pmatrix} - & \boldsymbol{v}_1 & - \end{pmatrix} + \cdots + \begin{pmatrix} | \\ \boldsymbol{u}_r \\ | \end{pmatrix} \begin{pmatrix} - & \boldsymbol{v}_r & - \end{pmatrix}$$

$$\boldsymbol{A} = \sigma_1 \boldsymbol{u}_1 \boldsymbol{v}_1^T + \cdots + \sigma_n \boldsymbol{u}_n \boldsymbol{v}_n^T$$

VOLTEMOS AO NOSSO EXEMPLO...

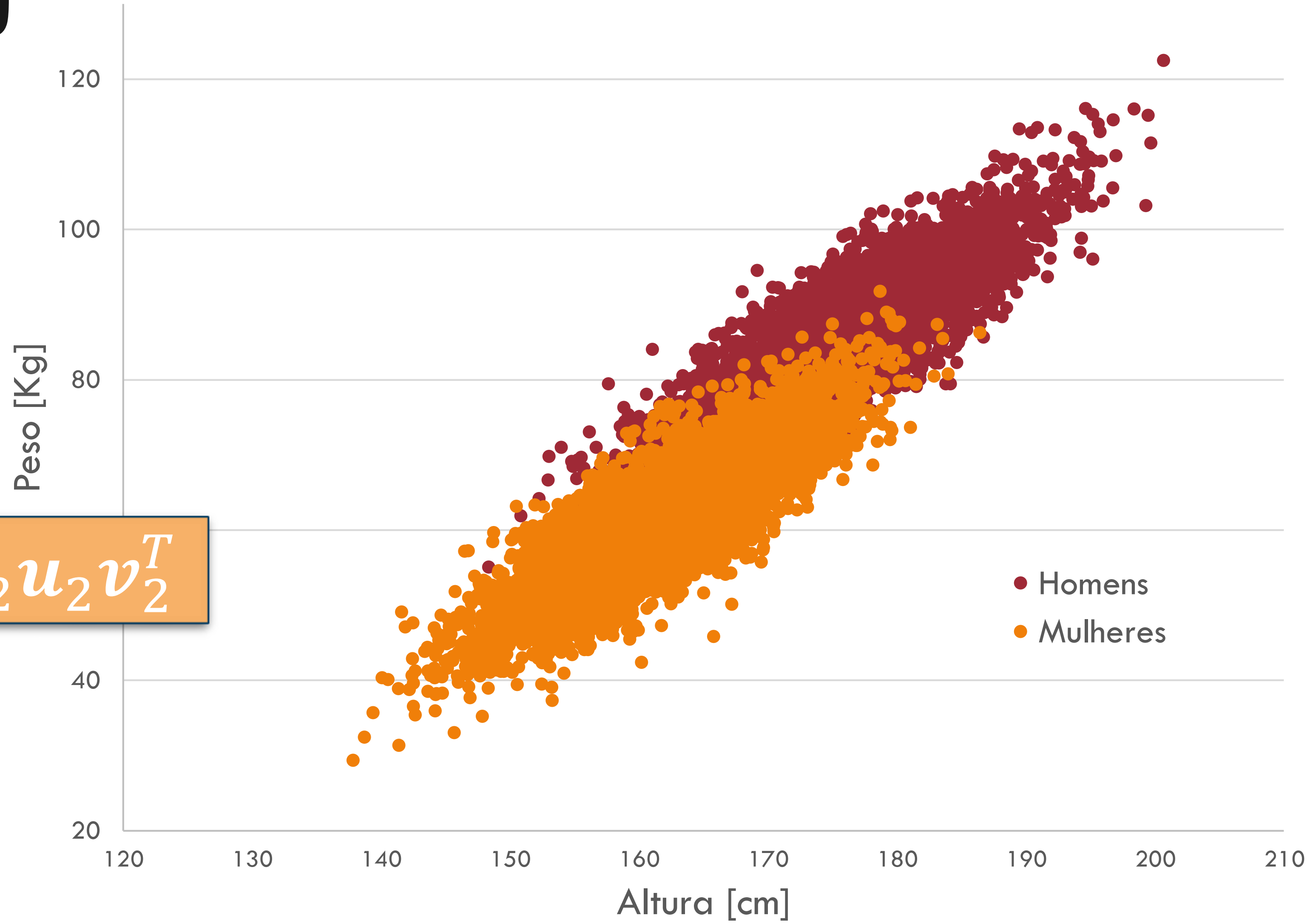
$$\boldsymbol{A} = \begin{pmatrix} 3 & 2 & 2 \\ 2 & 3 & -2 \end{pmatrix}$$

$$\boldsymbol{V^T} = \begin{pmatrix} 1/\sqrt{2} & 1/\sqrt{2} & 0 \\ 1/\sqrt{18} & -1/\sqrt{18} & 4/\sqrt{18} \\ 2/3 & -2/3 & -1/3 \end{pmatrix} \qquad \boldsymbol{U} = \begin{pmatrix} 1/\sqrt{2} & 1/\sqrt{2} \\ 1/\sqrt{2} & -1/\sqrt{2} \end{pmatrix} \qquad \boldsymbol{S} = \begin{pmatrix} 5 & 0 & 0 \\ 0 & 3 & 0 \end{pmatrix}$$

$$\boldsymbol{A} = 5\begin{pmatrix} 1/\sqrt{2} \\ 1/\sqrt{2} \end{pmatrix}\begin{pmatrix} 1/\sqrt{2} & 1/\sqrt{2} & 0 \end{pmatrix} + 3\begin{pmatrix} 1/\sqrt{2} \\ 1/\sqrt{2} \end{pmatrix}\begin{pmatrix} 1/\sqrt{18} & -1/\sqrt{18} & 4/\sqrt{18} \end{pmatrix}$$

$$\boxed{\boldsymbol{A} = \sigma_1 \boldsymbol{u}_1 \boldsymbol{v}_1^T + \cdots + \sigma_r \boldsymbol{u}_r \boldsymbol{v}_r^T}$$

SEGUNDA LIÇÃO



$$\boldsymbol{A} = \sigma_1 \boldsymbol{u}_1 \boldsymbol{v}_1^T + \cdots + \sigma_2 \boldsymbol{u}_2 \boldsymbol{v}_2^T$$

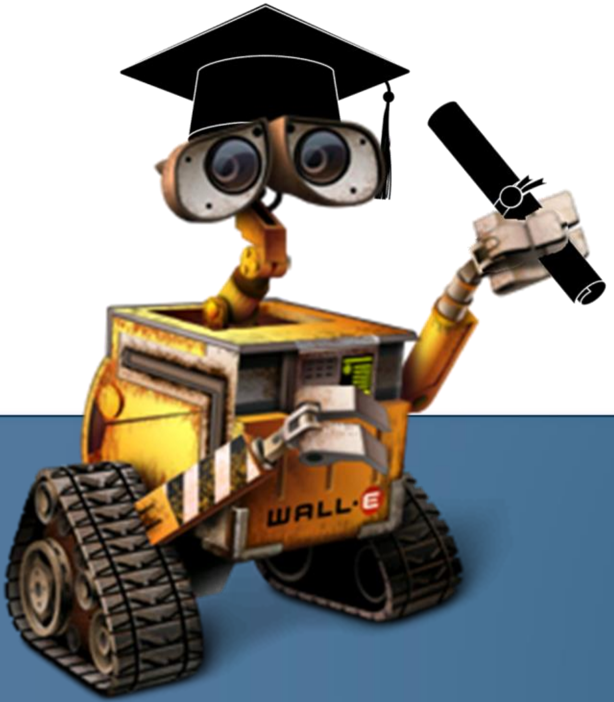
ENTREGA NO MOODLE

1. Leiam o texto do link abaixo:

https://blogs.sas.com/content/iml/2017/08/30/svd-low-rank-approximation.html

2. Com os dados de altura e peso de homens e mulheres, que você pode retirar do link https://www.kaggle.com/mustafaali96/weight-height/version/1,

- Ache a matriz $\boldsymbol{A}$
- Ache os valores singulares
- Ache $\boldsymbol{V}, \boldsymbol{S}, \boldsymbol{U}$
- Escreva a matriz na forma, $\boldsymbol{A} = \sigma_1 \boldsymbol{u}_1 \boldsymbol{v}_1^T + \cdots + \sigma_2 \boldsymbol{u}_2 \boldsymbol{v}_2^T$
- Inclua gênero como dado de entrada.
- Discuta suas conclusões.



INVERSA E DADOS REAIS

$$\boldsymbol{Ax} = \boldsymbol{b}$$
$$\boldsymbol{x} = \boldsymbol{A}^{-1}\boldsymbol{b}$$

Mas nem todas as matrizes são invertíveis. Além disso, em ML, é improvável encontrar uma solução exata, com a certa presença de ruído nos dados. Nosso objetivo é encontrar o modelo que melhor se ajusta aos dados.

PSEUDO INVERSA MOORE-PENROSE

Pseudo inversa de A:

$$\boldsymbol{A}^{+}$$

de forma que

$$x \approx \boldsymbol{A}^{+}\boldsymbol{b}$$

Restrição de A $\rightarrow$ $\|\boldsymbol{A}\boldsymbol{A}^{+} - \boldsymbol{I}_n\|^2$

$$\boldsymbol{A}\boldsymbol{x} = \boldsymbol{b}$$
$$(\boldsymbol{U}, \boldsymbol{S}, \boldsymbol{V}) \leftarrow svd(\boldsymbol{A})$$
$$\boldsymbol{A}^{+} = \mathbf{V}\boldsymbol{S}^{+}\boldsymbol{U}^{T}$$
$$\boldsymbol{x} = \boldsymbol{A}^{+}\boldsymbol{b}$$

onde $\boldsymbol{S}^{+}$ é o recíproco $\frac{1}{x_i}$ dos elementos não nulos de $\boldsymbol{S}$

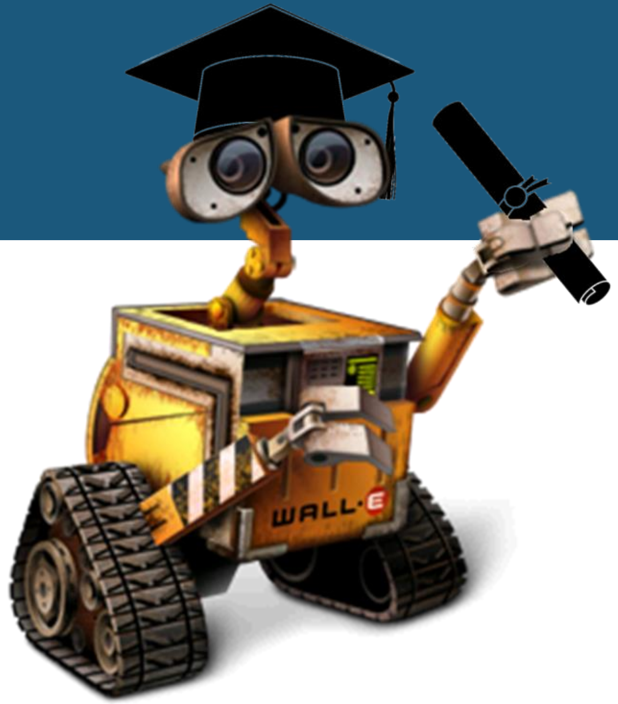
$$\begin{bmatrix} -3 & 0 & 0 \\ 0 & 4 & 0 \\ 0 & 0 & 0 \end{bmatrix}^{+} = \begin{bmatrix} -1/3 & 0 & 0 \\ 0 & -1/4 & 0 \\ 0 & 0 & 0 \end{bmatrix}$$

EXEMPLO

$$\boldsymbol{A} = \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix}$$ (singular, inversa não existe)

$$\boldsymbol{A} = \boldsymbol{U}\boldsymbol{S}\boldsymbol{V}^T = \frac{1}{\sqrt{2}}\begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix}\begin{bmatrix} 2 & 0 \\ 0 & 0 \end{bmatrix}\frac{1}{\sqrt{2}}\begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix}$$

$$\boldsymbol{A}^+ = \boldsymbol{V}\boldsymbol{S}^+\boldsymbol{U}^T = \frac{1}{\sqrt{2}}\begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix}\begin{bmatrix} 1/2 & 0 \\ 0 & 0 \end{bmatrix}\frac{1}{\sqrt{2}}\begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix} = \frac{1}{4}\begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix}$$



ACABOU...