

Revisão de Álgebra Linear e Cálculo

Afshine AMIDI e Shervine AMIDI

13 de Outubro de 2018

Traduzido por Gabriel Fonseca. Revisado por Leticia Portella.

Notações gerais

□ **Vetor** – Indicamos por $x \in \mathbb{R}^n$ um vetor com n elementos, onde $x_i \in \mathbb{R}$ é o $i^{\text{ésimo}}$ elemento:

$$x = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix} \in \mathbb{R}^n$$

□ **Matriz** – Indicamos por $A \in \mathbb{R}^{m \times n}$ uma matriz com m linhas e n colunas, onde $A_{i,j} \in \mathbb{R}$ é o elemento localizado na $i^{\text{ésima}}$ linha e $j^{\text{ésima}}$ coluna:

$$A = \begin{pmatrix} A_{1,1} & \cdots & A_{1,n} \\ \vdots & & \vdots \\ A_{m,1} & \cdots & A_{m,n} \end{pmatrix} \in \mathbb{R}^{m \times n}$$

Observação: o vetor x definido acima pode ser visto como uma matriz $n \times 1$ e é mais particularmente chamado de vetor coluna.

□ **Matriz identidade** – A matriz identidade $I \in \mathbb{R}^{n \times n}$ é uma matriz quadrada com uns na sua diagonal e zeros nas demais posições:

$$I = \begin{pmatrix} 1 & 0 & \cdots & 0 \\ 0 & \ddots & & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & 1 \end{pmatrix}$$

Observação: para todas as matrizes $A \in \mathbb{R}^{n \times n}$, nós temos $A \times I = I \times A = A$.

□ **Matriz diagonal** – Uma matriz diagonal $D \in \mathbb{R}^{n \times n}$ é uma matriz quadrada com valores não nulos na sua diagonal e zeros nas demais posições:

$$D = \begin{pmatrix} d_1 & 0 & \cdots & 0 \\ 0 & \ddots & & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & d_n \end{pmatrix}$$

Observação: nós também indicamos D como $\text{diag}(d_1, \dots, d_n)$.

Operações de matriz

□ **Vetor-vetor** – Há dois tipos de produtos vetoriais:

- Produto interno: para $x, y \in \mathbb{R}^n$, temos:

$$x^T y = \sum_{i=1}^n x_i y_i \in \mathbb{R}$$

- Produto tensorial: para $x \in \mathbb{R}^m, y \in \mathbb{R}^n$, temos:

$$xy^T = \begin{pmatrix} x_1 y_1 & \cdots & x_1 y_n \\ \vdots & & \vdots \\ x_m y_1 & \cdots & x_m y_n \end{pmatrix} \in \mathbb{R}^{m \times n}$$

□ **Matriz-vetor** – O produto de uma matriz $A \in \mathbb{R}^{m \times n}$ e um vetor $x \in \mathbb{R}^n$ é um vetor de tamanho \mathbb{R}^m , de tal modo que:

$$Ax = \begin{pmatrix} a_{r,1}^T x \\ \vdots \\ a_{r,m}^T x \end{pmatrix} = \sum_{i=1}^n a_{c,i} x_i \in \mathbb{R}^m$$

onde $a_{r,i}^T$ são vetores linhas e $a_{c,j}$ vetores colunas de A , e x_i são os elementos de x .

□ **Matriz-matriz** – O produto das matrizes $A \in \mathbb{R}^{m \times n}$ e $B \in \mathbb{R}^{n \times p}$ é uma matriz de tamanho $\mathbb{R}^{m \times p}$, de tal modo que:

$$AB = \begin{pmatrix} a_{r,1}^T b_{c,1} & \cdots & a_{r,1}^T b_{c,p} \\ \vdots & & \vdots \\ a_{r,m}^T b_{c,1} & \cdots & a_{r,m}^T b_{c,p} \end{pmatrix} = \sum_{i=1}^n a_{c,i} b_{r,i}^T \in \mathbb{R}^{m \times p}$$

onde $a_{r,i}^T, b_{r,i}^T$ são vetores linhas e $a_{c,j}, b_{c,j}$ vetores colunas de A e B respectivamente.

□ **Transposta** – A transposta de uma matriz $A \in \mathbb{R}^{m \times n}$, indicada por A^T , é tal que suas linhas são trocadas por suas colunas:

$$\forall i, j, \quad A_{i,j}^T = A_{j,i}$$

Observação: para matrizes A, B , temos $(AB)^T = B^T A^T$.

□ **Inversa** – A inversa de uma matriz quadrada inversível A é indicada por A^{-1} e é uma matriz única de tal modo que:

$$AA^{-1} = A^{-1}A = I$$

Observação: nem todas as matrizes quadrada são inversíveis. Também, para matrizes A, B , temos $(AB)^{-1} = B^{-1}A^{-1}$.

□ **Traço** – O traço de uma matriz quadrada A , indicado por $\text{tr}(A)$, é a soma dos elementos de sua diagonal:

$$\text{tr}(A) = \sum_{i=1}^n A_{i,i}$$

Observação: para matrizes A, B , temos $\text{tr}(A^T) = \text{tr}(A)$ e $\text{tr}(AB) = \text{tr}(BA)$.

□ **Determinante** – A determinante de uma matriz quadrada $A \in \mathbb{R}^{n \times n}$, indicada por $|A|$ ou $\det(A)$ é expressa recursivamente em termos de $A_{\setminus i, \setminus j}$, a qual é a matriz A sem a sua $i^{\text{ésima}}$ linha e $j^{\text{ésima}}$ coluna, como se segue:

$$\det(A) = |A| = \sum_{j=1}^n (-1)^{i+j} A_{i,j} |A_{\setminus i, \setminus j}|$$

Observação: A é inversível se e somente se $|A| \neq 0$. Além disso, $|AB| = |A||B|$ e $|A^T| = |A|$.

Propriedades da matriz

□ **Decomposição simétrica** – Uma dada matriz A pode ser expressa em termos de suas partes simétricas e assimétricas como a seguir:

$$A = \underbrace{\frac{A + A^T}{2}}_{\text{Simétrica}} + \underbrace{\frac{A - A^T}{2}}_{\text{Assimétrica}}$$

□ **Norma** – Uma norma é uma função $N : V \rightarrow [0, +\infty[$ onde V é um vetor espaço, e de tal modo que para todo $x, y \in V$, nós temos:

- $N(x + y) \leq N(x) + N(y)$
- $N(ax) = |a|N(x)$ para a escalar
- se $N(x) = 0$, então $x = 0$

Para $x \in V$, as mais comumente utilizadas normas estão resumidas na tabela abaixo:

Norma	Notação	Definição	Caso de uso
Manhattan, L^1	$\ x\ _1$	$\sum_{i=1}^n x_i $	LASSO
Euclidean, L^2	$\ x\ _2$	$\sqrt{\sum_{i=1}^n x_i^2}$	Ridge
p -norme, L^p	$\ x\ _p$	$\left(\sum_{i=1}^n x_i^p\right)^{\frac{1}{p}}$	Inégalité de Hölder
Infini, L^∞	$\ x\ _\infty$	$\max_i x_i $	Convergence uniforme

□ **Dependência linear** – Um conjunto de vetores é dito ser linearmente dependente se um dos vetores no conjunto puder ser definido como uma combinação linear dos demais.

Observação: se nenhum vetor puder ser escrito dessa maneira, então os vetores são ditos serem linearmente independentes.

□ **Posto da matriz** – O posto de uma dada matriz A é indicada por $\text{rank}(A)$ e é a dimensão do vetor espaço gerado por suas colunas. Isso é equivalente ao número máximo de colunas linearmente independentes de A .

□ **Matriz positiva semi-definida** – Uma matriz $A \in \mathbb{R}^{n \times n}$ é positiva semi-definida (PSD) e é indicada por $A \succeq 0$ se tivermos:

$$A = A^T \quad \text{e} \quad \forall x \in \mathbb{R}^n, \quad x^T A x \geq 0$$

Observação: de forma similar, uma matriz A é dita ser positiva definida, e é indicada por $A \succ 0$ se ela é uma matriz (PSD) que satisfaz todo vetor x não nulo, $x^T A x > 0$.

□ **Autovalor, autovetor** – Dada uma matriz $A \in \mathbb{R}^{n \times n}$, λ é dita ser um autovalor de A se existe um vetor $z \in \mathbb{R}^n \setminus \{0\}$, chamado autovetor, nós temos:

$$Az = \lambda z$$

□ **Teorema spectral** – Seja $A \in \mathbb{R}^{n \times n}$. Se A é simétrica, então A é diagonalizável por uma matriz ortogonal $U \in \mathbb{R}^{n \times n}$. Indicando $\Lambda = \text{diag}(\lambda_1, \dots, \lambda_n)$, nós temos:

$$\exists \Lambda \text{ diagonal}, \quad A = U \Lambda U^T$$

□ **Decomposição em valor singular** – Para uma dada matriz A de dimensões $m \times n$, a decomposição em valor singular (SVD) é uma técnica de fatorização que garante a existência de matrizes unitária $U \in \mathbb{R}^{m \times m}$, diagonal $\Sigma \in \mathbb{R}^{m \times n}$ e unitária $V \in \mathbb{R}^{n \times n}$, de tal modo que:

$$A = U \Sigma V^T$$

Cálculo com matriz

□ **Gradiente** – Seja $f : \mathbb{R}^{m \times n} \rightarrow \mathbb{R}$ uma função e $A \in \mathbb{R}^{m \times n}$ uma matriz. O gradiente de f a respeito a A é a matriz $m \times n$, indicada por $\nabla_A f(A)$, de tal modo que:

$$\left(\nabla_A f(A) \right)_{i,j} = \frac{\partial f(A)}{\partial A_{i,j}}$$

Observação: o gradiente de f é somente definido quando f é uma função que retorna um escalar.

□ **Hessiano** – Seja $f : \mathbb{R}^n \rightarrow \mathbb{R}$ uma função e $x \in \mathbb{R}^n$ um vetor. O hessiano de f a respeito a x uma matriz simétrica $n \times n$, indicada por $\nabla_x^2 f(x)$, de tal modo que:

$$\left(\nabla_x^2 f(x) \right)_{i,j} = \frac{\partial^2 f(x)}{\partial x_i \partial x_j}$$

Observação: o hessiano de f é somente definido quando f é uma função que retorna um escalar.

□ **Operações com gradiente** – Para matrizes A, B, C , as seguintes propriedades de gradiente valem a pena ter em mente:

$$\nabla_A \text{tr}(AB) = B^T \quad \nabla_A^T f(A) = (\nabla_A f(A))^T$$

$$\nabla_A \text{tr}(ABA^T C) = CAB + C^T A B^T \quad \nabla_A |A| = |A| (A^{-1})^T$$