Revisão de Álgebra Linear e Cálculo

Afshine Amidi e Shervine Amidi

13 de Outubro de 2018

Notações gerais

❒ **Vetor** – Indicamos por $x \in \mathbb{R}^n$ um vetor com n elementos, onde $x_i \in \mathbb{R}$ é o $i^{ésimo}$ 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 elementos localizado na $i^{ésima}$ linha e $j^{ésima}$ 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 defindo 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:

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

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:

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

Observação: nós também indicamos D como diag(d_1,...,d_n).

Operações de matriz

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

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

$$\boxed{x^Ty = \sum_{i=1}^n x_iy_i \in \mathbb{R}}$$

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

$$\boxed{xy^T = \begin{pmatrix} x_1y_1 & \cdots & x_1y_n \\ \vdots & & \vdots \\ x_my_1 & \cdots & x_my_n \end{pmatrix} \in \mathbb{R}^{m\times n}}$$

❒ **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 $\mathbb{R}^n$, de tal modo que:

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

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 $\mathbb{R}^{n\times p}$, de tal modo que:

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

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:

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

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

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

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

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 tr(A), é a soma dos elementos de sua diagonal:

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

Observação: para matrizes A,B, temos tr(A^T) = tr(A) e tr(AB) = 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 $A_{\backslash i,\backslash j}$, a qual é a matriz A sem a sua $i^{ésima}$ linha e $j^{ésima}$ coluna, como se segue:

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

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:

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

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

- $N(x+y)\leqslant N(x)+N(y)$
- $N(ax)=|a|N(x)$ 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	$\vert\vert x\vert\vert_1$	$\displaystyle\sum_{i=1}^n\vert x_i\vert$	LASSO
Euclidean, L^2	$\vert\vert x\vert\vert_2$	$\displaystyle\sqrt{\sum_{i=1}^n x_i^2}$	Ridge
p-norme, L^p	$\vert\vert x\vert\vert_p$	$\displaystyle\left(\sum_{i=1}^n x_i^p\right)^{\frac{1}{p}}$	Inégalité de Hölder
Infini, L^∞	$\vert\vert x\vert\vert_\infty$	$\underset{i}{\max}\ \vert x_i\vert$	Convergence uniforme

❒ **Dependência linear** – Um conjunto de vetores é dito ser linearmente dependete 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 $\textrm{rank}(A)$ 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:

$$\boxed{A=A^T}\quad\textrm{e}\quad\boxed{\forall x\in\mathbb{R}^n,\quad x^TAx\geqslant 0}$$

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, $x^TAx>0$.

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

$$\boxed{Az=\lambda z}$$

❒ **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 $\Lambda=\textrm{diag}(\lambda_1,...,\lambda_n)$, nós temos:

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

❒ **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 $m\times m$, diagonal Σ $m\times n$ e unitária V $n\times n$, de tal modo que:

$$\boxed{A=U\Sigma V^T}$$

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 $\nabla_Af(A)$, de tal modo que:

$$\boxed{\Big(\nabla_Af(A)\Big)_{i,j}=\frac{\partial f(A)}{\partial A_{i,j}}}$$

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 $\nabla_x^2f(x)$, de tal modo que:

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

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

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

$$\boxed{\nabla_A\textrm{tr}(AB)=B^T}\quad\boxed{\nabla_{A^T}f(A)=\left(\nabla_Af(A)\right)^T}$$

$$\boxed{\nabla_A\textrm{tr}(ABA^TC)=CAB+C^TAB^T}\quad\boxed{\nabla_A|A|=|A|(A^{-1})^T}$$