Fazendo contas com matrizes e vetores

Esse é um capítulo sobre contas e algumas definições básicas que serão essenciais durante o curso. Embora pareça maçante (e muitas vezes seja mesmo), precisamos nos certificar de que entendemos as definições e operações básicas antes de prosseguirmos. Não subestime esse capítulo: saber fazer uma conta de várias maneiras diferentes ou mesmo escrever de várias maneiras diferentes muitas vezes é o segredo para conseguir demonstrar um teorema ou desenhar um algoritmo mais eficiente.

Neste capítulo, consideraremos $\overrightarrow{v}$ e $\overrightarrow{w}$ vetores com entradas

$$\overrightarrow{v} = \begin{bmatrix} v_1 \\ v_2 \\ \vdots \\ v_n \end{bmatrix} \in \mathbb{R}^n \text{ e } \overrightarrow{w} = \begin{bmatrix} w_1 \\ w_2 \\ \vdots \\ w_n \end{bmatrix} \in \mathbb{R}^n,$$

além de A e B matrizes com $m \times n$ e $n \times p$ entradas[4], ou seja,

[4] Ou seja, m linhas e n colunas

$$A = (a_{ij})_{m \times n} = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \ddots & \ddots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{bmatrix}$$

e

$$B = (b_{ij})_{n \times p} = \begin{bmatrix} b_{11} & b_{12} & \cdots & b_{1p} \\ b_{21} & b_{22} & \cdots & b_{2p} \\ \vdots & \ddots & \ddots & \vdots \\ b_{n1} & b_{n2} & \cdots & b_{np} \end{bmatrix}.$$

Somando e multiplicando

Podemos somar vetores e também multiplicar por escalares realizando as operações coordenada a coordenada. Se $\alpha \in \mathbb{R}$, então

$$\alpha \overrightarrow{v} + \overrightarrow{w} = \begin{bmatrix} \alpha v_1 + w_1 \\ \alpha v_2 + w_2 \\ \vdots \\ \alpha v_n + w_n \end{bmatrix}.$$

Também podemos multiplicar uma matriz por um escalar $\alpha \in \mathbb{R}$ realizando a multiplicação coordenada a coordenada, ou seja

$$\alpha A = (\alpha a_{ij})_{m\times n} = \begin{bmatrix} \alpha a_{11} & \alpha a_{12} & \cdots & \alpha a_{1n} \\ \alpha a_{21} & \alpha a_{22} & \cdots & \alpha a_{2n} \\ \vdots & \ddots & \ddots & \vdots \\ \alpha a_{m1} & \alpha a_{m2} & \cdots & \alpha a_{mn} \end{bmatrix}.$$

Como A tem n colunas e $\overrightarrow{v}$ tem n coordenadas podemos multiplicar $A\overrightarrow{v}$ e obter um vetor de m coordenadas da forma

$$A\overrightarrow{v} = \begin{bmatrix} a_{11}v_1 + a_{12}v_2 + \cdots + a_{1n}v_n \\ a_{21}v_1 + a_{22}v_2 + \cdots + a_{2n}v_n \\ \vdots \\ a_{m1}v_1 + a_{m2}v_2 + \cdots + a_{mn}v_n \end{bmatrix} \in \mathbb{R}^m.$$

A multiplicação de duas matrizes A e B, denotada por AB, é uma operação definida quando o número de colunas de A é igual ao número de linhas de B. Se A é uma matriz $m \times n$ e B é uma matriz $n \times p$, então o produto AB resultará em uma matriz $m \times p$. A entrada (i, j) da matriz resultante AB é obtida multiplicando cada elemento da i-ésima linha de A pelos elementos correspondentes da j-ésima coluna de B e somando esses produtos. Matematicamente, a entrada (i, j) de $C = AB$ é dada por [5]:

$$c_{ij} = a_{i1}b_{1j} + a_{i2}b_{2j} + \ldots + a_{in}b_{nj} = \sum_{k=1}^{n} a_{ik}b_{kj}.$$

[5] Talvez seja preciso gastar um tempinho entendendo o somatório abaixo (somatório é o nome desse Σ grande com índice k indo de $k = 1$ até $k = n$).

É interessante notar que o produto de $A\overrightarrow{v}$ pode ser visto como um produto de matrizes. Basta interpretar $\overrightarrow{v}$ como uma matriz de n linhas e 1 coluna.

Por fim, note que **o produto de matrizes não é comutativo!**[6] Ou seja, pode ocorrer que $AB \neq BA$ mesmo quando $n = m = p$ e o produto está bem definido. Por outro lado, **o produto de matriz é associativo**, ou seja, se A, B e C são matrizes e a multiplicação ABC faz sentido, então $ABC = (AB)C = A(BC)$.

[6] Tome, por exemplo, $A = \begin{bmatrix} 1 & 1 \\ 0 & 0 \end{bmatrix}$ e $B = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}$

A transposta e a inversa de uma matriz

A transposta de uma matriz A, denotada por A^T, é obtida trocando suas linhas por colunas e vice-versa. Se A é uma matriz $m \times n$, então a matriz transposta A^T terá dimensões $n \times m$. Logo:

$$A^T = (a_{ji})_{n\times m} = \begin{bmatrix} a_{11} & a_{21} & \cdots & a_{m1} \\ a_{12} & a_{22} & \cdots & a_{m2} \\ \vdots & \vdots & \ddots & \vdots \\ a_{1n} & a_{2n} & \cdots & a_{mn} \end{bmatrix}.$$

Note que enquanto o produto por A mapeia vetores de $\mathbb{R}^n$ para $\mathbb{R}^m$, o produto por A^T faz o contrário: mapeia vetores de $\mathbb{R}^m$ para $\mathbb{R}^n$.

Como o produto entre uma matriz A e um vetor $\overrightarrow{v} \in \mathbb{R}^n$ mapeia $\overrightarrow{v}$ em um vetor $A\overrightarrow{v} \in \mathbb{R}$, é natural pensar na matriz que realiza a operação inversa. Dizemos que uma matriz C ($n \times m$):

- é inversa pela direita de A se $AC = I_m$;
- é inversa pela esquerda de A se $CA = I_n$;
- é inversa de A se é inversa pela esquerda e pela direita.

Acima usamos que I_n é a matriz identidade

$$I_n = \begin{bmatrix} 1 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 1 \end{bmatrix}_{n\times n}.$$

Nem todas as matrizes tem inversa, nem inversa pela direita ou pela esquerda. Quando A tem inversa denotamos a inversa de A por

$$A^{-1}.$$

De fato, saber se uma matriz tem ou não inversa é um problema muito próximo do problema de resolver sistemas lineares[7]. Portanto, é o propósito de boa parte do nosso curso.

[7] Por que?

Algumas propriedades

Seja B inversa pela esquerda de A e C inversa pela direita de A, então

$$B = BI_m = B(AC) = (BA)C = I_nC = C.$$

Logo, se A tem inversa pela direita e pela esquerda as inversas são iguais. Ou seja, a inversa A^{-1} de uma matriz invertível é única.

Além disso, como veremos no futuro, apenas matrizes quadradas (com $n = m$) podem possuir inversa. Por enquanto, vamos fingir que não sabemos disso.

As operações que discutimos satisfazem as seguintes propriedades:

- $(AB)^T = B^T A^T$;
- Quando A e B possuem inversa, então AB é invertível e $(AB)^{-1} = B^{-1}A^{-1}$;
- Quando a inversa de A existe, então a inversa da inversa de A é A, em outras palavras, $(A^{-1})^{-1} = A$;
- Se $\alpha \neq 0$ e A tem inversa, então $(\alpha A)^{-1} = \frac{1}{\alpha} A^{-1}$;
- Se A é invertível, inversa e transposta comutam, ou seja, $(A^T)^{-1} = (A^{-1})^T$;

Vá em frente, tente provar essas propriedades. Aqui vamos provar apenas a última propriedade. Basta mostrar que

$$A^T(A^{-1})^T = I_n \text{ e } (A^{-1})^T A^T = I_m.$$

Podemos usar a primeira propriedade da lista, mostrando que

$$A^T(A^{-1})^T = (A^{-1}A)^T = I_n^T = I_n$$

e

$$(A^{-1})^T A^T = (AA^{-1})^T = I_m^T = I_m.$$

A norma Euclideana e o produto escalar

Uma propriedade muito importante na definição física de vetor é o seu módulo ou tamanho, que aqui conheceremos como norma Euclideana. A norma Euclideana de um vetor $\overrightarrow{v}$ é denotada por $\|\overrightarrow{v}\|_2$ e é definida como[8]

$$\|\overrightarrow{v}\|_2 = \sqrt{v_1^2 + v_2^2 + \cdots + v_n^2}.$$

[8] Note que a definição se parece muito com o Teorema de Pitágoras. Isso não é uma coincidência.

Definimos também o produto interno entre dois vetores $\overrightarrow{v}, \overrightarrow{w} \in \mathbb{R}^n$ como

$$\langle \overrightarrow{v}, \overrightarrow{w} \rangle = v_1 w_1 + v_2 w_2 + \cdots + v_n w_n = \sum_{i=1}^{n} v_i w_i.$$

O produto interno tem uma interpretação natural em termos da lei dos cossenos. Quaisquer dois vetores $\overrightarrow{v} \in \mathbb{R}^n$ e $\overrightarrow{v} \in \mathbb{R}^n$ quando posicionados numa mesma origem definem um plano (e isso independe de n!). Nesse plano esses vetores definem um ângulo θ.

Podemos formar um triângulo com lados de tamanhos $\|\vec{v}\|_2$, $\|\vec{w}\|_2$ e $\|\vec{w}-\vec{v}\|_2$. Além disso, o ângulo θ entre $\vec{v}$ e $\vec{w}$ aparece no triângulo.

A lei dos cossenos diz que um triângulo de vértices PQR satisfaz

$$|QR|^2 = |PQ|^2 + |PR|^2 - 2|PQ||PR| \cos\left(\widehat{RPQ}\right).$$

Aplicando a lei dos cossenos no triangulo da Figura 8 temos

$$\|\vec{w}-\vec{v}\|_2^2 = \|\vec{v}\|_2^2 + \|\vec{w}\|_2^2 - 2\|\vec{v}\|_2\|\vec{w}\|_2 \cos\theta,$$

além disso, vale que (faça as contas)

$$\|\vec{w}-\vec{v}\|_2^2 = \|\vec{v}\|_2^2 + \|\vec{w}\|_2^2 - 2\langle\vec{v},\vec{w}\rangle.$$

Logo,

$$\langle\vec{v},\vec{w}\rangle = \|\vec{v}\|_2\|\vec{w}\|_2 \cos\theta. \tag{2}$$

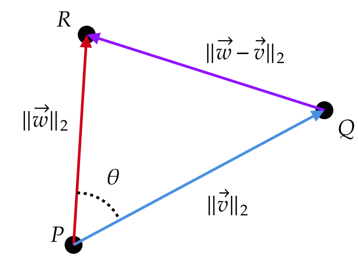

Figura 8: Dados vetores $\vec{v}$ e $\vec{v}$ podemos formar um triângulo relacionando as normas Euclideanas e o ângulo θ entre os vetores.

A identidade (2) fornece uma interpretação geométrica muito útil para aplicações. Assuma que $\|\vec{v}\|_2$ e $\|\vec{v}\|_2$ são fixas, mas que podemos escolher $\vec{v}$ e $\vec{w}$ de maneira a variar o ângulo θ: quando o produto escalar é alto então $\vec{v}$ e $\vec{w}$ estão próximos; quando o produto escalar é próximo de zero então $\vec{v}$ e $\vec{w}$ são quase perpendiculares; quando o produto escalar é negativo e com valor absoluto alto então $\vec{v}$ e $\vec{w}$ apontam em sentidos opostos. Ou seja, **o produto escalar mede a similaridade entre dois vetores**.[9]

[9] Uma curiosidade: a regressão linear, um dos métodos clássicos de aprendizado de máquina, usa o produto escalar como medida de similaridade.

O produto escalar e a norma Euclideana estão relacionados por

$$\|\vec{v}\|_2 = \sqrt{\langle\vec{v},\vec{v}\rangle}$$

e

$$\langle\vec{v},\vec{w}\rangle = \frac{\|\vec{v}+\vec{w}\|_2^2 - \|\vec{v}\|_2^2 - \|\vec{w}\|_2^2}{2}.$$

Além disso, valem as seguintes propriedades (prove):

- O produto escalar é simétrico, ou seja, $\langle\vec{v},\vec{w}\rangle = \langle\vec{w},\vec{v}\rangle$;
- Além disso, para $\vec{v},\vec{w},\vec{u} \in \mathbb{R}^n$ e $\alpha \in \mathbb{R}$,

$$\langle\alpha\vec{v}+\vec{w},\vec{u}\rangle = \alpha\langle\vec{v},\vec{u}\rangle + \langle\vec{w},\vec{u}\rangle.$$

Vale observar também que as propriedades vistas aqui valem em maior generalidade com as definições certas. Veremos no futuro que a norma Euclideana é apenas um exemplo de norma e que o produto escalar entre vetores é apenas um exemplo de produto interno. Além disso, as propriedades vistas vão continuar valendo em maior generalidade.

O traço de uma matriz

Definimos também mais uma operação relevante, o traço, que será definido apenas para matrizes quadradas. Definimos

$$\mathrm{tr}(A) = \sum_{i=1}^{n} a_{ii}.$$

Além disso, o traço satisfaz as seguintes propriedades:

- O traço é linear. Seja α um escalar e A e B matrizes $n \times n$, então:

$$\mathrm{tr}(\alpha A + B) = \alpha\, \mathrm{tr}(A) + \mathrm{tr}(B).$$

- O traço do produto não depende da ordem:

$$\mathrm{tr}(AB) = \mathrm{tr}(BA).$$

Exercícios

Exercício 1. *Mostre que se existe um vetor* $\overrightarrow{v} \neq \overrightarrow{0}$ *tal que* $A\overrightarrow{v} = \overrightarrow{0}$ *então* A *não possui inversa. Nesse caso, é possível que* A *tenha inversa pela direita? E pela esquerda?*

Exercício 2. *Sejam* $\overrightarrow{v}, \overrightarrow{w} \in \mathbb{R}^n$ *vetores e* V, W *as matrizes* $nx1$ *correspondentes aos vetores. Mostre que*

(i) $\langle \overrightarrow{v}, \overrightarrow{w} \rangle = V^T W = \mathrm{tr}(VW^T)$.

(ii) Se $A = (a_{ij})_{n \times n}$, *então*

$$\langle \overrightarrow{v}, A\overrightarrow{w} \rangle = V^T A W = \sum_{i,j \in [n]} a_{ij} v_i w_j = W^T A^T V = \langle A^T \overrightarrow{v}, \overrightarrow{w} \rangle.$$

Exercício 3. *Uma das mais conhecidas arquiteturas para redes neurais é o multilayer perceptron, cuja arquitetura baseia-se em compor multiplicações por matrizes com ativações não-lineares. Tome* A *uma matriz* $m \times n$ *e* B *uma matriz* $p \times m$. *Defina* $\sigma : \mathbb{R}^m \mapsto \mathbb{R}^m$ *como sendo a aplicação* $x \mapsto \max(0, x)$ *coordenada a coordenada, ou seja,* σ *zera as entradas negativas e preserva as entradas positivas. Chamamos* σ *de função de ativação*[10]. *Uma rede* $\varphi : \mathbb{R}^n \mapsto \mathbb{R}^p$ *com duas camadas lineares é dada pelo mapa*

$$\varphi\left(\overrightarrow{v}\right) = B\sigma\left(A\overrightarrow{v}\right),$$

onde as matrizes A *e* B *normalmente são aprendidas treinando a rede neural. Escolha* m, p, A *e* B *de maneira que* φ *seja a função identidade em* $\mathbb{R}^n$. *Justifique sua escolha.*

[10] Essa escolha de σ em específico é conhecida como ReLU.