第二章 向量和矩阵基础

第 2 讲 向量与矩阵的概念和运算

黄定江

DaSE @ ECNU

djhuang@dase.ecnu.edu.cn

2.1.1 用向量表示数据引例：文本的词袋模型表示

例 1

下面是纽约时报网络版在 *2010* 年 *12* 月 *7* 日的四则新闻提要：

(a) Suit Over Targeted Killing in Terror Case Is Dismissed. A federal judge on Tuesday dismissed a lawsuit that sought to block the United States from attempting to kill an American citizen, Anwar Al-Awlaki, who has been accused of aiding Al Qaeda.
(b) In Tax Deal With G.O.P, a Portent for the Next 2 Years. President Obama made clear that he was willing to alienate his liberal base in the interest of compromise. Tax Deal suggests new path for Obama. President Obama agreed to a tentative deal to extend the Bush tax cuts, part of a package to keep jobless aid and cut pay roll taxes.

2.1.1 用向量表示数据引例：文本的词袋模型表示

(c) Obama Urges China to Check North Koreans. In a frank discussion, President Obama urged China's president to put the North Korean government on a tighter leash after a series of provocations.

(d) Top Test Scores From Shanghai Stun Educators. With China's debut in international standardized testing Shanghai students have surprised experts by outscoring counterparts in dozens of other countries.

2.1.1 用向量表示数据引例：文本的词袋模型表示

- 假设一本字典 $V = \{aid, kill, deal, president, tax, china\}$，我们想知道每则新闻提要中的单词在字典中出现的频率。
- 根据词袋模型，首先将新闻中的单词进行简化，如去除复数变为单数（例如将 (b) 中 Years 改为 Year）；现在时改为动词（例如将 (a) 中 Killing 改为 kill）。
- 非常容易可以看出，在新闻 (a) 中 aid 共出现了 1 次，kill 共出现了 2 次，而字典 V中的其它单词并没有出现，通过一个一元数组表示这一结果，即
$$\boldsymbol{a} = (1, 2, 0, 0, 0, 0).$$
- 将数组 $\boldsymbol{a}$ 归一化（$\boldsymbol{a}$ 中每个单词在 V中出现次数除以所有单词在 V中总共出现的次数），便可以得到这则新闻提要在字典 $\boldsymbol{V}$ 中出现的相对频率，
$$\boldsymbol{a}' = (\frac{1}{3}, \frac{2}{3}, 0, 0, 0, 0).$$

2.1.1 用向量表示数据引例：文本的词袋模型表示

- 类似的，将其它三则新闻提要也进行同样的处理，则可分别得三个数组

$$\begin{aligned}\boldsymbol{b}^{'} &= (\frac{1}{10}, 0, \frac{3}{10}, \frac{1}{5}, \frac{2}{5}, 0),\\ \boldsymbol{c}^{'} &= (0, 0, 0, \frac{1}{2}, 0, \frac{1}{2}),\\ \boldsymbol{d}^{'} &= (0, 0, 0, 0, 0, 1).\end{aligned}$$

这样我们就把上述每一个新闻提要按照词频表示成一个一元六维的数组，这些数组是由一些具有意义的数值构成的，我们可以给它抽象出来，赋予新的定义，也即向量。

2.1.2 向量的定义

在给出向量的形式化定义之前，我们先给数组中的数值定义一个所属的范围，以方便向量进行代数运算，也即数域。

定义 1

设 $\mathbb{K}$ 是由一些数组成的集合，如果 $\mathbb{K}$ 中包含 0 与 1，并且 $\mathbb{K}$ 中任意两个数的和，差，积，商（除数不为零）仍在 $\mathbb{K}$ 中，则称 $\mathbb{K}$ 是一个**数域**。

有理数集，实数集和复数集都是数域，他们分别称为有理数域 $\mathbb{Q}$，实数域 $\mathbb{R}$ 和复数域 $\mathbb{C}$.

2.1.2 向量的定义

定义 2

由数域 $\mathbb{K}$ 中的 n 个数组成的有序数组 $(a_1, a_2, \cdots, a_n)$ 称为 $\mathbb{K}$ 上的 n 维**向量**，记为 $\boldsymbol{a} = (a_1, a_2, \cdots, a_n)$，其中第 i 个数 a_i 称为 $\boldsymbol{a}$ 的第 i 个分量.

当 $\boldsymbol{a}$ 的每个分量是实数，也即 $a_i \in \mathbb{R}$ 时，则称 $\boldsymbol{a}$ 为**实向量**。所有 n 维实向量构成的集合记为$\mathbb{R}^n$。
当 $\boldsymbol{a}$ 的每个分量是复数，也即 $a_i \in \mathbb{C}$ 时，则称 $\boldsymbol{a}$ 为**复向量**。所有 n 维复向量构成的集合记为$\mathbb{C}^n$。

上述定义的是行向量。在实际使用中，我们通常使用列向量。我们可以使用**转置**把列向量转换成行向量，或者把行向量转换成列向量。

$$\begin{pmatrix} a_1 \\ a_2 \\ a_3 \\ \vdots \\ a_n \end{pmatrix} = \left(a_1, a_2, a_3, \ldots, a_n\right)^{\mathrm{T}}; \left(a_1, a_2, a_3, \ldots, a_n\right) = \begin{pmatrix} a_1 \\ a_2 \\ a_3 \\ \vdots \\ a_n \end{pmatrix}^{\mathrm{T}}$$

以后我们在未加注明的情况下提到的向量指的都是列向量。

2.1.2 向量的定义

根据上述定义，例1中的四个数组用向量表示即为：

$$
\begin{aligned}
\boldsymbol{a}^{'} &= (\frac{1}{3}, \frac{2}{3}, 0, 0, 0, 0)^{\mathrm{T}}. \\
\boldsymbol{b}^{'} &= (\frac{1}{10}, 0, \frac{3}{10}, \frac{1}{5}, \frac{2}{5}, 0)^{\mathrm{T}}, \\
\boldsymbol{c}^{'} &= (0, 0, 0, \frac{1}{2}, 0, \frac{1}{2})^{\mathrm{T}}, \\
\boldsymbol{d}^{'} &= (0, 0, 0, 0, 0, 1)^{\mathrm{T}}.
\end{aligned}
$$

2.1.2 向量的定义

在科学和工程中遇到的向量可以分为以下三种：

(1) 物理向量：泛指既有幅值，又有方向的物理量，如速度、加速度、位移等。

(2) 几何向量：为了将物理向量可视化，常用带方向的（简称“有向”）线段表示之。这种有向线段称为**几何向量**。例如，$\boldsymbol{v} = \overrightarrow{\boldsymbol{AB}}$ 表示的有向线段，其起点为 $\boldsymbol{A}$，终点为 $\boldsymbol{B}$。

(3) 代数向量：几何向量可以用代数形式表示。例如，若平面上的几何向量 $\boldsymbol{v} = \overrightarrow{\boldsymbol{AB}}$ 的起点坐标 $A = (a_1, a_2)$，终点坐标 $\boldsymbol{B} = (b_1, b_2)$，则该几何向量可以表示为代数形式 $\boldsymbol{v} = \begin{pmatrix} b_1 - a_1 \\ b_2 - a_2 \end{pmatrix}$。这种用代数形式表示的几何向量称为**代数向量**。

2.1.2 向量的定义

下图归纳了向量的分类：

$$\text{向量}\begin{cases}\text{物理向量}\\ \text{代数向量}\begin{cases}\text{常数向量}\\ \text{函数向量}\\ \text{随机向量}\end{cases}\\ \text{几何向量}\end{cases}$$

2.1.2 向量的定义

根据元素取值种类的不同，代数向量又可分为以下三种：

(1) 常数向量：向量的元素全部为实常数或者复常数，如 $\boldsymbol{a} = [1, 5, 4]^{\mathrm{T}}$ 等。

(2) 函数向量：向量的元素包含了函数，如 $\boldsymbol{x} = [1, x^2, \cdots, x^n]^{\mathrm{T}}$ 等。

(3) 随机向量：向量的元素为随机变量或随机过程，如 $\boldsymbol{x}(n) = [x_1(n), \cdots, x_m(n)]^{\mathrm{T}}$，其中 $x_1(n), \cdots, x_m(n)$ 是 m 个随机过程或随机信号。

我们将在概率基础部分会讨论随机向量。

2.1.3 用矩阵表示数据引例：文本向量集的表示

例 2

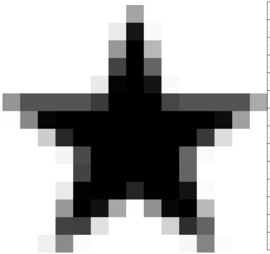
在例1中，每则新闻提要都由一个 6 维向量所表示，那么这四则新闻提要组成的新闻集可以由 4 个这样的 6 维向量组成的向量集所表示，换言之，这个新闻集可以按列组成一个 6×4 的二元数组，即

$$\boldsymbol{A} = \begin{pmatrix} \frac{1}{3} & \frac{1}{10} & 0 & 0 \\ \frac{2}{3} & 0 & 0 & 0 \\ 0 & \frac{3}{10} & 0 & 0 \\ 0 & \frac{1}{5} & \frac{1}{2} & 0 \\ 0 & \frac{2}{5} & 0 & 0 \\ 0 & 0 & \frac{1}{2} & 1 \end{pmatrix}$$

2.1.3 用矩阵表示数据引例：灰度图像的表示

例 3

在计算机中，如果只保留图像的灰度，那么我们可以使用一个二元数组存储该图像，其中数组中的元素是图像中相应像素的强度值（介于 *0* 至 *255* 之间）。下面左图是具有 *15* 个水平像素和 *14* 个垂直像素的五角星图像，右图是对应的二元数组。



255	255	255	255	255	255	255	164	255	255	255	255	255	255	255
255	255	255	255	255	255	234	22	250	255	255	255	255	255	255
255	255	255	255	255	255	151	0	184	255	255	255	255	255	255
255	255	255	255	255	255	66	0	99	255	255	255	255	255	255
255	255	255	255	255	233	4	0	20	250	255	255	255	255	255
149	84	84	84	84	59	0	0	0	70	84	84	84	84	170
253	136	6	0	0	0	0	0	0	0	0	0	13	160	255
255	255	205	40	0	0	0	0	0	0	0	57	221	255	255
255	255	255	245	74	0	0	0	0	0	101	251	255	255	255
255	255	255	255	67	0	0	0	0	0	100	255	255	255	255
255	255	255	234	3	0	0	37	0	0	19	251	255	255	255
255	255	255	152	0	17	167	254	145	8	0	185	255	255	255
255	255	255	67	63	225	255	255	255	210	46	100	255	255	255
255	255	240	139	252	255	255	255	255	255	247	132	252	255	255

图 1: 图像的表示

2.1.4 矩阵的定义

定义 3

由数域 $\mathbb{K}$ 中的 $m \times n$ 个数 $a_{ij}(i = 1, 2, \cdots, m; j = 1, 2, \cdots, n)$ 排成 m 行、n 列的表

$$\boldsymbol{A} = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \cdots & & & \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{pmatrix}, \tag{1}$$

称为 $\mathbb{K}$ 上的$m \times n$ **矩阵**, 记为 $\boldsymbol{A} = (a_{ij})_{m \times n}$ 或 $\boldsymbol{A}_{m \times n}$, 表中的每一个数都称为矩阵 $\boldsymbol{A}$ 的一个元素. 若 $m = n$, 则 $n \times n$ 矩阵 $\boldsymbol{A} = (a_{ij})_{m \times n}$ 也称为 **n 阶方阵**。

2.1.4 矩阵的定义

- 矩阵是一个二元数组，而向量是矩阵的一种特殊形式，当 $m=1$ 或者 $n=1$ 时，矩阵退化为向量，其中 $1\times n$ 的矩阵称之为行向量，而 $m\times 1$ 的矩阵称之为列向量。
- 当 $\boldsymbol{A}$ 的每个元素是实数，也即 $a_{ij}\in\mathbb{R}$ 时，则称 $\boldsymbol{A}$ 为**实矩阵**。所有 n 维实矩阵构成的集合记为$\mathbb{R}^{m\times n}$。
- 当 $\boldsymbol{A}$ 的每个元素是复数，也即 $a_{ij}\in\mathbb{C}$ 时，则称 $\boldsymbol{A}$ 为**复矩阵**。所有 n 维复矩阵构成的集合记为$\mathbb{C}^{m\times n}$。

2.1.4 矩阵的定义

- 在例2中 6×4 的二元数组被称为一个词项 -文档矩阵，这个矩阵的每一列表示一个新闻提要的词频向量，每一行表示每个词项在各个新闻提要中出现的权重。
- 在例3中 15×14 的二元数组被称为一个灰度图像矩阵，这个矩阵的每一行代表水平像素，每一列代表垂直像素。

2.1.4 数据的表示总结

- 一元数组：向量
- 二元数组：矩阵
- 三元数组及以上的高维数组：?
- 非数组型结构化目标：?

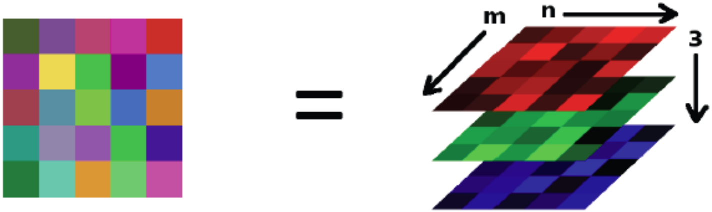


图 2: 彩色图像的张量表示

2.1.4 数据的表示总结

更好的表示

- 寻找原始特征向量的低维近似：主成分分析和奇异值分解
- 寻找原始特征向量的高维表示：利用特征映射构造原始特征向量的高维非线性组合

2.2.1 向量的线性运算

向量之间的基本代数运算有两种：加法和数乘，统称为向量的线性运算。

定义 4

设向量 $\boldsymbol{a}=(a_1,a_2,\cdots,a_n)$,$\boldsymbol{b}=(b_1,b_2,\cdots,b_n)$，则向量 $(a_1+b_1,a_2+b_2,\cdots,a_n+b_n)$ 称为向量 $\boldsymbol{a}$ 与向量 $\boldsymbol{b}$ 的和，记作 $\boldsymbol{a}+\boldsymbol{b}$，即 $\boldsymbol{a}+\boldsymbol{b}=(a_1+b_1,a_2+b_2,\cdots,a_n+b_n)$.

定义 5

设向量 $\boldsymbol{a}=(a_1,a_2,\cdots,a_n)$, k 是数域 $\mathbb{K}$ 中的数，则向量 $(ka_1,ka_2,\cdots,ka_n)$ 称为数 k 与向量 $\boldsymbol{a}$ 的乘积，简称**数乘**，记作 $k\boldsymbol{a}$，即 $k\boldsymbol{a}=(ka_1,ka_2,\cdots,ka_n)$.

定理 1

设 $\boldsymbol{a}, \boldsymbol{b}, \boldsymbol{c}$ 是 $\mathbb{K}$ 上的 n 维向量，λ, μ 是数域 $\mathbb{K}$ 中的数，则向量的加法和数乘运算满足下列运算规律：

1. $\boldsymbol{a}+\boldsymbol{b}=\boldsymbol{b}+\boldsymbol{a}$,
2. $(\boldsymbol{a}+\boldsymbol{b})+\boldsymbol{c}=\boldsymbol{a}+(\boldsymbol{b}+\boldsymbol{c})$,
3. $\boldsymbol{a}+\boldsymbol{0}=\boldsymbol{a}$,
4. $\boldsymbol{a}+(-\boldsymbol{a})=\boldsymbol{0}$,
5. $1\boldsymbol{a}=\boldsymbol{a}$,
6. $\lambda(\mu\boldsymbol{a})=(\lambda\mu)\boldsymbol{a}$,
7. $(\lambda+\mu)\boldsymbol{a}=\lambda\boldsymbol{a}+\mu\boldsymbol{a}$,
8. $\lambda(\boldsymbol{a}+\boldsymbol{b})=\lambda\boldsymbol{a}+\lambda\boldsymbol{b}$.

2.2.2 矩阵的加法

定义 6

设 $\boldsymbol{A}=(a_{ij})_{m\times n}, \boldsymbol{B}=(b_{ij})_{m\times n}$ 是两个 $m\times n$ 矩阵，则矩阵 $\boldsymbol{C}=(c_{ij})_{m\times n}=(a_{ij}+b_{ij})_{m\times n}$ 称为 $\boldsymbol{A}$ 与 $\boldsymbol{B}$ 的和，记作 $\boldsymbol{C}=\boldsymbol{A}+\boldsymbol{B}$。

$$\boldsymbol{A}+\boldsymbol{B}=\begin{pmatrix} a_{11}+b_{11} & \cdots & a_{1n}+b_{1n} \\ \vdots & & \vdots \\ a_{m1}+b_{m1} & \cdots & a_{mn}+b_{mn} \end{pmatrix}$$

$\boldsymbol{A}$ 的负矩阵：$-\boldsymbol{A}=(-a_{ij})_{m\times n}$。

$\boldsymbol{A}$ 与 $\boldsymbol{B}$ 的差：$\boldsymbol{A}-\boldsymbol{B}=\boldsymbol{A}+(-\boldsymbol{B})=(a_{ij}-b_{ij})_{m\times n}$。

2.2.2 矩阵的加法

定理 2

设元素全为零的矩阵称为**零矩阵**，记作 $\boldsymbol{O}_{m\times n}$，或简记作 $\boldsymbol{O}$。对任意的 $m\times n$ 矩阵 $\boldsymbol{A},\boldsymbol{B},\boldsymbol{C}$，矩阵的加法满足下列规律：

(1) 交换律 $\boldsymbol{A}+\boldsymbol{B}=\boldsymbol{B}+\boldsymbol{A}$,

(2) 结合律 $(\boldsymbol{A}+\boldsymbol{B})+\boldsymbol{C}=\boldsymbol{A}+(\boldsymbol{B}+\boldsymbol{C})$,

(3) $\boldsymbol{A}+\boldsymbol{O}=\boldsymbol{A}$,

(4) $\boldsymbol{A}+(-\boldsymbol{A})=\boldsymbol{O}$。

2.2.3 矩阵的乘法

定义 7

设 $\boldsymbol{A}=(a_{ij})_{m\times r}, \boldsymbol{B}=(b_{ij})_{r\times n}$，那么矩阵 $\boldsymbol{C}=(c_{ij})_{m\times n}$，其中

$c_{ij}=a_{i1}b_{1j}+a_{i2}b_{2j}+\cdots+a_{ir}b_{rj}=\sum_{k=1}^{r}a_{ik}b_{kj}\ (i=1,2,\cdots,m;j=1,2,\cdots,n)$，称为 $\boldsymbol{A}$ 与 $\boldsymbol{B}$ 的乘积，记作 $\boldsymbol{C}=\boldsymbol{A}\boldsymbol{B}$。

定理 3

矩阵的乘积满足下列规律:

(1) 结合律 $(\boldsymbol{A}\boldsymbol{B})\boldsymbol{C}=\boldsymbol{A}(\boldsymbol{B}\boldsymbol{C})$,

(2) 左分配律 $\boldsymbol{A}(\boldsymbol{B}+\boldsymbol{C})=\boldsymbol{A}\boldsymbol{B}+\boldsymbol{A}\boldsymbol{C}$, 右分配律 $(\boldsymbol{B}+\boldsymbol{C})\boldsymbol{A}=\boldsymbol{B}\boldsymbol{A}+\boldsymbol{C}\boldsymbol{A}$。

2.2.3 矩阵的乘法

一般情况下矩阵乘法不满足交换律，也即 $\boldsymbol{AB} \neq \boldsymbol{BA}$。

例 4

令矩阵 $\boldsymbol{A} = \begin{pmatrix} 1 & 2 & 3 \\ 3 & 2 & 1 \end{pmatrix} \in \mathbb{R}^{2\times3}, \boldsymbol{B} = \begin{pmatrix} 0 & 2 \\ 1 & -1 \\ 0 & 1 \end{pmatrix} \in \mathbb{R}^{3\times2}$

$$\boldsymbol{AB} = \begin{pmatrix} 1 & 2 & 3 \\ 3 & 2 & 1 \end{pmatrix} \begin{pmatrix} 0 & 2 \\ 1 & -1 \\ 0 & 1 \end{pmatrix} = \begin{pmatrix} 2 & 3 \\ 2 & 5 \end{pmatrix} \in \mathbb{R}^{2\times2}$$

$$\boldsymbol{BA} = \begin{pmatrix} 0 & 2 \\ 1 & -1 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & 2 & 3 \\ 3 & 2 & 1 \end{pmatrix} = \begin{pmatrix} 6 & 4 & 2 \\ -2 & 0 & 2 \\ 3 & 2 & 1 \end{pmatrix} \in \mathbb{R}^{3\times3}$$

2.2.3 矩阵的乘法

例 5

主对角线上的元素全是 *1*, 其余元素全是 *0* 的 *n* 阶方阵

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

称为 n 阶 **单位矩阵**, 记为 $\boldsymbol{I}_n$ 或简记为 $\boldsymbol{I}$。显然，对于 n 阶矩阵 $\boldsymbol{A}$, $\boldsymbol{A}\boldsymbol{I} = \boldsymbol{I}\boldsymbol{A} = \boldsymbol{A}$。

2.2.3 矩阵的乘法

定义 8

设 $\boldsymbol{A}=(a_{ij})_{m\times n}$ 是 $m\times n$ 矩阵，$\boldsymbol{x}=(x_1,x_2,\cdots,x_n)^{\mathrm{T}}$ 是 n 维列向量，令 $\boldsymbol{b}=(b_1,b_2,\cdots,b_m)$，其中 $b_i=a_{i1}x_1+a_{i2}x_2+\cdots+a_{in}x_n=\sum_{k=1}^{n}a_{ik}x_k(i=1,\cdots,m)$，则 $\boldsymbol{b}$ 称为矩阵 $\boldsymbol{A}$ 与向量 $\boldsymbol{x}$ 的乘积，记作 $\boldsymbol{b}=\boldsymbol{A}\boldsymbol{x}$。

$$\begin{pmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & & \vdots \\ a_{m1} & \cdots & a_{mn} \end{pmatrix}\begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix}=\begin{pmatrix} b_1 \\ \vdots \\ b_n \end{pmatrix}$$

2.2.4 矩阵的数量乘法

定义 9

设 $\boldsymbol{A}=(a_{ij})_{m\times n},\lambda\in\mathbb{K}$，则 λ 与 $\boldsymbol{A}$ **的数量乘积或标量乘积**定义为 $\lambda\boldsymbol{A}=(\lambda a_{ij})_{m\times n}$。

定理 4

设 $\boldsymbol{A}=(a_{ij})_{m\times r},\boldsymbol{B}=(b_{ij})_{r\times n}$，$\lambda,\mu\in\mathbb{K}$，则矩阵的数量乘积满足下列规律：

(1) $1\boldsymbol{A}=\boldsymbol{A}$,

(2) $(\lambda\mu)\boldsymbol{A}=\lambda(\mu\boldsymbol{A})$,

(3) $(\lambda+\mu)\boldsymbol{A}=\lambda\boldsymbol{A}+\mu\boldsymbol{A}$,

(4) $\lambda(\boldsymbol{A}+\boldsymbol{B})=\lambda\boldsymbol{A}+\lambda\boldsymbol{B}$,

(5) $\lambda(\boldsymbol{A}\boldsymbol{B})=(\lambda\boldsymbol{A})\boldsymbol{B}=\boldsymbol{A}(\lambda\boldsymbol{B})$。

2.2.5 分块矩阵的乘法

设 $\boldsymbol{A}=(a_{ik})_{m\times r}$,$\boldsymbol{B}=(b_{kj})_{r\times n}$, 把 $\boldsymbol{A}$,$\boldsymbol{B}$ 分成一些小矩阵;

$$\boldsymbol{A}=\begin{pmatrix} \boldsymbol{A}_{11} & \boldsymbol{A}_{12} & \cdots & \boldsymbol{A}_{1l} \\ \boldsymbol{A}_{21} & \boldsymbol{A}_{22} & \cdots & \boldsymbol{A}_{2l} \\ \cdots & \cdots & \cdots & \cdots \\ \boldsymbol{A}_{s1} & \boldsymbol{A}_{s2} & \cdots & \boldsymbol{A}_{sl} \end{pmatrix}, \boldsymbol{B}=\begin{pmatrix} \boldsymbol{B}_{11} & \boldsymbol{B}_{12} & \cdots & \boldsymbol{B}_{1t} \\ \boldsymbol{B}_{21} & \boldsymbol{B}_{22} & \cdots & \boldsymbol{B}_{2t} \\ \cdots & \cdots & \cdots & \cdots \\ \boldsymbol{B}_{l1} & \boldsymbol{B}_{l2} & \cdots & \boldsymbol{B}_{lt} \end{pmatrix},$$

其中每个 $\boldsymbol{A}_{ij}$ 是 $m_i\times r_j$ 矩阵，B_{ij} 是 $r_i\times n_j$ 矩阵。像这种把一个矩阵分为若干小矩阵，而把每个小矩阵当作数一样来处理的做法，称为矩阵的分块。

2.2.5 分块矩阵的乘法

若 $\boldsymbol{A}$ 的列的分法与 $\boldsymbol{B}$ 的行的分法一致，则

$$\boldsymbol{C}=\boldsymbol{A}\boldsymbol{B}=\begin{pmatrix} \boldsymbol{C}_{11} & \boldsymbol{C}_{12} & \cdots & \boldsymbol{C}_{1t} \\ \boldsymbol{C}_{21} & \boldsymbol{C}_{22} & \cdots & \boldsymbol{C}_{2t} \\ \cdots & \cdots & \cdots & \cdots \\ \boldsymbol{C}_{s1} & \boldsymbol{C}_{s2} & \cdots & \boldsymbol{C}_{st} \end{pmatrix},$$

其中 $\boldsymbol{C}_{pq}=\sum_{k=1}^{l}\boldsymbol{A}_{pk}\boldsymbol{B}_{kq}(p=1,2,\cdots,s;q=1,2,\cdots,t)$

例 6

设矩阵 $\boldsymbol{A}$ 和 $\boldsymbol{B}$ 为

$$\boldsymbol{A}=\begin{pmatrix}1&0&0&0\\0&1&0&0\\-1&2&1&0\\1&1&0&1\end{pmatrix}=\begin{pmatrix}\boldsymbol{I}_2&\boldsymbol{O}\\\boldsymbol{A}_1&\boldsymbol{I}_2\end{pmatrix}\quad \boldsymbol{B}=\begin{pmatrix}1&0&3&2\\-1&2&0&1\\1&0&4&1\\-1&-1&2&0\end{pmatrix}=\begin{pmatrix}\boldsymbol{B}_{11}&\boldsymbol{B}_{12}\\\boldsymbol{B}_{21}&\boldsymbol{B}_{22}\end{pmatrix}$$

其中，$\boldsymbol{I}_2$ 表示 2×2 的单位矩阵，$\boldsymbol{O}$ 表示零矩阵，而

$$\boldsymbol{A}_1=\begin{pmatrix}-1&2\\1&1\end{pmatrix},\boldsymbol{B}_{11}=\begin{pmatrix}1&0\\-1&2\end{pmatrix},\boldsymbol{B}_{12}=\begin{pmatrix}3&2\\0&1\end{pmatrix},$$

例续

$$\boldsymbol{B}_{21}=\begin{pmatrix}1 & 0\\ -1 & -1\end{pmatrix},\boldsymbol{B}_{22}=\begin{pmatrix}4 & 1\\ 2 & 0\end{pmatrix},$$

在计算 $\boldsymbol{AB}$ 时，把 $\boldsymbol{A},\boldsymbol{B}$ 都看成是由这些小矩阵组成的，即按 2 维矩阵来计算，于是有

$$\boldsymbol{AB}=\begin{pmatrix}\boldsymbol{I}_2 & \boldsymbol{O}\\ \boldsymbol{A}_1 & \boldsymbol{I}_2\end{pmatrix}\begin{pmatrix}\boldsymbol{B}_{11} & \boldsymbol{B}_{12}\\ \boldsymbol{B}_{21} & \boldsymbol{B}_{22}\end{pmatrix}$$

$$=\begin{pmatrix}\boldsymbol{B}_{11} & \boldsymbol{B}_{12}\\ \boldsymbol{A}_1\boldsymbol{B}_{11}+\boldsymbol{B}_{21} & \boldsymbol{A}_1\boldsymbol{B}_{12}+\boldsymbol{B}_{22}\end{pmatrix}=\begin{pmatrix}1 & 0 & 3 & 2\\ -1 & 2 & 0 & 1\\ -2 & 4 & 1 & 1\\ -1 & 1 & 5 & 3\end{pmatrix}$$

初等矩阵

定义 10

所谓数域 $\mathbb{K}$ 上矩阵的初等行 (列) 变换是指下列三种变换:
(1) 以 $\mathbb{K}$ 中一个非零的数乘矩阵的某一行 (列);
(2) 把矩阵的某一行 (列) 的 c 倍加到另一行 (列), 这里 c 是 $\mathbb{K}$ 中任意一个数;
(3) 互换矩阵中两行 (列) 的位置.

定义 11

由单位矩阵 $\boldsymbol{I}$ 经过一次初等行 (列) 变换得到的矩阵称为**初等矩阵**.

定理 5

对一个 $m \times n$ 的矩阵 $\boldsymbol{A}$ 作一初等行变换就相当于在 $\boldsymbol{A}$ 左边乘上相应的 $m \times m$ 的初等矩阵；对 $\boldsymbol{A}$ 作一初等列变换就相当于在 $\boldsymbol{A}$ 右边乘上相应的 $n \times n$ 的初等矩阵。

例 7

设矩阵 $\boldsymbol{A}$

$$\boldsymbol{A} = \begin{pmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 9 \end{pmatrix}$$

对矩阵 $\boldsymbol{A}$ 作一初等行变换（互换第 1 和第 3 行的位置），则有

$$\begin{pmatrix} 0 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & 0 \end{pmatrix} \begin{pmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 9 \end{pmatrix} = \begin{pmatrix} 7 & 8 & 9 \\ 4 & 5 & 6 \\ 1 & 2 & 3 \end{pmatrix}$$

2.2.6 逆

定义 12

设 $\boldsymbol{A}$ 是数域 $\mathbb{K}$ 上的 $n\times n$ 矩阵，如果存在 $\mathbb{K}$ 上的 $n\times n$ 矩阵 $\boldsymbol{B}$，使得 $\boldsymbol{AB}=\boldsymbol{BA}=\boldsymbol{I}$, 则称 $\boldsymbol{A}$ 为**可逆矩阵**，简称 $\boldsymbol{A}$ 可逆，而 $\boldsymbol{B}$ 则称为 $\boldsymbol{A}$ 的**逆矩阵**，记作 $\boldsymbol{A}^{-1}$。

并不是每个 $n\times n$ 矩阵 $\boldsymbol{A}$ 都可逆，如果 $\boldsymbol{A}$ 可逆，我们称 $\boldsymbol{A}$ 是可逆的或正规的或非奇异的，否则称 $\boldsymbol{A}$ 是奇异的或非可逆的。

定理 6

n 阶矩阵 $\boldsymbol{A}$ 可逆的充分必要条件是它能表示成一些初等矩阵 $\boldsymbol{Q}_1,\boldsymbol{Q}_2,\cdots,\boldsymbol{Q}_m$ 的乘积:

$$\boldsymbol{A}=\boldsymbol{Q}_1\boldsymbol{Q}_2\cdots\boldsymbol{Q}_m.$$

2.2.6 逆

例 8

考虑一个 2×2 矩阵：$\boldsymbol{A}=\begin{pmatrix}1 & 2\\ 3 & 4\end{pmatrix}$，则

$$(\boldsymbol{A}\quad \boldsymbol{I})=\begin{pmatrix}1 & 2 & 1 & 0\\ 3 & 4 & 0 & 1\end{pmatrix}\to\begin{pmatrix}1 & 2 & 1 & 0\\ 0 & -2 & -3 & 1\end{pmatrix}\to$$
$$\begin{pmatrix}1 & 0 & -2 & 1\\ 0 & -2 & -3 & 1\end{pmatrix}\to\begin{pmatrix}1 & 0 & -2 & 1\\ 0 & 1 & \frac{3}{2} & -\frac{1}{2}\end{pmatrix}\to(\boldsymbol{I}\quad \boldsymbol{A}^{-1})$$

即

$$\boldsymbol{A}^{-1}=\begin{pmatrix}-2 & 1\\ \frac{3}{2} & -\frac{1}{2}\end{pmatrix}$$

2.2.7 矩阵的转置

定义 13

设 $\boldsymbol{A}=(a_{ij})_{m\times n}$，把 $\boldsymbol{A}$ 的行、列互换所得到的矩阵称为 $\boldsymbol{A}$ 的**转置**，记作 $\boldsymbol{A}^{\mathrm{T}}$ 或 $\boldsymbol{A}^{'}$，即 $\boldsymbol{A}^{\mathrm{T}}=\boldsymbol{A}^{'}=(a_{ji})_{n\times m}$。

对于一个方阵 $\boldsymbol{A}$，$\boldsymbol{A}^{\mathrm{T}}$ 相当于我们把 $\boldsymbol{A}$ 沿着它的对角线做一个镜像, 而且如果 $\boldsymbol{A}$ 是对称的，则 $\boldsymbol{A}=\boldsymbol{A}^{\mathrm{T}}$。

2.2.7 矩阵的转置

逆和转置的一些性质

- $\boldsymbol{A}\boldsymbol{A}^{-1} = \boldsymbol{I} = \boldsymbol{A}^{-1}\boldsymbol{A}$
- $(\boldsymbol{A}\boldsymbol{B})^{-1} = \boldsymbol{B}^{-1}\boldsymbol{A}^{-1}$
- $(\boldsymbol{A}+\boldsymbol{B})^{-1} \neq \boldsymbol{A}^{-1} + \boldsymbol{B}^{-1}$
- $(\boldsymbol{A}+\boldsymbol{B})^{\mathrm{T}} = \boldsymbol{A}^{\mathrm{T}} + \boldsymbol{B}^{\mathrm{T}}$
- $(\boldsymbol{A}\boldsymbol{B})^{\mathrm{T}} = \boldsymbol{B}^{\mathrm{T}}\boldsymbol{A}^{\mathrm{T}}$
- $(\boldsymbol{A}^{-1})^{\mathrm{T}} = (\boldsymbol{A}^{\mathrm{T}})^{-1} =: \boldsymbol{A}^{-\mathrm{T}}$

2.2.8 更多例子：用矩阵表示线性方程组

例 9

某公司生产产品 $N_1,\ldots,N_n$，其需要的资源分别为 $R_1,\ldots,R_m$。为了生产一单位的 N_j 需要 a_{ij} 单位的 R_i，其中 $i=1,\ldots,m; j=1,\ldots,n$。目的是找到最佳的生产计划，也即如果有 b_i 单位的 R_i 可供使用，那么应该生产多少 (设为 x_j) 单位的产品 N_j 使得恰好用尽资源。如果我们生产 $x_1,\ldots,x_n$ 单位的对应产品，我们一共需要 $a_{i1}x_1+\cdots+a_{in}x_n$ 单位资源 R_i。最优生产计划 $(x_1,\ldots,x_n)\in\mathbb{R}^n$，因此它必须满足方程组

$$\begin{aligned} a_{11}x_1+\cdots+a_{1n}x_n &= b_1 \\ &\vdots \\ a_{m1}x_1+\cdots+a_{mn}x_n &= b_m \end{aligned}$$

2.2.8 更多例子：用矩阵表示线性方程组

上述方程组可以写成矩阵的形式

$$\begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \vdots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix} = \begin{pmatrix} b_1 \\ b_2 \\ \vdots \\ b_m \end{pmatrix}$$

用一个紧凑形式表示即为

$$\boldsymbol{A}\boldsymbol{x} = \boldsymbol{b} \tag{2}$$

我们称(2)为线性方程组的一般形式。

例 10

求

$$\begin{pmatrix} 1 & 2 \\ 3 & 4 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} 3 \\ 7 \end{pmatrix}$$

我们记增广矩阵为

$$\begin{pmatrix} 1 & 2 & 3 \\ 3 & 4 & 7 \end{pmatrix}$$

我们使用行初等变换法求解这个线性方程组。即

$$\begin{pmatrix} 1 & 2 & 3 \\ 3 & 4 & 7 \end{pmatrix} \to \begin{pmatrix} 1 & 2 & 3 \\ & -2 & -2 \end{pmatrix} \to \begin{pmatrix} 1 & 2 & 3 \\ & 1 & 1 \end{pmatrix} \to \begin{pmatrix} 1 & & 1 \\ & 1 & 1 \end{pmatrix}$$

我们将左边化为单位阵，最右边的一列即为解。所以解

$$\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} 1 \\ 1 \end{pmatrix}$$

解法 2

上面，我们介绍了矩阵的逆。如果系数矩阵可逆，则可以在方程两边同时左乘系数矩阵的逆 $\boldsymbol{A}^{-1}$ 即

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

对于上面那个线性方程组，他的系数矩阵的逆我们已经求过了。

$$\begin{pmatrix} 1 & 2 \\ 3 & 4 \end{pmatrix}^{-1} = \begin{pmatrix} -2 & 1 \\ \frac{3}{2} & -\frac{1}{2} \end{pmatrix}$$

那么

$$\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} -2 & 1 \\ \frac{3}{2} & -\frac{1}{2} \end{pmatrix} \begin{pmatrix} 3 \\ 7 \end{pmatrix} = \begin{pmatrix} 1 \\ 1 \end{pmatrix}$$