

作者	生姜 DrGinger
脚本	生姜 DrGinger
视频	崔崔 CuiCui
开源学习资源	https://github.com/Visualize-ML
平台	https://www.youtube.com/@DrGinger_Jiang https://space.bilibili.com/3546865719052873 https://space.bilibili.com/513194466

03

Matrix Multiplication

矩阵乘法

内积、外积、列向量组合、行向量组合 ...

3.1 矩阵乘法第一视角



本节你将掌握的核心技能：

- ▶ 行列对齐理解乘法形状匹配： AB 相乘，将 A 行对齐放左侧，将 B 列对齐放上方，可直观判断 AB 的合法性及结果形状。
- ▶ 内积视角计算元素：矩阵乘法中每个元素是 A 的一行与 B 的一列内积。
- ▶ 向量内积可转化为矩阵乘法： 1×1 矩阵相当于标量。
- ▶ 向量外积生成矩阵：构造秩一矩阵（下一节展开讲解）。

上一章已经介绍过矩阵乘法，本章专门介绍理解矩阵乘法的不同视角。

矩阵乘法的第一视角是一种直观理解矩阵乘法的方式，通过将矩阵 A 和矩阵 B 以特定的方式对齐，帮助我们更好地理解矩阵乘法的形状要求，以及每个元素计算过程。

行、列对齐

如图 1 所示，矩阵 A 是一个 $m \times p$ 的矩阵，矩阵 B 是一个 $p \times n$ 的矩阵。

将矩阵 A 放在结果矩阵 C 的左侧，行对齐；同时将矩阵 B 放在 C 的上面，列对齐。这种对齐方式帮助我们清晰地看到矩阵乘法的形状规则。

将 A 放在 C 的左侧，意味着 A 的行数 m 决定了 C 的行数 m 。

将 B 放在 C 的上面，意味着 B 的列数 n 决定了 C 的列数 n 。

因此，矩阵乘法 $A @ B$ 结果矩阵 C 的形状是 $m \times n$ 。

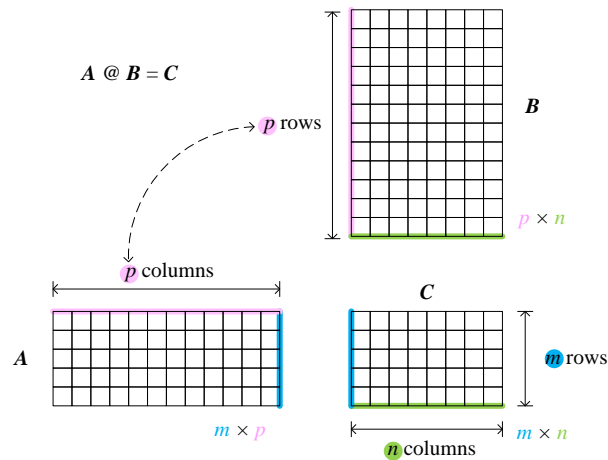


图 1. A 放在 AB 的左侧，行对齐； B 放在 AB 的上面，列对齐

这种对齐方式还揭示了矩阵乘法的核心规则： A 的列数必须等于 B 的行数，否则无法对齐。

如图 2 所示，对齐后， C 的每个元素 $c_{i,j}$ 是 A 的第 i 行与 B 的第 j 列对应元素相乘并求和的结果。

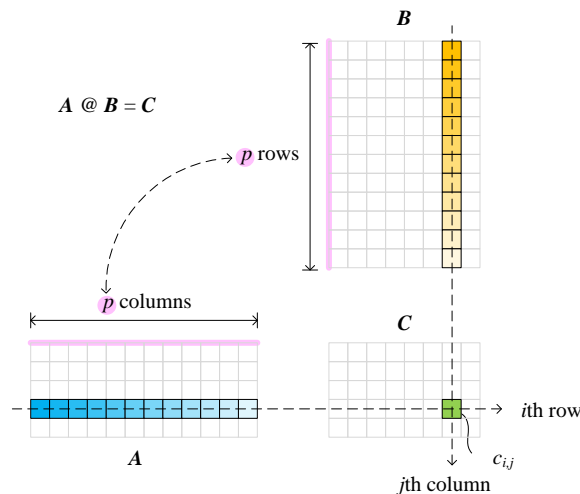


图 2. 计算 AB 的第 i 行、第 j 列元素： A 放在 AB 的左侧，行对齐； B 放在 AB 的上面，列对齐

这种空间对齐的视角不仅帮助我们理解矩阵乘法的形状，还让我们更直观地感受到矩阵乘法是如何将两个矩阵的信息通过特定的方式整合在一起的。

计算每个元素

前文提过，矩阵 A 写成一组行向量，每个行向量记作 $a^{(i)}$ ；矩阵 B 写成一组列向量，每个列向量记作 b_j 。我们可以把矩阵乘法 AB 写成若干个矩阵乘法（内积视角）。

如图 3 所示，把矩阵乘积 $C = AB$ 的第 i 行、第 j 列元素 $c_{i,j}$ 写成一个矩阵乘法

$$c_{i,j} = a^{(i)} @ b_j \quad (1)$$

当然，我们也可以用向量内积计算 $c_{i,j}$ ，即

$$c_{i,j} = a^{(i)T} \cdot b_j \quad (2)$$

因此，矩阵乘法第一视角也叫内积展开。

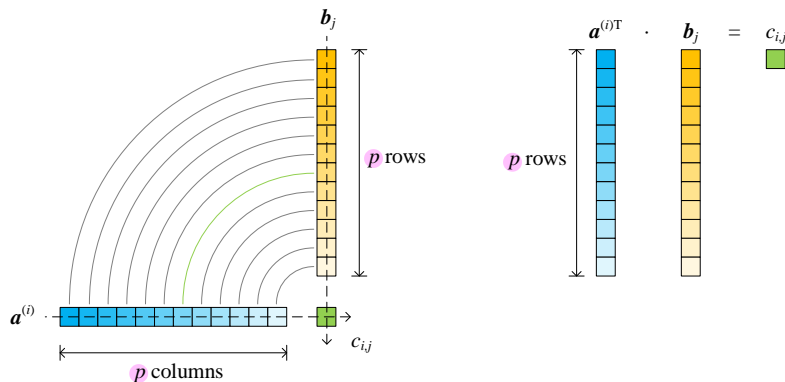


图 3. 内积展开视角看矩阵乘积 $a^{(i)} @ b_j = c_{i,j}$

内积

(1) 还告诉我们两个等维数的列向量内积也可以写成矩阵乘法形式。

比如，如图 4 所示， n 维向量 a 、 b 的内积，等价于

$$a \cdot b = \langle a, b \rangle = a^T b = b^T a \quad (3)$$

这便是本书前文介绍的向量**内积** (inner product)。

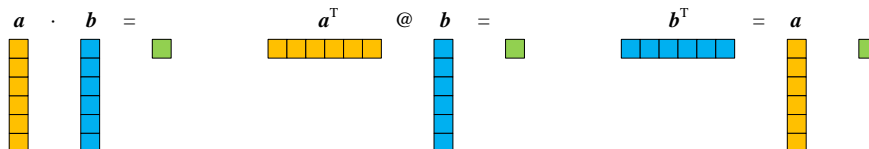


图 4. 向量内积也可以写成矩阵乘法形式

举个例子，给定等维数向量

$$\mathbf{a} = \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix}, \quad \mathbf{b} = \begin{bmatrix} 4 \\ 5 \\ 6 \end{bmatrix} \quad (4)$$

用矩阵乘法 $\mathbf{a}^T \mathbf{b}$ 计算两者内积

$$\mathbf{a}^T @ \mathbf{b} = \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix}^T @ \begin{bmatrix} 4 \\ 5 \\ 6 \end{bmatrix} = [1 \quad 2 \quad 3] @ \begin{bmatrix} 4 \\ 5 \\ 6 \end{bmatrix} = 32 \quad (5)$$

用矩阵乘法 $\mathbf{b}^T \mathbf{a}$ 计算两者内积

$$\mathbf{b}^T @ \mathbf{a} = \begin{bmatrix} 4 \\ 5 \\ 6 \end{bmatrix}^T @ \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix} = [4 \quad 5 \quad 6] @ \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix} = 32 \quad (6)$$

从矩阵形状来看， $(1 \times 3) @ (3 \times 1)$ 夹在中间的 (3) 被“消去”，矩阵乘法结果的形状为 (1×1) 。前文提过， 1×1 矩阵可以视作标量。

如图 5 所示，列向量 \mathbf{a} 长度（大小、 L^2 范数、模、欧几里得范数）的平方既可以通过内积求得，也可以用矩阵乘法计算，即

$$\|\mathbf{a}\|^2 = \|\mathbf{a}\|_2^2 = \mathbf{a} \cdot \mathbf{a} = \langle \mathbf{a}, \mathbf{a} \rangle = \mathbf{a}^T \mathbf{a} \quad (7)$$

这样来看， \mathbf{a} 长度（大小、 L^2 范数、模、欧几里得范数）就是上式值的平方根

$$\|\mathbf{a}\| = \|\mathbf{a}\|_2 = \sqrt{\mathbf{a} \cdot \mathbf{a}} = \sqrt{\langle \mathbf{a}, \mathbf{a} \rangle} = \sqrt{\mathbf{a}^T \mathbf{a}} \quad (8)$$

图 5. 向量长度的平方也可以写成矩阵乘法形式

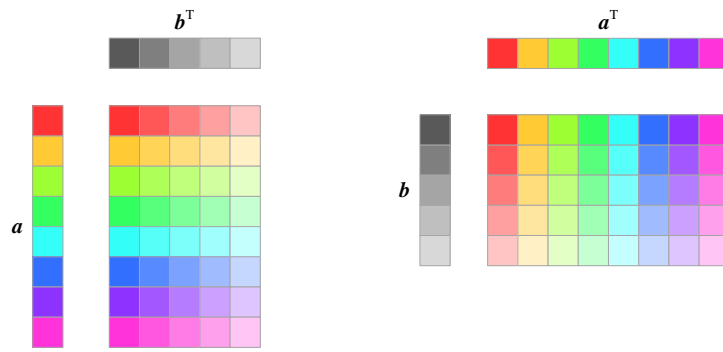
用矩阵乘法计算 (4) 中向量 \mathbf{a} 的长度（大小、 L^2 范数、模、欧几里得范数）

$$\|\mathbf{a}\| = \|\mathbf{a}\|_2 = \sqrt{\mathbf{a}^T \mathbf{a}} = \sqrt{[1 \quad 2 \quad 3] @ \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix}} = \sqrt{14} \quad (9)$$

外积

和内积相对的就是**外积** (outer product)。

如果说两个向量的内积结果为标量，那么两个向量的外积结果就是矩阵，具体如图 6 所示。这种特殊的矩阵乘法形态会帮助我们理解下一节的“矩阵乘法第二视角”。

图 6. 向量 a 、 b 的两个外积

举个例子，给定两个列向量（不要求等维数）

$$a = \begin{bmatrix} 1 \\ 2 \\ 3 \\ 4 \\ 5 \end{bmatrix}, \quad b = \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix} \quad (10)$$

计算第一个外积 $a @ b^T$

$$a @ b^T = \begin{bmatrix} 1 \\ 2 \\ 3 \\ 4 \\ 5 \end{bmatrix} @ \begin{bmatrix} 1 & 2 & 3 \end{bmatrix} = \begin{bmatrix} 1 & 2 & 3 \\ 2 & 4 & 6 \\ 3 & 6 & 9 \\ 4 & 8 & 12 \\ 5 & 10 & 15 \end{bmatrix} \quad (11)$$

从矩阵形状来看， $(5 \times 1) @ (1 \times 3)$ 夹在中间的 (1) 被“消去”，矩阵乘法结果的形状为 (5×3) 。

再计算第二个外积 $b @ a^T$

$$b @ a^T = \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix} @ \begin{bmatrix} 1 & 2 & 3 & 4 & 5 \end{bmatrix} = \begin{bmatrix} 1 & 2 & 3 \\ 2 & 4 & 6 \\ 3 & 6 & 9 \\ 4 & 8 & 12 \\ 5 & 10 & 15 \end{bmatrix} \quad (12)$$

从矩阵形状来看， $(3 \times 1) @ (1 \times 5)$ 夹在中间的 (1) 被“消去”，矩阵乘法结果的形状为 (3×5) 。

第一个例子

回到本书前文用过的矩阵乘法的例子

$$A @ B = \begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \end{bmatrix} @ \begin{bmatrix} 1 & 4 \\ 2 & 5 \\ 3 & 6 \end{bmatrix} = \begin{bmatrix} 14 & 32 \\ 32 & 77 \end{bmatrix} \quad (13)$$

A 写成行向量, B 写成列向量后, $A @ B$ 可以写成

$$\begin{aligned}
 A @ B &= \begin{bmatrix} a^{(1)} \\ a^{(2)} \end{bmatrix} \begin{bmatrix} b_1 & b_2 \end{bmatrix} = \begin{bmatrix} a^{(1)}b_1 & a^{(1)}b_2 \\ a^{(2)}b_1 & a^{(2)}b_2 \end{bmatrix} \\
 &= \begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \end{bmatrix} @ \begin{bmatrix} \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix} & \begin{bmatrix} 4 \\ 5 \\ 6 \end{bmatrix} \end{bmatrix} = \begin{bmatrix} [1 \ 2 \ 3] @ \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix} & [1 \ 2 \ 3] @ \begin{bmatrix} 4 \\ 5 \\ 6 \end{bmatrix} \\ [4 \ 5 \ 6] @ \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix} & [4 \ 5 \ 6] @ \begin{bmatrix} 4 \\ 5 \\ 6 \end{bmatrix} \end{bmatrix} \\
 &= \begin{bmatrix} 14 & 32 \\ 32 & 77 \end{bmatrix}
 \end{aligned} \tag{14}$$

如图 7 所示, 利用矩阵乘法第一视角, 大家可能已经发现上式中包含四个矩阵乘法运算, 即。

$$\begin{cases} c_{1,1} = a^{(1)} @ b_1 \\ c_{1,2} = a^{(1)} @ b_2 \\ c_{2,1} = a^{(2)} @ b_1 \\ c_{2,2} = a^{(2)} @ b_2 \end{cases} \tag{15}$$

比如, $a^{(1)} @ b_1$ 是一个矩阵乘法; 形状为 1×3 行向量 $a^{(1)}$ 和形状为 3×1 列向量 b_1 乘积结果为 1×1 矩阵, 相当于一个标量。

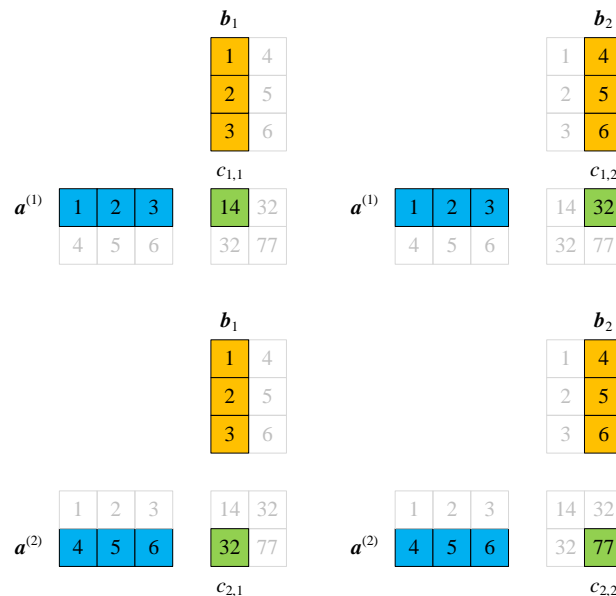


图 7. A 放在 AB 的左侧, 行对齐; B 放在 AB 的上面, 列对齐

第二个例子

再看第二个例子, 计算矩阵乘积 $B @ A$

$$D = B @ A = \begin{bmatrix} 1 & 4 \\ 2 & 5 \\ 3 & 6 \end{bmatrix} @ \begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \end{bmatrix} = \begin{bmatrix} 17 & 22 & 27 \\ 22 & 29 & 36 \\ 27 & 36 & 45 \end{bmatrix} \quad (16)$$

B 写成行向量, A 写成列向量后, $B @ A$ 可以写成

$$\begin{aligned} B @ A &= \begin{bmatrix} b^{(1)} \\ b^{(2)} \\ b^{(3)} \end{bmatrix} \begin{bmatrix} a_1 & a_2 & a_3 \end{bmatrix} = \begin{bmatrix} b^{(1)}a_1 & b^{(1)}a_2 & b^{(1)}a_3 \\ b^{(2)}a_1 & b^{(2)}a_2 & b^{(2)}a_3 \\ b^{(3)}a_1 & b^{(3)}a_2 & b^{(3)}a_3 \end{bmatrix} \\ &= \begin{bmatrix} [1 & 4] \\ [2 & 5] \\ [3 & 6] \end{bmatrix} @ \begin{bmatrix} \begin{bmatrix} 1 \\ 4 \end{bmatrix} & \begin{bmatrix} 2 \\ 5 \end{bmatrix} & \begin{bmatrix} 3 \\ 6 \end{bmatrix} \end{bmatrix} = \begin{bmatrix} [1 & 4] @ \begin{bmatrix} 1 \\ 4 \end{bmatrix} & [1 & 4] @ \begin{bmatrix} 2 \\ 5 \end{bmatrix} & [1 & 4] @ \begin{bmatrix} 3 \\ 6 \end{bmatrix} \\ [2 & 5] @ \begin{bmatrix} 1 \\ 4 \end{bmatrix} & [2 & 5] @ \begin{bmatrix} 2 \\ 5 \end{bmatrix} & [2 & 5] @ \begin{bmatrix} 3 \\ 6 \end{bmatrix} \\ [3 & 6] @ \begin{bmatrix} 1 \\ 4 \end{bmatrix} & [3 & 6] @ \begin{bmatrix} 2 \\ 5 \end{bmatrix} & [3 & 6] @ \begin{bmatrix} 3 \\ 6 \end{bmatrix} \end{bmatrix} \\ &= \begin{bmatrix} 17 & 22 & 27 \\ 22 & 29 & 36 \\ 27 & 36 & 45 \end{bmatrix} \end{aligned} \quad (17)$$

图 8 展示如何用矩阵乘法第一视角理解 $B @ A$ 的每个元素的计算规则。

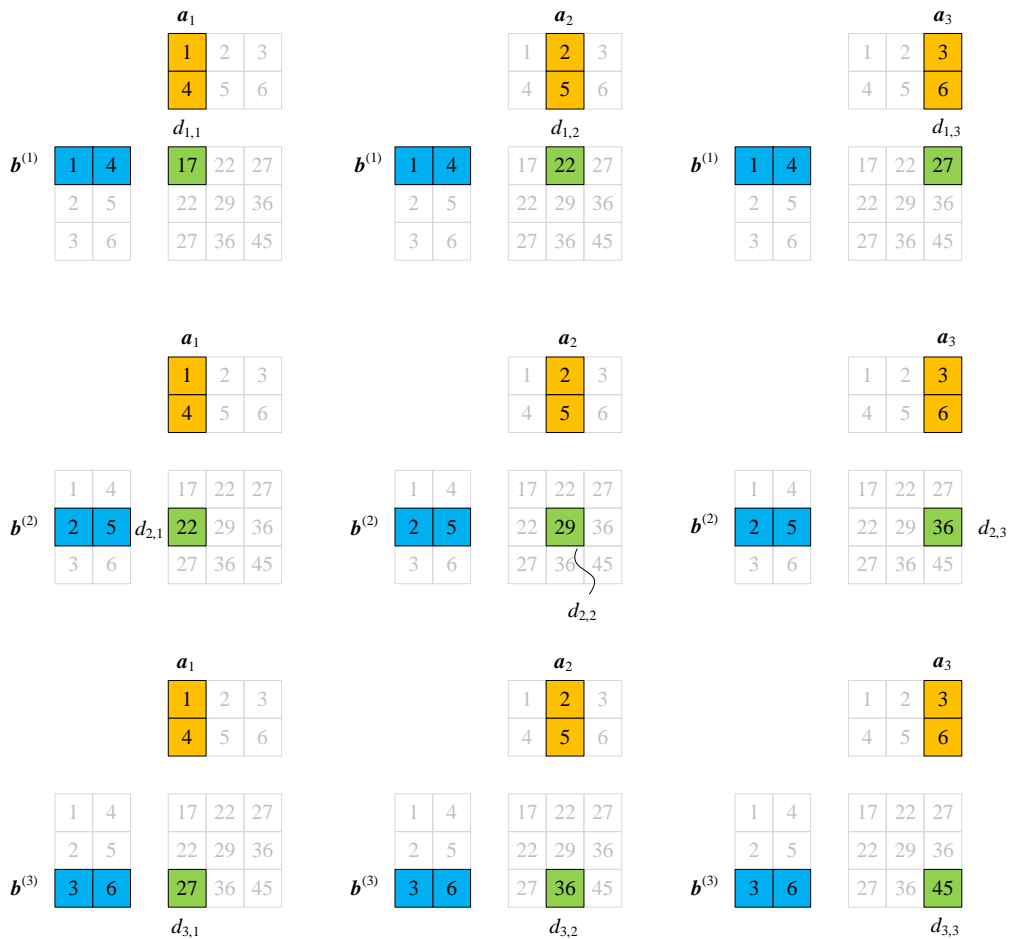


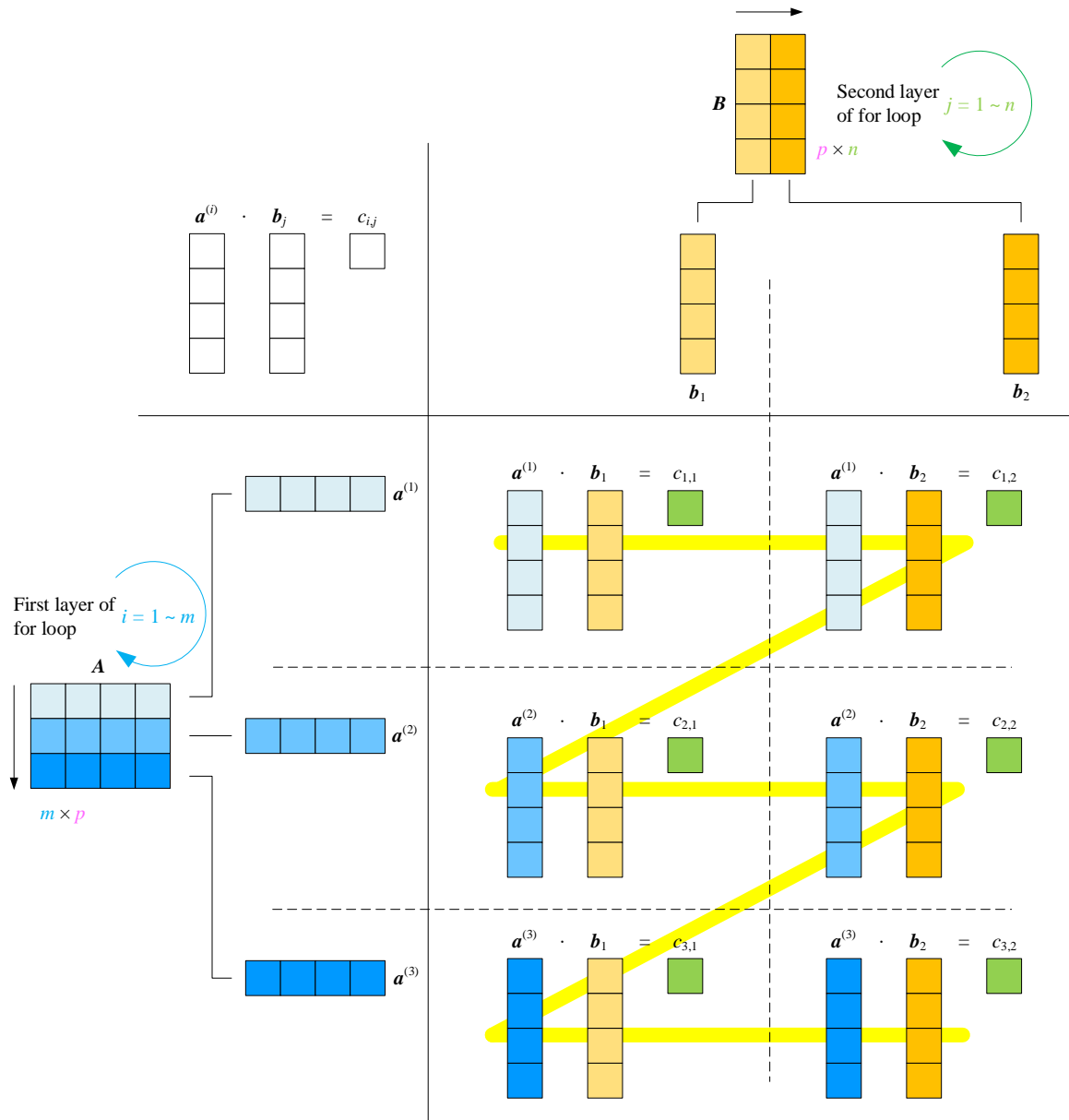
图 8. 计算每个元素： B 放在 BA 的左侧，行对齐； A 放在 BA 的上面，列对齐

LA_03_01_01.ipynb 以矩阵乘法第一视角计算上述矩阵乘法。

自定义 Python 函数计算矩阵乘法：内积视角

代码 1 自定义 Python 函数计算矩阵乘法，这段代码采用的本节介绍的内积视角。下面聊聊其中关键语句。注意，上一章介绍过的语句，本节略过，建议大家回顾。

- a** 用 `def` 定义了一个名为 `matrix_multiplication_inner` 自定义函数，它需要两个输入，分别是矩阵 A 和矩阵 B 。
- b** 开启一个循环。意思是“从第 0 行开始，一直到第 $m-1$ 行”，每次拿出矩阵 A 的一行来处理。 i 是当前的行号。
- c** 是第二层循环。意思是“对每一列都做一次处理”，从第 0 列到第 $n-1$ 列。 j 是当前的列号。其中，用 `numpy.dot()` 函数把 `row_i_A` 和 `col_j_B` 做“内积”。内积就是一行和一列相乘并相加的结果，是矩阵乘法中单个元素的值。

图 9. 矩阵乘法 AB 规则，内积视角

代码 1. 自定义 Python 函数计算矩阵乘法，内积视角 | LA_03_01_02.ipynb

```

## 初始化
import numpy as np

## 自定义函数，矩阵乘法第一视角
def matrix_multiplication_inner(A, B):

    # 获取矩阵 A 和 B 的形状
    m, p_A = A.shape
    p_B, n = B.shape

    # 检测矩阵形状是否符合矩阵乘法规则
    if p_A != p_B:
        raise ValueError('Dimensions do not match')

    # 初始化结果矩阵 C，形状 (m, n)，初始值设为 0
    C = np.zeros((m, n))

    # 计算每个元素
    for i in range(m):
        for j in range(n):
            row_i_A = A[i, :]          # A 的第 i 行（行向量）
            col_j_B = B[:, j]          # B 的第 j 列（列向量）
            C[i, j] = np.dot(row_i_A, col_j_B) # 计算内积

    return C

## 矩阵乘法
A = np.array([[1, 2, 3],
              [4, 5, 6]])
B = A.T

## 矩阵乘法
matrix_multiplication_inner(A, B)
matrix_multiplication_inner(B, A)

```



请大家用 DeepSeek/ChatGPT 等工具完成本节如下习题。

Q1. 给定列向量 $\mathbf{a} = \begin{bmatrix} 1 \\ 2 \end{bmatrix}$ ，用矩阵乘法第一视角方法摆放并计算 $\mathbf{a}^T @ \mathbf{a}$ 、 $\mathbf{a} @ \mathbf{a}^T$ 。

Q2. 给定列向量 $\mathbf{a} = \begin{bmatrix} 1 \\ 2 \end{bmatrix}$ 和 $\mathbf{b} = \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix}$ ，用矩阵乘法第一视角方法摆放并计算 $\mathbf{a} @ \mathbf{b}^T$ 、 $\mathbf{b} @ \mathbf{a}^T$ 。

Q3. 给定 $\mathbf{A} = \begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix}$ 和 $\mathbf{b} = \begin{bmatrix} 1 \\ 2 \end{bmatrix}$ ，用矩阵乘法第一视角方法摆放并计算 $\mathbf{A} @ \mathbf{b}$ 、 $\mathbf{b}^T @ \mathbf{A}^T$ 。

Q4. 给定 $A = \begin{bmatrix} 1 & 2 \\ 3 & 4 \\ 5 & 6 \end{bmatrix}$ ，用矩阵乘法第一视角方法摆放并计算 $A @ A^T$ 、 $A^T @ A$ 。

Q5. 修改代码 1，调换内外两层 for 循环。