矩阵和向量空间

模型里，数据的基本表示形式

小胖

目录

如果有人不相信数学是简单的，那是因为他们没有意识到人生有多复杂

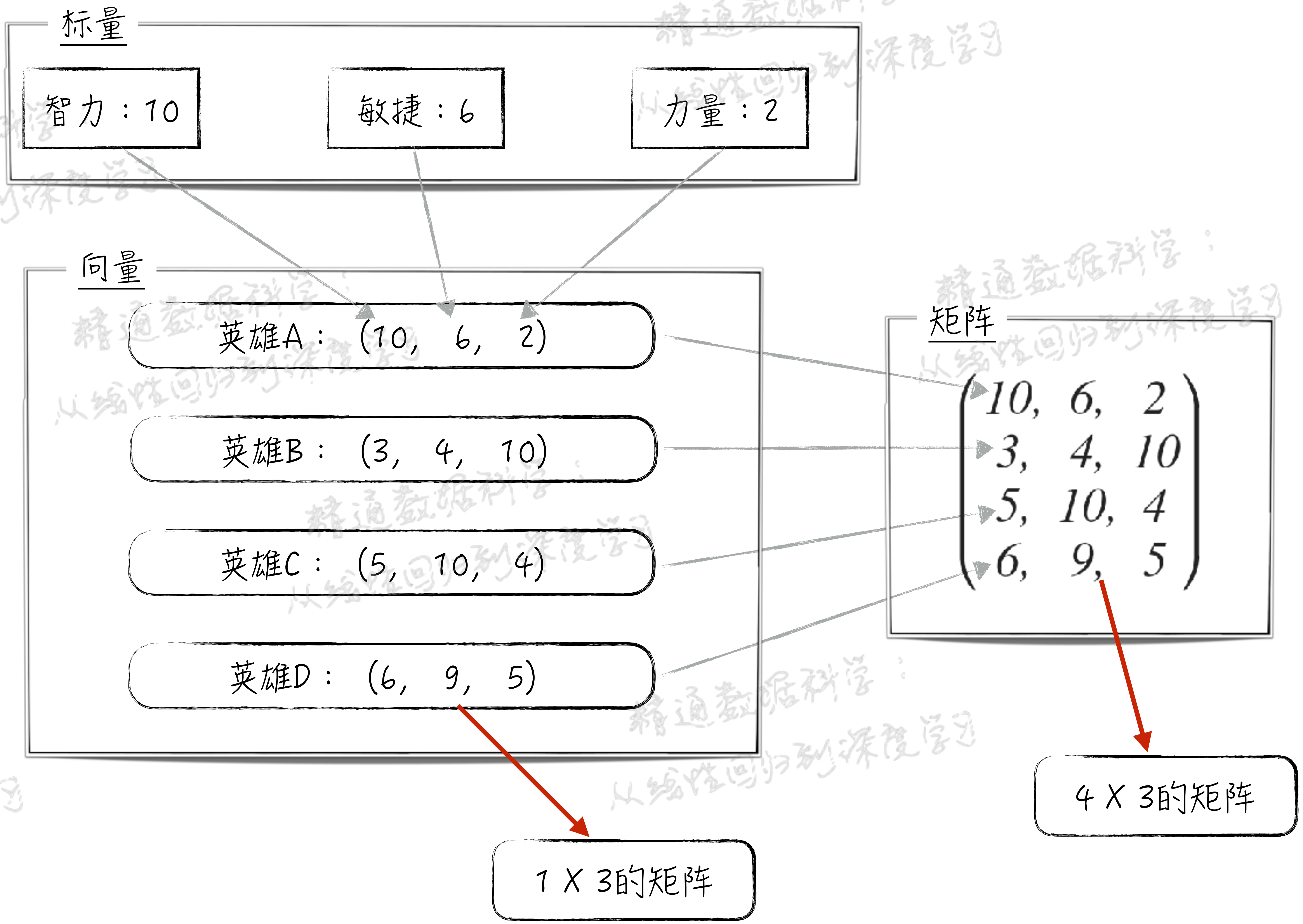
矩阵的定义

标量、向量和矩阵

想象一款网络游戏

- 它用智力、敏捷和力量这三个属性来客户英雄

DOTA 2

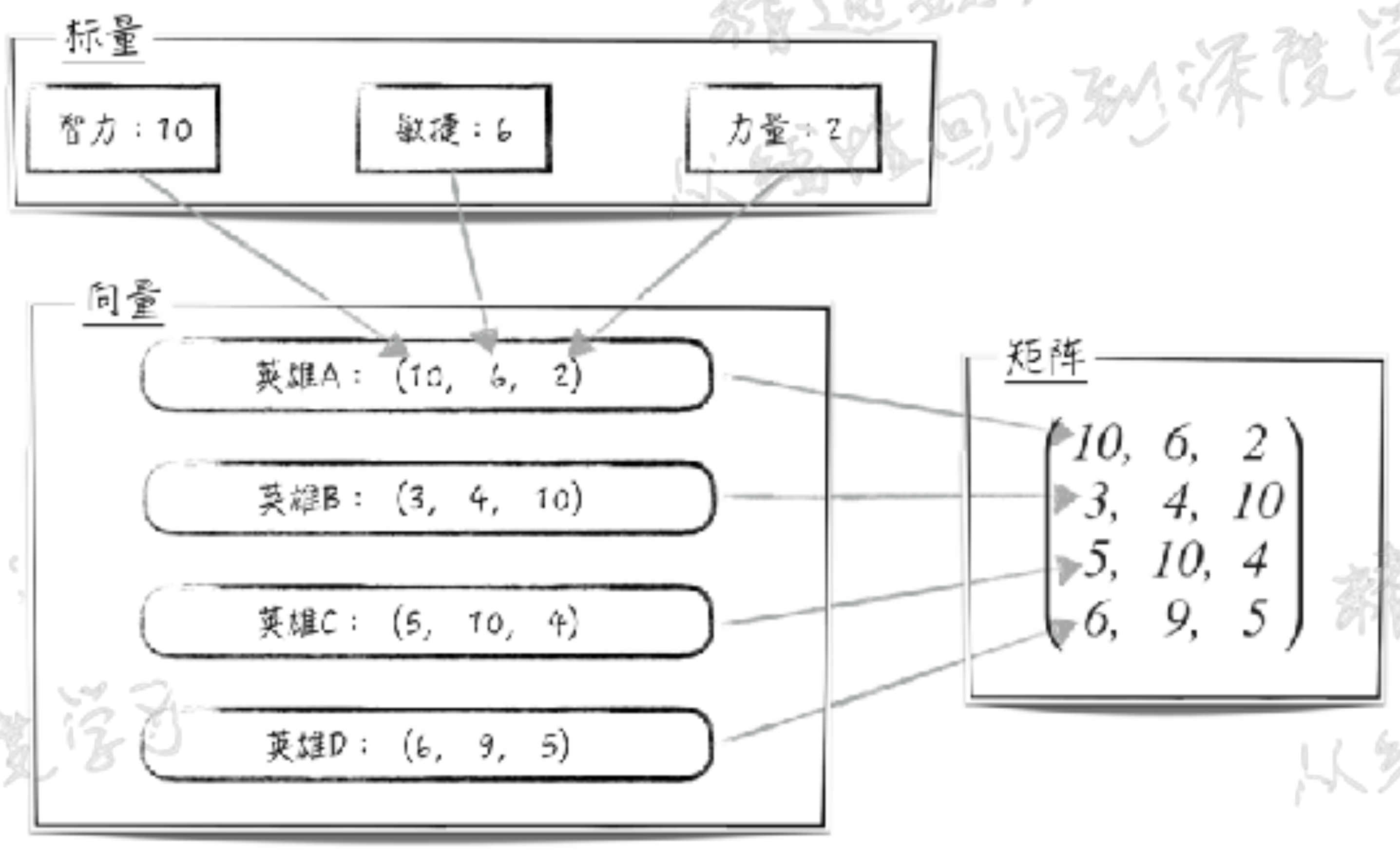


矩阵的定义

标量、向量和矩阵

在数学上，通常

- 使用 $x_{i,j}$ 来表示标量
- 使用X_i来表示向量
- 使用X 来表示n x m的矩阵



标量、向量和矩阵的关系如下

在数学上，通常

- 使用$R^{n \times m}$ 来表示n x m 的矩阵全体

$$X_i = (x_{i,1}, x_{i,2}, \dots, x_{i,m})$$

$$X = \begin{pmatrix} X_1 \\ X_2 \\ \vdots \\ X_n \end{pmatrix} = (x_{i,j}) \in \mathbf{R}^{n \times m}$$

矩阵的定义

特殊矩阵

单位矩阵

$$I_n = \begin{pmatrix} 1 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \cdots & 1 \end{pmatrix} = \left(1_{\{i=j\}}\right) \in \mathbf{R}^{n\times n}$$

对角矩阵

$$diag(d_1, d_2, \ldots, d_n) = \begin{pmatrix} d_1 & 0 & \cdots & 0 \\ 0 & d_2 & \cdots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \cdots & d_n \end{pmatrix} \in \mathbf{R}^{n\times n}$$

三角矩阵

$$L = \begin{pmatrix} l_{1,1} & 0 & \cdots & 0 \\ l_{2,1} & l_{2,2} & \cdots & 0 \\ \vdots & \vdots & & \vdots \\ l_{n,1} & l_{n,2} & \cdots & l_{n,n} \end{pmatrix} \in \mathbf{R}^{n\times n} \qquad U = \begin{pmatrix} u_{1,1} & u_{1,2} & \cdots & u_{1,n} \\ 0 & u_{2,2} & \cdots & u_{2,n} \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \cdots & u_{n,n} \end{pmatrix} \in \mathbf{R}^{n\times n}$$

目录

矩阵的运算

加法、减法

矩阵可以看作是一些实数的某种排列，因此，
我们希望将实数的四则运算延续到矩阵上。

两个矩阵的行数和列数相同

$$X = (x_{i,j}) \in \mathbf{R}^{n \times m};\ Y = (y_{i,j}) \in \mathbf{R}^{n \times m}$$

矩阵的加减法：
要求两个矩阵的形状是一样的

$$X \pm Y = \begin{pmatrix} x_{1,1} \pm y_{1,1} & \cdots & x_{1,m} \pm y_{1,m} \\ \vdots & & \vdots \\ x_{n,1} \pm y_{n,1} & \cdots & x_{n,m} \pm y_{n,m} \end{pmatrix} = (x_{i,j} + y_{i,j}) \in \mathbf{R}^{n \times m}$$

矩阵加法满足
加法交换律和结合律

$$X + Y = Y + X$$
$$X + Y + Z = X + (Y + Z)$$

矩阵的运算

矩阵的乘法

有关矩阵的乘法可以分为两种：

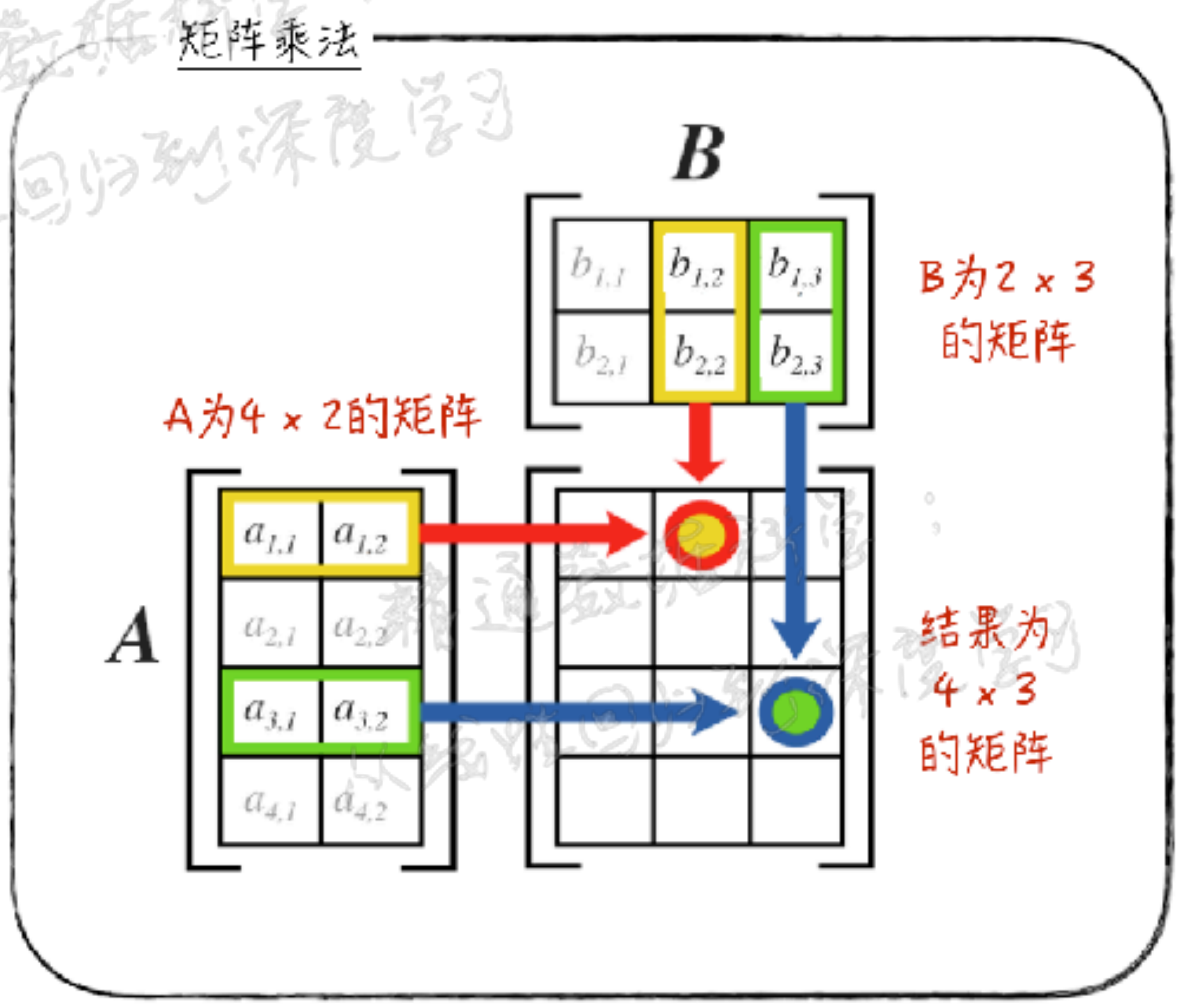
- 矩阵与数字的乘法
- 矩阵与矩阵的乘法

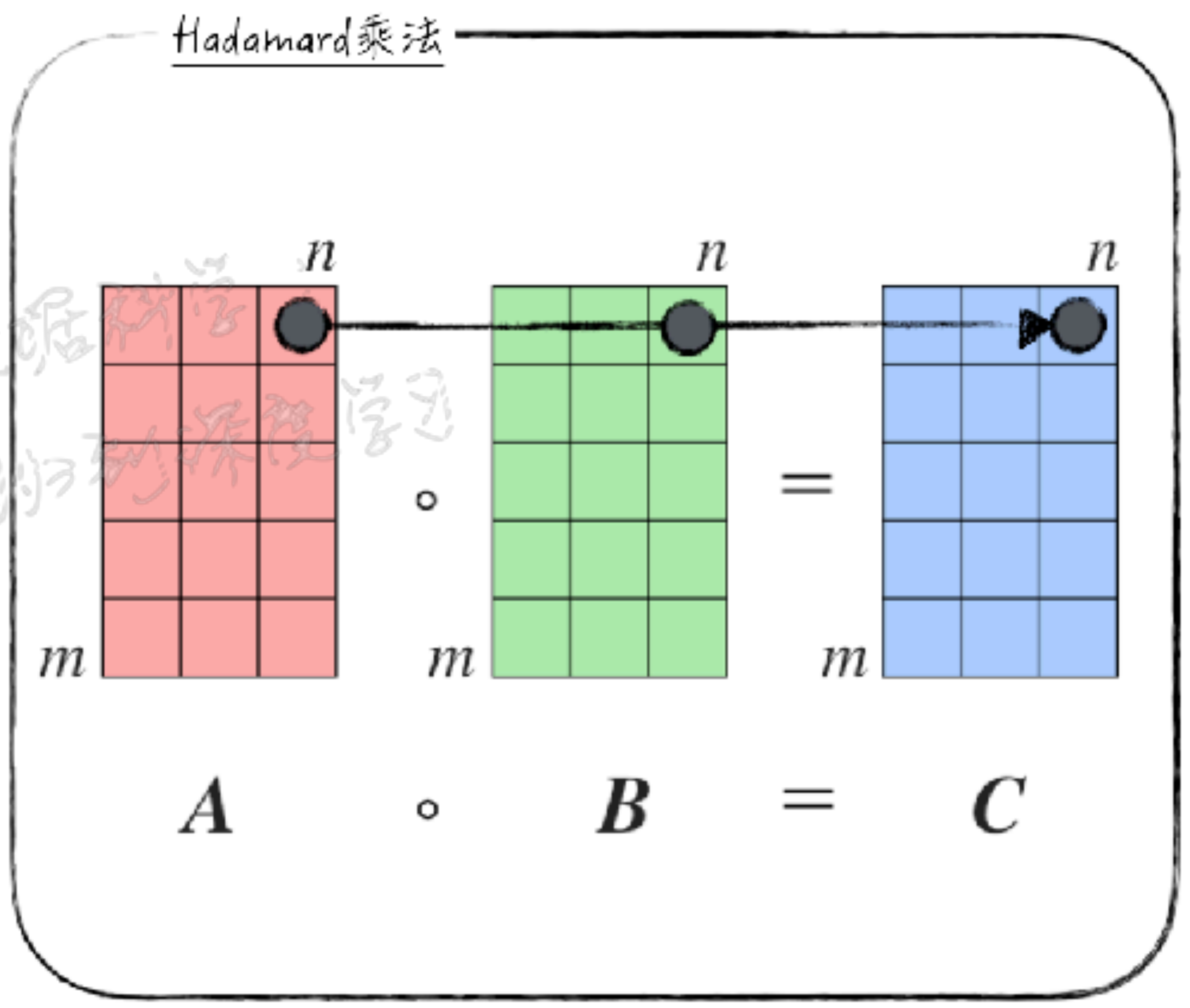
矩阵与数字的乘法很简单

$$kX = \begin{pmatrix} kx_{1,1} & \dots & kx_{1,m} \\ \vdots & & \vdots \\ kx_{n,1} & \dots & kx_{n,m} \end{pmatrix} = (kx_{i,j}) \in \mathbf{R}^{n \times m}$$

矩阵与矩阵的乘法有两种：

- 矩阵乘法
- Hadamard乘法





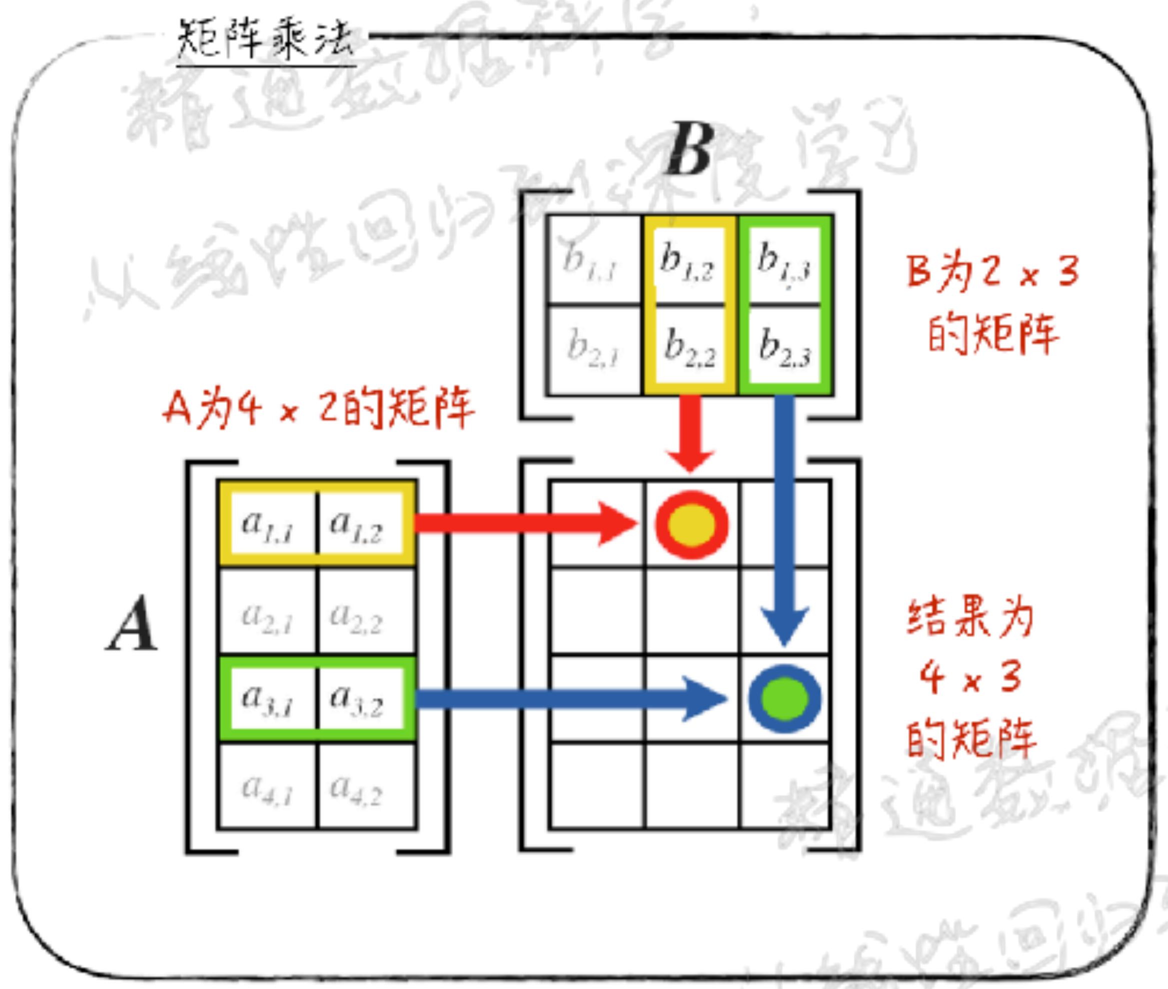
矩阵的运算

矩阵的乘法

矩阵乘法要求
第一个矩阵的列数 = 第二个矩阵的行数

$$A = (a_{i,j}) \in \mathbf{R}^{n \times p};\ B = (b_{i,j}) \in \mathbf{R}^{p \times m}$$

$$AB = \left(\sum_{r=1}^{p} a_{i,r} b_{r,j}\right) \in \mathbf{R}^{n \times m}$$



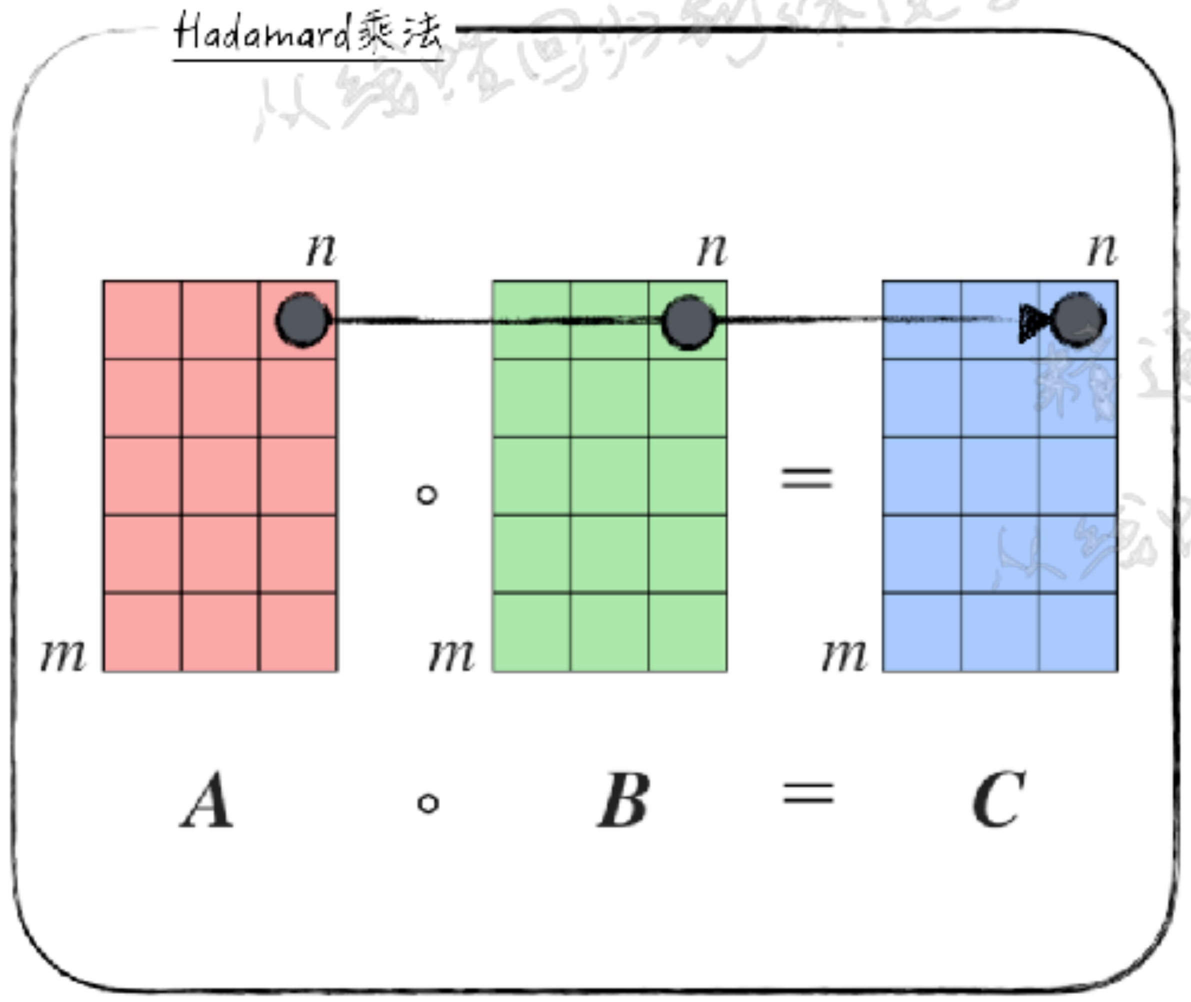
矩阵乘法满足
结合率和分配律

$$(AB)C = A(BC)$$

$$A(B + C) = AB + AC$$

矩阵的Hadamard乘法要求
两个矩阵的形状一样

$$A \circ B = (a_{i,j} b_{i,j}) \in \mathbf{R}^{n \times m}$$



矩阵乘法可以极大地
简化模型的表示和计算

$$\begin{cases} y_1 = a x_1 + b \\ y_2 = a x_2 + b \end{cases}$$

令$X = \begin{pmatrix} x_1 & 1 \\ x_2 & 1 \end{pmatrix}$, $\beta = \begin{pmatrix} a \\ b \end{pmatrix}$, $Y = \begin{pmatrix} y_1 \\ y_2 \end{pmatrix}$

$$Y = X\beta$$

矩阵的运算

逆矩阵和转置

逆矩阵的定义只针对n x n的方阵

如果对于方阵M，存在方阵N满足

$$\boldsymbol{MN} = \boldsymbol{NM} = \boldsymbol{I}_n$$

则称N为M的逆矩阵，记为

$$\boldsymbol{N} = \boldsymbol{M}^{-1}$$

这是课后习题

容易证明，逆矩阵如果存在，则逆矩唯一。并且有如下的性质：

$$(\boldsymbol{M}^{-1})^{-1} = \boldsymbol{M}$$

$$(k\boldsymbol{M})^{-1} = \frac{1}{k}\boldsymbol{M}^{-1}$$

$$(\boldsymbol{MN})^{-1} = \boldsymbol{N}^{-1}\boldsymbol{M}^{-1}$$

对于一个n x m的矩阵，它的转置是一个m x n的矩阵

具体的定义十分简单

$$\boldsymbol{X} = (x_{i,j}) \in \mathbf{R}^{n \times m}$$

$$\boldsymbol{X}^{\mathrm{T}} = (x_{j,i}) \in \mathbf{R}^{m \times n}$$

容易证明

$$(\boldsymbol{X}^{\mathrm{T}})^{\mathrm{T}} = \boldsymbol{X}$$

$$(\boldsymbol{X} + \boldsymbol{Y})^{\mathrm{T}} = \boldsymbol{X}^{\mathrm{T}} + \boldsymbol{Y}^{\mathrm{T}}$$

$$(k\boldsymbol{X})^{\mathrm{T}} = k\boldsymbol{X}^{\mathrm{T}}$$

$$(\boldsymbol{XY})^{\mathrm{T}} = \boldsymbol{Y}^{\mathrm{T}}\boldsymbol{X}^{\mathrm{T}}$$

$$(\boldsymbol{X}^{\mathrm{T}})^{-1} = (\boldsymbol{X}^{-1})^{\mathrm{T}}$$

定义对称矩阵：

$$X^T = X$$

目录

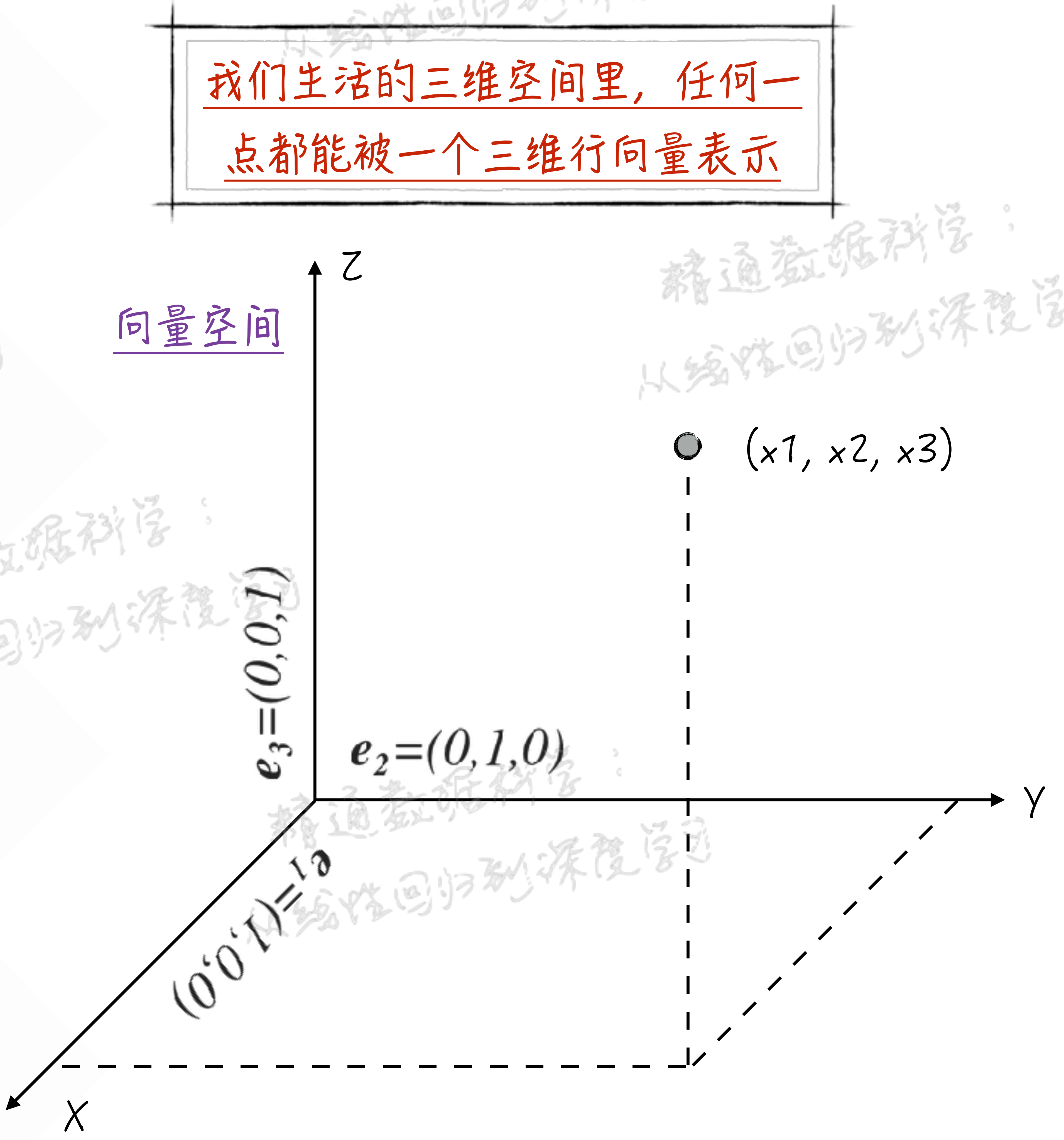
向量空间

内积定义

抛开严谨的数学定义，向量空间可以认为是在线性变换下保持封闭的空间，比如我们生活的三维空间

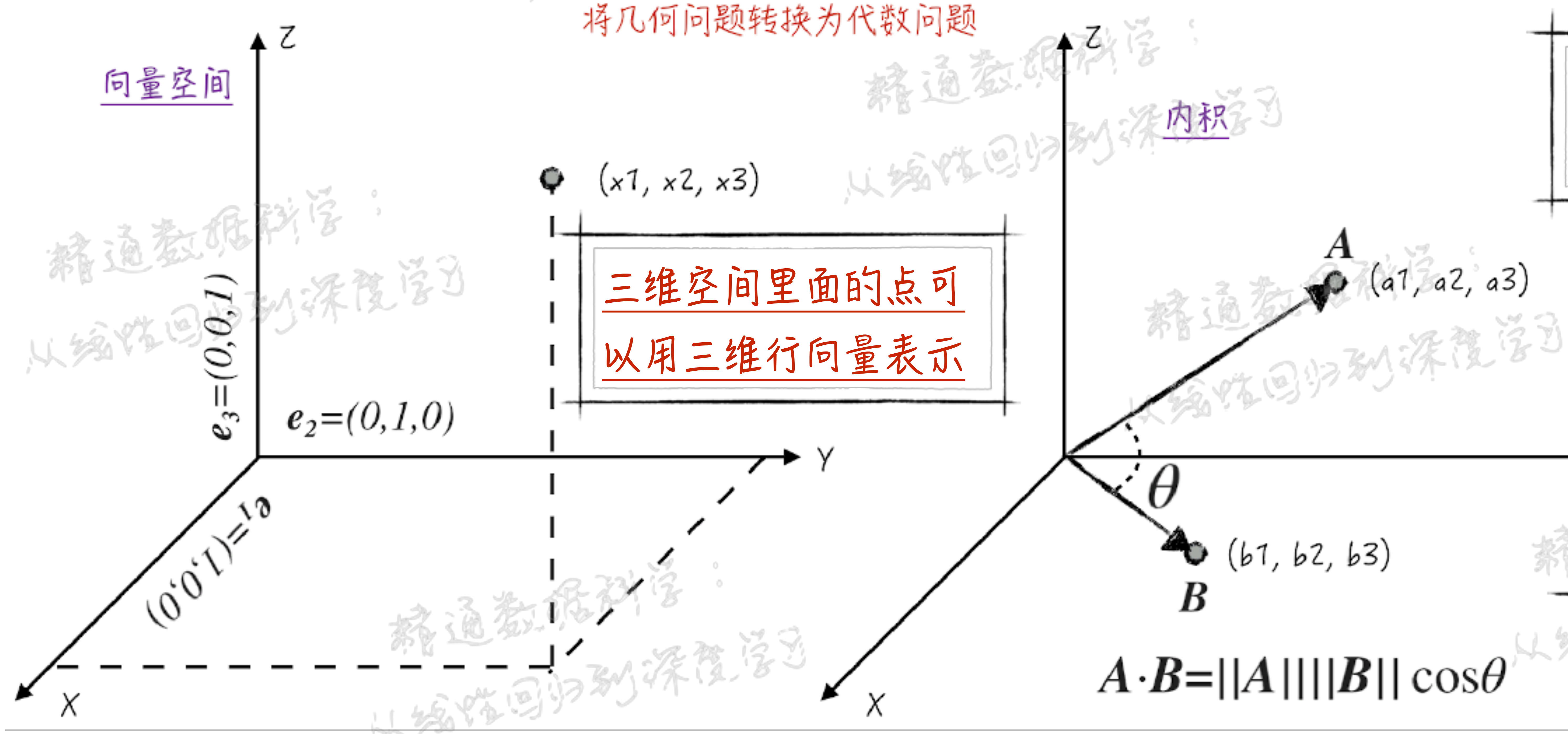
任意一个三维行向量都能表示成 $e_1=(1,0,0)$、$e_2=(0,1,0)$ 和 $e_3=(0,0,1)$ 的线性组合

- 这三个向量在一起组成了三维行向量空间的一组**正交基**
- 任意三个相互正交的三维行向量都是三维行向量空间的一组**正交基**
- 这个结论可以推广到n维行向量空间



向量空间

内积定义



内积等于零 → 两向量正交

定义两个向量的内积为

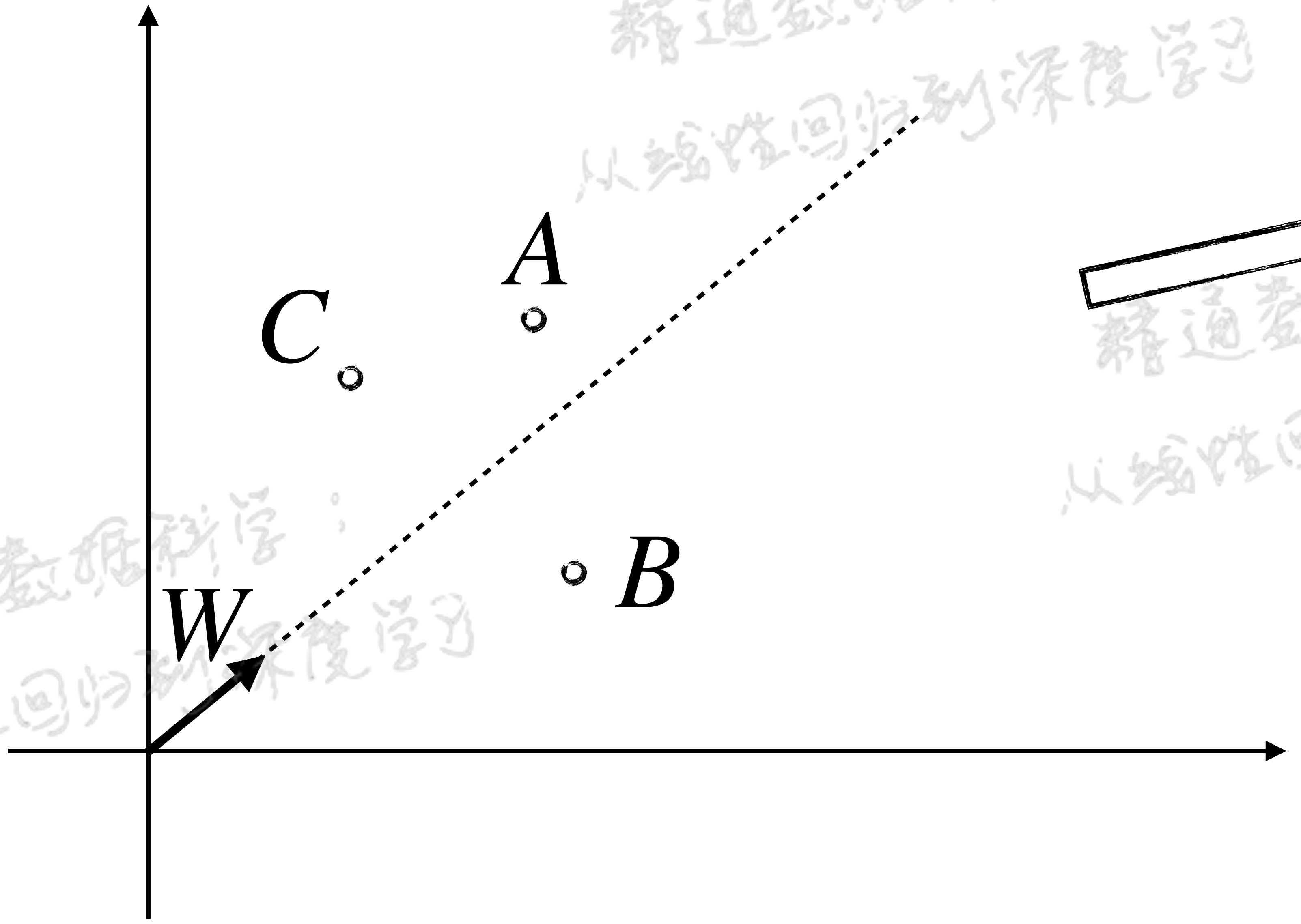
$$A \cdot B = a_1 b_1 + a_2 b_2 + a_3 b_3$$

$$A \cdot B = AB^{\mathrm{T}}$$

如果B的长度为1 → A和B的内积等于A在B上的投影长度

向量空间

特征向量和特征值



$$X=\begin{pmatrix} A \\ B \\ C \end{pmatrix} \xrightarrow{\text{A、B、C到直线W的投影平方和}} ||Xw^T||^2$$

$$\boldsymbol{v} = \text{argmax}_{\|\boldsymbol{w}\|=1} \|\boldsymbol{X}\boldsymbol{w}^{\mathrm{T}}\|^2 = \text{argmax}_{\|\boldsymbol{w}\|=1} \boldsymbol{w}\boldsymbol{X}^{\mathrm{T}}\boldsymbol{X}\boldsymbol{w}^{\mathrm{T}}$$

现在假设二维空间中有一些点，希望找到一条直线，使得这些点到这条直线的距离平方和最小

- 距离平方和最小意味着**投影平方和最大**

注意到 $K = X^T X$ 是一个对称矩阵，为了解决上面这个问题，我们需要引入特征向量（eigenvector）和特征值（eigenvalue）这两个概念

向量空间

特征向量和特征值

对于一个矩阵M，如果下面的公式成立，
则特征向量和特征值定义如下：

特征值

$$\boldsymbol{M}\boldsymbol{w}^{\mathrm{T}} = \lambda \boldsymbol{w}^{\mathrm{T}}$$

特征向量

w^T表示一个列向量

对于n阶的对称矩阵，存在n个相互正交的特征矩阵。

矩阵 $K = X^T X$ 最大特征值对应的特征向量就是我们要找的v

$$\boldsymbol{v} = \operatorname{argmax}_{\|\boldsymbol{w}\| = 1} \|\boldsymbol{X}\boldsymbol{w}^{\mathrm{T}}\|^2 = \operatorname{argmax}_{\|\boldsymbol{w}\| = 1} \boldsymbol{w}\boldsymbol{X}^{\mathrm{T}}\boldsymbol{X}\boldsymbol{w}^{\mathrm{T}}$$

最大特征值对应的特征向量

这是课后习题

THANK YOU