机器学习与深度学习

——线性代数入门

Personal Website: https://www.miaopeng.info/
Email: miaopeng@stu.scu.edu.cn
Github: https://github.com/MMeowhite
Youtube: https://www.youtube.com/@pengmiao-bmm

目录章节

CONTENTS

▶ 向量与矩阵：引言

- 假设我们有如下的线性方程组：

$$
\begin{aligned}
1x + 2y + 3z &= 10 \\
2x + 3y + 4z &= 3 \\
3x - 6y - 2z &= 5
\end{aligned}
$$

- 引入向量与矩阵的概念之后，可以写成如下的形式：

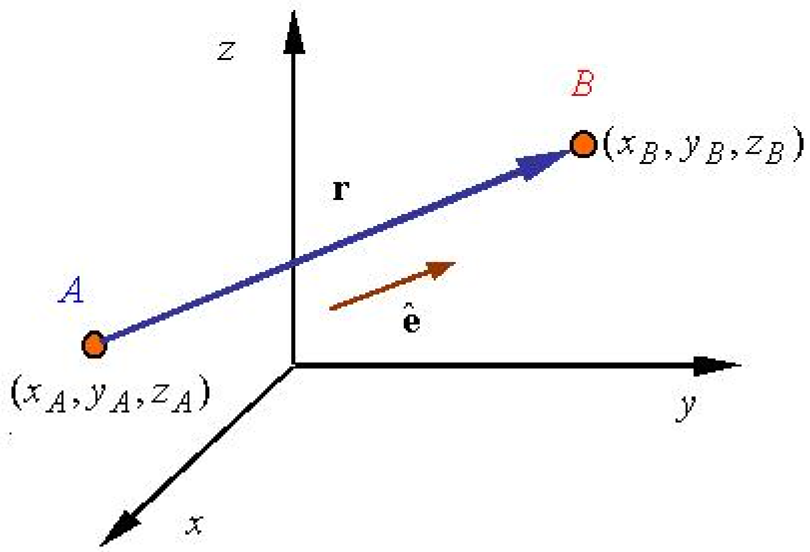
$$
\begin{pmatrix} 1 & 2 & 3 \\ 2 & 3 & 4 \\ 3 & -6 & -2 \end{pmatrix} \cdot \begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} 10 \\ 3 \\ 5 \end{pmatrix}
$$

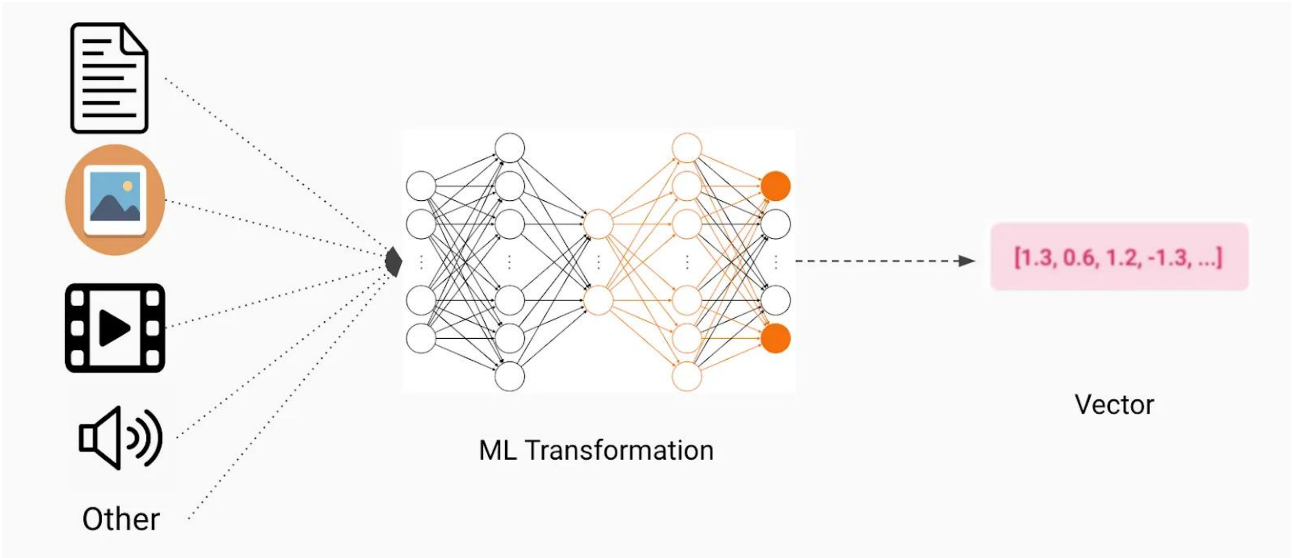
- 如果系数矩阵 可逆的（即存在逆矩阵，相当于“分母不为零”的情况），那么线性方程组的解可以写为：

$$
\begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} 1 & 2 & 3 \\ 2 & 3 & 4 \\ 3 & -6 & -2 \end{pmatrix}^{-1} \cdot \begin{pmatrix} 10 \\ 3 \\ 5 \end{pmatrix}
$$

▶ 什么是向量？

- ➢ 向量是**有大小和方向的量**，可以理解为一个“箭头”或一组有序数值（坐标）。
- ➢ 向量可以表示空间中的点或方向，也可以表示数据特征（如图像像素、词向量）。





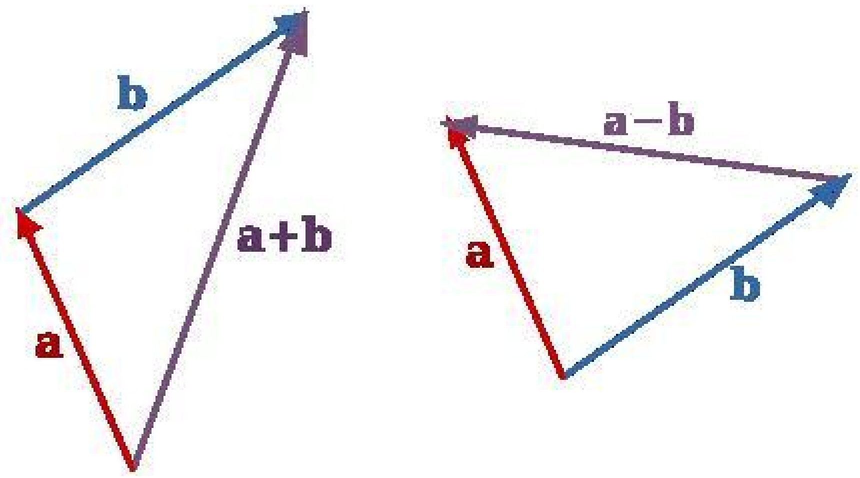
向量的核心意义在于运算：它可以相加、缩放、点积/叉积，用来描述空间关系和变化。

▶ 向量的运算：加法、减法

➢ 定义：对应分量的相加/相减【**注意维度一定要相同**】：

$$\vec{a}+\vec{b}=(a_1+a_2, b_1+b_2, \cdots)$$

➢ 几何意义：平移 & 平行四边形法则（两个箭头拼成一个箭头）。



$$\vec{a}=\begin{pmatrix}1\\2\\3\end{pmatrix},\quad \vec{b}=\begin{pmatrix}4\\-1\\2\end{pmatrix}$$

$$\vec{a}+\vec{b}=\begin{pmatrix}1+4\\2+(-1)\\3+2\end{pmatrix}=\begin{pmatrix}5\\1\\5\end{pmatrix}\qquad \vec{a}-\vec{b}=\begin{pmatrix}1-4\\2-(-1)\\3-2\end{pmatrix}=\begin{pmatrix}-2\\3\\1\end{pmatrix}$$

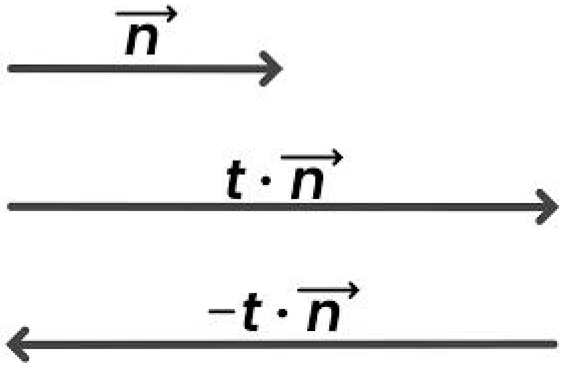
应用：词向量运算（king - man + woman ≈ queen）；特征组合（将不同特征向量合并）。

▶ 向量的运算：数乘（标量乘法）

➢ 定义：向量与常数k相乘，改变长度

$$k\vec{a} = (ka_1, ka_2, \cdots)$$

➢ 几何意义：同方向伸缩或反向翻转（k<0 时方向相反）。



$$\vec{a} = \begin{pmatrix} 3 \\ 3 \\ 3 \end{pmatrix}$$

$$2 \cdot \vec{a} = 2 \cdot \begin{pmatrix} 3 \\ 3 \\ 3 \end{pmatrix} = \begin{pmatrix} 3 \cdot 2 \\ 3 \cdot 2 \\ 3 \cdot 2 \end{pmatrix} = \begin{pmatrix} 6 \\ 6 \\ 6 \end{pmatrix}$$

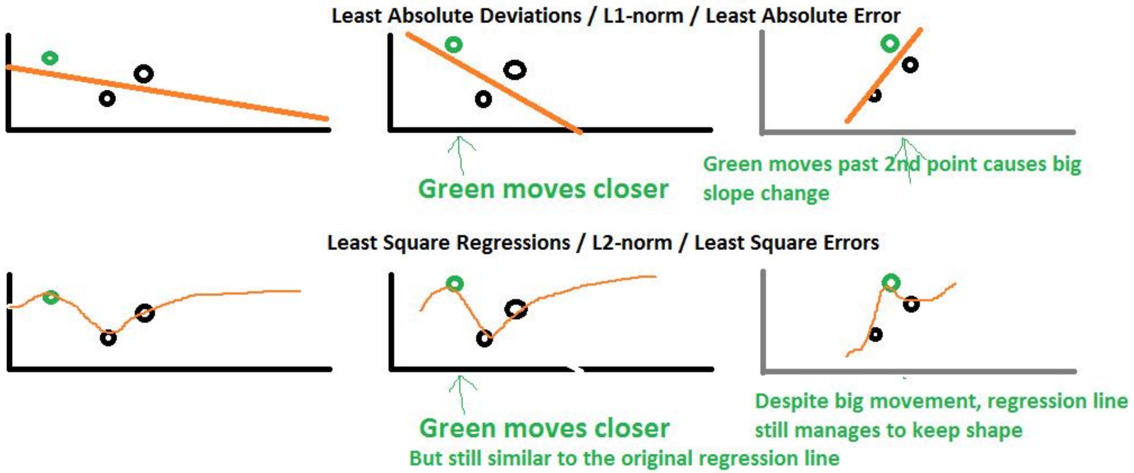
应用：学习率调整：梯度方向不变，但步长（长度）改变。

▶ 向量的运算：范数（长度）

- 定义：L2范数【最常见的“欧几里得长度”】与L1范数【稀疏化常用】

$$||x||_2 = \sqrt{\sum x_i^2}, \quad ||x||_1 = \sum |x_i|$$

- 几何意义：衡量向量“有多大”。



$$\vec{a} = \begin{pmatrix} 3 \\ 3 \\ 3 \end{pmatrix}$$

$$||\vec{a}||_2 = \sqrt{1^2 + 2^2 + 3^2} = \sqrt{14}, \quad ||\vec{a}||_1 = |1 + 2 + 3| = 6$$

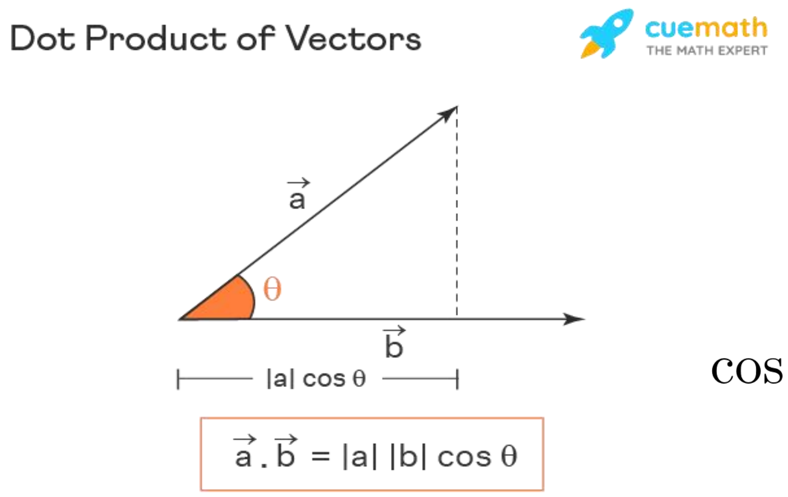
应用：正则化（L1: Lasso，L2: Ridge）；向量归一化（特征标准化）。

▶ 向量的运算：点积（内积）

➢ 定义：两个同维向量对应元素相乘后再求和，结果是一个标量：

$$\vec{a} \cdot \vec{b} = \sum (a_i b_i)$$

➢ 几何意义：｜a｜｜b｜cosθ，**衡量方向相似性**。直观表示**向量投影的长度乘以另一向量的长度。**



$$\vec{a} = \begin{pmatrix} 1 \\ 2 \\ 2 \end{pmatrix}, \quad \vec{b} = \begin{pmatrix} 2 \\ 4 \\ 4 \end{pmatrix}$$

$$\cos\theta = \frac{\vec{a} \cdot \vec{b}}{||a||||b||} = \frac{1 \cdot 2 + 2 \cdot 4 + 2 \cdot 4}{\sqrt{1^2 + 2^2 + 2^2} \cdot \sqrt{2^2 + 4^2 + 4^2}} = \frac{18}{18} = 1 \rightarrow \theta = \frac{\pi}{2}$$

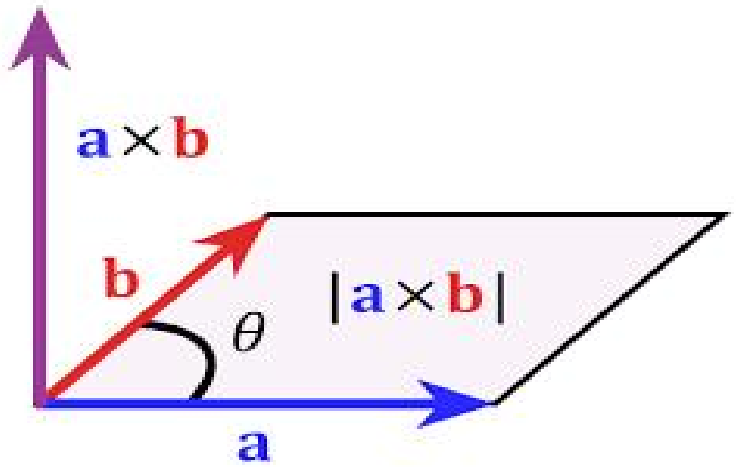
应用：相似度计算（如余弦相似度）；神经元加权求和（权重向量 · 输入向量）。

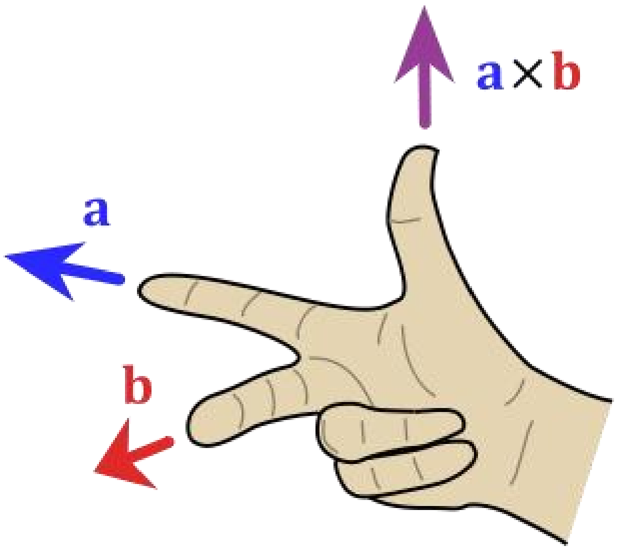
▶ 向量的运算：叉乘

➢ 定义：（a×b）定义在三维空间中，结果是一个垂直于两向量平面的向量。

$$\vec{a} \times \vec{b} = \begin{vmatrix} \mathbf{i} & \mathbf{j} & \mathbf{k} \\ a_1 & a_2 & a_3 \\ b_1 & b_2 & b_3 \end{vmatrix} = \begin{pmatrix} a_2 b_3 - a_3 b_2 \\ a_3 b_1 - a_1 b_3 \\ a_1 b_2 - a_2 b_1 \end{pmatrix}$$

➢ 大小：等于两向量夹角的正弦乘积与两向量长度的积：｜a x b｜=｜a｜｜b｜sinθ；方向：右手定则确定。



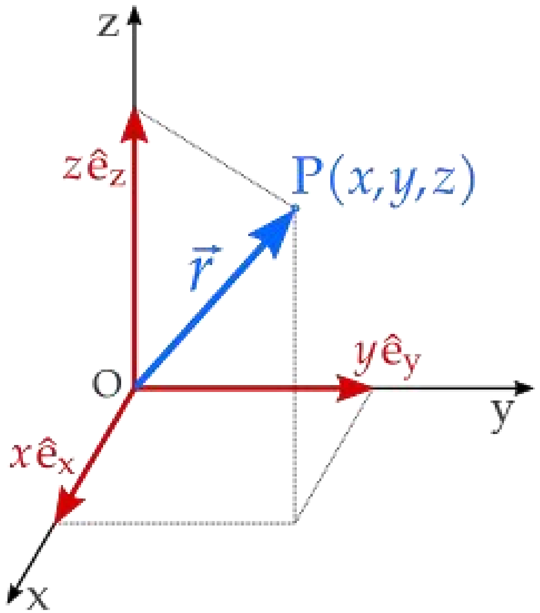


应用：法向量计算；几何特征提取；旋转与坐标变换等。

▶ 向量的几何意义

➢ 向量可以看作从原点出发指向某个点的箭头，既有方向又有大小。

➢ 向量既可以表示空间中一个点的位置【如果把原点 (0,0,0) 当作参考系，一个向量 (x,y,z) 就能精确地描述一个点在三维空间中的位置】，也可以表示移动或作用的方向【如果我们只关心从 A 到 B 的“变化”，那向量就表示一种方向和长度，不依赖于起点在哪】。



- 如果x = 3, y = 2, z = 1，点(3, 2, 1) 就是从原点走 3 个单位到 x 方向，再走 2 个单位到 y 方向，最终再走1个单位到z方向即可到达那个点。
- 向量(3,2,1)可以看作“沿 x 方向走 3，沿 y 方向走 2，沿 z 方向走 1”的移动，这在平移、速度、力等场景下非常常见。

改变向量的长度（缩放）或方向（旋转），就是在做线性变换。

▶ 什么是矩阵？

- 矩阵是一个**按行和列排列的数字表格**，可以看作**多个向量的组合**。
- 矩阵可以表示**数据**（如一批样本的特征）、也可以表示**线性变换（如旋转、缩放、投影）**。

$$A = \begin{array}{cc} & \begin{array}{ccc} & \text{列} & \end{array} \\ \text{行} & \begin{pmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 9 \end{pmatrix} \end{array}$$

$$\vec{v} = \begin{pmatrix} x \\ y \end{pmatrix}, \vec{R} = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}$$

$$\vec{v}_{\text{Rotation}} = \vec{R} \cdot \vec{v} = \begin{pmatrix} y \\ -x \end{pmatrix}$$

向量的核心意义在于运算：它可以相加、缩放、点积/叉积，用来描述空间关系和变化。

▶ 特殊矩阵：单位矩阵、对角矩阵、稀疏矩阵

- 单位矩阵I是对角线上全是1，其余元素全是0的方阵，相当于矩阵乘法中的“1”。
- 对角矩阵D只有对角线上的元素可能非零，其它位置均为0，简化了计算和存储。
- 稀疏矩阵S大部分元素为0，专门设计的存储和算法能高效处理这种矩阵。

$$I_3 = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

单位矩阵I

$$D = \begin{pmatrix} 4 & 0 & 0 \\ 0 & 5 & 0 \\ 0 & 0 & 6 \end{pmatrix}$$

对角矩阵D

$$S = \begin{pmatrix} 0 & 0 & 7 \\ 0 & 0 & 0 \\ 3 & 0 & 0 \end{pmatrix}$$

稀疏矩阵S

特殊矩阵如单位矩阵、对角矩阵和稀疏矩阵，分别在保持元素不变、简化计算以及提高存储和运算效率中发挥关键作用。

▶ 矩阵的运算：加法、减法

➢ 矩阵加法或减法就是把两个同型矩阵对应位置的元素逐一相加或相减，结果仍是同型矩阵。

➢ 注意：只有维度（行数、列数）完全相同的矩阵才能相加或相减，否则运算无意义。

$$A = \begin{pmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{pmatrix}$$

$$B = \begin{pmatrix} b_{11} & b_{12} & b_{13} \\ b_{21} & b_{22} & b_{23} \\ b_{31} & b_{32} & b_{33} \end{pmatrix}$$

$$C = A + B = \begin{pmatrix} a_{11}+b_{11} & a_{12}+b_{12} & a_{13}+b_{13} \\ a_{21}+b_{21} & a_{22}+b_{22} & a_{23}+b_{23} \\ a_{31}+b_{31} & a_{32}+b_{32} & a_{33}+b_{33} \end{pmatrix}$$

矩阵加减法是对同型矩阵逐元素相加或相减，结果仍为同型矩阵。

▶ 矩阵的运算：数乘

- 矩阵数乘是把矩阵中每个元素同时乘以同一个标量，结果仍是同型矩阵。
- 数乘可以看作对矩阵表示的向量或线性变换进行整体缩放，标量为负数时还会改变方向，在几何上体现为翻转加缩放。

$$A = \begin{pmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{pmatrix} \qquad A' = k \cdot A = \begin{pmatrix} ka_{11} & ka_{12} & ka_{13} \\ ka_{21} & ka_{22} & ka_{23} \\ ka_{31} & ka_{32} & ka_{33} \end{pmatrix}$$

矩阵数乘是对矩阵中所有元素进行相同的缩放运算，从而整体放大或缩小矩阵所表示的向量或变换，但不改变其形态与结构。

▶ 矩阵的运算：矩阵乘法

- 矩阵乘法通过**行向量与列向量的点积**得到结果元素，**要求前一个矩阵的列数等于后一个矩阵的行数**。
- **乘法不可交换，但可以结合和分配，用于复合线性变换（如旋转后再缩放）。**

$$C = A \times B, \quad C_{ij} = \sum_{k=1}^{n} A_{ik} B_{kj}$$

$$A = \begin{pmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{pmatrix} \qquad B = \begin{pmatrix} b_{11} & b_{12} & b_{13} \\ b_{21} & b_{22} & b_{23} \\ b_{31} & b_{32} & b_{33} \end{pmatrix}$$

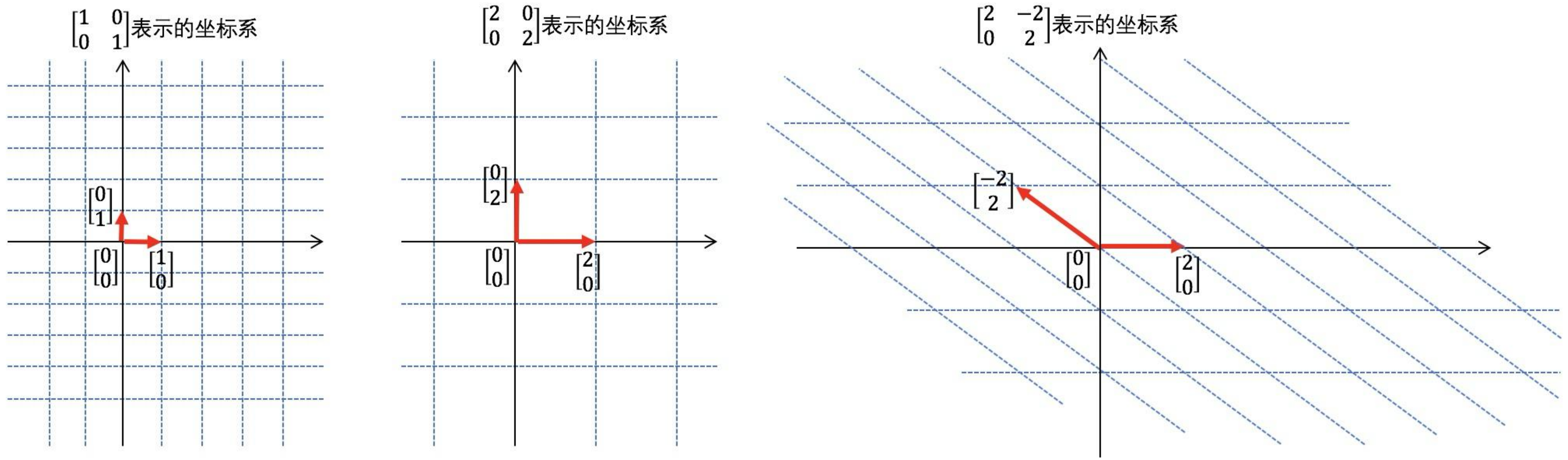
$$C = A \times B = \begin{pmatrix} a_{11}b_{11} + a_{12}b_{21} + a_{13}b_{31} & a_{11}b_{12} + a_{12}b_{22} + a_{13}b_{32} & a_{11}b_{13} + a_{12}b_{23} + a_{13}b_{33} \\ a_{21}b_{11} + a_{22}b_{21} + a_{23}b_{31} & a_{21}b_{12} + a_{22}b_{22} + a_{23}b_{32} & a_{21}b_{13} + a_{22}b_{23} + a_{23}b_{33} \\ a_{31}b_{11} + a_{32}b_{21} + a_{33}b_{31} & a_{31}b_{12} + a_{32}b_{22} + a_{33}b_{32} & a_{31}b_{13} + a_{32}b_{23} + a_{33}b_{33} \end{pmatrix}.$$

矩阵乘法在几何上表示连续线性变换的组合，是理解矩阵意义的核心。

▶ 矩阵的几何意义

➢ 矩阵可以看作一种线性变换，把空间中的向量映射到另一个向量，常表现为**旋转**、**缩放**、**剪切或投影**等几何操作，即**矩阵是对空间的变换**。

$$\begin{pmatrix}0 & 1\\ 1 & 0\end{pmatrix}\cdot 2 \rightarrow \begin{pmatrix}2 & 0\\ 0 & 2\end{pmatrix},\quad \begin{pmatrix}2 & 0\\ 0 & 2\end{pmatrix}\cdot\begin{pmatrix}0 & 2\\ 2 & -2\end{pmatrix} \rightarrow \begin{pmatrix}2 & -2\\ 0 & -2\end{pmatrix}.$$



改变向量的长度（缩放）或方向（旋转），就是在做线性变换。

▶ 矩阵的运算：转置

- 矩阵转置是将**矩阵的行与列互换**，记作A^T。
- 转置不会改变矩阵的维度类型（行列交换），但会改变元素的位置。

$$B = A^{\mathrm{T}}, \quad B_{ij} = A_{ji}$$

$$A = \begin{pmatrix} 2 & 3 & 6 \\ 4 & 6 & 1 \\ 9 & 8 & 7 \end{pmatrix}, \quad A^T = \begin{pmatrix} 2 & 4 & 9 \\ 3 & 6 & 8 \\ 6 & 1 & 7 \end{pmatrix}$$

在几何上，转置常用于将行向量与列向量互换，并在内积、对称矩阵等运算中广泛应用。

▶ 什么是线性变换？

- 定义：线性变换是**满足加法与数乘保持性的映射** T(x)=Ax【一个向量空间到另一个向量空间的映射，且保持加法运算和数量乘法运算】。矩阵 A 就是实现这种变换的“工具”。
- 可以将线性变换理解为一个特殊的函数，这个函数可以满足以下的性质：

$$T(x+y) = T(x) + T(y)$$

$$T(ax) = aT(x)$$

- 例如，函数f为线性变换，函数g和h就不为线性变换：

$$f(x,y,z) = (3x-y, 3z, 0, z-2x)$$

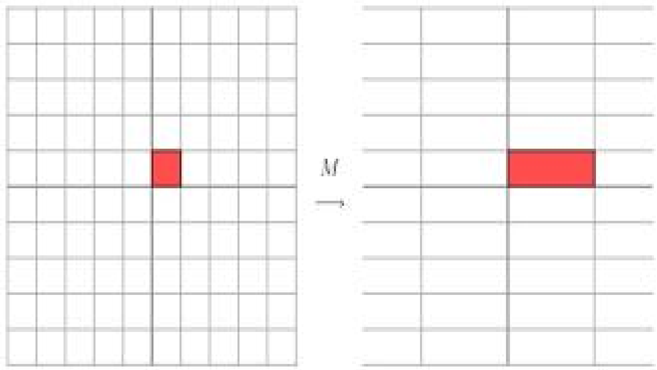
$$g(x,y,z) = (3x-y, 3z+2, 0, z-2x)$$

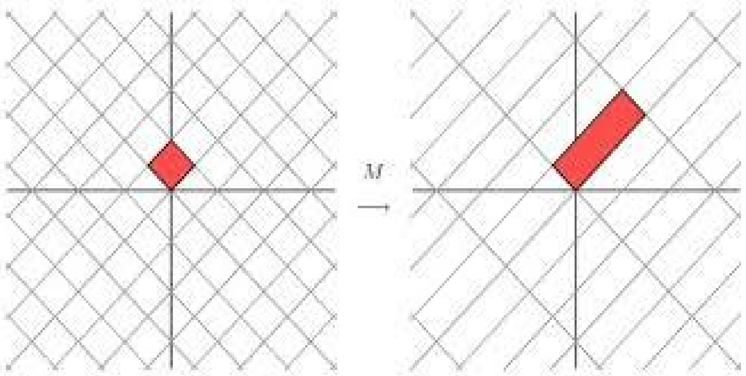
$$h(x,y,z) = (3x-y, 3xz, 0, z-2x)$$

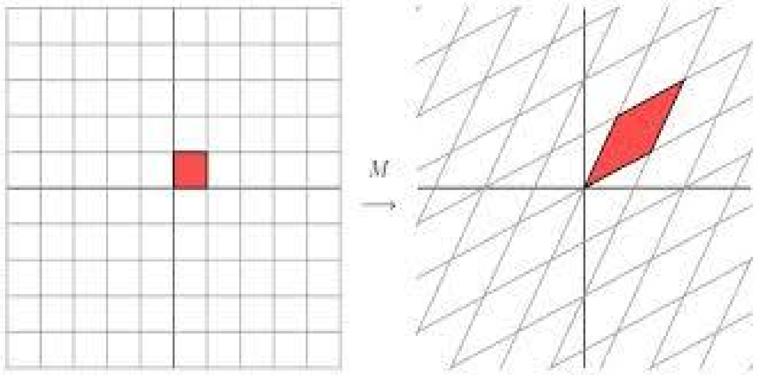
改变向量的长度（缩放）或方向（旋转），就是在做线性变换。

▶ 线性变换的几何意义

- 线性变换通过矩阵作用于向量，改变向量的大小和方向，体现空间的形状变化。
- 它们可以表示旋转、缩放、剪切、反射和投影等基本几何操作。



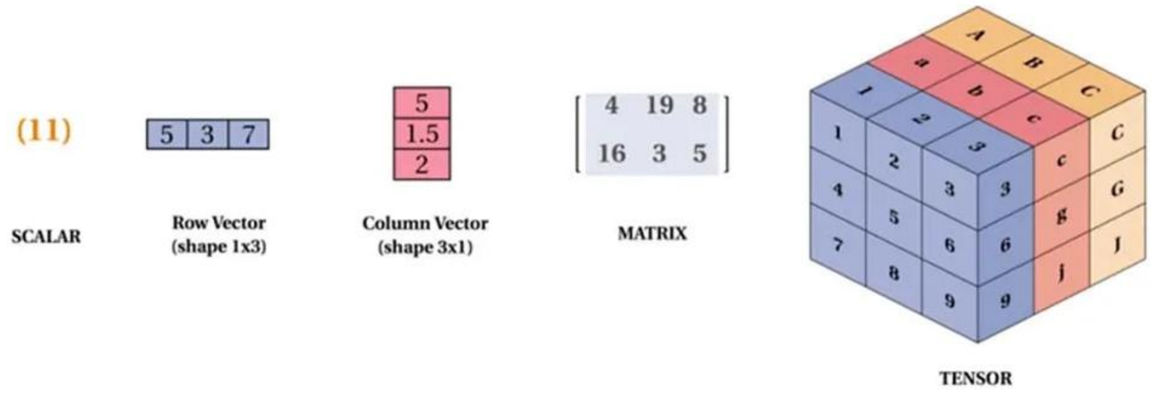




线性变换保持原点不变，并且对向量的加法和数乘运算保持一致性，保证结构的线性特性。

▶ 小结

- 向量：表示空间中的点或方向，常用于描述特征、样本或梯度；其运算包括加减（平移）、数乘（缩放）、点积（投影与相似性）和叉乘（面积与法向量）。
- 矩阵：是按行列排列的数值表，可表示线性变换或数据集合；其运算包括加减（同维度元素相加减）、数乘（整体缩放）、乘法（复合变换）和转置（行列互换）。
- 两者联系：1）向量是矩阵的特例：列向量可看作 n×1 的矩阵，行向量是 1×n 的矩阵；2）矩阵由向量构成：矩阵可以看作一组向量的集合，按行或列排列表示数据或基底；3）运算联系：**矩阵乘以向量，等价于对该向量进行线性变换**，这是机器学习中最常见的操作（如特征变换、权重作用）。



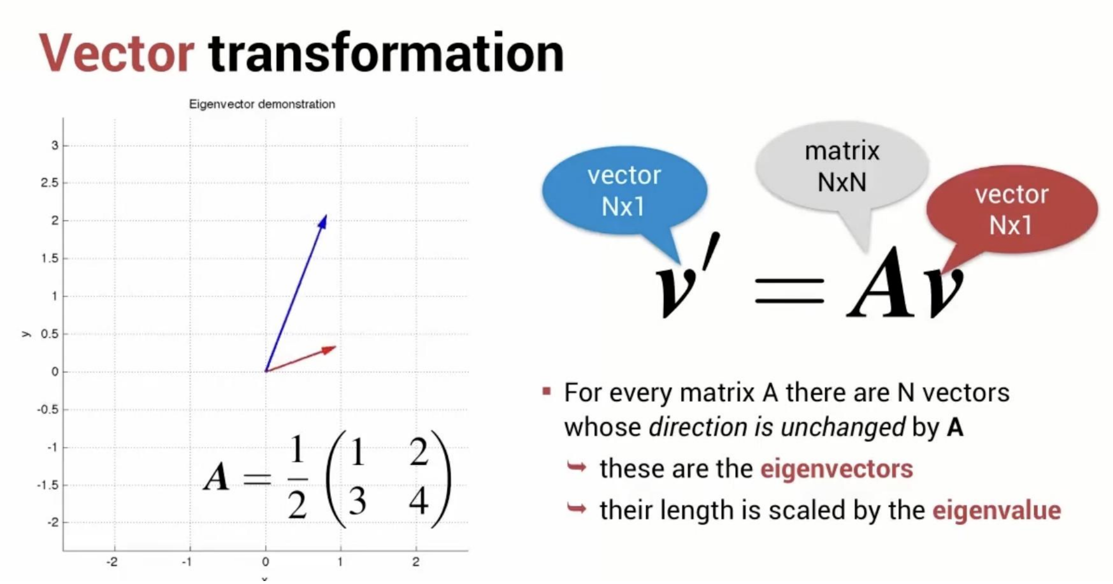
向量和矩阵运算在机器学习中支撑特征表示、数据变换、梯度计算和模型参数更新，是理解算法的核心工具。

目录章节

CONTENTS

▶ 引言

- 大多数矩阵作用在向量上会改变方向和长度，**但有一些特殊方向的向量，矩阵只会改变它们的长度（放大或缩小），方向不变。**
- 这些**方向就是特征向量，它被放大或缩小的倍数就是特征值**。它们共同揭示了矩阵在线性变换的**内在特征**。



- 方向不变是特征向量的核心意义：1）揭示线性变换的本质：任意矩阵可以看作“旋转 + 拉伸 + 投影”的组合，大部分向量在变换后都会扭曲方向，特征向量就是这个变换的“稳定方向”，它们描述了矩阵最核心的行为；2）降维与最优方向：在 PCA 中，特征向量指示数据变化最大的方向；选前几个特征向量就能捕捉主要信息。

矩阵像一台“变形机”，特征向量是“不会被扭曲方向”的特殊杆子，特征值就是它被拉伸或压缩的倍数。

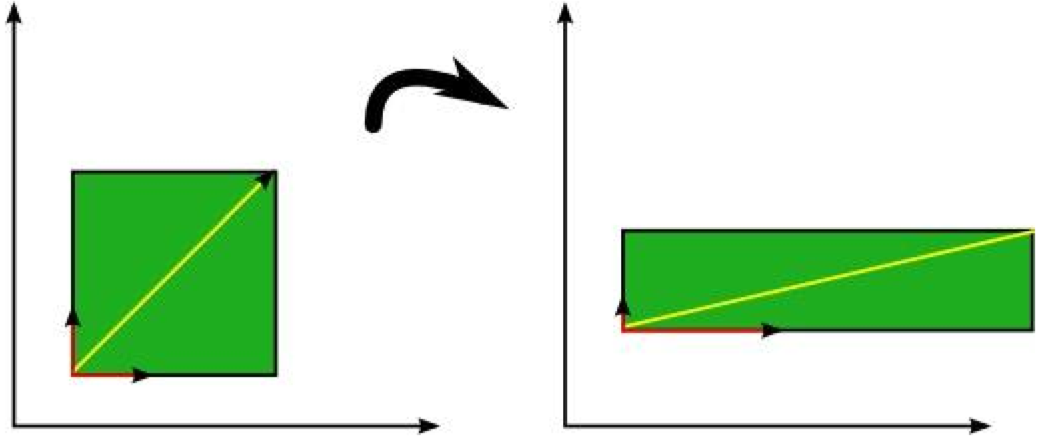
参考视频：https://robotacademy.net.au/lesson/eigenvalues-and-eigenvectors-recap/

▶ 什么是特征值和特征向量？

- ➢ 一个矩阵的特征向量v和特征值λ满足：

$$A\vec{v} = \lambda\vec{v}, \quad \vec{v} = \vec{0}$$

- ➢ 这些方向就是特征向量，它被放大或缩小的倍数就是特征值。**【注意A一定是nxn的方阵】**

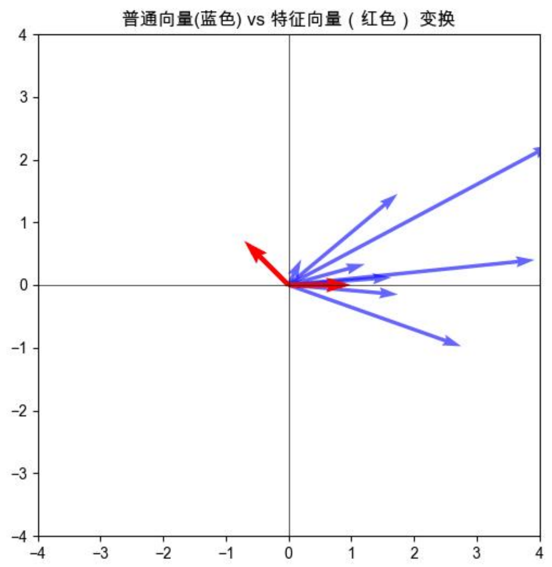


- 特征向量（红色）在施加线性变换（例如缩放）时方向不变，而其他向量（黄色）的方向会改变。

矩阵像一台“变形机”，特征向量是“不会被扭曲方向”的特殊杆子，特征值就是它被拉伸或压缩的倍数。

▶ 特征值和特征向量的几何意义

- 特征向量是**矩阵作用下方向保持不变的向量**（只会被拉伸或压缩，不会被扭曲或旋转）。
- 特征值表示该特征向量**被拉伸或压缩的倍数**，是矩阵在这一方向上的“伸缩因子”。



- 例如，普通向量（蓝色）在多次施加矩阵变换后后不断改变方向，而特征向量（红色）始终保持方向不变，仅被拉伸或压缩。

矩阵像一台“变形机”，特征向量是“不会被扭曲方向”的特殊杆子，特征值就是它被拉伸或压缩的倍数。

▶ 特征值和特征向量的计算

➢ （1）写出特征方程：

$$A\vec{v} = \lambda\vec{v} \rightarrow \det(A - \lambda I) = 0$$

- 这是一个关于λ的多项式方程，解它即可得到所有特征值λi。【I是与A相同维度的单位矩阵】

➢ （2）求特征向量：

- 对每个特征值λi，将其带入：

$$(A - \lambda_i I)\vec{v} = 0$$

- 解这个线性方程组，得到对应的特征向量v。

计算特征值和特征向量，就是先找出矩阵在某些特殊方向上只发生伸缩而不改变方向的比例（特征值），再确定这些对应保持方向不变的特殊向量（特征向量）。

▶ 特征值和特征向量的计算案例

➢ 设矩阵：

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

➢ 步骤1：求特征值

- 写出特征方程：

$$\det(A - \lambda I) = \begin{bmatrix} 4-\lambda & 2 \\ 1 & 3-\lambda \end{bmatrix} = (4-\lambda)(3-\lambda) - 2 = \lambda^2 - 7\lambda + 10 = 0$$

- 解得：

$$\lambda_1 = 5, \lambda_2 = 2$$

➢ 步骤2：求特征向量

- 讲两个特征值打入方程中：

$$(A - 5I)v = 0 \rightarrow \begin{pmatrix} 4-5 & 2-0 \\ 1-0 & 3-5 \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} = 0 \rightarrow \begin{pmatrix} -1 & 2 \\ 1 & -2 \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} = 0 \rightarrow v = \begin{pmatrix} 2 \\ 1 \end{pmatrix}$$

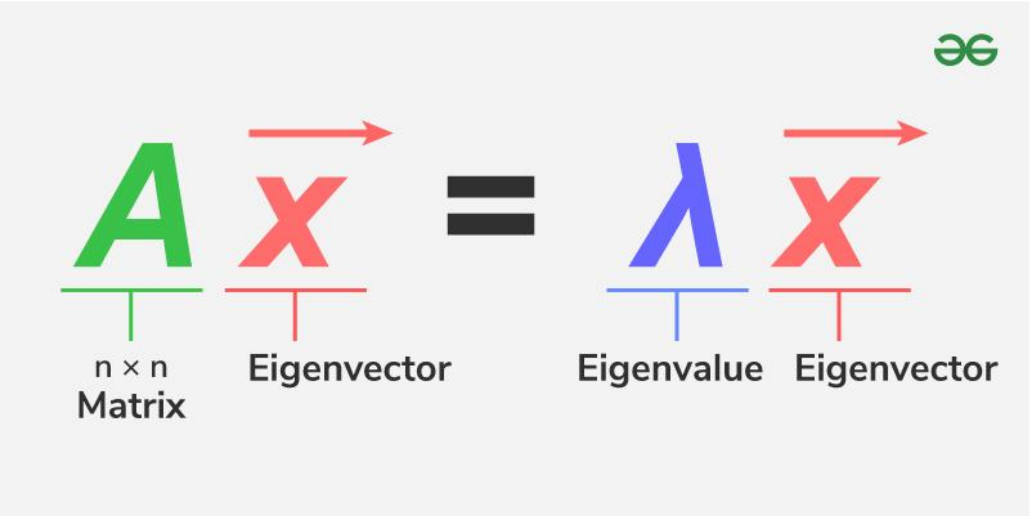
$$(A - 2I)v = 0 \rightarrow \begin{pmatrix} 4-2 & 2-0 \\ 1-0 & 3-2 \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} = 0 \rightarrow \begin{pmatrix} 2 & 2 \\ 1 & 1 \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} = 0 \rightarrow v = \begin{pmatrix} -1 \\ 1 \end{pmatrix}$$

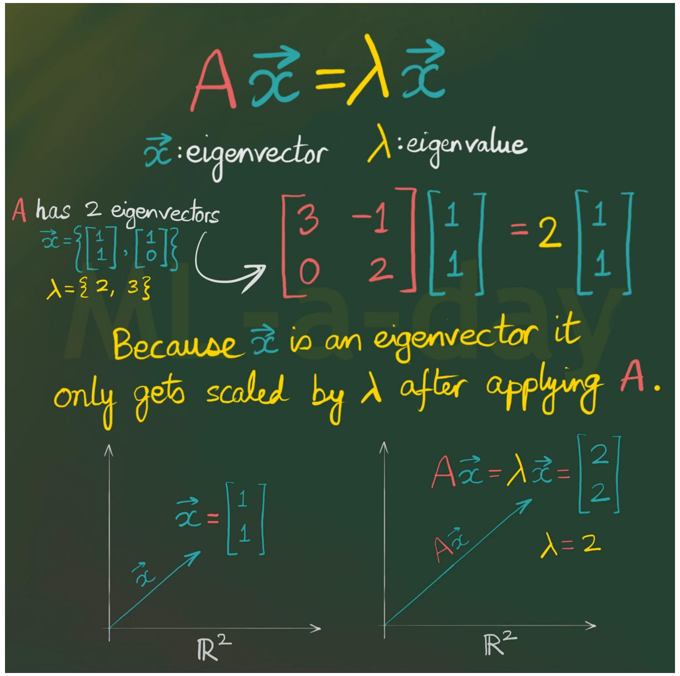
▶ 小结

➢ 特征值描述的是一个线性变换在某些特殊方向上对向量的伸缩比例，而特征向量是指在该变换作用下方向保持不变的非零向量。

➢ 它们通过解方程 $(A-\lambda I)v=0$ 获得，其中 λ 是特征值，v 是对应的特征向量，且解出的向量通常需要归一化以便比较和应用。

➢ 严格来说，特征值和特征向量的经典定义只适用于方阵【因为I必须和A同型，才能做减法】：**对于非方阵(m×n），无法直接定义特征值和特征向量，但可以通过奇异值分解（SVD）得到类似的概念（奇异值与奇异向量）。**



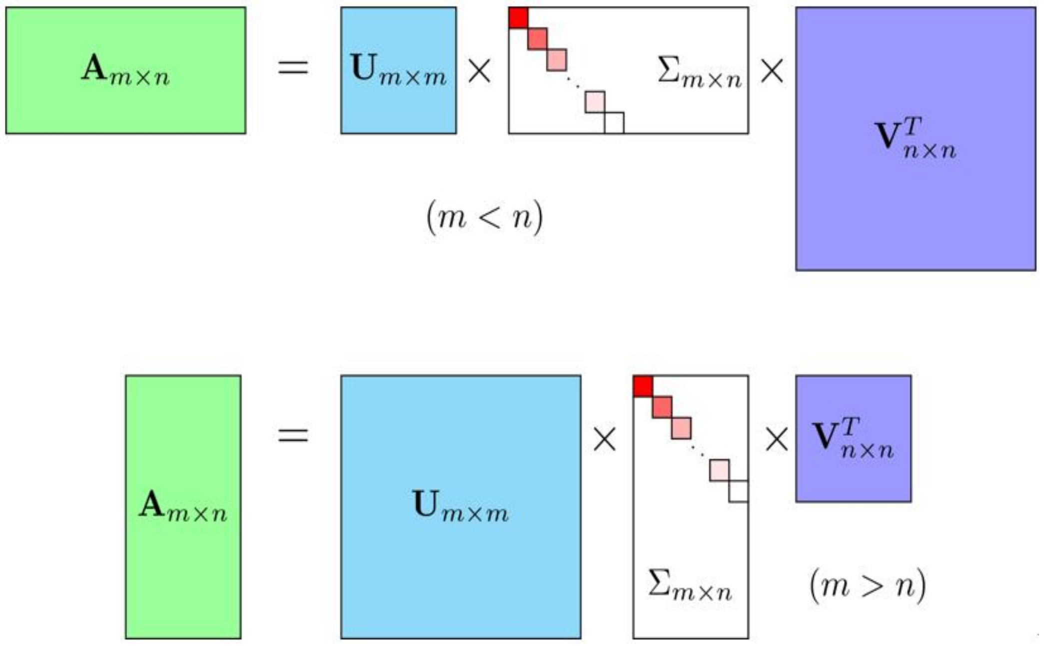


目录章节

CONTENTS

▶ 什么是奇异值和奇异向量？

- 奇异值：**任意矩阵（特征值/特征向量只能在方阵nxn矩阵中完成）**在**奇异值分解（SVD）**中得到的非负实数，表示该矩阵在特定方向上的伸缩幅度大小。
- 奇异向量：分为左右奇异向量，**右奇异向量表示输入空间中的正交基，左奇异向量表示输出空间中的正交基，它们共同描述矩阵如何作用于空间。**



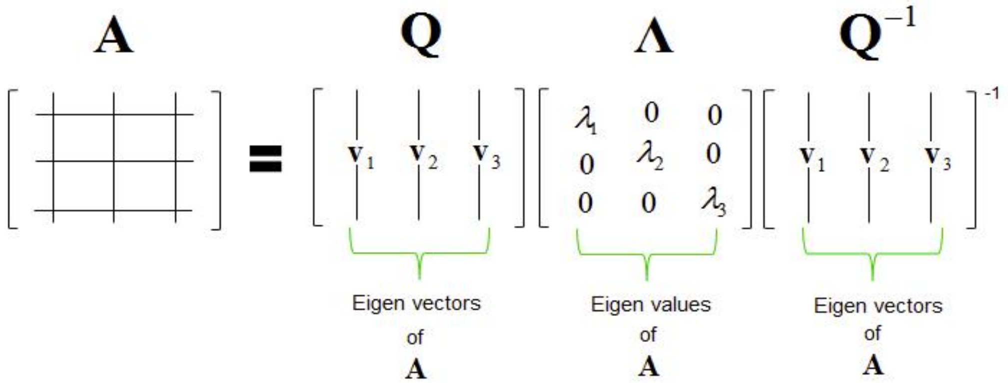
奇异值和奇异向量揭示了非方阵或任意矩阵的核心几何特性，是数据降维（如PCA）、信号压缩和噪声消除等机器学习应用的重要工具。

▶ 什么是奇异值分解？

- 奇异值分解是把任意矩阵 $A \in R^{m \times n}$，分解为三个矩阵的乘积：

$$A = U \Sigma V^T$$

- 其中 U 和 V 是正交矩阵，Σ 是对角矩阵（对角元为奇异值）。

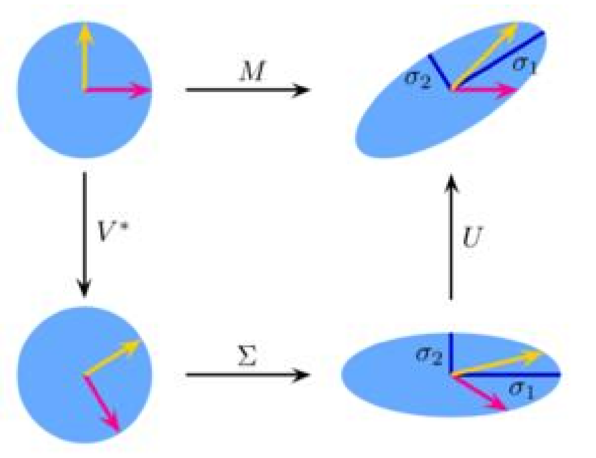


- 作用：这种分解将矩阵的作用分解为：先把向量旋转到一个新基（V），再按奇异值进行伸缩（Σ），最后再旋转到输出空间（U）。

奇异值分解将任意矩阵分解为三个矩阵的乘积，揭示其固有的旋转和缩放特性。

▶ 奇异值分解的几何意义

- 奇异值分解把一个任意矩阵 A 的作用分解为两次旋转（或正交变换）和一次按不同方向伸缩，即：先把输入坐标系旋转到合适方向（V），再沿这些方向按奇异值缩放（Σ），最后再旋转到输出坐标系（U）。
- 奇异值分解另一层几何含义为：**对于任何的一个矩阵，我们要找到一组两两正交单位向量序列，使得矩阵作用在此向量序列上后得到新的向量序列保持两两正交。**



$M = U \cdot \Sigma \cdot V^*$

- 任意实矩阵（方阵或非方阵）都可以看成：旋转（V） -> 缩放（Σ） -> 旋转（U）得来。
- 特殊情况：1）如果矩阵是方阵且对称，SVD退化为特征值分解（只需一次旋转+缩放）；2）如果矩阵奇异值有0，对应方向会被压扁（降维）。

奇异值代表矩阵在各个正交方向上的拉伸或压缩程度，奇异向量提供这些“主方向”的基。

▶ 怎么进行奇异值分解？

- **构造对称矩阵**：计算A^TA，得到一个(n x n矩阵)。
- **求右奇异向量V**：对A^TA进行特征分解，求出特征值λ_i及其对应的特征向量v_i, 这些特征向量ui构成矩阵V（列向量是右奇异向量）。
- **求奇异值**：将特征值开平方$\sigma_i=\lambda_i^{1/2}$，然后把这些奇异值从大到小排列，构成对角矩阵（m x n），大的奇异值对应'重要'的信息，小的奇异值对应'次要'的信息或噪声，保留前r个奇异值 = 保留最重要的信息。
- **求左奇异向量**：利用公式：

$$u_i = \frac{1}{\sigma_i} A u_i$$

计算每个左奇异向量u_i，这些u_i组成矩阵U（列向量是左奇异向量）。

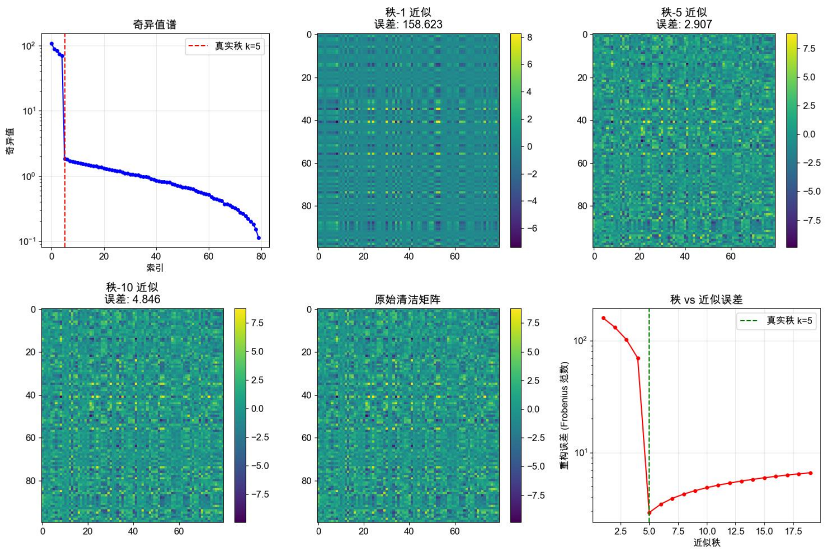
- 最终得到分解：$A=U\Sigma V^T$。

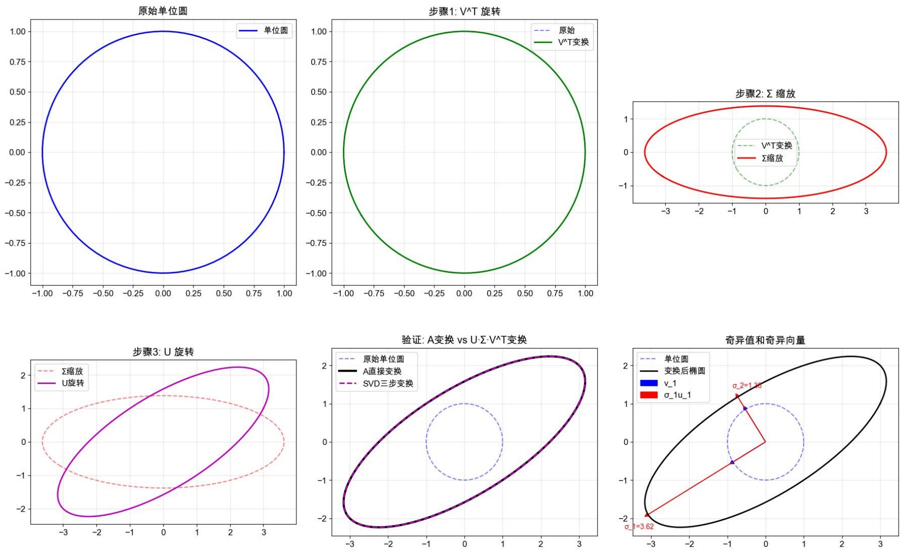
通过对A^TA 进行特征分解得到 V 和奇异值，再用 A_{vi}得到U，从而完成$A=U\Sigma V^T$的分解。

▶ 小结

➢ **奇异值衡量矩阵在各个方向上的“拉伸”强度，奇异向量则是对应这些拉伸的方向**；它们是对任意矩阵（不要求方阵）几何作用的核心刻画。

➢ **奇异值分解（SVD）**：任意的矩阵A都可以分解为**$A=U\Sigma V^T$**，其中 U 和 V 是正交矩阵，Σ 是对角矩阵，奇异值按大小排列，揭示矩阵的主方向与尺度。

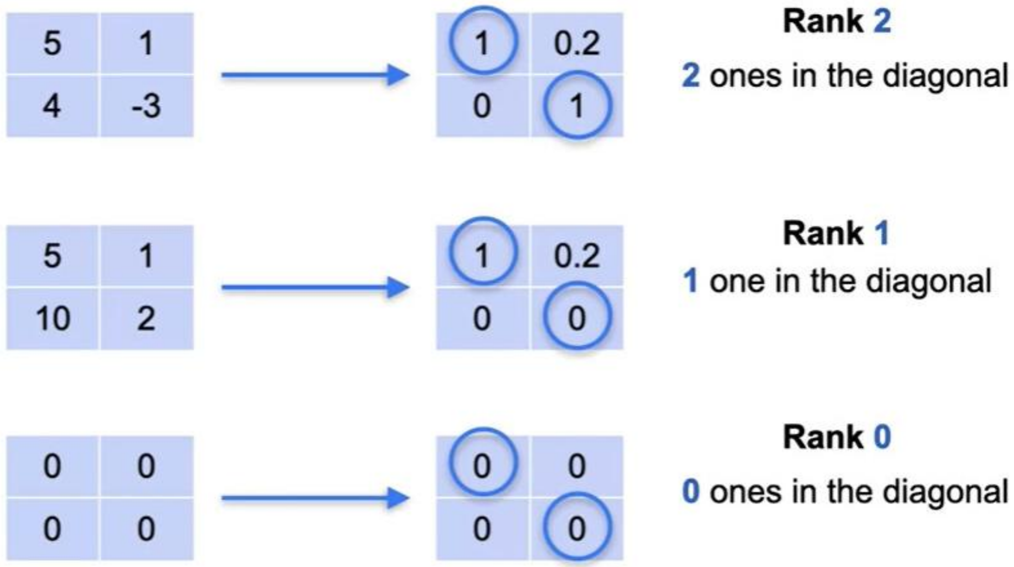




SVD 是理解线性变换几何意义的通用工具，广泛用于降维（PCA）、噪声消除、推荐系统、压缩与矩阵近似等机器学习和数据科学场景。

▶ 扩展：矩阵的秩（Rank）

➢ 设A是一个m x n 的矩阵，矩阵A的秩记做Rank(A)，定义为：在所有 A 的子矩阵（即由 A 中部分行和部分列组成的矩阵）中，最大非零行列式子矩阵的阶数，即存在的最大方阵子矩阵的阶数，使其行列式不等于零。



- 等价地，秩也可以定义为：A的所有行向量（或列向量）中，**线性无关的向量的最大数量**；或A的行空间（或列空间）的维度

- 直观理解：1）线性独立性的度量：矩阵的秩等于它所包含的最大线性无关行（或列）向量的个数，衡量了矩阵中真正“有用”的独立信息量；2）空间维度的描述：矩阵的秩等于它将向量映射后生成的子空间的维度（像空间的维度）

注：秩的大小范围：对于一个 m×n 的矩阵，秩 r 的取值范围是 $0 \leqslant r \leqslant \min(m,n)$。

矩阵秩就是矩阵中最大数量的线性无关行（或列）数目，反映矩阵的“非退化程度”及其所能表示的空间维度。

▶ 扩展：低秩近似与重构

- 通过SVD，可知任意矩阵A可以分解为**$A=U\Sigma V^T$**，如果我们只保留k个最大的奇异值（以及对应的奇异向量），得到一个近似矩阵：

$$A_k = U_k \Sigma_k V_k^T$$

这样就用少量信息（前 k 个方向）近似了原矩阵，实现压缩【低秩近似】。

- 当需要恢复时，直接用保留的奇异值与奇异向量重新计算：

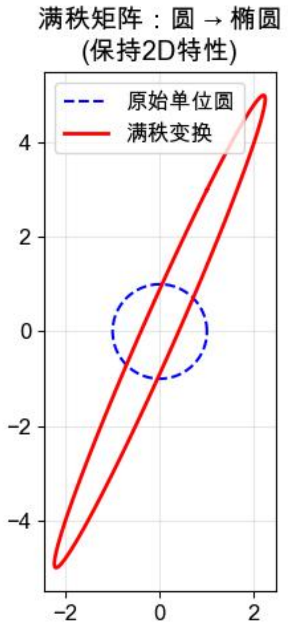
$$\hat{A} = U_k \Sigma_k V_k^T$$

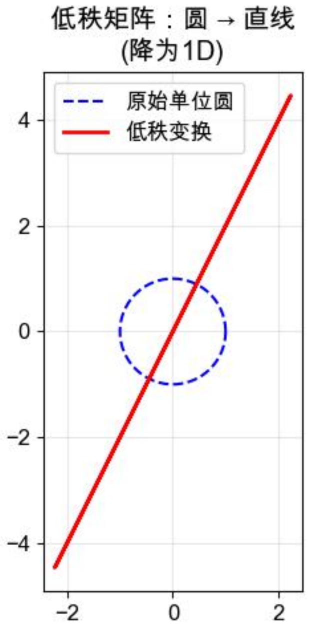
就可以得到一个近似的矩阵。如果k越大，恢复的矩阵越接近原始矩阵，如果k很小，压缩率高但失真也会增加。

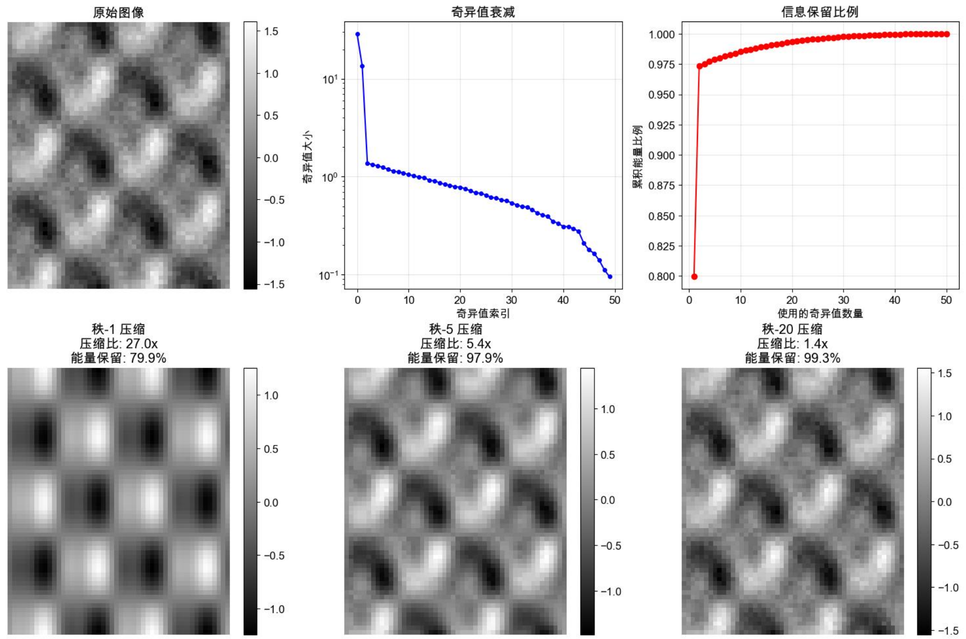
SVD 压缩就像“只保留最重要的方向信息”，还原时，通过这些主要方向重新“拼出”一个近似版本，尽量恢复原矩阵。

▶ 扩展：低秩近似与重构

- 直观理解：原矩阵就像存了所有像素，而 SVD 只存主要方向（奇异向量）和权重（奇异值）。而在还原时，用这些“主要方向”重新组合出一个近似矩阵，信息量大幅减少但结构尽量保留。
- 问题：一个矩阵变成三个矩阵，不是信息量变多了？如果只保留前k个奇异值（最大的能量方向），并且对应的 $U_k(m \times k)$、$V_k(n \times k)$，只保留前 k 列，原始矩阵存储量为mn，压缩后的矩阵存储量为mk + k + nk，比mn小很多【当k <<min(m, n)时】。







目录章节

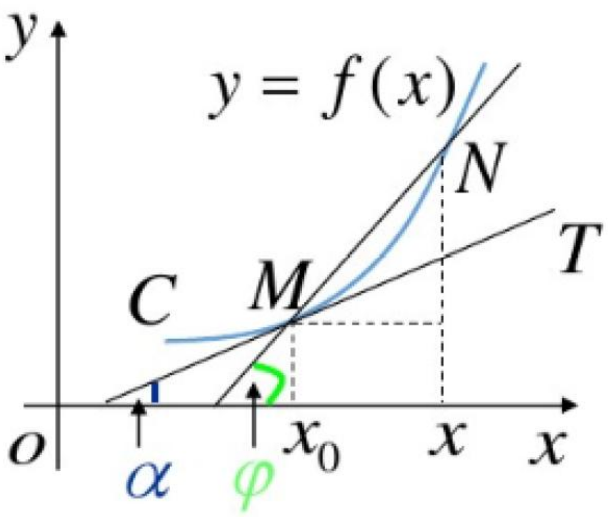
CONTENTS

▶ 标量函数的导数（标量对标量）

➢ 导数就是一个量变化时，函数值变化的速率，其公式定义为：

$$f'(x) = \frac{df}{dx}$$

● 几何意义：导数是函数曲线上某一点切线的斜率，反映函数在该点的瞬时变化率和变化方向。



导数描述一个数变化时函数值的变化率，是函数对输入的灵敏度，是后面梯度（向量）、矩阵导数的起点。

▶ 标量函数的导数（标量对向量）

➢ 定义：一个标量函数f(x1, x2, ..., xn)对每个变量的偏导数组成的向量（标量对向量）：

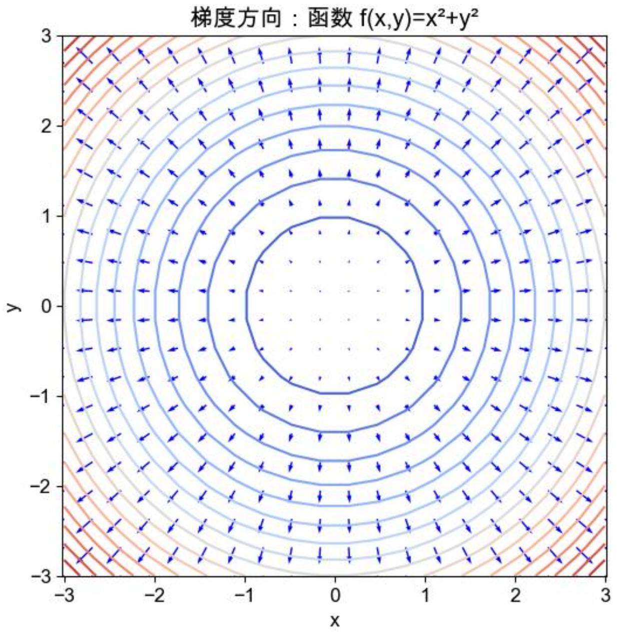
$$\nabla f = \begin{pmatrix} \frac{\partial f}{\partial x_1} \\ \frac{\partial f}{\partial x_2} \\ \vdots \\ \frac{\partial f}{\partial x_n} \end{pmatrix}$$

- 梯度就是函数在当前位置上升最快的方向，模长代表增长速度。

- 对于函数f(x, y)定义为圆，计算其梯度为：

$$f(x,y) = x^2 + y^2 \rightarrow \nabla f(x,y) = \begin{pmatrix} \frac{\partial(x^2+y^2)}{\partial x} \\ \frac{\partial(x^2+y^2)}{\partial y} \end{pmatrix} = \begin{pmatrix} 2x \\ 2y \end{pmatrix}$$

- 在几何上，梯度指向离圆心最远的方向。



代表梯度，用一个向量表示函数对每个分量的敏感度，方向指向函数增长最快的方向。

▶ 标量函数的导数（标量对矩阵）

➢ 定义：若f: R^{mxn} -> R，则定义标量对矩阵的导数定义为：

$$\frac{\partial f}{\partial X} \in \mathbb{R}^{m\times n}, (\frac{\partial f}{\partial X})_{ij} = \frac{\partial f}{\partial X_{ij}}$$

- 维度：和X相同，这就是矩阵梯度（Matrix Gradient）。展开式：$\frac{\partial f}{\partial X} = \begin{pmatrix} \frac{\partial f}{\partial X_{11}} & \frac{\partial f}{\partial X_{12}} & \cdots & \frac{\partial f}{\partial X_{1n}} \\ \frac{\partial f}{\partial X_{21}} & \frac{\partial f}{\partial X_{22}} & \cdots & \frac{\partial f}{\partial X_{2n}} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial f}{\partial X_{m1}} & \frac{\partial f}{\partial X_{m2}} & \cdots & \frac{\partial f}{\partial X_{mn}} \end{pmatrix}$

- 向量化理解：如果把矩阵X拉直：$x = \mathrm{vec}(X) \in \mathbb{R}^{m\times n}$

$$\frac{\partial f}{\partial X} = \mathrm{reshape}(\frac{\partial f}{\partial x}, (m, n))$$

- 也就是说：**标量对矩阵的梯度，本质就是对向量的梯度，只是按矩阵形状排布。**

代表矩阵梯度，用一个同形状的矩阵表示函数对每个元素的敏感度。

▶ 标量函数的导数（标量对矩阵）—— 常见求导

- （1）线性函数：

$$f(X) = \mathrm{tr}(A^T X) \to \frac{\partial f}{\partial X} = A$$

- （2）Frobenius范数平方：

$$f(X) = ||X||_F^2 = \mathrm{tr}(A^T A) \to \frac{\partial f}{\partial X} = 2X$$

- （3）二次型：

$$f(X) = \mathrm{tr}(X^T A X) \to \frac{\partial f}{\partial X} = 2(A + A^T)X$$

- （4）行列式 & 对数行列式：

$$f(X) = \det(X) \to \frac{\partial f}{\partial X} = \det(X) \cdot (X^{-1})^{\top}.$$

- （5）逆矩阵：

$$f(X) = \mathrm{tr}(AX^{-1}B) \to \frac{\partial f}{\partial X} = -(X^{-1}BAX^{-1})^T$$

▶ 向量函数的导数（向量对向量）

- 标量对标量的导数是斜率，标量对向量的导数是梯度。那么向量函数对向量变量呢？一个数已经不够用了，我们需要一个矩阵来描述它 —— 这就是 Jacobian 矩阵。
- 假设$y = f(x)$, $x \in R^n$, $y \in R^m$, Jacobian是一个 m x n 矩阵，它描述了输入每个分量对输出每个分量的影响：

$$J = \frac{\partial \vec{y}}{\partial \vec{x}} = \begin{pmatrix} \frac{\partial y_1}{\partial x_1} & \cdots & \frac{\partial y_1}{\partial x_n} \\ \vdots & \ddots & \vdots \\ \frac{\partial y_m}{\partial x_1} & \cdots & \frac{\partial y_m}{\partial x_n} \end{pmatrix}$$

- 解释：每一行 = 一个输出分量的梯度；每一列 = 所有输出对某个输入分量的敏感度。

- 几何意义：Jacobian 是多维导数的推广：它告诉我们在某点，函数如何把一个小小的输入变化映射到输出变化。可以理解为函数的‘局部线性近似’。

在机器学习里，Jacobian用于反向传播；在优化里，用它做非线性函数线性化近似（比如牛顿法）；在物理里，它描述坐标变换的缩放关系。

▶ 向量函数的导数（矩阵对向量）

➢ 设X ∈ $R^{m\times n}$, x ∈ R^p，函数X = f(x)。

- 输出是矩阵（m x n），输入是向量（p x 1）。
- 每个矩阵元素x_{ij}对每个向量分量x_k都有偏导。

$$\frac{\partial X}{\partial \vec{x}} \in \mathbb{R}^{(mn)\times p}$$

➢ 关键思路：由于矩阵不好直接放进导数里，所以把矩阵X按列展开形成一个大向量：

$$y = \mathrm{vec}(X) \in \mathbb{R}^{mn}$$

- 原问题就变成了向量函数对向量的导数。

➢ 最后这个大向量求Jacobian矩阵表示：

$$J = \frac{\partial y}{\partial x} = \frac{\partial \mathrm{vec}(X)}{\partial x} \in \mathbb{R}^{(mn)\times p}$$

- Jacobian 的第 (i+(j−1)m) 行就是矩阵元素 x_{ij}对向量分量的偏导：

$$J_{(i+(j-1)m),k} = \frac{\partial x_{ij}}{\partial x_k}$$

▶ 向量函数的导数（矩阵对矩阵）

- 在很多场景里，输入和输出都是矩阵，比如深度学习里的卷积核更新、图像处理中的仿射变换、对数行列式的梯度计算。**如何刻画输出矩阵每个元素对输入矩阵每个元素的变化？**
- 定义：设Y = f(X)，$X \in R^{m\times n}$，$Y \in R^{p\times q}$，目标是求四维张量：

$$\frac{\partial Y}{\partial X} = (\frac{\partial y_{ij}}{\partial x_{kl}}), i = 1...p, j = 1...q, k = 1...m, l = 1...n$$

- 关键思路：类似矩阵对向量的导数，仍然进行向量化后求Jacobian矩阵：

$$y = \mathrm{vec}(Y) \in \mathbb{R}^{pq}, x = \mathrm{vec}(X) \in \mathbb{R}^{mn} \to J = \frac{\partial y}{\partial x} \in \mathbb{R}^{(pq)\times(mn)}$$

- 几何意义：描述矩阵到矩阵的局部线性近似，告诉你输入矩阵每个元素微小变化如何影响输出矩阵每个元素。

应用：深度学习，卷积核、注意力矩阵的梯度计算；优化：对矩阵变量的函数求梯度（如 log detX）；统计/控制：协方差矩阵变换。

▶ 小结

输入输出	导数类型	形状或结构	直观理解	典型应用
标量对标量	标量导数	数字	最简单情况	变化率、曲线斜率
标量对向量	梯度向量	n x 1 向量	灵敏度向量	优化梯度、机器学习反向传播
标量对矩阵	矩阵梯度	m x n 矩阵	每个元素偏导数组成	矩阵优化、对数行列式、正则化
向量对向量	Jacobian矩阵	m x n 矩阵	每个输出对输入偏导	非线性变换、神经网络、动力学
矩阵对向量	Jacobian矩阵	(mn) x p	先向量化矩阵	放射变换梯度、控制线性化
矩阵对矩阵	Jacobian矩阵	(pq) x (mn) 矩阵	最高维，四维张量压成矩阵	卷积核梯度、矩阵方程优化

导数的本质是描述输出对输入的灵敏度，随着输入输出维度的增加，导数从数字扩展为向量、矩阵，甚至高维张量，通过向量化和Jacobian矩阵统一表达各种类型的导数关系。

目录章节

CONTENTS

▶ 总结

➢ 线性代数核心：

- ✓ 矩阵与向量：矩阵 = 线性变换，向量 = 被变换的对象。
- ✓ 特征值和特征向量：特征向量为方向不变的向量，拉伸/压缩系数为特征值。
- ✓ 奇异值分解：任意矩阵 = 旋转 + 缩放 + 旋转。

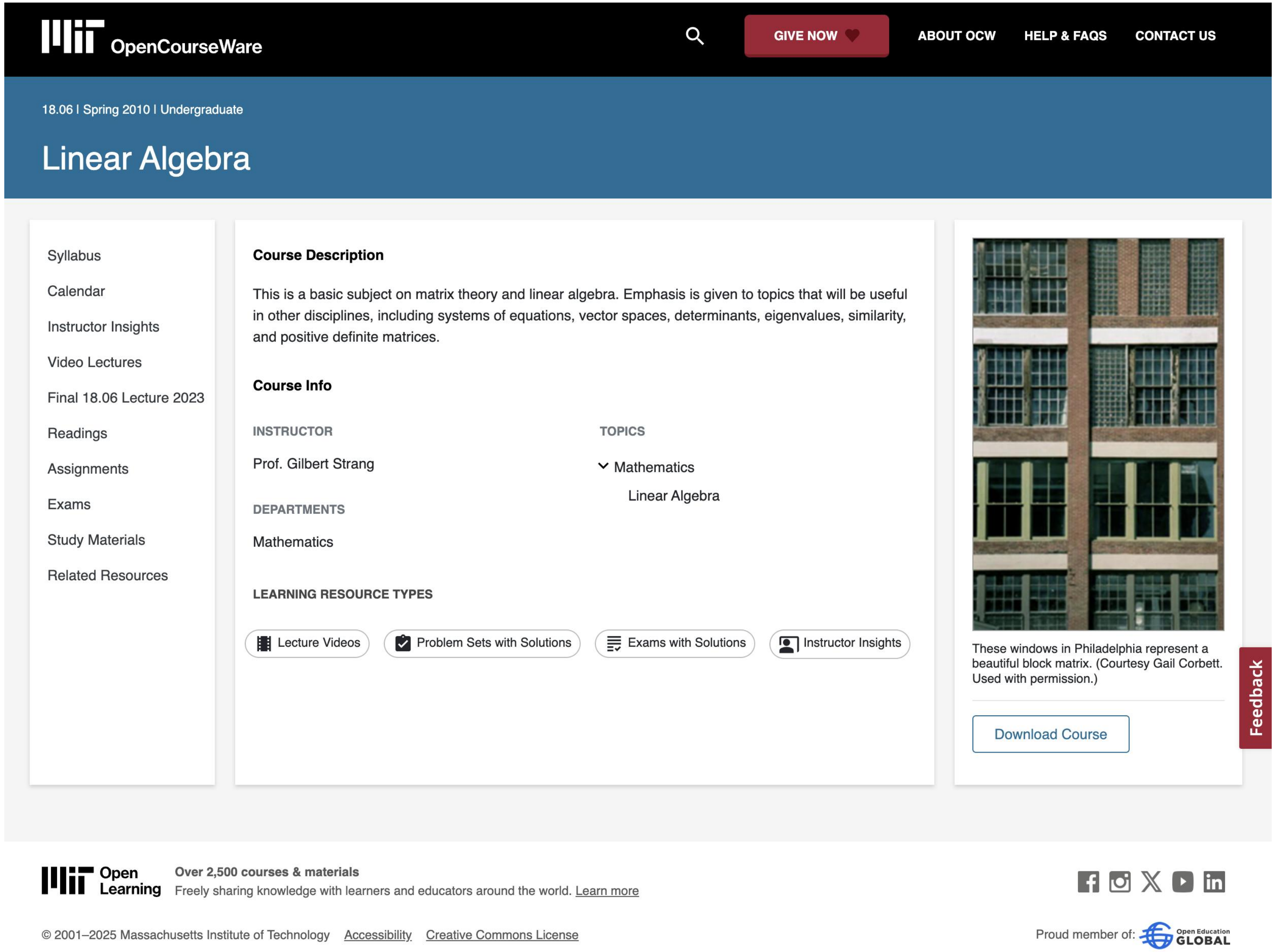
➢ 标量、向量、矩阵的导数：

- ✓ 标量对标量、向量和矩阵的导数描述函数输出的变化率，分别对应数字、梯度向量和矩阵梯度，反映函数对输入各元素的敏感度。
- ✓ 向量对向量的导数用Jacobian矩阵表示，矩阵对向量和矩阵对矩阵的导数则通过先向量化再计算Jacobian，将复杂的高维变化关系转化为矩阵形式。

线性代数揭示了向量与矩阵的结构与变换本质，而标量、向量、矩阵的导数则系统刻画了函数输出对输入的多维敏感度与变化规律。

▶ 扩展：学习资料


3Blue1Brown
232.8万 2.2万 02:40:39
-UP主汉语配音-【线性代数的本质】合集-转载于3Blue1B...
UP 婆婆町 · 2019-02-27


OpenCourseWare
GIVE NOW
ABOUT OCW
HELP & FAQS
CONTACT US
18.06 | Spring 2010 | Undergraduate
Linear Algebra
Syllabus
Calendar
Instructor Insights
Video Lectures
Final 18.06 Lecture 2023
Readings
Assignments
Exams
Study Materials
Related Resources
Course Description
This is a basic subject on matrix theory and linear algebra. Emphasis is given to topics that will be useful in other disciplines, including systems of equations, vector spaces, determinants, eigenvalues, similarity, and positive definite matrices.
Course Info
INSTRUCTOR
Prof. Gilbert Strang
DEPARTMENTS
Mathematics
TOPICS
Mathematics
Linear Algebra
LEARNING RESOURCE TYPES
Lecture Videos
Problem Sets with Solutions
Exams with Solutions
Instructor Insights
These windows in Philadelphia represent a beautiful block matrix. (Courtesy Gail Corbett. Used with permission.)
Download Course
Feedback
Open Learning
Over 2,500 courses & materials
Freely sharing knowledge with learners and educators around the world. Learn more
© 2001–2025 Massachusetts Institute of Technology
Accessibility
Creative Commons License
Proud member of: Open Education GLOBAL

感谢聆听

Personal Website: https://www.miaopeng.info/

Email: miaopeng@stu.scu.edu.cn

Github: https://github.com/MMeowhite

Youtube: https://www.youtube.com/@pengmiao-bmm