2 Vector Calculations 向量运算

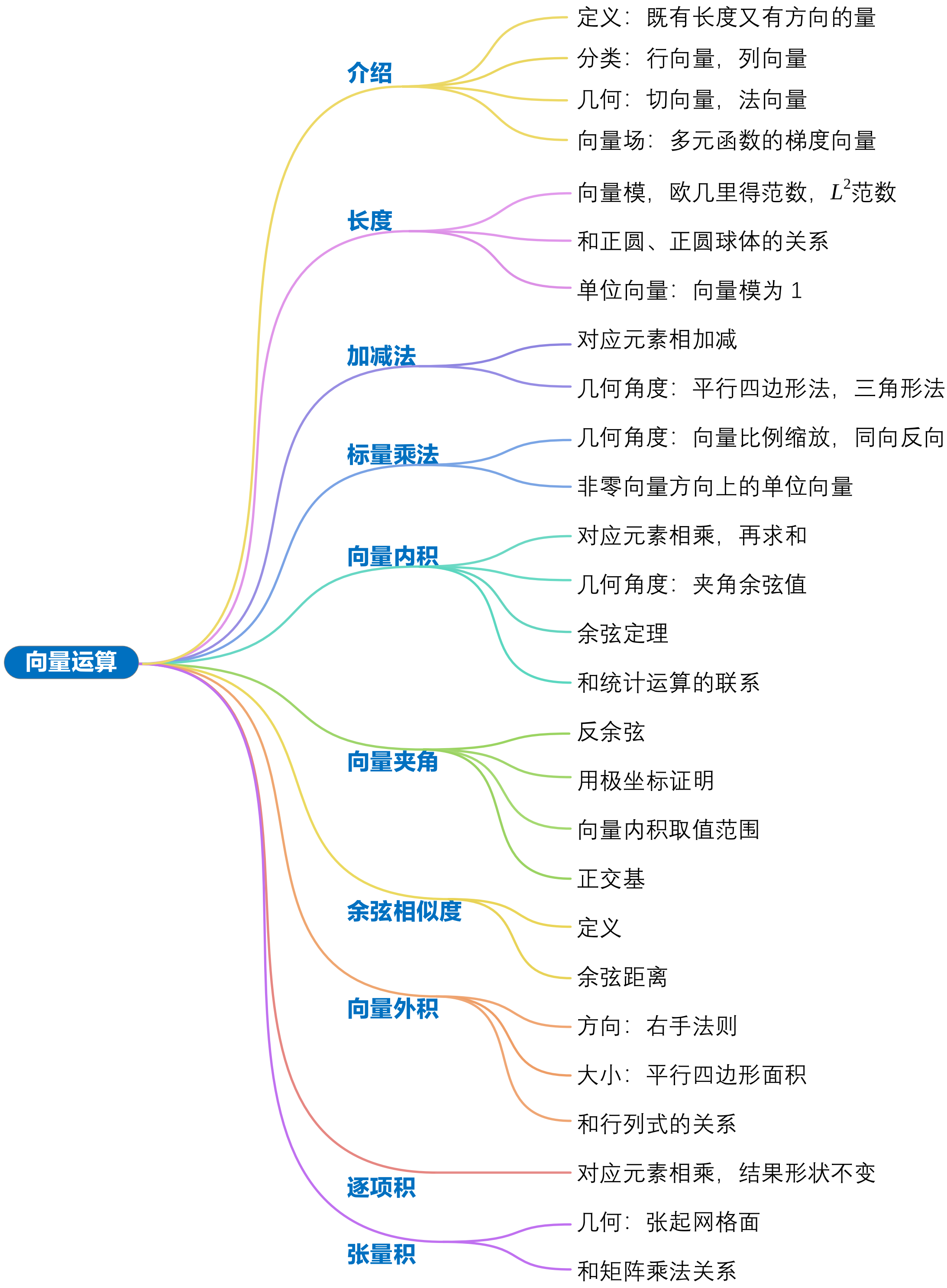
从几何和数据角度解释

几何——指向真理之乡，创造哲学之魂。

Geometry will draw the soul toward truth and create the spirit of philosophy.

—— 柏拉图 (Plato) | 古希腊哲学家 | 424/423 ~ 348/347 BC

- `matplotlib.pyplot.quiver()` 绘制箭头图
- `numpy.add()` 向量/矩阵加法
- `numpy.arccos()` 计算反余弦
- `numpy.array([[4,3]])` 构造行向量，注意双重方括号
- `numpy.array([[4,3]]).T` 行向量转置得到列向量，注意双重方括号
- `numpy.array([[4], [3]])` 构造列向量，注意双重方括号
- `numpy.array([4, 3])[:, None]` 构造列向量
- `numpy.array([4, 3])[:, numpy.newaxis]` 构造列向量
- `numpy.array([4, 3])[None, :]` 构造行向量
- `numpy.array([4, 3])[numpy.newaxis, :]` 构造行向量
- `numpy.array([4,3])` 构造一维数组，严格来说不是行向量
- `numpy.array([4,3]).reshape((-1, 1))` 构造列向量
- `numpy.array([4,3]).reshape((1, -1))` 构造行向量
- `numpy.array([4,3], ndmin=2)` 构造行向量
- `numpy.cross()` 计算列向量或行向量的向量积
- `numpy.dot()` 计算向量内积。值得注意的是，如果输入为一维数组，`numpy.dot()` 输出结果为向量内积；如果输入为矩阵，`numpy.dot()` 输出结果为矩阵乘积，相当于矩阵运算符`@`
- `numpy.linalg.norm()` 默认计算 L2 范数
- `numpy.multiply()` 计算向量逐项积
- `numpy.ones()` 生成全 1 向量/矩阵
- `numpy.outer()` 计算外积、张量积
- `numpy.r_[]` 将一系列数组合并；`'r'` 设定结果以行向量（默认）展示，比如 `numpy.r_[numpy.array([1,2]), 0, 0, numpy.array([4,5])]` 默认产生行向量
- `numpy.r_['c', [4,3]]` 构造列向量
- `numpy.subtract()` 向量/矩阵减法
- `numpy.vdot()` 计算两个向量的向量内积。如果输入是矩阵，矩阵会按照先行、后列顺序展开成向量之后，再计算向量内积
- `numpy.zeros()` 生成全 0 向量/矩阵
- `scipy.spatial.distance.cosine()` 计算余弦距离
- `zip(*)` 用于将可迭代的对象作为参数，将对象中对应的元素打包成一个个元组，然后返回由这些元组组成的列表。`*`代表解包，返回的每一个都是元祖类型，而并非是原来的数据类型



向量运算
介绍
定义：既有长度又有方向的量
分类：行向量，列向量
几何：切向量，法向量
向量场：多元函数的梯度向量
长度
向量模，欧几里得范数，L^2范数
和正圆、正圆球体的关系
单位向量：向量模为 1
加减法
对应元素相加减
几何角度：平行四边形法，三角形法
标量乘法
几何角度：向量比例缩放，同向反向
非零向量方向上的单位向量
向量内积
对应元素相乘，再求和
几何角度：夹角余弦值
余弦定理
和统计运算的联系
向量夹角
反余弦
用极坐标证明
向量内积取值范围
正交基
余弦相似度
定义
余弦距离
向量外积
方向：右手法则
大小：平行四边形面积
和行列式的关系
逐项积
对应元素相乘，结果形状不变
张量积
几何：张起网格面
和矩阵乘法关系

2.1 向量：多面手

几何视角

如图 1 所示，平面上，向量是**有方向的线段** (directed line segment)。**线段的长度代表向量的大小** (the length of the line segment represents the magnitude of the vector)。**箭头代表向量的方向** (the direction of the arrowhead indicates the direction of the vector)。

再次强调，本书中向量符号采用加粗、斜体、小写字母，比如 $\boldsymbol{a}$；矩阵符号则采用加粗、斜体、大写字母，比如 $\boldsymbol{A}$。

图 1 中，向量 $\boldsymbol{a}$ 的**起点** (initial point) 是原点 O，向量的**终点** (terminal point) 是 A。如果向量的起点和终点相同，向量则为**零向量** (zero vector)，可以表示为 $\boldsymbol{0}$。

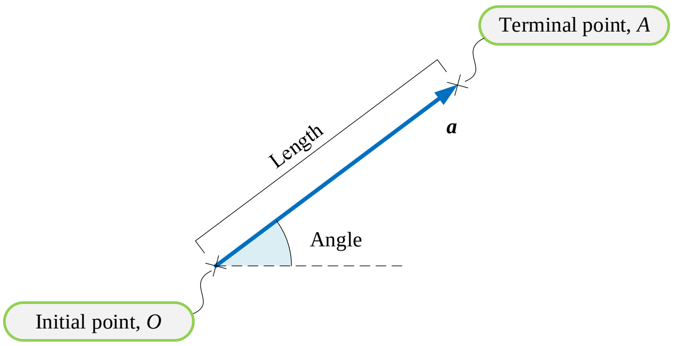


图 1. 向量起点、终点、大小和方向

图 2 给出的是几种向量的类型。

和起点无关的向量叫做**自由向量** (free vector)，如图 2 (a)。和起点有关的向量被称作，**固定向量** (fixed vector)，如图 2 (b) 和 (c)。称方向上沿着某一个特定直线的向量为**滑动向量** (sliding vector)，如图 2 (d)。

没有特别说明时，本书的向量一般是固定向量，且起点一般都在原点，除非特别说明。

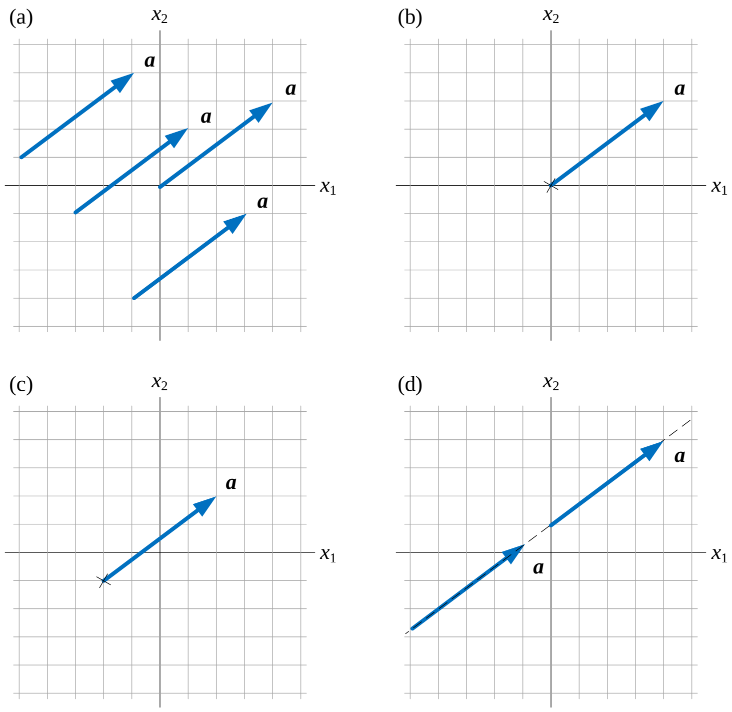


图 2. 几种向量类型

坐标点

从解析几何角度看，向量和坐标存在直接联系。

一般情况下直角坐标系中任意一点坐标可以通过**多元组** (tuple) 来表达。比如，图 3 (a) 所示平面直角坐标系上，A 点坐标为 (4, 3)，B 点坐标为 (−3, 4)。

图 3 (b) 所示，以原点 O 作为向量起点 A 为终点的向量 $\overrightarrow{OA}$ 对应向量 $\boldsymbol{a}$，而 $\overrightarrow{OB}$ 对应向量 $\boldsymbol{b}$。

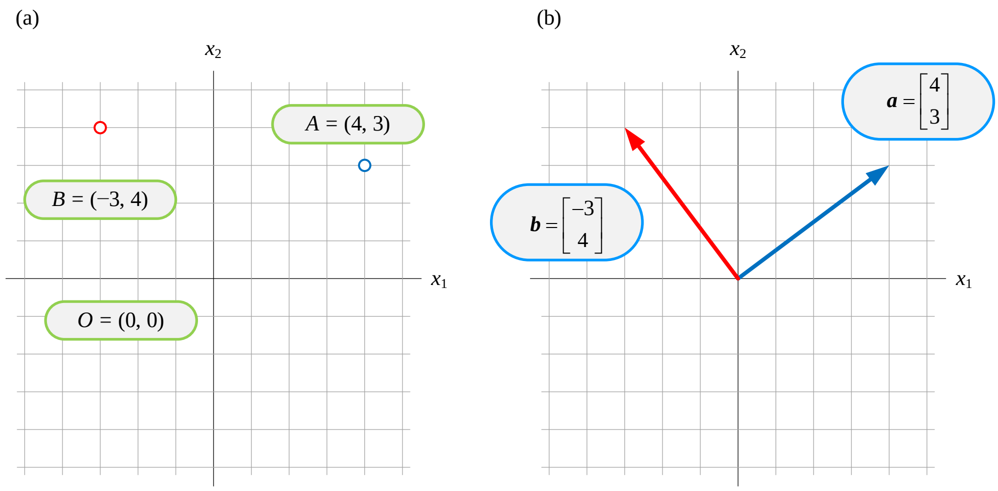


图 3. 平面坐标和向量关系

向量的元素也可以是未知量，比如 $\boldsymbol{x} = [x_1, x_2]^{\mathrm{T}}$、$\boldsymbol{x} = [x_1, x_2, \ldots, x_D]^{\mathrm{T}}$。

Bk4_Ch2_01.py 绘制图 3 (b) 所示向量。matplotlib.pyplot.quiver() 绘制箭头图。

继续丰富向量几何内涵

几何上，切线指的是一条刚好触碰到曲线上某一点的直线。曲线的法线则是垂直于曲线上一点的切线的直线。将向量引入切线、法线可以得到**切向量** (tangent vector) 和**法向量** (normal vector)。图 4 所示为直线和曲线某一点处的切向量和法向量，两个向量的起点都是**切点** (point of tangency)。

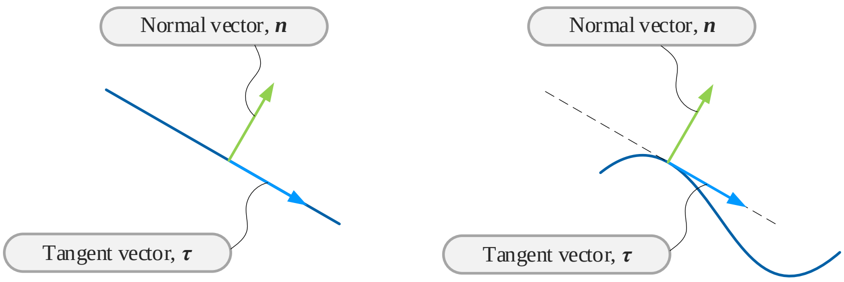


图 4. 切向量和法向量

梯度

自然界的风、水流、电磁场，在各自空间的每一个点上对应的物理量既有强度、也有方向。将这些既有大小又有方向的场抽象出来便得到**向量场** (vector field)。本书中，我们会使用向量场来描述函数在一系列排列整齐点的梯度向量。

图 5 (a) 所示为某个二元函数 $f(x_1, x_2)$ 对应的曲面。把图 5 (a) 比作一座山峰的话，在坡面上放置一个小球，松手瞬间小球运动的方向在 x_1x_2 平面上的投影就是梯度下降方向，也叫做下山方向；而它的反方向叫做**梯度向量** (gradient vector) 方向，也叫上山方向。图 5 (b) 所示为在 x_1x_2 平面上，二元函数 $f(x_1, x_2)$ 在不同点处的平面等高线和梯度向量。仔细观察，可以发现任意一点处梯度向量垂直于该点处等高线。

二元函数 $f(x_1, x_2)$ 梯度向量定义如下：

$$\operatorname{grad} f\left(x_1, x_2\right) = \nabla f\left(x_1, x_2\right) = \begin{bmatrix} \dfrac{\partial f}{\partial x_1} \\ \dfrac{\partial f}{\partial x_2} \end{bmatrix} \tag{1}$$

在 $f(x_1, x_2)$ 梯度向量中，我们看到了两个偏导数。

本书将在第 17 章回顾偏导数并讲解梯度向量。

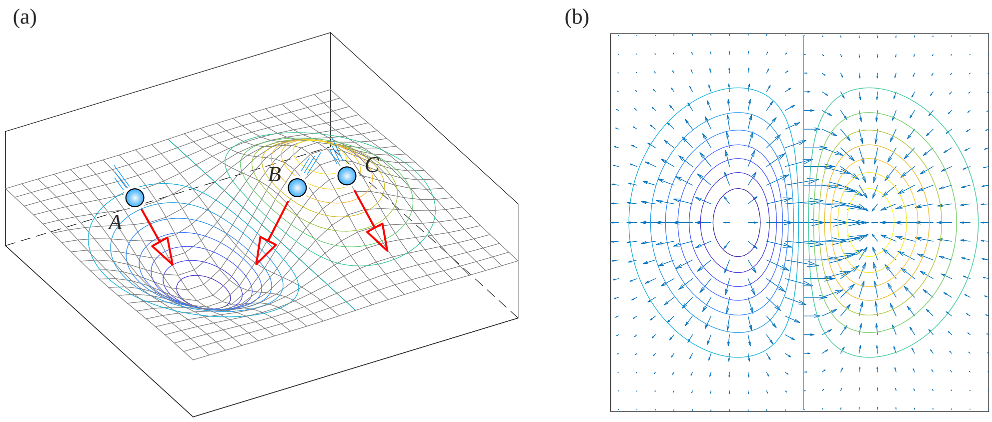


图 5. 梯度向量

2.2 行向量、列向量

上一章提到，向量要么一行多列、要么一列多行，因此向量可以看做是特殊的矩阵——**一维矩阵** (one-dimensional matrix)。一行多列的向量是**行向量** (row vector)，一列多行的向量叫**列向量** (column vector)。

一个矩阵可以视作由若干行向量或列向量整齐排列而成。如图 6 所示，数据矩阵 $\boldsymbol{X}$ 的每一行是一个行向量，代表一个观察值；$\boldsymbol{X}$ 的每一列为一个列向量，代表某个特征上的所有样本数据。

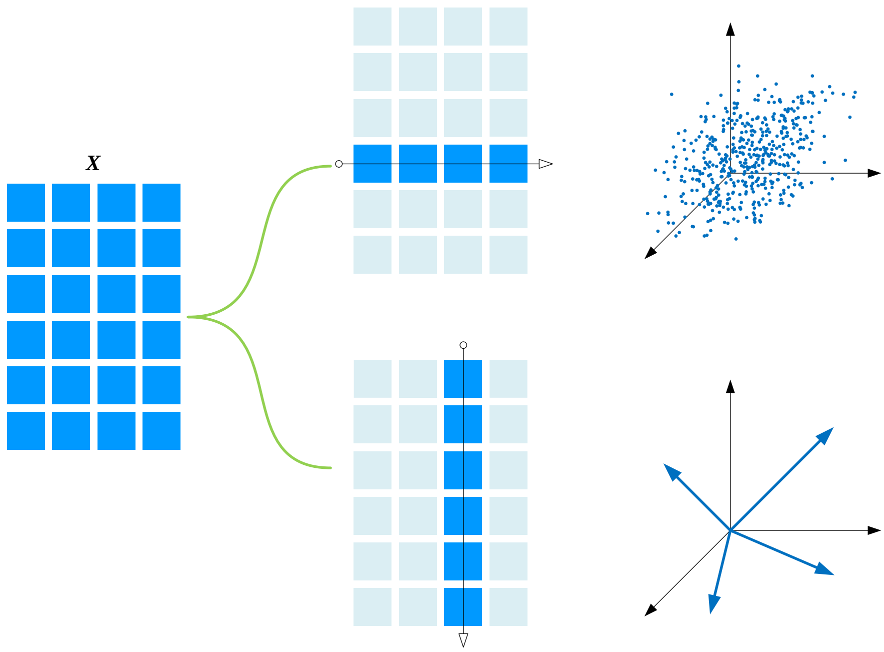


图 6. 观察数据矩阵的两个视角

行向量：一行多列，一个样本数据点

行向量将 n 个元素排成一行，形状为 $1 \times n$ (代表 1 行、n 列)。下式行向量 $\boldsymbol{a}$ 为 1 行 4 列：

$$\boldsymbol{a} = \begin{bmatrix} 1 & 2 & 3 & 4 \end{bmatrix} \tag{2}$$

如图 7 所示，行向量**转置** (transpose) 得到列向量，反之亦然。转置运算符号为正体上标 $^\mathrm{T}$。

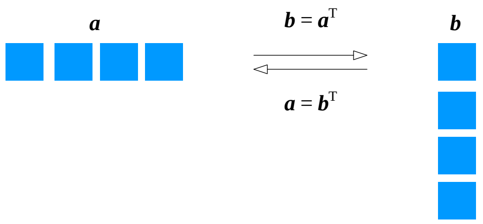


图 7. 行向量的转置是列向量

表 1 所示为利用 `Numpy` 构造行向量几种常见方法。可以用 `len(a)` 计算向量的长度，即向量中元素个数。

表 1. 用 Numpy 构造行向量

代码	注意事项
`a = numpy.array([4,3])`	严格地说，这种方法产生的并不是行向量；运行 `a.ndim` 发现 `a` 只有一个维度。因此，转置 `numpy.array([4,3]).T` 得到的仍然是一维数组，只不过默认展示方式为行向量。
`a = numpy.array([[4,3]])`	运行 `a.ndim` 发现 `a` 有二个维度，这个行向量转置 `a.T` 可以获得列向量。`a.T` 求 `a` 转置，等价于 `a.transpose()`。
`a = numpy.array([4,3], ndmin=2)`	`ndmin=2` 设定数据有两个维度，转置 `a.T` 可以获得列向量。
`a = numpy.r_['r', [4,3]]`	`numpy.r_[]` 将一系列数组合并；`'r'` 设定结果以行向量（默认）展示，比如 `numpy.r_[numpy.array([1,2]), 0, 0, numpy.array([4,5])]` 默认产生行向量。
`a = numpy.array([4,3]).reshape((1, -1))`	`reshape()` 按某种形式重新排列数据，`-1` 自动获取数组长度 `n`。
`a = numpy.array([4, 3])[None, :]`	按照 `[None, :]` 形式广播数组，`None` 代表 `numpy.newaxis`，增加新维度。
`a = numpy.array([4, 3])[numpy.newaxis, :]`	等同于上一例。

前文提过，数据矩阵 $\boldsymbol{X}$ 的每一行代表一个样本点。为了方便区分，$\boldsymbol{X}$ 的行向量序号采用“上标加括号”方式，比如 $\boldsymbol{x}^{(1)}$ 代表 $\boldsymbol{X}$ 的第一行行向量。

如图 8 所示，矩阵 $\boldsymbol{X}$ 可以写成一组行向量上下叠放：

$$\boldsymbol{X}=\begin{bmatrix}\boldsymbol{x}^{(1)}\\ \boldsymbol{x}^{(2)}\\ \vdots\\ \boldsymbol{x}^{(6)}\end{bmatrix} \tag{3}$$

再次强调，数据分析偏爱用行向量表达样本点。

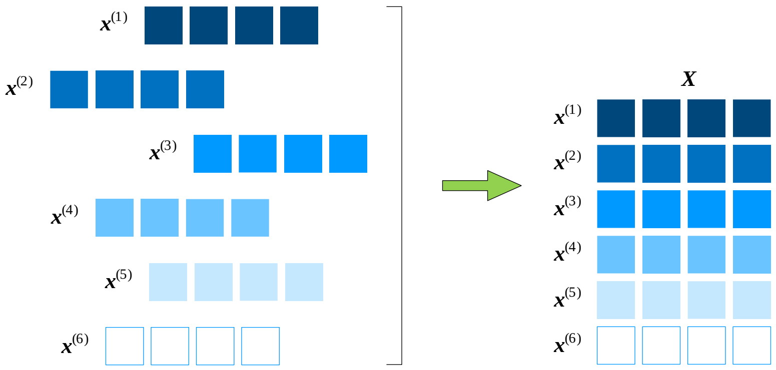


图 8. 矩阵由一系列行向量构造

列向量：一列多行，一个特征样本数据

列向量将 n 个元素排成一列，形状为 $n \times 1$ (即 n 行、1 列)。举个例子，下式中列向量 $\boldsymbol{b}$ 为 4 行 1 列：

$$\boldsymbol{b}=\begin{bmatrix}1\\ 2\\ 3\\ 4\end{bmatrix} \tag{4}$$

注意，不加说明时，本书中向量一般指的是列向量。

构造 $\boldsymbol{X}$ 的列向量序号则采用下标表达，比如 $\boldsymbol{x}_1$。如图 9 所示，矩阵 $\boldsymbol{X}$ 看做是 4 个等长列向量整齐排列得到：

$$\boldsymbol{X}=\begin{bmatrix}\boldsymbol{x}_1 & \boldsymbol{x}_2 & \boldsymbol{x}_3 & \boldsymbol{x}_4\end{bmatrix} \tag{5}$$

数据分析偏爱列向量表达特征，比如 $\boldsymbol{x}_j$ 代表第 j 个特征上的样本数据构成的列向量。因此，列向量又常称作**特征向量** (feature vector)。$\boldsymbol{x}_j$ 对应概率统计的随机变量 X_j，或者代数中的变量 x_j。

注意，此处特征向量不同于特征值分解 (eigen decomposition) 中的特征向量 (eigenvector)。

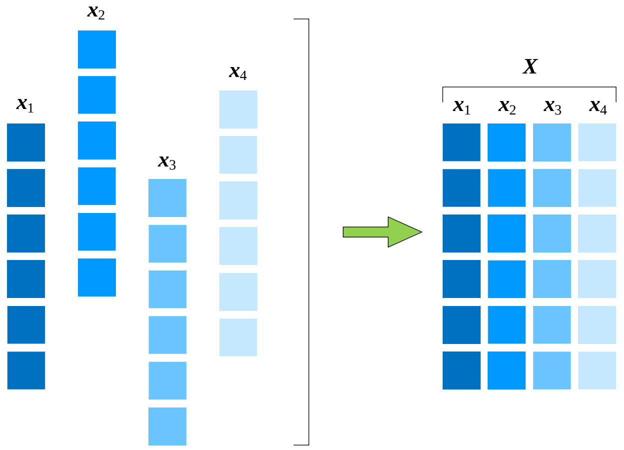


图 9. 矩阵由一排列向量构造

表 2 总结 Numpy 构造列向量几种常见方法。

表 2. 用 Numpy 构造列向量

代码	注意事项
`a = numpy.array([[4], [3]])`	运行 `a.ndim` 发现 `a` 有二个维度。`numpy.array([[4], [3]]).T` 获得行向量
`a = numpy.r_['c', [4,3]]`	`numpy.r_[]` 将一系列的数组合并。`'c'` 设定结果以列向量展示
`a = numpy.array([4,3]).reshape((-1, 1))`	`reshape()` 按某种形式重新排列数据；`-1` 自动获取数组长度 n
`a = numpy.array([4, 3])[:, None]`	按照 `[:, None]` 形式广播数组；`None` 代表 `numpy.newaxis`，增加新维度
`a = numpy.array([4, 3])[:, numpy.newaxis]`	等同于上一例

特殊列向量

全零列向量 (zero column vector) $\boldsymbol{0}$，是指每个元素均为 0 的列向量：

$$\boldsymbol{0} = \begin{bmatrix} 0 & 0 & \cdots & 0 \end{bmatrix}^{\mathrm{T}} \tag{6}$$

代码 `numpy.zeros((4,1))` 可以生成 4 × 1 全 0 列向量。多维空间中，原点也常记做零向量 $\boldsymbol{0}$。

全 1 列向量 (all-ones column vector) $\boldsymbol{1}$，是指每个元素均为 1 的列向量：

$$\mathbf{1} = \begin{bmatrix} 1 & 1 & \cdots & 1 \end{bmatrix}^{\mathrm{T}} \tag{7}$$

代码 `numpy.ones((4,1))` 可以生成 4 × 1 全 1 列向量。

全 1 列向量 $\mathbf{1}$ 在矩阵乘法中有特殊的地位，本书第 5、22 章将分别从矩阵乘法和统计两个角度讲解。

2.3 向量长度：模，欧氏距离，L^2 范数

向量长度 (length of a vector) 又叫做**向量模** (vector norm)、**欧几里得距离** (Euclidean distance)、**欧几里得范数** (Euclidean norm) 或 L^2 **范数** (L2-norm)。

给定向量 $\boldsymbol{a}$:

$$\boldsymbol{a} = \begin{bmatrix} a_1 & a_2 & \cdots & a_n \end{bmatrix}^{\mathrm{T}} \tag{8}$$

向量 $\boldsymbol{a}$ 的模为:

$$\|\boldsymbol{a}\| = \|\boldsymbol{a}\|_2 = \sqrt{a_1^2 + a_2^2 + \cdots a_n^2} = \left(\sum_{i=1}^{n} a_i^2\right)^{\frac{1}{2}} \tag{9}$$

观察上式，容易知道向量模非负，即 $\|\boldsymbol{a}\| \geq 0$ 。

注意，$\|\boldsymbol{a}\|_2$ 下角标 2，代表 L^2 范数。没有特殊说明，$\|\boldsymbol{a}\|$ 默认代表 L^2 范数。

L^2 范数是 L^p 范数的一种，本书第 3 章将介绍其他范数。

二维向量的模

特别地，对于如下二维向量 $\boldsymbol{a}$:

$$\boldsymbol{a} = \begin{bmatrix} a_1 & a_2 \end{bmatrix}^{\mathrm{T}} \tag{10}$$

二维向量指的是有两个元素的向量。

二维向量 $\boldsymbol{a}$ 的 L^2 范数为:

$$\|\boldsymbol{a}\| = \sqrt{a_1^2 + a_2^2} \tag{11}$$

图 3 (b) 中向量 $\boldsymbol{a}$ 和 $\boldsymbol{b}$ 的模可以这样计算得到:

$$
\begin{aligned}
\|\boldsymbol{a}\| &= \sqrt{4^2 + 3^2} = \sqrt{25} = 5 \\
\|\boldsymbol{b}\| &= \sqrt{(-3)^2 + 4^2} = \sqrt{25} = 5
\end{aligned}
\tag{12}
$$

二维向量 $\boldsymbol{a}$ 和横轴夹角可以通过反正切求解：

$$
\theta_a = \arctan\left(\frac{a_2}{a_1}\right) \tag{13}
$$

上述角度和直角坐标系直接关联，因此可以视作“绝对角度”。本章后续将介绍如何用向量内积求两个向量之间的“相对角度”。

Bk4_Ch2_02.py 计算图 3 (b) 中向量 $\boldsymbol{a}$ 和 $\boldsymbol{b}$ 模。函数 `numpy.linalg.norm()` 默认计算 L^2 范数，也可以用 `numpy.sqrt(np.sum(a**2))` 计算向量 $\boldsymbol{a}$ 的 L^2 范数。

等距线

值得一提的是，如果起点重合，和 $\|\boldsymbol{a}\|$ 等长的二维向量的终点位于同一个圆上，如图 10 (a) 所示。看到这里大家是否想到了本系列丛书《数学要素》第 7 章讲过的“等距线”。

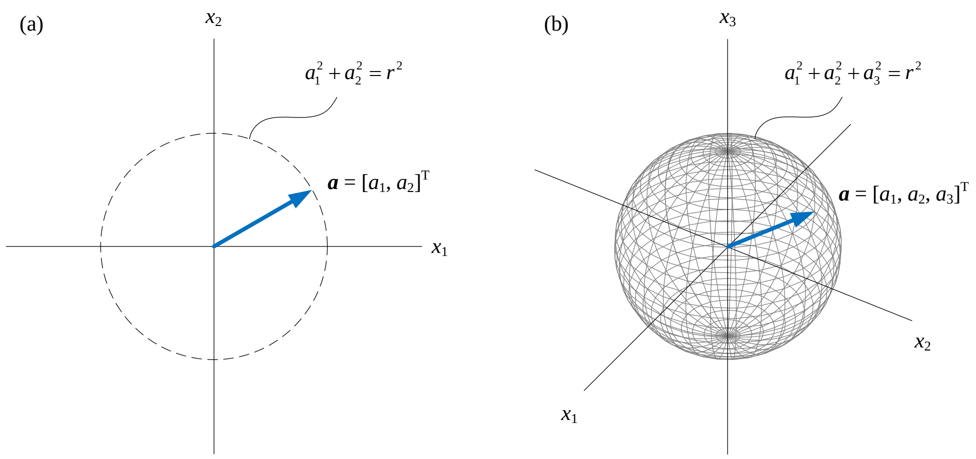


图 10. 等 L^2 范数向量

如图 11 所示，起点位于原点的二维向量 $\boldsymbol{x}$ 的模 $\|\boldsymbol{x}\|$ 取不同数值 c 时，我们可以得到一系列同心圆，对应的解析式如下：

$$
\|\boldsymbol{x}\| = \sqrt{x_1^2 + x_2^2} = c \tag{14}
$$

强调一点，$\boldsymbol{x}$ 是向量，既有大小、又有方向；而 $\|\boldsymbol{x}\|$ 是标量，代表“距离”。$\|\cdot\|$ 这个运算符是一种“向量 → 标量”的运算规则。

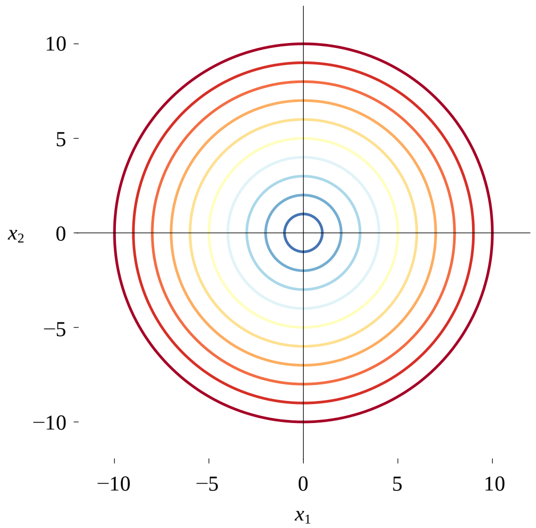


图 11. 起点为 $\boldsymbol{0}$、L^2 范数相等的向量终点位于一系列同心圆上

Bk4_Ch2_03.py 绘制图 11。

三维向量的模

类似地，给定三维向量 $\boldsymbol{a}$：

$$\boldsymbol{a} = \begin{bmatrix} a_1 & a_2 & a_3 \end{bmatrix}^{\mathrm{T}} \tag{15}$$

三维向量 $\boldsymbol{a}$ 的 L^2 范数为：

$$\|\boldsymbol{a}\| = \sqrt{a_1^2 + a_2^2 + a_3^2} \tag{16}$$

如图 10 (b) 所示，起点为原点的等长三维向量终点落在同一正圆球面上。

单位向量

长度为 1 的向量叫做**单位向量** (unit vector)。

非 $\boldsymbol{0}$ 向量 $\boldsymbol{a}$ 除以自身的模得到 **$\boldsymbol{a}$ 方向上的单位向量** (unit vector in the direction of vector $\boldsymbol{a}$)：

$$\hat{\boldsymbol{a}} = \frac{\boldsymbol{a}}{\|\boldsymbol{a}\|} \tag{17}$$

$\hat{\boldsymbol{a}}$ 读作 “vector a hat”。`a/numpy.linalg.norm(a)` 可以计算非 $\boldsymbol{0}$ 向量 $\boldsymbol{a}$ 方向上的单位向量。

图 12 (a) 所示平面直角坐标系，起点位于原点的单位向量 $\boldsymbol{x} = [x_1, x_2]^T$ 终点位于**单位圆** (unit circle) 上，对应的解析式为：

$$\|\boldsymbol{x}\| = \sqrt{x_1^2 + x_2^2} = 1 \quad \Rightarrow \quad x_1^2 + x_2^2 = 1 \tag{18}$$

这无数个单位向量 $\boldsymbol{x}$ 中，有两个单位向量最为特殊——$\boldsymbol{e}_1$ ($\boldsymbol{i}$) 和 $\boldsymbol{e}_2$ ($\boldsymbol{j}$)。如图 12 (b) 所示平面直角坐标系中，$\boldsymbol{e}_1$ 和 $\boldsymbol{e}_2$ 分别为沿着 x_1 (水平) 和 x_2 (竖直) 方向的单位向量：

$$\boldsymbol{e}_1 = \boldsymbol{i} = \begin{bmatrix} 1 \\ 0 \end{bmatrix}, \quad \boldsymbol{e}_2 = \boldsymbol{j} = \begin{bmatrix} 0 \\ 1 \end{bmatrix} \tag{19}$$

显然，$\boldsymbol{e}_1$ 和 $\boldsymbol{e}_2$ 相互垂直。

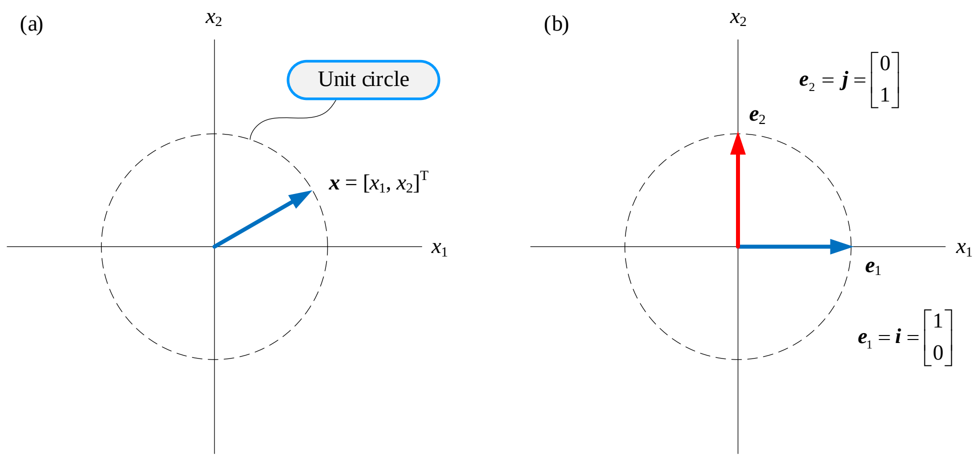


图 12. 单位向量

张成

图 3 (b) 给出向量 $\boldsymbol{a}$ 和 $\boldsymbol{b}$ 可以用 $\boldsymbol{e}_1$ 和 $\boldsymbol{e}_2$ 合成得到：

$$\begin{aligned} \boldsymbol{a} &= 4\boldsymbol{e}_1 + 3\boldsymbol{e}_2 \\ \boldsymbol{b} &= -3\boldsymbol{e}_1 + 4\boldsymbol{e}_2 \end{aligned} \tag{20}$$

(20) 用到的便是向量加减法，这是下一节要介绍的内容。

图 3 (b) 整个平面就是用 $\boldsymbol{e}_1$ 和 $\boldsymbol{e}_2$ 张成的。白话说，$\boldsymbol{e}_1$ 和 $\boldsymbol{e}_2$ 好比经纬度，可以定位地表任意一点。比如，$\mathbb{R}^2$ 平面上的任意一点 $\boldsymbol{x}$ 都可以写成：

$$\boldsymbol{x} = x_1\boldsymbol{e}_1 + x_2\boldsymbol{e}_2 \tag{21}$$

从集合角度来看，$\boldsymbol{x} \in \mathbb{R}^2$。

本书第 7 章将讲解张成 (span)、向量空间等概念。

三维直角坐标系

三维直角坐标系中，$\boldsymbol{e}_1$ ($\boldsymbol{i}$)、$\boldsymbol{e}_2$ ($\boldsymbol{j}$) 和 $\boldsymbol{e}_3$ ($\boldsymbol{k}$) 代表沿着 x_1、x_2 和 x_3 单位向量：

$$\boldsymbol{e}_1=\boldsymbol{i}=\begin{bmatrix}1\\0\\0\end{bmatrix},\quad \boldsymbol{e}_2=\boldsymbol{j}=\begin{bmatrix}0\\1\\0\end{bmatrix},\quad \boldsymbol{e}_3=\boldsymbol{k}=\begin{bmatrix}0\\0\\1\end{bmatrix} \tag{22}$$

如图 13 所示，$\boldsymbol{e}_1$ ($\boldsymbol{i}$)、$\boldsymbol{e}_2$ ($\boldsymbol{j}$) 和 $\boldsymbol{e}_3$ ($\boldsymbol{k}$) 两两相互垂直。

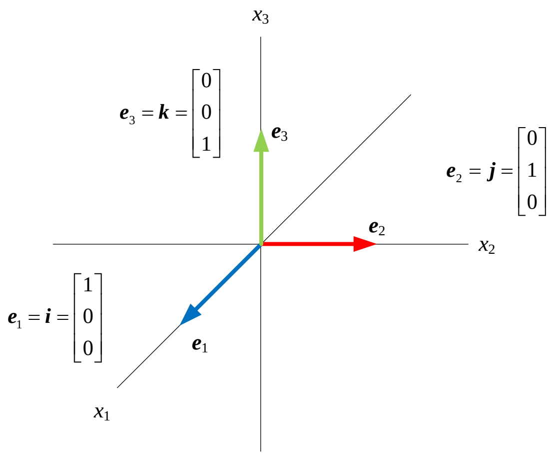


图 13. 三维空间单位向量

同理，图 13 这个三维空间是用 $\boldsymbol{e}_1$、$\boldsymbol{e}_2$、$\boldsymbol{e}_3$ 张成的。白话说，$\boldsymbol{e}_1$、$\boldsymbol{e}_2$、$\boldsymbol{e}_3$ 相当于经度、维度、海拔，定位能力从地表扩展到整个地球空间。

$\mathbb{R}^3$ 空间任意一点 $\boldsymbol{x}$ 可以写成：

$$\boldsymbol{x}=x_1\boldsymbol{e}_1+x_2\boldsymbol{e}_2+x_3\boldsymbol{e}_3 \tag{23}$$

此外，大家可能已经注意到，$\boldsymbol{e}_1$ 可以用不同的形式表达，比如：

$$\boldsymbol{e}_1=\begin{bmatrix}1\\0\end{bmatrix},\quad \boldsymbol{e}_1=\begin{bmatrix}1\\0\\0\end{bmatrix},\quad \boldsymbol{e}_1=\begin{bmatrix}1\\0\\0\\0\end{bmatrix},\quad \boldsymbol{e}_1=\begin{bmatrix}1\\0\\\vdots\\0\end{bmatrix} \tag{24}$$

上式中几个 $\boldsymbol{e}_1$ 虽然维度不同，但是本质上等价，它们代表不同维度空间中的 $\boldsymbol{e}_1$。这些 $\boldsymbol{e}_1$ 之间的关系是，从低维到高维或从高维到低维投影。

本书将在第 8、9、10 三章深入探讨投影这一重要线性代数工具。

2.4 加减法：对应位置元素分别相加减

从数据角度看，两个等长列向量相加，结果为对应位置元素分别相加，得到长度相同的列向量，比如下例：

$$\boldsymbol{a}+\boldsymbol{b}=\begin{bmatrix}-2\\5\end{bmatrix}+\begin{bmatrix}5\\-1\end{bmatrix}=\begin{bmatrix}-2+5\\5-1\end{bmatrix}=\begin{bmatrix}3\\4\end{bmatrix} \tag{25}$$

两个等长列向量相减，则是对应元素分别相减，得到等长列向量，比如：

$$\boldsymbol{a}-\boldsymbol{b}=\begin{bmatrix}-2\\5\end{bmatrix}-\begin{bmatrix}5\\-1\end{bmatrix}=\begin{bmatrix}-2-5\\5-(-1)\end{bmatrix}=\begin{bmatrix}-7\\6\end{bmatrix} \tag{26}$$

几何视角

从几何角度看，**向量加法** (vector addition) 结果可以用**平行四边形法则** (parallelogram method) 或**三角形法则** (triangle method) 获得，具体如图 14 所示。

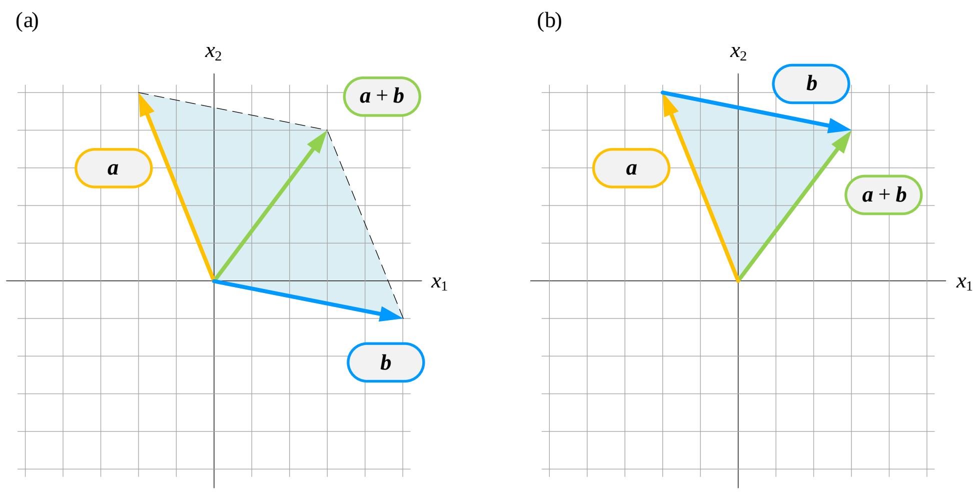


图 14. 几何角度看向量加法

向量减法 (vector subtraction) 可以写成向量加法。比如，向量 $\boldsymbol{a}$ 减去向量 $\boldsymbol{b}$，可以将向量 $\boldsymbol{b}$ 换向得到 $-\boldsymbol{b}$；然后再计算向量 $\boldsymbol{a}$ 与向量 $-\boldsymbol{b}$ 之和，即：

$$
\boldsymbol{a}-\boldsymbol{b}=\boldsymbol{a}+(-\boldsymbol{b})=\begin{bmatrix}-2\\5\end{bmatrix}+\underbrace{\begin{bmatrix}-5\\1\end{bmatrix}}_{-\boldsymbol{b}}=\begin{bmatrix}-7\\6\end{bmatrix} \tag{27}
$$

两个向量相同，当且仅当两者大小方向均相同。如果两个向量的模 (长度) 相同但是方向相反，两者互为反向量。若两个向量方向相同或相反，则称向量平行。

请大家注意以下向量加减法性质：

$$
\begin{aligned}
&\boldsymbol{a}+\boldsymbol{b}=\boldsymbol{b}+\boldsymbol{a}\\
&(\boldsymbol{a}+\boldsymbol{b})+\boldsymbol{c}=\boldsymbol{a}+(\boldsymbol{b}+\boldsymbol{c})\\
&\boldsymbol{a}+(-\boldsymbol{a})=\boldsymbol{0}
\end{aligned} \tag{28}
$$

⚠ 注意，向量 $\boldsymbol{a}$ 减去向量 $\boldsymbol{b}$，结果 $\boldsymbol{a}-\boldsymbol{b}$ 对应向量箭头指向 $\boldsymbol{a}$ 终点；相反，向量 $\boldsymbol{b}$ 减去向量 $\boldsymbol{a}$ 得到 $\boldsymbol{b}-\boldsymbol{a}$ 指向 $\boldsymbol{b}$ 终点。

Bk4_Ch2_04.py 计算本节向量加减法示例。

2.5 标量乘法：向量缩放

向量标量乘法 (scalar multiplication of vectors) 指的是标量和向量每个元素分别相乘，结果仍为向量。

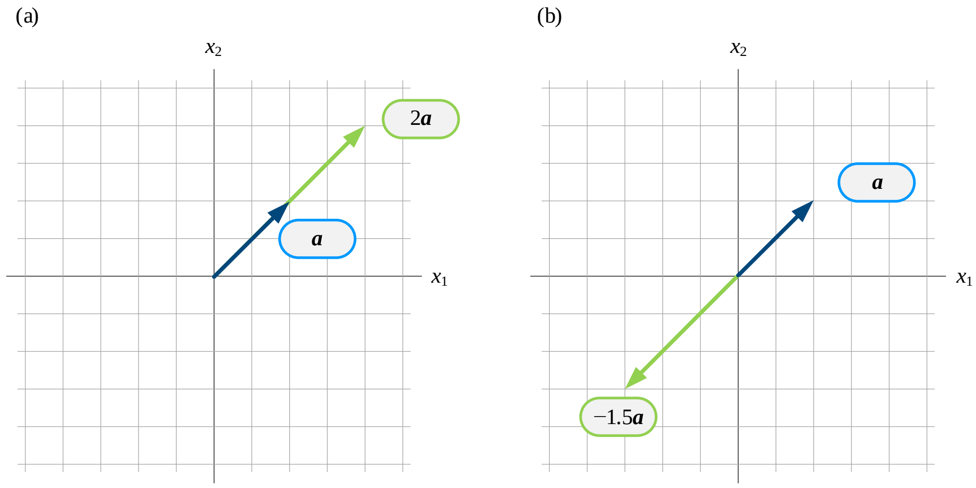


图 15. 向量标量乘法

从几何角度来看，标量乘法将原向量按标量比例缩放，结果中向量方向同向或反向，如图 15 所示。

Bk4_Ch2_05.py 完成图 15 中运算。

请大家注意以下向量标量乘法性质：

$$
\begin{aligned}
&(t+k)\boldsymbol{a}=t\boldsymbol{a}+k\boldsymbol{a}\\
&t(\boldsymbol{a}+\boldsymbol{b})=t\boldsymbol{a}+t\boldsymbol{b}\\
&t(k\boldsymbol{a})=tk\boldsymbol{a}\\
&1\boldsymbol{a}=\boldsymbol{a}\\
&-1\boldsymbol{a}=-\boldsymbol{a}\\
&0\boldsymbol{a}=\boldsymbol{0}
\end{aligned}
\tag{29}
$$

其中，t 和 k 为标量。请大家特别注意，0 乘向量 $\boldsymbol{a}$ 结果不是 0，而是零向量 $\boldsymbol{0}$，这个零向量的形状取决于向量 $\boldsymbol{a}$。

2.6 向量内积：结果为标量

向量内积 (inner product)，又叫**标量积** (scalar product)、**点积** (dot product)、点乘。注意，向量内积的运算结果为标量，而非向量。

给定如下 $\boldsymbol{a}$ 和 $\boldsymbol{b}$ 两个等长列向量：

$$
\begin{aligned}
\boldsymbol{a}&=\begin{bmatrix}a_1 & a_2 & \cdots & a_n\end{bmatrix}^{\mathrm{T}}\\
\boldsymbol{b}&=\begin{bmatrix}b_1 & b_2 & \cdots & b_n\end{bmatrix}^{\mathrm{T}}
\end{aligned}
\tag{30}
$$

列向量 $\boldsymbol{a}$ 和 $\boldsymbol{b}$ 的内积定义如下：

$$
\boldsymbol{a}\cdot\boldsymbol{b}=\langle\boldsymbol{a},\boldsymbol{b}\rangle=\sum_{i=1}^{n}a_ib_i=a_1b_1+a_2b_2+\ldots+a_nb_n \tag{31}
$$

(31) 也适用于两个等长行向量计算内积。注意，向量内积也是一种“向量 → 标量”的运算规则。

图 16 所示的两个列向量 $\boldsymbol{a}$ 和 $\boldsymbol{b}$ 的内积为：

$$
\boldsymbol{a}\cdot\boldsymbol{b}=\begin{bmatrix}4\\3\end{bmatrix}\cdot\begin{bmatrix}5\\-2\end{bmatrix}=4\times5+3\times(-2)=14 \tag{32}
$$

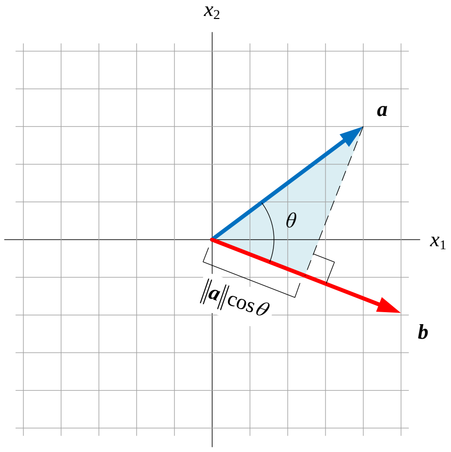


图 16. $\boldsymbol{a}$ 和 $\boldsymbol{b}$ 两个平面向量

Bk4_Ch2_06.py 计算上述向量内积。此外，还可以用 numpy.dot() 计算向量内积。值得注意的是，如果输入为一维数组，numpy.dot() 输出结果为内积。

如果输入为矩阵，numpy.dot() 输出结果为矩阵乘积，相当于矩阵运算符@，比如 Bk4_Ch2_07.py 给出例子。

numpy.vdot() 函数也可以计算两个向量内积。如果输入是矩阵，矩阵会按照先行后列顺序展开成向量之后，再计算向量内积。Bk4_Ch2_08.py 给出示例。

常用的向量内积性质如下：

$$\begin{aligned}&\boldsymbol{a}\cdot\boldsymbol{b}=\boldsymbol{b}\cdot\boldsymbol{a}\\&\boldsymbol{a}\cdot(\boldsymbol{b}+\boldsymbol{c})=\boldsymbol{a}\cdot\boldsymbol{b}+\boldsymbol{a}\cdot\boldsymbol{c}\\&(k\boldsymbol{a})\cdot(t\boldsymbol{b})=kt(\boldsymbol{a}\cdot\boldsymbol{b})\end{aligned}\tag{33}$$

请读者格外注意以下几个向量内积运算和 Σ 求和运算的关系：

$$\begin{aligned}&\boldsymbol{1}\cdot\boldsymbol{x}=x_1+x_2+\cdots x_n=\sum_{i=1}^{n}x_i\\&\boldsymbol{x}\cdot\boldsymbol{x}=x_1^2+x_2^2+\cdots x_n^2=\sum_{i=1}^{n}x_i^2\\&\boldsymbol{x}\cdot\boldsymbol{y}=x_1y_1+x_2y_2+\cdots x_ny_n=\sum_{i=1}^{n}x_iy_i\end{aligned}\tag{34}$$

其中，

$$\boldsymbol{x}=\begin{bmatrix}x_1 & x_2 & \cdots & x_n\end{bmatrix}^{\mathrm{T}},\quad \boldsymbol{1}=\begin{bmatrix}1 & 1 & \cdots & 1\end{bmatrix}^{\mathrm{T}},\quad \boldsymbol{y}=\begin{bmatrix}y_1 & y_2 & \cdots & y_n\end{bmatrix}^{\mathrm{T}}\tag{35}$$

几何视角

如图 16 所示，从几何角度看，向量内积相当于两个向量的模 (L^2 范数) 与它们之间夹角余弦值三者之积：

$$\boldsymbol{a}\cdot\boldsymbol{b}=\|\boldsymbol{a}\|\|\boldsymbol{b}\|\cos\theta \tag{36}$$

注意，上式中 θ 代表向量 $\boldsymbol{a}$ 和 $\boldsymbol{b}$ 的“相对夹角”。

此外，向量内积还可以从投影 (projection) 角度来解释，这是本书第 9 章要介绍的内容。

$\boldsymbol{a}$ 的 L^2 范数也可以通过向量内积求得：

$$\|\boldsymbol{a}\|_2=\|\boldsymbol{a}\|=\sqrt{\boldsymbol{a}\cdot\boldsymbol{a}}=\sqrt{\langle\boldsymbol{a},\boldsymbol{a}\rangle} \tag{37}$$

(37) 左右等式平方得到：

$$\|\boldsymbol{a}\|_2^2=\|\boldsymbol{a}\|^2=\boldsymbol{a}\cdot\boldsymbol{a}=\langle\boldsymbol{a},\boldsymbol{a}\rangle \tag{38}$$

上式相当于“距离的平方”。

柯西-施瓦茨不等式

观察 (36)，我们可以发现 $\cos\theta$ 的取值范围为 $[-1, 1]$，因此 $\boldsymbol{a}$ 和 $\boldsymbol{b}$ 内积取值范围如下：

$$-\|\boldsymbol{a}\|\|\boldsymbol{b}\|\le\boldsymbol{a}\cdot\boldsymbol{b}\le\|\boldsymbol{a}\|\|\boldsymbol{b}\| \tag{39}$$

图 17 所示为 7 个不同向量夹角状态。

$\theta = 0°$时，$\cos\theta = 1$，$\boldsymbol{a}$ 和 $\boldsymbol{b}$ 同向，此时向量内积最大；$\theta = 180°$时，$\cos\theta = -1$，$\boldsymbol{a}$ 和 $\boldsymbol{b}$ 反向，此时向量内积最小。

平面上，非零向量 $\boldsymbol{a}$ 和 $\boldsymbol{b}$ 垂直，$\boldsymbol{a}$ 和 $\boldsymbol{b}$ 夹角为 90°，两者向量内积为 0：

$$\boldsymbol{a}\cdot\boldsymbol{b}=\|\boldsymbol{a}\|\|\boldsymbol{b}\|\cos 90°=0 \tag{40}$$

多维向量 $\boldsymbol{a}$ 和 $\boldsymbol{b}$ 向量内积为 0，我们称 $\boldsymbol{a}$ 和 $\boldsymbol{b}$ **正交** (orthogonal)。本书上一章提到，正交是线性代数的概念，是垂直的推广。

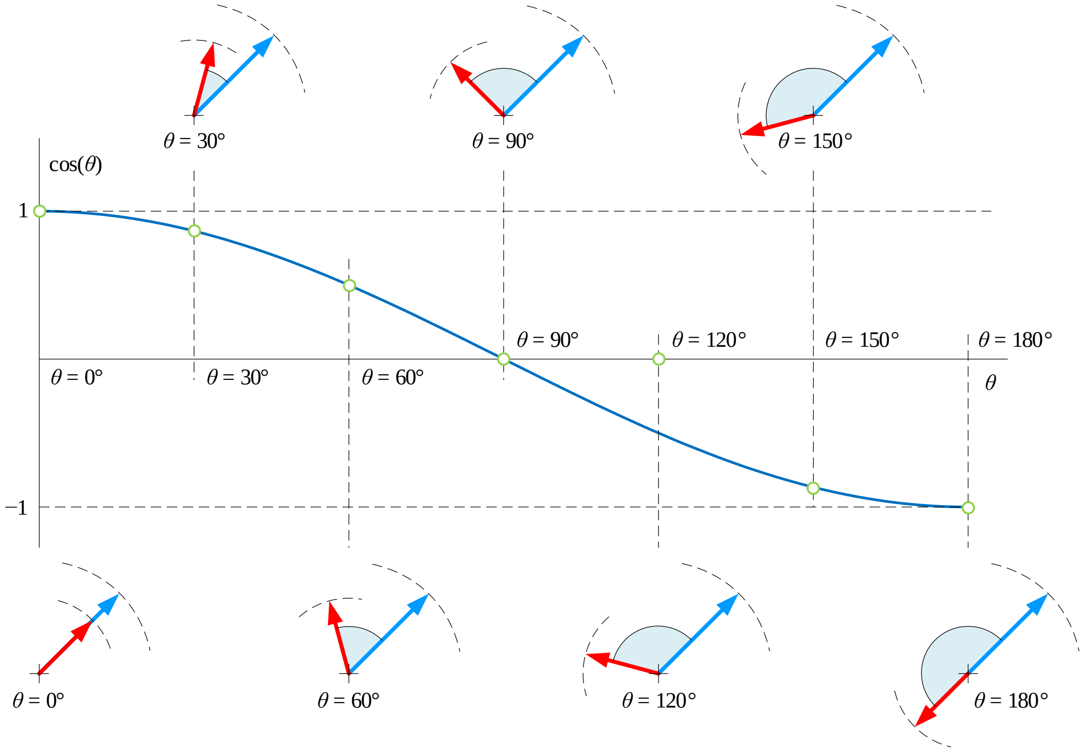


图 17. 向量夹角

有了以上分析，我们就可以引入一个重要的不等式——柯西-施瓦茨不等式 (Cauchy–Schwarz inequality)：

$$(\boldsymbol{a}\cdot\boldsymbol{b})^2 \le \|\boldsymbol{a}\|^2\|\boldsymbol{b}\|^2 \tag{41}$$

即：

$$|\boldsymbol{a}\cdot\boldsymbol{b}| \le \|\boldsymbol{a}\|\|\boldsymbol{b}\| \tag{42}$$

$|\boldsymbol{a}\cdot\boldsymbol{b}|$ 代表 $\boldsymbol{a}$ 和 $\boldsymbol{b}$ 向量内积绝对值。

用尖括号来表达向量内积，(41) 可以写成：

$$\langle\boldsymbol{a},\boldsymbol{b}\rangle^2 \le \langle\boldsymbol{a},\boldsymbol{a}\rangle\langle\boldsymbol{b},\boldsymbol{b}\rangle \tag{43}$$

即：

$$|\langle\boldsymbol{a},\boldsymbol{b}\rangle| \le \|\boldsymbol{a}\|\|\boldsymbol{b}\| \tag{44}$$

在 $\mathbb{R}^n$ 空间中，上述不等式等价于：

$$\left(\sum_{i=1}^{n} a_i b_i\right)^2 \le \left(\sum_{i=1}^{n} a_i^2\right)\left(\sum_{i=1}^{n} b_i^2\right) \tag{45}$$

余弦定理

回忆丛书第一本书讲解的**余弦定理** (law of cosines)：

$$c^2 = a^2 + b^2 - 2ab\cos\theta \tag{46}$$

其中，a、b 和 c 为图 18 所示三角形的三边的边长。下面，我们用余弦定理来用余弦定理推导 (36)。

如图 18 所示，将三角形三个边视作向量，将三个向量长度代入 (46)，可以得到：

$$\|\boldsymbol{c}\|^2 = \|\boldsymbol{a}\|^2 + \|\boldsymbol{b}\|^2 - 2\|\boldsymbol{a}\|\|\boldsymbol{b}\|\cos\theta \tag{47}$$

向量 $\boldsymbol{a}$ 和 $\boldsymbol{b}$ 之差为向量 $\boldsymbol{c}$：

$$\boldsymbol{c} = \boldsymbol{a} - \boldsymbol{b} \tag{48}$$

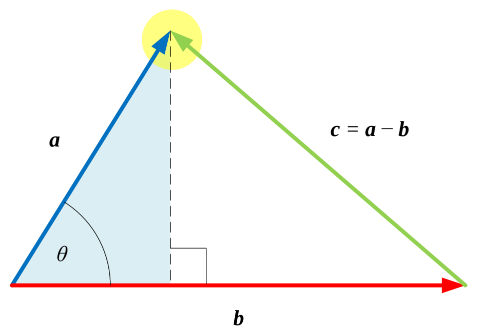


图 18. 余弦定理

(48) 等式左右分别和自身计算向量内积，得到如下等式：

$$\boldsymbol{c}\cdot\boldsymbol{c} = (\boldsymbol{a}-\boldsymbol{b})\cdot(\boldsymbol{a}-\boldsymbol{b}) \tag{49}$$

整理得到：

$$\begin{aligned}\boldsymbol{c}\cdot\boldsymbol{c} &= (\boldsymbol{a}-\boldsymbol{b})\cdot(\boldsymbol{a}-\boldsymbol{b}) = \boldsymbol{a}\cdot\boldsymbol{a} + \boldsymbol{b}\cdot\boldsymbol{b} - \boldsymbol{a}\cdot\boldsymbol{b} - \boldsymbol{b}\cdot\boldsymbol{a} \\ &= \boldsymbol{a}\cdot\boldsymbol{a} + \boldsymbol{b}\cdot\boldsymbol{b} - 2\boldsymbol{a}\cdot\boldsymbol{b}\end{aligned} \tag{50}$$

利用 (38)，(50) 可以写作：

$$\|\boldsymbol{c}\|^2 = \|\boldsymbol{a}\|^2 + \|\boldsymbol{b}\|^2 - 2\boldsymbol{a}\cdot\boldsymbol{b} \tag{51}$$

比较 (47) 和 (51)，可以得到：

$$\boldsymbol{a}\cdot\boldsymbol{b}=\|\boldsymbol{a}\|\|\boldsymbol{b}\|\cos\theta \tag{52}$$

在概率统计、数据分析、机器学习等领域，向量内积无处不在。下面举几个例子。

在多维空间中，给定 A 和 B 坐标如下：

$$A(a_1,a_2,\ldots,a_n),\quad B(b_1,b_2,\ldots,b_n) \tag{53}$$

计算 A 和 B 两点的距离 AB：

$$\begin{aligned} AB &= \sqrt{(a_1-b_1)^2+(a_2-b_2)^2+\cdots+(a_n-b_n)^2} \\ &= \sqrt{\sum_{i=1}^{n}(a_i-b_i)^2} \end{aligned} \tag{54}$$

用起点位于原点的向量 $\boldsymbol{a}$ 和 $\boldsymbol{b}$ 分别代表 A 和 B 点，AB 距离就是 $\boldsymbol{a}-\boldsymbol{b}$ 的 L_2 范数，也就是欧几里得距离：

$$AB=\|\boldsymbol{a}-\boldsymbol{b}\|=\sqrt{(\boldsymbol{a}-\boldsymbol{b})\cdot(\boldsymbol{a}-\boldsymbol{b})}=\sqrt{\boldsymbol{a}\cdot\boldsymbol{a}+\boldsymbol{b}\cdot\boldsymbol{b}-2\boldsymbol{a}\cdot\boldsymbol{b}} \tag{55}$$

回忆《概率统计》一册中介绍的样本方差公式，具体如下：

$$\operatorname{var}(X)=\frac{1}{n-1}\sum_{i=1}^{n}(x_i-\mu)^2 \tag{56}$$

注意，对于总体方差，上式分母中 $n-1$ 改为 n。

令 $\boldsymbol{x}$ 为，

$$\boldsymbol{x}=\begin{bmatrix} x_1 & x_2 & \cdots & x_n \end{bmatrix}^{\mathrm{T}} \tag{57}$$

(56) 可以写成：

$$\operatorname{var}(X)=\frac{(\boldsymbol{x}-\mu)\cdot(\boldsymbol{x}-\mu)}{n-1} \tag{58}$$

根据广播原则，$\boldsymbol{x}-\mu$ 相当于向量 $\boldsymbol{x}$ 的每一个元素分别减去 μ。

回忆总样本协方差公式：

$$\operatorname{cov}(X,Y)=\frac{1}{n-1}\sum_{i=1}^{n}(x_i-\mu_X)(y_i-\mu_Y) \tag{59}$$

同样，对于总体协方差，上式分母中 $n-1$ 改为 n。

同样利用向量内积运算法则，上式可以写成：

$$\operatorname{cov}(X,Y)=\frac{(\boldsymbol{x}-\mu_X)\cdot(\boldsymbol{y}-\mu_Y)}{n-1} \tag{60}$$

本书第 22 章将从线性代数角度再和大家探讨概率统计相关内容。

2.7 向量夹角：反余弦

根据 (36)，可以得到向量 $\boldsymbol{a}$ 和 $\boldsymbol{b}$ 夹角余弦值：

$$\cos\theta=\frac{\boldsymbol{a}\cdot\boldsymbol{b}}{\|\boldsymbol{a}\|\|\boldsymbol{b}\|} \tag{61}$$

通过反余弦，可以得到向量 $\boldsymbol{a}$ 和 $\boldsymbol{b}$ 夹角：

$$\theta=\arccos\left(\frac{\boldsymbol{a}\cdot\boldsymbol{b}}{\|\boldsymbol{a}\|\|\boldsymbol{b}\|}\right) \tag{62}$$

arccos() 为反余弦函数，即从余弦值获得弧度。需要时，可以进一步将弧度转化为角度。再次强调，(62) 代表向量 $\boldsymbol{a}$ 和 $\boldsymbol{b}$ 之间的"相对角度"。而 $\boldsymbol{a}$ 和 $\boldsymbol{e}_1$、$\boldsymbol{b}$ 和 $\boldsymbol{e}_1$ 的夹角可以视作"绝对夹角"。

图 16 中向量 $\boldsymbol{a}$ 和 $\boldsymbol{b}$ 夹角弧度值和角度值可以通过 `Bk4_Ch2_09.py` 计算。

极坐标

下面，我们将向量放在极坐标中解释向量夹角余弦值。给定向量 $\boldsymbol{a}$ 和 $\boldsymbol{b}$ 坐标如下：

$$\boldsymbol{a}=\begin{bmatrix}a_1\\a_2\end{bmatrix},\quad \boldsymbol{b}=\begin{bmatrix}b_1\\b_2\end{bmatrix} \tag{63}$$

向量 $\boldsymbol{a}$ 和 $\boldsymbol{b}$ 在极坐标中各自的角度为 θ_a 和 θ_b。角度 θ_a 和 θ_b 的正弦和余弦可以通过下式计算得到：

$$\begin{cases}\cos\theta_a=\dfrac{a_1}{\|\boldsymbol{a}\|}, & \sin\theta_a=\dfrac{a_2}{\|\boldsymbol{a}\|}\\[2ex] \cos\theta_b=\dfrac{b_1}{\|\boldsymbol{b}\|}, & \sin\theta_b=\dfrac{b_2}{\|\boldsymbol{b}\|}\end{cases} \tag{64}$$

θ_a 和 θ_b 就相当于绝对角度。

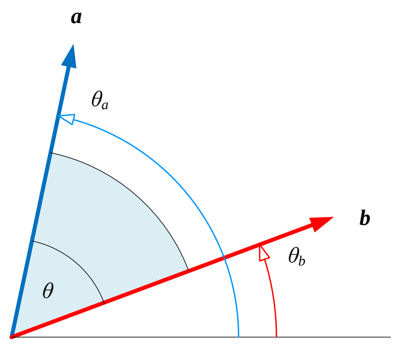


图 19. 极坐标中解释向量夹角

根据角的余弦和差恒等式，$\cos(\theta)$ 可以由 θ_a 和 θ_b 正、余弦构造：

$$\begin{aligned}\cos(\theta) &= \cos(\theta_b - \theta_a) = \cos(\theta_b)\cos(\theta_a) + \sin(\theta_b)\sin(\theta_a) \\ &= \frac{a_1}{\|\boldsymbol{a}\|}\frac{b_1}{\|\boldsymbol{b}\|} + \frac{a_2}{\|\boldsymbol{a}\|}\frac{b_2}{\|\boldsymbol{b}\|} = \frac{a_1 b_1 + a_2 b_2}{\|\boldsymbol{a}\|\|\boldsymbol{b}\|}\end{aligned} \tag{65}$$

将 (64) 代入 (65) 得到：

$$\cos\theta = \frac{a_1}{\|\boldsymbol{a}\|}\frac{b_1}{\|\boldsymbol{b}\|} + \frac{a_2}{\|\boldsymbol{a}\|}\frac{b_2}{\|\boldsymbol{b}\|} = \frac{\overbrace{a_1 b_1 + a_2 b_2}^{\boldsymbol{a}\cdot\boldsymbol{b}}}{\|\boldsymbol{a}\|\|\boldsymbol{b}\|} \tag{66}$$

相信大家已经在上式分母中看到向量内积。

单位向量

本章前文介绍过某一向量方向上的单位向量这个概念，单位向量为我们提供了观察向量夹角余弦值的另外一个视角。

给定两个非 $\boldsymbol{0}$ 向量 $\boldsymbol{a}$ 和 $\boldsymbol{b}$，首先计算它们各自方向上的单位向量：

$$\hat{\boldsymbol{a}} = \frac{\boldsymbol{a}}{\|\boldsymbol{a}\|}, \quad \hat{\boldsymbol{b}} = \frac{\boldsymbol{b}}{\|\boldsymbol{b}\|} \tag{67}$$

两个单位向量的内积就是夹角的余弦值：

$$\hat{\boldsymbol{a}}\cdot\hat{\boldsymbol{b}} = \frac{\boldsymbol{a}}{\|\boldsymbol{a}\|}\cdot\frac{\boldsymbol{b}}{\|\boldsymbol{b}\|} = \cos\theta \tag{68}$$

正交单位向量

本章前文介绍的平面直角坐标系中 $\boldsymbol{e}_1$ 和 $\boldsymbol{e}_2$ 分别代表为沿着 x_1 和 x_2 单位向量。它们相互正交，也就是向量内积为 0：

$$\boldsymbol{e}_1 \cdot \boldsymbol{e}_2 = \langle \boldsymbol{e}_1, \boldsymbol{e}_2 \rangle = \begin{bmatrix} 1 \\ 0 \end{bmatrix} \cdot \begin{bmatrix} 0 \\ 1 \end{bmatrix} = 0 \tag{69}$$

在一个平面上，单位向量 $\boldsymbol{e}_1$、$\boldsymbol{e}_2$ 相互垂直，它俩“张起”的方方正正的网格，就是标准直角坐标系，具体如图 20 (a) 所示。

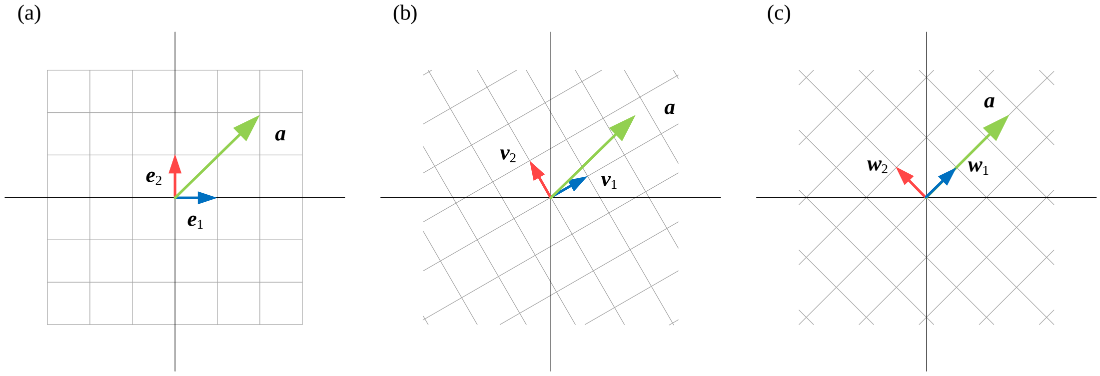


图 20. 向量 $\boldsymbol{a}$ 在三个不同的正交直角坐标系中位置

而平面上，成对正交单位向量有无数组，比如图 21 所示平面两组正交单位向量：

$$\boldsymbol{v}_1 \cdot \boldsymbol{v}_2 = \begin{bmatrix} \frac{\sqrt{3}}{2} \\ \frac{1}{2} \end{bmatrix} \cdot \begin{bmatrix} -\frac{1}{2} \\ \frac{\sqrt{3}}{2} \end{bmatrix} = 0, \quad \boldsymbol{w}_1 \cdot \boldsymbol{w}_2 = \begin{bmatrix} \frac{\sqrt{2}}{2} \\ \frac{\sqrt{2}}{2} \end{bmatrix} \cdot \begin{bmatrix} -\frac{\sqrt{2}}{2} \\ \frac{\sqrt{2}}{2} \end{bmatrix} = 0 \tag{70}$$

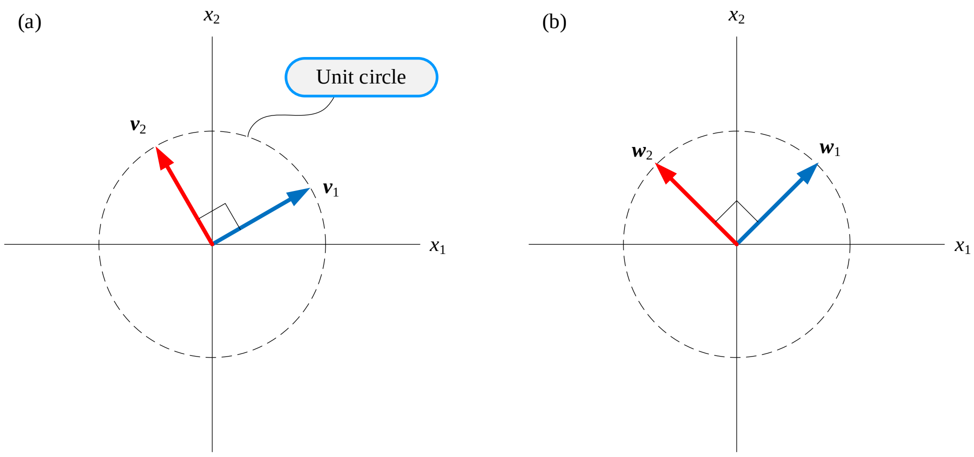


图 21. 两组正交单位向量

$\boldsymbol{v}_1$、$\boldsymbol{v}_2$ 构造如图 20 (b) 所示直角坐标系。类似地，$\boldsymbol{w}_1$、$\boldsymbol{w}_2$ 也可以构造如图 20 (c) 所示直角坐标系。也就是一个 $\mathbb{R}^2$ 平面上可以存在无数个直角坐标系。

比较图 20 三幅子图，同一个向量 $\boldsymbol{a}$ 在三个直角坐标系中有不同的坐标值。向量 $\boldsymbol{a}$ 在图 20 (a) 所示直角坐标系的坐标值很容易确定为 (2, 2)。目前我们还没有掌握足够的数学工具来计算向量 $\boldsymbol{a}$ 在图 20 (b) 和 (c) 两个直角坐标系中的坐标值。这个问题要留到本书第 7 章来解决。

[$\boldsymbol{e}_1$, $\boldsymbol{e}_2$]、[$\boldsymbol{v}_1$, $\boldsymbol{v}_2$]、[$\boldsymbol{w}_1$, $\boldsymbol{w}_2$] 都叫做 $\mathbb{R}^2$ 的规范正交基 (orthonormal basis)，而 [$\boldsymbol{e}_1$, $\boldsymbol{e}_2$] 有自己特别的名字——标准正交基 (standard basis)。而且大家很快就会发现 [$\boldsymbol{e}_1$, $\boldsymbol{e}_2$] 旋转一定角度可以得到[$\boldsymbol{v}_1$, $\boldsymbol{v}_2$]、[$\boldsymbol{w}_1$, $\boldsymbol{w}_2$]。本书第 7 章将深入介绍相关概念。

2.8 余弦相似度和余弦距离

机器学习中有一个重要的概念，叫做**余弦相似度** (cosine similarity)。余弦相似度用向量夹角的余弦值度量样本数据的相似性。

用 $k(\boldsymbol{x}, \boldsymbol{q})$ 来表达 $\boldsymbol{x}$ 和 $\boldsymbol{q}$ 两个列向量的余弦相似度，定义如下：

$$k(\boldsymbol{x},\boldsymbol{q}) = \frac{\boldsymbol{x}\cdot\boldsymbol{q}}{\|\boldsymbol{x}\|\|\boldsymbol{q}\|} = \frac{\boldsymbol{x}^{\mathrm{T}}\boldsymbol{q}}{\|\boldsymbol{x}\|\|\boldsymbol{q}\|} \tag{71}$$

上一节我们介绍过，如果两个向量方向相同，则夹角 θ 余弦值 $\cos(\theta) = 1$。若两个向量方向完全相反，夹角 θ 余弦值 $\cos(\theta) = -1$。

因此，余弦相似度取值范围在 [−1, +1] 之间。此外，大家是否在余弦相似度中看到相关性系数的影子？

下面再介绍**余弦距离** (cosine distance)。余弦距离定义基于余弦相似度，用 $d(\boldsymbol{x}, \boldsymbol{q})$ 来表达 $\boldsymbol{x}$ 和 $\boldsymbol{q}$ 两个列向量的余弦距离，具体定义如下：

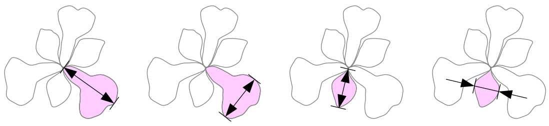
$$d(\boldsymbol{x},\boldsymbol{q}) = 1 - k(\boldsymbol{x},\boldsymbol{q}) = 1 - \frac{\boldsymbol{x}\cdot\boldsymbol{q}}{\|\boldsymbol{x}\|\|\boldsymbol{q}\|} \tag{72}$$

本章前文介绍的欧几里得距离，即 L^2 范数，是一种最常见的距离度量。本节介绍的余弦距离也是一种常见的距离度量。

本书下一章，以及《概率统计》、《机器学习》将逐步介绍常见距离度量，“距离”的内涵会不断丰富。

鸢尾花例子

图 22 给出鸢尾花四个样本数据。$\boldsymbol{x}^{(1)}$ 和 $\boldsymbol{x}^{(2)}$ 两个样本对应的鸢尾花都是 setosa 这一亚属。$\boldsymbol{x}^{(51)}$ 样本对应的鸢尾花为 versicolor 这一亚属；$\boldsymbol{x}^{(101)}$ 样本对应的鸢尾花为 virginica 这一亚属。



	Sepal length	Sepal width	Petal length	Petal width	
$\boldsymbol{x}^{(1)}$, 1	5.1	3.5	1.4	0.2	setosa
$\boldsymbol{x}^{(2)}$, 2	4.9	3	1.4	0.2	setosa
$\boldsymbol{x}^{(51)}$, 51	7	3.2	4.7	1.4	versicolor
$\boldsymbol{x}^{(101)}$, 101	6.3	3.3	6	2.5	virginica

图 22. 鸢尾花的四个样本数据

计算 $\boldsymbol{x}^{(1)}$ 和 $\boldsymbol{x}^{(2)}$ 两个向量余弦距离：

$$
\begin{aligned}
& d\left(\boldsymbol{x}^{(1)}, \boldsymbol{x}^{(2)}\right)=1-k\left(\boldsymbol{x}^{(1)}, \boldsymbol{x}^{(2)}\right) \\
& =1-\frac{5.1\times 4.9+3.5\times 3 + 1.4\times 1.4 + 0.2\times 0.2}{\sqrt{5.1^2+3.5^2 + 1.4^2 + 0.2^2}\times\sqrt{4.9^2+3^2 + 1.4^2 + 0.2^2}} \\
& =1-\frac{37.49}{6.34507\times 5.9169} \\
& =1-0.99857=0.00142
\end{aligned}
\tag{73}
$$

同理，可以计算得到 $\boldsymbol{x}^{(1)}$ 和 $\boldsymbol{x}^{(51)}$，$\boldsymbol{x}^{(1)}$ 和 $\boldsymbol{x}^{(101)}$ 两各余弦距离：

$$
\begin{aligned}
& d\left(\boldsymbol{x}^{(1)}, \boldsymbol{x}^{(51)}\right)=0.07161 \\
& d\left(\boldsymbol{x}^{(1)}, \boldsymbol{x}^{(101)}\right)=0.13991
\end{aligned}
\tag{74}
$$

可以发现，$\boldsymbol{x}^{(1)}$ 和 $\boldsymbol{x}^{(2)}$ 两朵同属于 setosa 亚属的鸢尾花，余弦距离较近，也就是较为相似。

$\boldsymbol{x}^{(1)}$ 和 $\boldsymbol{x}^{(101)}$ 分别属于 setota 和 virginica 亚属，余弦距离较远，也就是不相似。

给大家留个思考题，鸢尾花数据有 150 个数据点，任意两个数据点可以计算得到一个余弦相似度。因此成对余弦相似度有 11175 个，大家想想该怎么便捷计算、存储这些数据呢？

`Bk4_Ch2_10.py` 可以完成上述计算。感兴趣的读者可以修改代码计算 $\boldsymbol{x}^{(51)}$ 和 $\boldsymbol{x}^{(101)}$ 的余弦距离，并结合样本标签分析结果。

2.9 向量积：结果为向量

向量积 (vector product) 也叫**叉乘** (cross product) 或外积，向量积结果为向量。也就是说，向量积一种“向量 → 向量”的运算规则。

$\boldsymbol{a}$ 和 $\boldsymbol{b}$ 向量积，记做 $\boldsymbol{a} \times \boldsymbol{b}$。$\boldsymbol{a} \times \boldsymbol{b}$ 作为一个向量，我们需要了解它的方向和大小两个成分。

方向

如图 23 所示，$\boldsymbol{a} \times \boldsymbol{b}$ 方向分别垂直于向量 $\boldsymbol{a}$ 和 $\boldsymbol{b}$，即 $\boldsymbol{a} \times \boldsymbol{b}$ 垂直于向量 $\boldsymbol{a}$ 和 $\boldsymbol{b}$ 构成平面。

向量 $\boldsymbol{a}$ 和 $\boldsymbol{b}$ 以及 $\boldsymbol{a} \times \boldsymbol{b}$ 三者关系可以用右手法则判断，如图 24 所示。图 24 这幅图中，我们可以看到 $\boldsymbol{a} \times \boldsymbol{b}$ 和 $\boldsymbol{b} \times \boldsymbol{a}$ 方向相反。

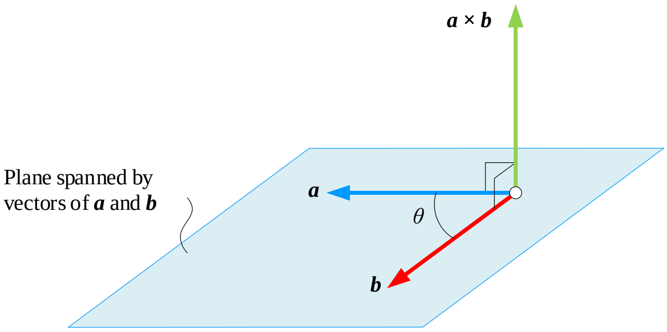


图 23. $\boldsymbol{a} \times \boldsymbol{b}$ 垂直于向量 $\boldsymbol{a}$ 和 $\boldsymbol{b}$ 构成平面

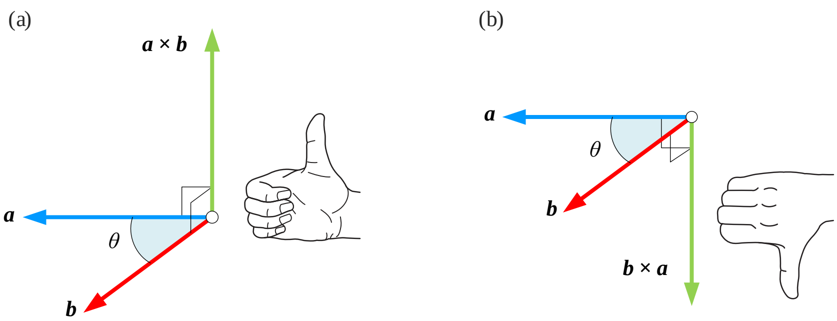


图 24. 向量叉乘右手定则

大小

$\boldsymbol{a} \times \boldsymbol{b}$ 模，也就是 $\boldsymbol{a} \times \boldsymbol{b}$ 向量积大小，通过下式获得：

$$\|\boldsymbol{a}\times\boldsymbol{b}\| = \|\boldsymbol{a}\|\|\boldsymbol{b}\|\sin(\theta) \tag{75}$$

其中 θ 为向量 $\boldsymbol{a}$ 和 $\boldsymbol{b}$ 夹角。如图 25 所示，从几何角度，向量积的模 $\|\boldsymbol{a}\times\boldsymbol{b}\|$ 相当于图中平行四边形的面积。

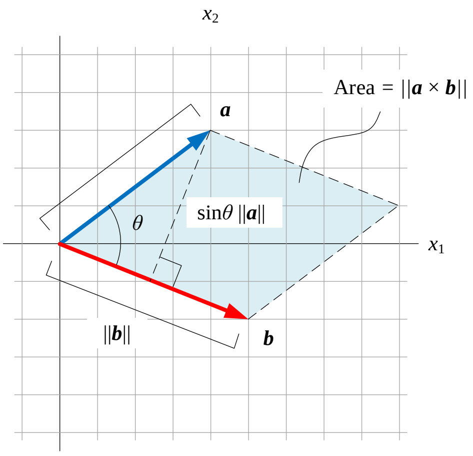


图 25. $\boldsymbol{a}\times\boldsymbol{b}$ 向量积的几何含义

正交向量之间的叉乘

如图 26 (a) 所示，空间直角坐标系中三个正交向量 $\boldsymbol{e}_1$ ($\boldsymbol{i}$) (x_1 轴正方向)、$\boldsymbol{e}_2$ ($\boldsymbol{j}$) (x_2 轴正方向) 和 $\boldsymbol{e}_3$ ($\boldsymbol{k}$) (x_3 轴正方向) 向量叉乘关系存在如下关系：

$$\boldsymbol{i}\times\boldsymbol{j}=\boldsymbol{k},\quad \boldsymbol{j}\times\boldsymbol{k}=\boldsymbol{i},\quad \boldsymbol{k}\times\boldsymbol{i}=\boldsymbol{j} \tag{76}$$

图 26 (b) 展示以上三个等式中 $\boldsymbol{i}$、$\boldsymbol{j}$ 和 $\boldsymbol{k}$ 前后顺序关系。若调换 (76) 叉乘元素顺序，结果反向，对应以下三个运算式：

$$\boldsymbol{j}\times\boldsymbol{i}=-\boldsymbol{k},\quad \boldsymbol{k}\times\boldsymbol{j}=-\boldsymbol{i},\quad \boldsymbol{i}\times\boldsymbol{k}=-\boldsymbol{j} \tag{77}$$

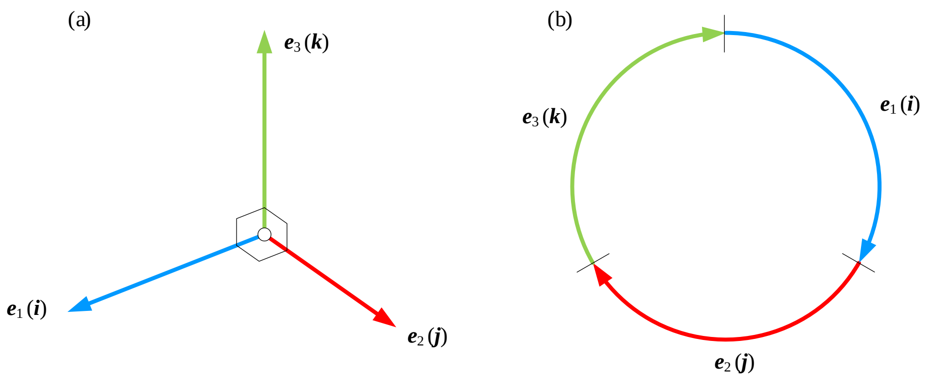


图 26. 三维空间正交单位向量基底之间关系

特别的，向量与自身叉乘等于 $\boldsymbol{0}$ 向量，比如：

$$\boldsymbol{i} \times \boldsymbol{i} = \boldsymbol{0}, \quad \boldsymbol{j} \times \boldsymbol{j} = \boldsymbol{0}, \quad \boldsymbol{k} \times \boldsymbol{k} = \boldsymbol{0} \tag{78}$$

下列为叉乘运算常见性质：

$$\begin{aligned}
&\boldsymbol{a} \times \boldsymbol{a} = \boldsymbol{0} \\
&\boldsymbol{a} \times (\boldsymbol{b} + \boldsymbol{c}) = \boldsymbol{a} \times \boldsymbol{b} + \boldsymbol{a} \times \boldsymbol{c} \\
&(\boldsymbol{a} + \boldsymbol{b}) \times \boldsymbol{c} = \boldsymbol{a} \times \boldsymbol{b} + \boldsymbol{a} \times \boldsymbol{c} \\
&\boldsymbol{a} \times (\boldsymbol{b} \times \boldsymbol{c}) \neq (\boldsymbol{a} \times \boldsymbol{b}) \times \boldsymbol{c} \\
&k(\boldsymbol{a} \times \boldsymbol{b}) = k(\boldsymbol{a}) \times \boldsymbol{b} = \boldsymbol{a} \times (k\boldsymbol{b}) \\
&\boldsymbol{a} \cdot (\boldsymbol{b} \times \boldsymbol{c}) = (\boldsymbol{a} \times \boldsymbol{b}) \cdot \boldsymbol{c}
\end{aligned} \tag{79}$$

任意两个向量的叉乘

在三维直角坐标系中，用 $\boldsymbol{i}$、$\boldsymbol{j}$ 和 $\boldsymbol{k}$ 表达向量 $\boldsymbol{a}$ 和 $\boldsymbol{b}$：

$$\begin{aligned}
\boldsymbol{a} &= a_1\boldsymbol{i} + a_2\boldsymbol{j} + a_3\boldsymbol{k} \\
\boldsymbol{b} &= b_1\boldsymbol{i} + b_2\boldsymbol{j} + b_3\boldsymbol{k}
\end{aligned} \tag{80}$$

整理向量 $\boldsymbol{a}$ 和 $\boldsymbol{b}$ 叉乘，如下：

$$\begin{aligned}
\boldsymbol{a} \times \boldsymbol{b} &= (a_1\boldsymbol{i} + a_2\boldsymbol{j} + a_3\boldsymbol{k}) \times (b_1\boldsymbol{i} + b_2\boldsymbol{j} + b_3\boldsymbol{k}) \\
&= a_1b_1(\boldsymbol{i} \times \boldsymbol{i}) + a_1b_2(\boldsymbol{i} \times \boldsymbol{j}) + a_1b_3(\boldsymbol{i} \times \boldsymbol{k}) \\
&\quad + a_2b_1(\boldsymbol{j} \times \boldsymbol{i}) + a_2b_2(\boldsymbol{j} \times \boldsymbol{j}) + a_2b_3(\boldsymbol{j} \times \boldsymbol{k}) \\
&\quad + a_3b_1(\boldsymbol{k} \times \boldsymbol{i}) + a_3b_2(\boldsymbol{k} \times \boldsymbol{j}) + a_3b_3(\boldsymbol{k} \times \boldsymbol{k}) \\
&= (a_2b_3 - a_3b_2)\boldsymbol{i} + (a_3b_1 - a_1b_3)\boldsymbol{j} + (a_1b_2 - a_2b_1)\boldsymbol{k}
\end{aligned} \tag{81}$$

$\boldsymbol{a}$ 和 $\boldsymbol{b}$ 叉乘还可以通过行列式求解，我们将在本书第 4 章讲解。

举个例子

下面结合代码计算 $\boldsymbol{a}$ 和 $\boldsymbol{b}$ 两个向量叉乘：

$$\boldsymbol{a} = \begin{bmatrix} -2 \\ 1 \\ 1 \end{bmatrix}, \quad \boldsymbol{b} = \begin{bmatrix} 1 \\ -2 \\ -1 \end{bmatrix} \tag{82}$$

$\boldsymbol{a} \times \boldsymbol{b}$ 结果如下：

$$\boldsymbol{a} \times \boldsymbol{b} = \begin{bmatrix} 1 \\ -1 \\ 3 \end{bmatrix} \tag{83}$$

Bk4_Ch2_11.py 计算得到 (83)。其中，numpy.cross() 函数可以用来计算列向量和行向量的向量积。

2.10 逐项积：对应元素分别相乘

元素乘积 (element-wise multiplication)，也称为**阿达玛乘积** (Hadamard product) 或**逐项积** (piecewise product)。逐项积指的是两个形状相同的矩阵，对应元素相乘得到同样形状的矩阵。向量是一种特殊矩阵，阿达玛乘积也适用于向量。图 27 给出的是从数据角度看向量逐项积运算。

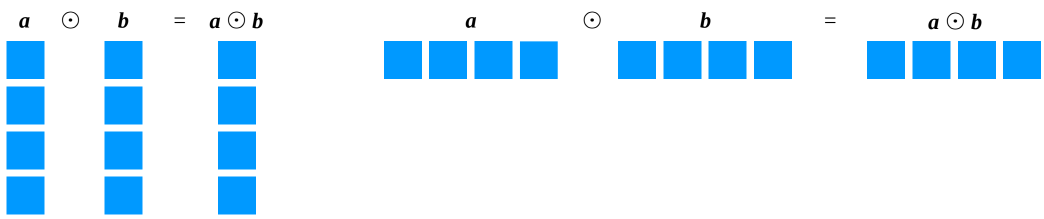


图 27. 向量逐项积运算

给定如下 $\boldsymbol{a}$ 和 $\boldsymbol{b}$ 两个列向量：

$$\begin{aligned} \boldsymbol{a} &= \begin{bmatrix} a_1 & a_2 & \cdots & a_n \end{bmatrix}^{\mathrm{T}} \\ \boldsymbol{b} &= \begin{bmatrix} b_1 & b_2 & \cdots & b_n \end{bmatrix}^{\mathrm{T}} \end{aligned} \tag{84}$$

列向量 $\boldsymbol{a}$ 和 $\boldsymbol{b}$ 的逐项积定义如下：

$$\boldsymbol{a} \odot \boldsymbol{b} = \begin{bmatrix} a_1 b_1 & a_2 b_2 & \cdots & a_n b_n \end{bmatrix}^{\mathrm{T}} \tag{85}$$

逐项积是一种"向量 → 向量"的运算规则。

Bk4_Ch2_12.py 计算行向量逐项积。

2.11 向量张量积：张起网格面

张量积 (tensor product) 又叫**克罗内克积** (Kronecker product)，两个列向量 $\boldsymbol{a}$ 和 $\boldsymbol{b}$ 张量积 $\boldsymbol{a}\otimes\boldsymbol{b}$ 定义如下：

$$\boldsymbol{a}\otimes\boldsymbol{b}=\begin{bmatrix}a_1\\a_2\\\vdots\\a_n\end{bmatrix}_{n\times1}\otimes\begin{bmatrix}b_1\\b_2\\\vdots\\b_m\end{bmatrix}_{m\times1}=\boldsymbol{a}\boldsymbol{b}^{\mathrm{T}}=\begin{bmatrix}a_1\\a_2\\\vdots\\a_n\end{bmatrix}\begin{bmatrix}b_1\\b_2\\\vdots\\b_m\end{bmatrix}^{\mathrm{T}}=\begin{bmatrix}a_1b_1&a_1b_2&\cdots&a_1b_m\\a_2b_1&a_2b_2&\cdots&a_2b_m\\\vdots&\vdots&\ddots&\vdots\\a_nb_1&a_nb_2&\cdots&a_nb_m\end{bmatrix}_{n\times m}\tag{86}$$

向量张量积是一种“向量 → 矩阵”的运算规则。

⚠ 注意，(86) 上式中 $\boldsymbol{a}\boldsymbol{b}^{\mathrm{T}}$ 为向量 $\boldsymbol{a}$ 和 $\boldsymbol{b}^{\mathrm{T}}$ 的矩阵乘法。本书第 4、5、6 三章要从不同角度讲解矩阵乘法。

向量 $\boldsymbol{a}$ 和其自身张量积 $\boldsymbol{a}\otimes\boldsymbol{a}$ 结果为方阵：

$$\boldsymbol{a}\otimes\boldsymbol{a}=\begin{bmatrix}a_1\\a_2\\\vdots\\a_n\end{bmatrix}_{n\times1}\otimes\begin{bmatrix}a_1\\a_2\\\vdots\\a_n\end{bmatrix}_{n\times1}=\boldsymbol{a}\boldsymbol{a}^{\mathrm{T}}=\begin{bmatrix}a_1\\a_2\\\vdots\\a_n\end{bmatrix}\begin{bmatrix}a_1\\a_2\\\vdots\\a_n\end{bmatrix}^{\mathrm{T}}=\begin{bmatrix}a_1a_1&a_1a_2&\cdots&a_1a_n\\a_2a_1&a_2a_2&\cdots&a_2a_n\\\vdots&\vdots&\ddots&\vdots\\a_na_1&a_na_2&\cdots&a_na_n\end{bmatrix}\tag{87}$$

请大家注意张量积一些常见性质：

$$\begin{aligned}&(\boldsymbol{a}\otimes\boldsymbol{a})^{\mathrm{T}}=\boldsymbol{a}\otimes\boldsymbol{a}\\&(\boldsymbol{a}\otimes\boldsymbol{b})^{\mathrm{T}}=\boldsymbol{b}\otimes\boldsymbol{a}\\&(\boldsymbol{a}+\boldsymbol{b})\otimes\boldsymbol{v}=\boldsymbol{a}\otimes\boldsymbol{v}+\boldsymbol{b}\otimes\boldsymbol{v}\\&\boldsymbol{v}\otimes(\boldsymbol{a}+\boldsymbol{b})=\boldsymbol{v}\otimes\boldsymbol{a}+\boldsymbol{v}\otimes\boldsymbol{b}\\&t(\boldsymbol{a}\otimes\boldsymbol{b})=(t\boldsymbol{a})\otimes\boldsymbol{b}=\boldsymbol{a}\otimes(t\boldsymbol{b})\\&(\boldsymbol{a}\otimes\boldsymbol{b})\otimes\boldsymbol{v}=\boldsymbol{a}\otimes(\boldsymbol{b}\otimes\boldsymbol{v})\end{aligned}\tag{88}$$

几何视角

图 28 所示为从几何图像角度解释向量的张量积。向量 $\boldsymbol{a}$ 和 $\boldsymbol{b}$ 相当于两个维度上的支撑框架，两者的张量积则“张起”一个网格面 $\boldsymbol{a}\otimes\boldsymbol{b}$。

当我们关注 $\boldsymbol{b}$ 方向时，网格面沿同一方向的每一条曲线都类似 $\boldsymbol{b}$，唯一的差别是高度上存在一定比例的缩放，这个缩放比例就是 a_i。a_i 是向量 $\boldsymbol{a}$ 中的某一个元素。

同理，观察 $\boldsymbol{a}$ 方向的网格面，每一条曲线都类似 $\boldsymbol{a}$。向量 $\boldsymbol{b}$ 的某一元素 b_j 提供曲线高度的缩放系数。

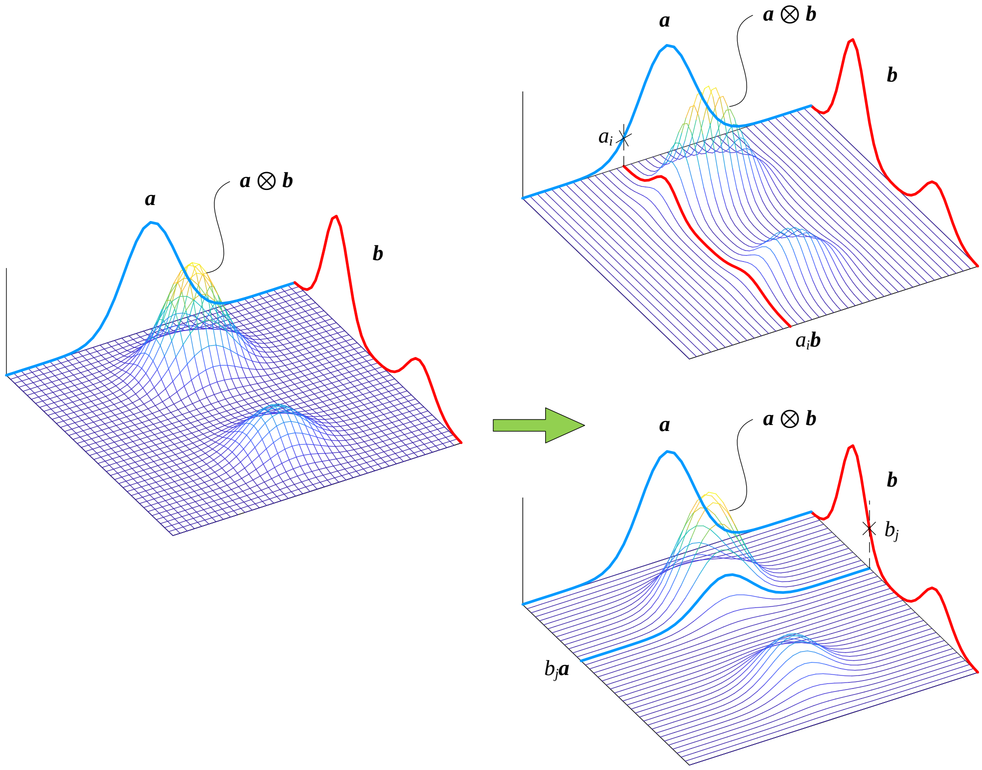


图 28. 从几何角度解释向量张量积

举个例子

给定列向量 $\boldsymbol{a}$ 和 $\boldsymbol{b}$ 分别为：

$$\begin{aligned}\boldsymbol{a} &= \begin{bmatrix}0.5 & -0.7 & 1 & 0.25 & -0.6 & -1\end{bmatrix}^{\mathrm{T}} \\ \boldsymbol{b} &= \begin{bmatrix}-0.8 & 0.5 & -0.6 & 0.9\end{bmatrix}^{\mathrm{T}}\end{aligned} \tag{89}$$

图 29 所示为张量积 $\boldsymbol{a} \otimes \boldsymbol{b}$ 结果热图，形状为 6×4。

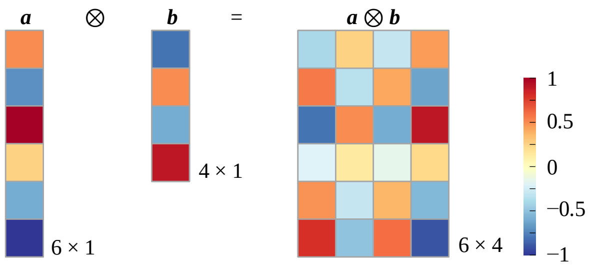


图 29. 张量积 $\boldsymbol{a} \otimes \boldsymbol{b}$ 热图

观察 (86)，利用矩阵乘法展开，发现 $\boldsymbol{a} \otimes \boldsymbol{b}$ 可以写成两种形式：

$$
\boldsymbol{a} \otimes \boldsymbol{b} = \begin{bmatrix} a_1 \boldsymbol{b}^{\mathrm{T}} \\ a_2 \boldsymbol{b}^{\mathrm{T}} \\ \vdots \\ a_n \boldsymbol{b}^{\mathrm{T}} \end{bmatrix}_{n \times 1}
$$

$$
\boldsymbol{a} \otimes \boldsymbol{b} = \begin{bmatrix} b_1 \boldsymbol{a} & b_2 \boldsymbol{a} & \cdots & b_1 \boldsymbol{a} \end{bmatrix} \tag{90}
$$

上式中，第一种形式相当于，$\boldsymbol{b}^{\mathrm{T}}$ 先按不同比例 (a_i) 缩放得到 $a_i \boldsymbol{b}^{\mathrm{T}}$，再上下叠加。

第二种形式相当于，$\boldsymbol{a}$ 先按不同比例 (b_j) 缩放得到 $b_j \boldsymbol{a}$，再左右排列。如果对于 (90) 这种矩阵乘法展开方式感到陌生，请大家读完第 4 ~ 6 章后再回头看这部分内容。

如图 30 (a) 所示，$\boldsymbol{a} \otimes \boldsymbol{b}$ 的每一列都和 $\boldsymbol{a}$“相似”，也就是说它们之间呈现倍数关系。

类似地，如图 30 (b) 所示，$\boldsymbol{a} \otimes \boldsymbol{b}$ 等价于 $\boldsymbol{a}\boldsymbol{b}^{\mathrm{T}}$，因此 $\boldsymbol{a} \otimes \boldsymbol{b}$ 每一行都和 $\boldsymbol{b}^{\mathrm{T}}$“相似”，也呈现倍数关系。

本书第 7 章会聊到向量的秩 (rank)，大家就会知道 $\boldsymbol{a} \otimes \boldsymbol{b}$ 的秩为 1，就是因为行、列这种“相似”。

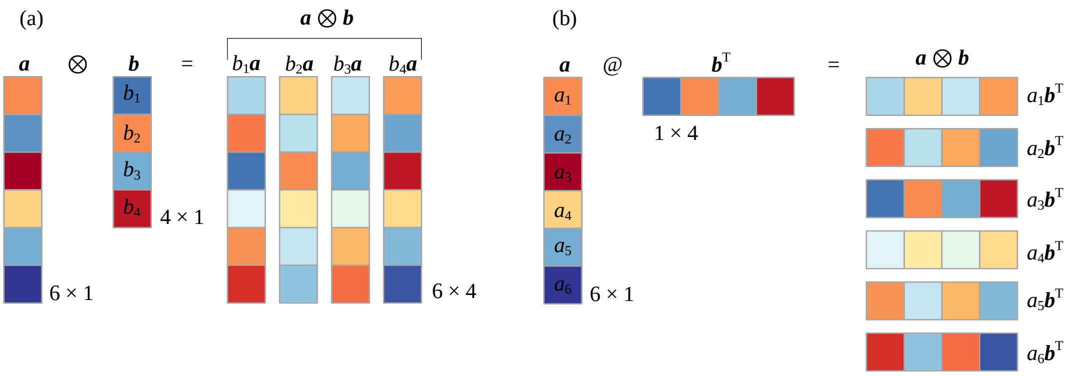


图 30. $\boldsymbol{a} \otimes \boldsymbol{b}$ 的列、行存在的相似

图 31 (a) 所示为张量积 $\boldsymbol{a} \otimes \boldsymbol{a}$ 结果热图，形状为 6 × 6 方阵。图 31 (b) 所示为张量积 $\boldsymbol{b} \otimes \boldsymbol{b}$ 结果热图，形状为 4 × 4 对称方阵。显然，$\boldsymbol{a} \otimes \boldsymbol{a}$ 和 $\boldsymbol{b} \otimes \boldsymbol{b}$ 都是对称矩阵。

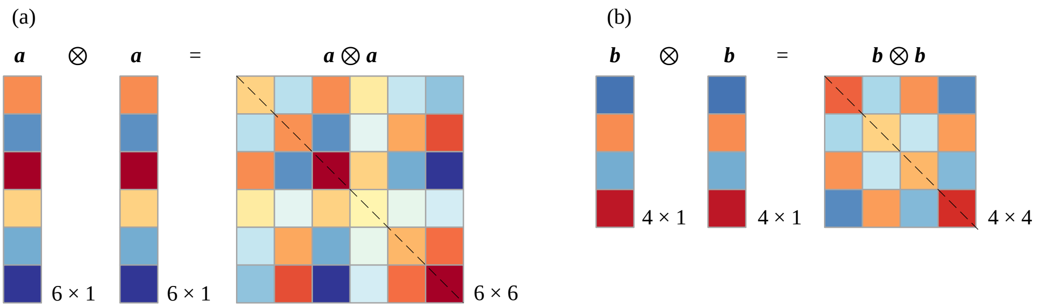


图 31. $\boldsymbol{a} \otimes \boldsymbol{a}$ 和 $\boldsymbol{b} \otimes \boldsymbol{b}$ 向量张量积

Bk4_Ch2_13.py 绘制图 29、图 30、图 31。

《概率统计》将介绍，如果两个离散随机变量 X 和 Y 独立，联合概率 $p_{X,Y}(x,y)$ 等于 $p_X(x)$ 和 $p_Y(y)$ 这两个边缘概率质量函数 PMF 乘积：

$$\underbrace{p_{X,Y}(x,y)}_{\text{Joint}} = \underbrace{p_X(x)}_{\text{Marginal}} \cdot \underbrace{p_Y(y)}_{\text{Marginal}} \tag{91}$$

如图 32 所示，$p_X(x)$ 和 $p_Y(y)$ 可以分别用火柴梗图可视化，而 $p_{X,Y}(x,y)$ 用二维火柴梗图展示。

从线性代数角度，当 x 和 y 分别取不同值时，$p_X(x)$ 和 $p_Y(y)$ 相当于两个向量，而 $p_{X,Y}(x,y)$ 相当于矩阵。X 和 Y 独立时，$p_{X,Y}(x,y)$ 值的矩阵就是 $p_Y(y)$ 和 $p_X(x)$ 两个向量的张量积。

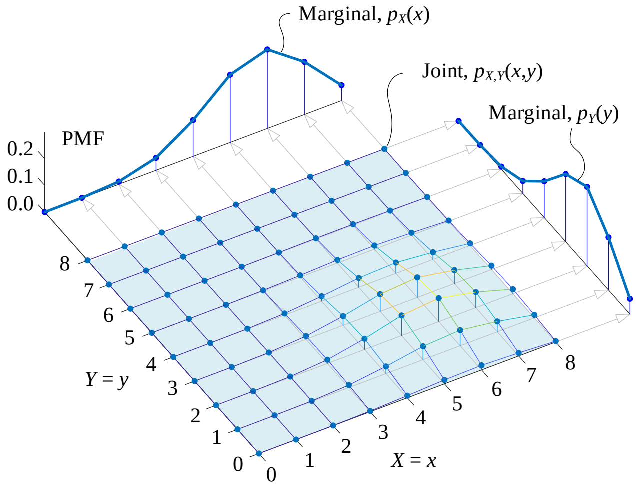


图 32. 离散随机变量独立条件下，联合概率 $p_{X,Y}(x,y)$ 等于 $p_Y(y)$ 和 $p_X(x)$ 乘积

本章聊了聊向量常见运算。学完本章，希望大家看到任何向量和向量运算，可以试着从几何、数据两个角度来思考。

从几何角度，向量是既有长度又有方向的量。从数据角度，表格数据就是矩阵。而矩阵的每一行向量是一个样本点，每一列向量代表一个特征。

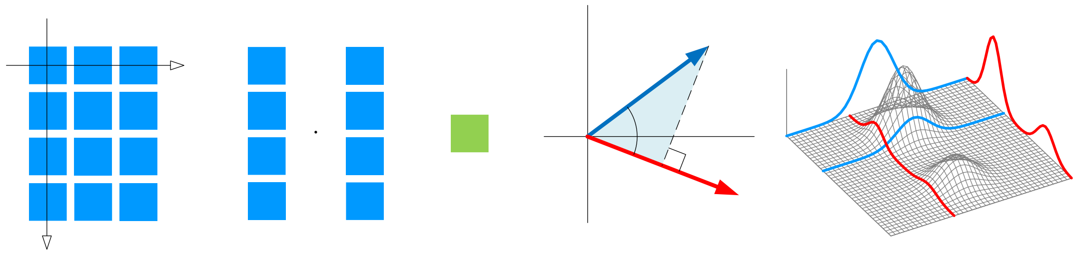


图 33. 总结本章重要内容的四副图

向量有两个元素——长度和方向。向量的长度就是向量的模，向量之间的相对角度可以用向量内积来求解。

提到向量模、L^2 范数、欧几里得距离，希望大家能够联想到正圆、正圆球。本书第 3 章还要介绍更多范数以及它们对应的几何图像。

向量内积的结果是个标量，请大家格外注意向量内积和矩阵乘法联系，以及和 Σ 求和运算之间的关系。

从几何视角看向量内积特别重要，请大家格外关注向量夹角余弦值、余弦定理、余弦相似度、余弦距离，以及本书后续要讲的标量投影、向量投影、协方差、相关性系数等数学概念之间的关系。

向量的外积结果还是个向量，这个向量垂直于原来两个向量构成的平面。

几何视角下，张量积像是张起一个网格面。张量积在机器学习和数据科学算法中应用特别广泛，有关这个运算的性质我们会慢慢展开讲解。