

# 人工智能的数学基础

第1章 特征向量与矩阵分析 (4学时)	第6章 线性分析与卷积 (4学时)
第2章 相似性度量 (2学时)	第7章 正则化与范数 (4学时)
第3章 函数与范函分析 (2学时)	第8章 最优化理论 (4学时)
第4章 条件概率与贝叶斯 (4学时)	第9章 核函数映射 (4学时)
第5章 信息论与熵 (2学时)	第10章 性能评估与度量 (2学时)

补充：符号逻辑，图论

## 第一章 特征向量<sup>[1]</sup>与矩阵分析

### 向量基础

由单一数值构成的对待研究对象的量化评价，称作**标量**。标量的定义与其代表的数据类型强相关。一个用于描述某个对象的多维度特征的有序集合称为**特征向量**。用 $\mathbf{x}$ 标注。由各个特征可能的取值张成的空间，称作“特征空间”。显然，特征空间限制了特征向量的取值范围。

$$\mathbf{x} = [x_1, x_2, x_3, \dots, x_d]$$

给定任一向量，其包含**大小**<sup>[2]</sup>与**方向**两类信息。

$$\mathbf{x}_1 = [x_{1,1}, x_{1,2}, x_{1,3}, \dots, x_{1,d}] \quad \mathbf{x}_2 = [x_{2,1}, x_{2,2}, x_{2,3}, \dots, x_{2,d}]$$

相应地，只有分别在各分量位置处取相同值时，两个向量才相等（意味两个向量在空间中同一点）。

$$x_{1,i} = x_{2,i}$$

几类特殊的向量：

零向量： $\mathbf{o} = [0, 0, 0, 0, \dots, 0]$

单位向量： $\mathbf{e} = \frac{\mathbf{x}}{\|\mathbf{x}\|}$

向量的转置：略

## 向量运算

**向量加法**：向量加法是指将两个同维度的向量按照对应分量相加，得到一个新向量的运算。

**向量数乘**：数乘是指用一个标量（实数）乘以向量，每个分量都乘以该标量

**单位元**：在某种运算下，与任何元素结合都保持该元素不变。在线性代数中，向量的单位元是零向量。

**逆元**：对于向量加法，一个向量 $\mathbf{u}$ 的逆元是它的**负向量**，记作 $-\mathbf{u}$

**零元**：一个广义术语，泛指具有“归零”或“湮灭”性质的元素。在向量加法中，单位元是零向量 $\mathbf{0}$ ，在标量乘法中，零向量是“零元”，即 $0 \cdot \mathbf{x} = \mathbf{0}$ ，标量 $0$ 是标量域中的零元，但不是向量空间的“零元”。

**向量内积**：对于两个维度相同的向量，它们的内积是其对应分量乘积之和，结果是一个标量（长度仅有一个维度向量，转置等于自身）。（满足交换律）

内积结果为 $0$ ，说明两个向量正交，如果这两个向量均为单位向量，则称这种正交为**标准正交**。

若两个内积向量均已单位化，则向量内积可以作为两个**向量相似程度的判据**，单位向量的内积即为余弦相似度。长度确定情况下，内积越接近长度的乘积，则向量在方向上越相似。

用于改写给定函数的泰勒展开式+内积向量方向正好相反时，内积结果取最小值=梯度降

内积运算是行向量与列向量的乘积

**分类平面（决策边界）**：一个（或一组）用于将特征空间划分成不同类别区域的边界。与法向量同侧的向量为正，异侧为负。

设有平面方程： $ag + bh + cr + d = 0$

扩展特征向量 $\mathbf{x}_c = [g, h, r, 1]$  权重向量 $\mathbf{w} = [a, b, c, d]$

则该平面方程可以改写为向量形式 $\mathbf{w} \cdot \mathbf{x}_c^T = 0$

几何意义： $\mathbf{w} \cdot \mathbf{x}_c^T = 0$ 是 $ag + bh + cr + d = 0$ 的等价向量表述形式，它们描述了同一个分类平面。权重向量 $\mathbf{w}$ 的前三个分量 $[a, b, c]$ 构成了分类平面的法向量，指出平面方向，即决策得分（ $\mathbf{w} \cdot \mathbf{x}^T$ ）为正的一侧。

**向量外积**：向量外积是列向量与行向量的矩阵乘积。其运算结果是一个矩阵，该矩阵的每个元素都是列向量的一个分量与行向量的一个分量进行数乘（标量乘法）的结果。

**分量乘法（Hadamard积）**：分量乘法用 $\odot$ 来表示，对于两个维度相同的向量： $\mathbf{w} =$

$[w_1, w_2, w_3, \dots, w_d]$   $\mathbf{x} = [x_1, x_2, x_3, \dots, x_d]$  它们的Hadamard积（分量乘法）定义为对应分量相乘，结果是一个新的同维向量：

$$\mathbf{w} \odot \mathbf{x} = [w_1x_1, w_2x_2, w_3x_3, \dots, w_dx_d]$$

一般地，特征分量对于分类或评分结果的贡献度不一定相同。Hadamard积提供了一种为每个特征分量施加不同权重的直接方式，从而体现各分量的差异化影响效果。

## 向量线性相关性

任意向量，总能通过同一向量空间中其余向量得到。限定只包含**数乘与加法运算**——线性运算。

取向量空间中一组向量 $\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3, \dots, \mathbf{x}_n$ ，当且仅当标量值 $a_1, a_2, \dots, a_n$ 均等于 $0$ ， $a_1\mathbf{x}_1 + a_2\mathbf{x}_2 + \dots + a_n\mathbf{x}_n = \mathbf{0}$ 才成立，则称 $\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3, \dots, \mathbf{x}_n$ 为**线性无关**的，反之称这组向量为**线性相关**的，（如二维线性相关的两个向量共线，3维线性相关的两个向量共面）。

线性相关意味着其中至少有一个向量可以表示为其他向量的线性组合（即数乘与加法运算）。也就是说

其中任意一个向量**不能**写成由其它向量的数乘与加法运算构成的线性组合时，该向量组为线性无关向量组。

如果对于向量空间中的一个线性无关向量组，不存在空间中另一个向量可以加入该向量组并保持向量组的线性无关，则该向量组为**极大线性无关组**能生成整个向量空间的向量集合，称为该空间的**生成组**。如果一个生成组中不包含任何冗余的向量（即去掉其中任何一个向量都无法再生成整个空间），则称它为**极小生成组**。

互不线性相关的最大集合是构成对应向量空间的向量的最小集合——**基向量**，基向量等价于同时满足极大线性无关组与极小生成组的向量。

## 矩阵

**矩阵的定义**：一个  $m \times n$  矩阵 (Matrix) 是一个由数字（或更一般地，来自某个域的元素，如实数、复数）排成的  $m$  行 (Row)、 $n$  列 (Column) 的矩形阵列。

矩阵可以理解为对向量进行的线性变换。

向量是特殊的矩阵。

行数与列数对应相等的矩阵称作**同行矩阵**。同型且对应元素相等，则矩阵相等

行数与列数相等的矩阵，称作**方阵**。

零阵：

$$\mathbf{O} = \begin{bmatrix} 0 & 0 & \cdots & 0 \\ 0 & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 0 \end{bmatrix}$$

对角阵 $\text{diag}(\mathbf{x})$ :

$$\mathbf{D} = \begin{bmatrix} d_1 & 0 & \cdots & 0 \\ 0 & d_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & d_n \end{bmatrix}$$

$$\mathbf{D} = \text{diag}(d_1, d_2, \dots, d_n)$$

单位阵 ( $\mathbf{E}$ ):

$$\mathbf{E} = \begin{bmatrix} 1 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 1 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 1 \end{bmatrix}$$

矩阵的基本运算:

矩阵加法/减法: 只有同维度的矩阵才能相加减。规则是对应元素相加减。

标量乘法: 用一个数(标量)乘以矩阵中的每一个元素。

矩阵乘法: 这是最重要但也最需要理解的运算。

一个 $m \times n$ 的矩阵 $A$ 和一个 $n \times p$ 的矩阵 $B$ 可以相乘,结果是一个 $m \times p$ 的矩阵 $C$ 。

$C$ 中第 $i$ 行第 $j$ 列的元素 $c_{ij}$ , 等于 $A$ 的第 $i$ 行 $B$ 的第 $j$ 列的点积(内积)。

重要: 矩阵乘法不满足交换律, 即 $AB \neq BA$ 。

矩阵的转置:

一个 $m \times n$ 矩阵 $A$ 的转置, 记作 $A^T$ 或 $A'$ , 是一个 $n \times m$ 矩阵。

矩阵加法与数乘统称为矩阵的线性运算。

若无特殊说明, 矩阵内积乘法时, 不再特殊标注左操作数的列数与右操作数的行数, 而是**默认二者相等**。

矩阵的乘法满足结合律和分配律, 但是不满足交换律。

$AO = OA = O$  零元

$AE = EA = E$  单位元

$AB = BA = E$  逆元

矩阵的幂:

$A^{k+l} = A^k A^l$  (前提:  $A$ 是方阵)

$(AB)^k \neq A^k B^l$

$(AB)^T = B^T A^T$

矩阵内积等于向量外积的和——**内积的外积展开**。

矩阵乘法(基于内积)可以分解为多个外积矩阵的求和, 矩阵乘积 $AB$ (由内积计算)等于所有列向量和行向量外积的和。

设 $A$ 是 $m \times n$ 矩阵, 矩阵 $B$ 是 $n \times p$ 矩阵。它们的乘积 $C = AB$ 是一个 $m \times p$ 矩阵。

外积展开形式:

$$AB = \sum_{k=1}^n a_k b_k^T$$

其中:

$a_k$ 是 $A$ 的第 $k$ 列(一个 $m \times 1$ 列向量),

$b_k^T$ 是 $B$ 的第 $k$ 行(一个 $1 \times p$ 行向量),

$a_k b_k^T$  是一个  $m \times p$  矩阵（即外积）。

**矩阵元素相乘：**矩阵的元素相乘（Element-wise Multiplication），也称为 Hadamard 积（Hadamard Product）或 Schur 积（Schur Product），是一种二元运算。

核心规则：两个同维度的矩阵（即行数和列数完全相同）才能进行元素相乘。结果是一个新的同维度矩阵，其每个元素的值是原始两个矩阵对应位置元素的乘积。通常用  $\odot$  符号表示。

## 矩阵的特征值与特征向量

**乘法形式：左乘与右乘**<sup>[3]</sup>

矩阵与向量的乘法有两种等价的表示形式，取决于将向量视为列向量还是行向量。

右乘（列向量视角）

$$y = Ax$$

矩阵  $A$  右乘于列向量  $x$ ，得到新的列向量  $y$ 。（这是线性代数和科学技算中最常用的形式）

左乘（行向量视角）

$$y^T = x^T A, \text{ 得到新的列向量 } y^T。$$

**矩阵变换的几何效果：**矩阵对向量的乘法是一种**线性变换**，其效果取决于矩阵的形状。

非方阵 ( $m \times n, m \neq n$ )

变换是一种投影或升降维操作。变换后的新向量  $y$  与原向量  $x$  的维度不同 ( $m \neq n$ )，因此直接比较“长度” (参考注释2) 意义不大。

方阵 ( $n \times n$ )

变换是在同一空间内的变换（输入输出维度相同）。

几何效果可理解为对原向量  $x$  进行旋转和缩放的组合，从而得到新向量  $y$ 。

**特征向量与特征值：变换中的“不变量”**

对于方阵  $A$ ，存在一些特殊的向量，在线性变换中保持方向不变。

定义（右特征向量/值）：

若存在一个非零列向量  $v$  和一个标量  $\lambda$ ，使得：

$$Av = \lambda v$$

则称：

$v$  是矩阵  $A$  的一个特征向量。

$\lambda$  是该特征向量对应的特征值。

几何解释：

特征向量  $v$  在经过矩阵  $A$  的变换后，方向保持不变。

特征值  $\lambda$  量化了变换过程中沿该方向缩放的大小。

$|\lambda| > 1$ ：被拉伸

$|\lambda| = 1$ ：长度不变

$|\lambda| < 1$ ：被压缩

$\lambda < 0$  : 方向反向

左特征向量/值同理

重要性质:

1. 非零缩放: 任何非零标量 $k$ 乘以特征向量( $kv$ )后, 仍然是同一个特征值 $\lambda$ 对应的特征向量。
2. 单位特征向量: 为了比较的方便, 常将特征向量标准化为单位长度 (模长为1)。
3. 计算简化: 特征向量将矩阵乘法运算 $Av$ 简化为数乘运算 $\lambda v$ , 极大地简化了分析和计算。

**特征降维 (待深入学习, 目前理解程度低):** 特征降维是一种通过数学变换, 将高维数据映射到一个低维空间的技术, 其目标是保留数据中最重要的结构信息 (如方差、区分度), 同时摒弃冗余、噪声和无关细节。

特征向量指明了最重要的方向, 特征值告诉我们这些方向有多重要。

1. 注: 此处特征向量不等于线性代数中的特征向量, 在机器学习和数据科学中, 特征向量通常指一个样本的多维特征表示。↩
2. 我们常常使用两点间距离的平方和(欧氏距离)表示大小, 即 $\|\mathbf{x}\| = \sqrt{x_1^2 + x_2^2 + x_3^2 + \dots + x_d^2}$ 或 $\sqrt{\sum_{i=1}^d x_i^2}$ 但是请注意, 大小不要局限于欧氏距离 ↩
3. 两种形式描述的变换本质相同, 只是表示惯例的差异。通常优先使用右乘列向量的形式。↩