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2章 相似性度量（2学时）
第3章 函数与范函分析（2学时）
第4章 条件概率与贝叶斯（4学时）
第5章 信息论与熵（2学时）
第6章 线性分析与卷积（4学时）
第7章 正则化与范数（4学时）
第8章 最优化理论（4学时）
第9章 核函数映射（4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[1]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（Feature Vector，请注意与线性代数中的特征向量Eigenvector区分）。用[ ]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。

$$\boldsymbol{x} = [x_1, x_2, x_3, ..., x_d]$$

给定任一向量，其包含大小[2]（其模长，由范数Norm定义）与方向两类信息。

$\boldsymbol{x_1} = [x_{1,1}, x_{1,2}, x_{1,3}, ..., x_{1,d}]$ $\boldsymbol{x_2} = [x_{2,1}, x_{2,2}, x_{2,3}, ..., x_{2,d}]$
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：$\boldsymbol{o} = [0, 0, 0, 0, ..., 0]$
单位向量：$\boldsymbol{e} = \frac{\boldsymbol{x}}{\|\boldsymbol{x}\|}$
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$\boldsymbol{u}$的逆元是它的**负向量**，记作$-\boldsymbol{u}$
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 0,在标量乘法中，零向量是“零元”，即 $0\cdot\boldsymbol{x}=\mathbf{0}$,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$\boldsymbol{x_c}=[g,h,r,1]$　权重向量$\boldsymbol{w}=[a,b,c,d]$
则该平面方程可以改写为向量形式$\boldsymbol{w}\cdot\boldsymbol{x_c^T}=\mathbf{0}$
几何意义：$\boldsymbol{w}\cdot\boldsymbol{x_c^T}=0$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$\boldsymbol{w}$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（$\boldsymbol{w}\cdot\boldsymbol{x^T}$）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⊙来表示，对于两个维度相同的向量：$\boldsymbol{w}=[w_1,w_2,w_3,...,w_d]$　$\boldsymbol{x}=[x_1,x_2,x_3,...,x_d]$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：
$\boldsymbol{w}\odot\boldsymbol{x}=[w_1x_1,w_2x_2,w_3x_3,...,w_dx_d]$
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$\boldsymbol{x_1},\boldsymbol{x_2},\boldsymbol{x_3},...,\boldsymbol{x_n}$，当且仅当标量值$a_1,a_2,...,a_n$均等于0，$a_1\boldsymbol{x}_1+a_2\boldsymbol{x}_2+\cdots+a_n\boldsymbol{x}_n=\mathbf{0}$才成立,则称$\boldsymbol{x_1},\boldsymbol{x_2},\boldsymbol{x_3},...,\boldsymbol{x_n}$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 × 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：

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

对角阵$diag(\boldsymbol{x})$:

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

$$\boldsymbol{D}=\operatorname{diag}(d_1,d_2,\ldots,d_n)$$

单位阵（E）：

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

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×n的矩阵A和一个n×p的矩阵B可以相乘,结果是一个m×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×n矩阵A的转置，记作A^T或A'，是一个n×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⊙符号表示。

矩阵的特征值与特征向量

乘法形式：左乘与右乘[3]

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）

$\boldsymbol{y = Ax}$

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）

$\boldsymbol{y^T = x^T A}$，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λ，使得：

$$Av = \lambda v$$

则称：

v是矩阵A的一个特征向量。

λ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λ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λ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λv，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。

用特征降维简化运算

降维目标：降维后各维度方差尽可能大，保证不同维度之间的相关性为0（基向量正交）

💡 *也就是协方差尽可能小*

💡 *课上这里比较跨越，我们先来了解一下协方差的知识*

协方差

协方差是衡量两个随机变量（例如变量 X和 Y）之间线性相关关系的方向和强度的统计量。

若存在两个随机变量X和Y，其期望值（均值）分别为μ_x和μ_y，则它们的协方差定义为：

$$cov(X,Y) = E[(X-\mu_x)(Y-\mu_y)]$$

对于一组含有n个样本的数据，其样本协方差的计算公式为：

$$S_{xy} = \frac{1}{n-1}\sum_{i=1}^{n}(x_i-\bar{x})(y_i-\bar{y})$$

几何解释：

将每个数据点(x_i, y_i)视为一个高维空间中的向量（在二维情况下就是一个点）。

公式中的$(x_i-\bar{x})$和$(y_i-\bar{y})$表示该点相对于数据中心$(\bar{x},\bar{y})$的**偏差**。

协方差本质上是计算这些偏差向量在四个象限中的分布情况，并求其平均乘积。

协方差 > 0 (正)：数据点主要分布在第一、三象限。表明当一个变量取值高于其均值时，另一个变量也倾向于高于其均值。两个变量之间存在正相关趋势。其散点图呈“右上-左下”的椭圆形分布。

协方差 < 0 (负)：数据点主要分布在第二、四象限。表明当一个变量取值高于其均值时，另一个变量倾向于低于其均值。两个变量之间存在负相关趋势。其散点图呈“左上-右下”的椭圆形分布。

协方差 = 0：数据点在各象限均匀分布，无明显的线性趋势。两个变量线性不相关。

从协方差到协方差矩阵

当处理多个变量（p个）时，我们将所有变量两两之间的协方差排列成一个矩阵，称为协方差矩阵$\sum$（或S）

协方差矩阵

对于一个以向量 $\mathbf{X} = [X_1, X_2, ..., X_p]^T$ 表示的随机变量，其协方差矩阵是一个方阵，其中的每个元素 Σ_{ij} 是变量 X_i 和 X_j 之间的协方差。

$$\Sigma = \begin{bmatrix} \text{cov}(X_1, X_1) & \text{cov}(X_1, X_2) & \cdots & \text{cov}(X_1, X_p) \\ \text{cov}(X_2, X_1) & \text{cov}(X_2, X_2) & \cdots & \text{cov}(X_2, X_p) \\ \vdots & \vdots & \ddots & \vdots \\ \text{cov}(X_p, X_1) & \text{cov}(X_p, X_2) & \cdots & \text{cov}(X_p, X_p) \end{bmatrix}$$

重要性质

1. **方阵与对称性**: 协方差矩阵是一个对称方阵 $(p \times p)$，因为 $\text{cov}(X_i, X_j) = \text{cov}(X_j, X_i)$。
2. **对角线元素**: 对角线上的元素 $\Sigma_{ii} = \text{cov}(X_i, X_i) = \text{Var}(X_i)$ 是变量 X_i 自身的方差。
3. **半正定性**: 协方差矩阵是一个半正定矩阵，这意味着其所有特征值均 ≥ 0。这一性质至关重要。
   协方差矩阵的特征分解：揭示主方向
   协方差矩阵作为方阵，可以进行特征分解，其几何意义非常深刻。
   定义（特征分解）：

$$\Sigma \mathbf{v}_i = \lambda_i \mathbf{v}_i$$

   其中：

- $\mathbf{v}_i$ 是矩阵 Σ 的一个**特征向量**。
- λ_i 是该特征向量对应的**特征值**。

几何解释:

- 协方差矩阵 Σ 可以看作一个对数据点云进行**线性变换**的操作。
- **特征向量 $\mathbf{v}_i$** 指明了数据分布的主要**方向**（主方向）。
- **特征值 λ_i** 量化了数据在对应特征向量方向上的**离散程度（方差）**。λ_i 越大，表示数据在该方向上的伸展越长，包含的信息越多。
  重要性质

1. **主成分分析 (PCA) 的基础**: PCA的本质就是求取协方差矩阵的特征值和特征向量。最大的特征值对应的特征方向就是**第一主成分**，即数据方差最大的方向。
2. **标准化**: 特征向量通常被**标准化**为单位长度，以便比较不同方向上的方差（特征值）。
3. **正交性**: 由于协方差矩阵是对称矩阵，其特征向量之间是**相互正交**的，构成了数据空间的一组正交基。
   总结与应用

协方差: 衡量两个变量间的线性相关性。
协方差矩阵: 系统性地表示多个变量两两之间的协方差关系，是描述数据集形状（分布方向）的核心。
特征值/特征向量: 揭示了数据集内在的、不相关的**主方向**（特征向量）及其在各方向上的**扩展幅度**（特征值）。
应用:
主成分分析 (PCA)、线性判别分析 (LDA)、卡尔曼滤波、马氏距离等众多统计与机器学习算法都建立在协方差矩阵及其性质之上。

💡 *本段有些晦涩，不妨再来研究一下主成分分析法(PCA)*

PCA (主成分分析)

一种降维与特征提取方法。旨在找到一组新的正交坐标轴（**主成分**），以重新表述数据集，使得数据在新坐标轴上的投影**方差**依次最大化。

💡 *想象坐标轴上分布着散乱的点，每个点代表一个样本，PCA正是让这些点尽可能落在新坐标轴的方法。*

关键步骤与原理

1. 数据预处理：中心化
   将原始数据矩阵 $\boldsymbol{X}$ (m 个样本, n 个特征) 中心化，得到均值向量为 $\mathbf{0}$ 的新矩阵 $\boldsymbol{X'}$。

$$\boldsymbol{X'} = \boldsymbol{X} - \boldsymbol{\mu}$$

💡 *相当于移动原坐标轴的原点到散乱的点群中心，中心化其实就是换了一个中心*

2. 构建协方差矩阵
   计算中心化数据 $\boldsymbol{X'}$ 的协方差矩阵 $\boldsymbol{\Sigma}$，以衡量特征间的共同变化趋势。

$$\boldsymbol{\Sigma} = \frac{1}{m-1}\boldsymbol{X'}^T\boldsymbol{X'}$$

- $\boldsymbol{\Sigma}$ 是一个 $n \times n$ 的实对称方阵。

3. 特征分解：寻找主成分
   对 $\boldsymbol{\Sigma}$ 进行特征分解，求解特征方程：

$$\boldsymbol{\Sigma v} = \lambda \boldsymbol{v}$$

- **特征向量 $\boldsymbol{v}_i$**：方向，即**主成分**。指向数据方差最大的方向。
- **特征值 λ_i**：大小，量化了数据在对应主成分方向上的**方差**。

4. 排序与选择
   将特征对按特征值**从大到小**排序：

$$(\lambda_1, \boldsymbol{v}_1), (\lambda_2, \boldsymbol{v}_2), ..., (\lambda_n, \boldsymbol{v}_n)$$

- $\boldsymbol{v}_1$ 为第一主成分（方差最大），$\boldsymbol{v}_2$ 为第二主成分，依此类推。

5. 降维变换（投影）
   选取前 k 个主成分组成投影矩阵 $\boldsymbol{W}$：

$$\boldsymbol{W} = [\boldsymbol{v}_1, \boldsymbol{v}_2, ..., \boldsymbol{v}_k]$$

   将中心化数据 $\boldsymbol{X'}$ **右乘**于 $\boldsymbol{W}$，得到降维后的数据 $\boldsymbol{Y}$ ($m \times k$)：

$$\boldsymbol{Y} = \boldsymbol{X'W}$$

重要性质

- **正交性**：主成分（特征向量）之间相互正交。
- **方差解释度**：前 k 个主成分的方差贡献率为：

$$\frac{\sum_{i=1}^{k} \lambda_i}{\sum_{i=1}^{n} \lambda_i}$$

- **最优性**：在最小化重构误差和最大化投影方差的意义下，PCA是最优的降维方法。

本质
对数据协方差矩阵进行特征分解，通过保留最大方差对应的特征向量方向，实现数据的有效降维与信息保留。

矩阵的秩

一个 $m \times n$ 的矩阵 $\boldsymbol{A}$ 可以表示为：

$$\boldsymbol{A} = \begin{bmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & \ddots & \vdots \\ a_{m1} & \cdots & a_{mn} \end{bmatrix}_{m \times n}$$

- 其**第 i 行行向量**记为：$\boldsymbol{A}_{i,:}$
- 其**第 j 列列向量**记为：$\boldsymbol{A}_{:,j}$

矩阵 $\boldsymbol{A}$ 的**列向量组**（或**行向量组**）的**极大线性无关组**中所包含向量的个数，定义为矩阵的**秩**，记作 $R(\boldsymbol{A})$ 或 $\mathrm{rank}(\boldsymbol{A})$。

行满秩: $R(\boldsymbol{A}) = m$ (所有行向量线性无关)
列满秩: $R(\boldsymbol{A}) = n$ (所有列向量线性无关)
满秩方阵: $\boldsymbol{A}$ 为 $n \times n$ 方阵且 $R(\boldsymbol{A}) = n$ (又称**非奇异矩阵**，可逆)
欠秩矩阵: $R(\boldsymbol{A}) < \min(m, n)$ (方阵欠秩又称**奇异矩阵**，不可逆)

💡 *秩揭示了数据中**独立信息**的多少。秩越低，数据中的冗余度就越高。*

数据意义的解释:
若矩阵的行代表**观测对象**，列代表**观测特征**。

- **行欠秩**意味着部分观测对象可由其它对象的线性组合来表示（**样本冗余**）。
- **列欠秩**意味着某些观测特征可由其它特征的线性组合来表示（**特征冗余**或**多重共线性**）。

初等变换

给定任意一组维度相同的向量，如何求其最大线性无关组中向量的个数呢?
一种核心方法是对矩阵进行**初等变换**。

对于矩阵的**行向量**，初等变换包括三类操作：

1. 行对调: 交换两行的位置。
2. 非零数乘: 用一个非零常数 k 乘以某一行的所有元素。
3. 倍加行: 把某一行的 k 倍加到另一行上。

💡 *以上操作同样适用于**列变换**，只需将“行”替换为“列”。*

矩阵等价:
若矩阵 $\boldsymbol{A}$ 经过有限次初等变换变成矩阵 $\boldsymbol{B}$，则称矩阵 $\boldsymbol{A}$ 与 $\boldsymbol{B}$ **等价**，记作 $\boldsymbol{A} \sim \boldsymbol{B}$。

初等变换具有以下性质：

- **可逆性**: 三种变换均是可逆的，且其逆变换是同一类型的初等变换。
- **自反性**: $\boldsymbol{A} \sim \boldsymbol{A}$
- **对称性**: 若 $\boldsymbol{A} \sim \boldsymbol{B}$，则 $\boldsymbol{B} \sim \boldsymbol{A}$
- **传递性**: 若 $\boldsymbol{A} \sim \boldsymbol{B}$ 且 $\boldsymbol{B} \sim \boldsymbol{C}$，则 $\boldsymbol{A} \sim \boldsymbol{C}$

所有与矩阵 $\boldsymbol{A}$ 等价的矩阵构成一个**等价类**。**初等变换不改变矩阵的秩**，因此同一等价类中的所有矩阵秩相等。

初等矩阵与标准形

💡 *一个矩阵对应一种线性变换。初等变换是线性变换。*

对**单位矩阵** $\boldsymbol{E}$ 施以**一次**初等变换得到的矩阵，称为**初等矩阵**。

- 对矩阵 $\boldsymbol{A}$ 实施一次**初等行变换**，等价于在其**左侧**乘以对应的初等矩阵。

- 对矩阵 $\boldsymbol{A}$ 实施一次**初等列变换**，等价于在其**右侧**乘以对应的初等矩阵。

通过初等变换，任何矩阵都可以化为以下形式：

1. **行阶梯形**: 矩阵的零行位于底部，每个非零行的主元（第一个非零元）所在的列严格递增。

$$\begin{bmatrix} 1 & 2 & 5 & 3 \\ 0 & 1 & 2 & 6 \\ 0 & 0 & 0 & -3 \end{bmatrix}$$

2. **行最简形**: 在行阶梯形基础上，每个主元均为 1，且主元所在列的其它元素均为 0。

$$\begin{bmatrix} 1 & 0 & 1 & 0 \\ 0 & 1 & 2 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

3. **标准形**: 最简形式，左上角是单位阵，其余元素为 0。$\boldsymbol{D} = \begin{bmatrix} \boldsymbol{E}_r & \boldsymbol{O} \\ \boldsymbol{O} & \boldsymbol{O} \end{bmatrix}$，其中 r 就是矩阵的秩。

$$\begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix}$$

左逆与右逆

对于矩阵 $\boldsymbol{A}_{m\times n}$，若存在矩阵 $\boldsymbol{B}_{n\times m}$，使得：

$$\boldsymbol{B}_{n\times m}\boldsymbol{A}_{m\times n} = \boldsymbol{E}_{n\times n}$$

则称 $\boldsymbol{B}$ 是 $\boldsymbol{A}$ 的**左逆**，$\boldsymbol{A}$ 是 $\boldsymbol{B}$ 的**右逆**。

若存在矩阵 $\boldsymbol{B}_{n\times m}$，使得：

$$\boldsymbol{A}_{m\times n}\boldsymbol{B}_{n\times m} = \boldsymbol{E}_{m\times m}$$

则称 $\boldsymbol{B}$ 是 $\boldsymbol{A}$ 的**右逆**，$\boldsymbol{A}$ 是 $\boldsymbol{B}$ 的**左逆**。

💡 *除非 $m = n$，否则左逆和右逆对应的恒等变换的阶数不同。*

逆矩阵

若 $\boldsymbol{A}$ 是 $n \times n$ 方阵，且存在同阶方阵 $\boldsymbol{B}$ 使得：

$$\boldsymbol{AB} = \boldsymbol{BA} = \boldsymbol{E}_{n\times n}$$

则称矩阵 $\boldsymbol{A}$ 和 $\boldsymbol{B}$ **可逆**，且互为**逆矩阵**。记作 $\boldsymbol{B}=\boldsymbol{A}^{-1}$, $\boldsymbol{A}=\boldsymbol{B}^{-1}$。

性质:

1. 若矩阵 $\boldsymbol{A}$ 可逆，则其逆矩阵是**唯一**的。
2. $\boldsymbol{A}$ 可逆的充要条件是 $\boldsymbol{A}$ 为满秩方阵（非奇异矩阵）。
3. 若 $\boldsymbol{A}^{-1}=\boldsymbol{B}$，则 $(\boldsymbol{A}^{-1})^{-1}=\boldsymbol{B}^{-1}=\boldsymbol{A}$。
4. 若同阶方阵 $\boldsymbol{A}$ 和 $\boldsymbol{B}$ 均可逆，则 $(\boldsymbol{AB})^{-1}=\boldsymbol{B}^{-1}\boldsymbol{A}^{-1}$。
5. 若 $\boldsymbol{A}$ 可逆，则 $\boldsymbol{A}^T$ 亦可逆，且 $(\boldsymbol{A}^T)^{-1}=(\boldsymbol{A}^{-1})^T$。

行列式 (Determinant)

仅方阵有行列式，记作 $\det(\boldsymbol{A})$ 或 $|\boldsymbol{A}|$。

余子式 (M_{ij}): 在 n 阶行列式中，划去元素 a_{ij} 所在的行和列，剩下的 $(n-1)^2$ 个元素按原来的排列构成的 $n-1$ 阶行列式，称为 a_{ij} 的余子式。
代数余子式 (A_{ij}): $A_{ij}=(-1)^{i+j}M_{ij}$

行列式可以按行或按列展开：

- **按第 i 行展开**: $\det(\boldsymbol{A})=\sum_{j=1}^{n}a_{ij}A_{ij}=\sum_{j=1}^{n}(-1)^{i+j}a_{ij}M_{ij}$
- **按第 j 列展开**: $\det(\boldsymbol{A})=\sum_{i=1}^{n}a_{ij}A_{ij}=\sum_{i=1}^{n}(-1)^{i+j}a_{ij}M_{ij}$

伴随矩阵 (Adjugate Matrix)

矩阵 $\boldsymbol{A}$ 的伴随矩阵 $\boldsymbol{A}^*$ 由其**代数余子式**构成：

$$\boldsymbol{A}^*=\begin{bmatrix}A_{11} & A_{21} & \cdots & A_{n1}\\ A_{12} & A_{22} & \cdots & A_{n2}\\ \vdots & \vdots & \ddots & \vdots\\ A_{1n} & A_{2n} & \cdots & A_{nn}\end{bmatrix}$$

💡 *注意伴随矩阵中代数余子式的排列是**转置**的（第 i 行第 j 列的元素是 A_{ji}）。*

伴随矩阵的一个重要性质是：

$$\boldsymbol{A}\boldsymbol{A}^*=\boldsymbol{A}^*\boldsymbol{A}=\det(\boldsymbol{A})\boldsymbol{E}$$

逆矩阵公式

由伴随矩阵的性质，可得到求逆矩阵的一个直接公式：

$$\boldsymbol{A}^{-1} = \frac{1}{\det(\boldsymbol{A})}\boldsymbol{A}^*$$

此公式仅适用于 $\det(\boldsymbol{A}) \neq 0$ 的情况，即 $\boldsymbol{A}$ 可逆。

将大矩阵分割成若干个小矩阵（子块）进行运算，可以简化计算和理论分析。
设矩阵 $\boldsymbol{A}_{m\times n}$ 和 $\boldsymbol{B}_{m\times n}$ 按同样方式分块：

$$\boldsymbol{A} = \begin{bmatrix} \boldsymbol{A}^{(1,1)}_{m_1\times n_1} & \boldsymbol{A}^{(1,2)}_{m_1\times n_2} & \cdots & \boldsymbol{A}^{(1,q)}_{m_1\times n_q} \\ \boldsymbol{A}^{(2,1)}_{m_2\times n_1} & \boldsymbol{A}^{(2,2)}_{m_2\times n_2} & \cdots & \boldsymbol{A}^{(2,q)}_{m_2\times n_q} \\ \vdots & \vdots & \ddots & \vdots \\ \boldsymbol{A}^{(p,1)}_{m_p\times n_1} & \boldsymbol{A}^{(p,2)}_{m_p\times n_2} & \cdots & \boldsymbol{A}^{(p,q)}_{m_p\times n_q} \end{bmatrix}, \quad \boldsymbol{B} = \begin{bmatrix} \boldsymbol{B}^{(1,1)}_{m_1\times n_1} & \boldsymbol{B}^{(1,2)}_{m_1\times n_2} & \cdots & \boldsymbol{B}^{(1,q)}_{m_1\times n_q} \\ \boldsymbol{B}^{(2,1)}_{m_2\times n_1} & \boldsymbol{B}^{(2,2)}_{m_2\times n_2} & \cdots & \boldsymbol{B}^{(2,q)}_{m_2\times n_q} \\ \vdots & \vdots & \ddots & \vdots \\ \boldsymbol{B}^{(p,1)}_{m_p\times n_1} & \boldsymbol{B}^{(p,2)}_{m_p\times n_2} & \cdots & \boldsymbol{B}^{(p,q)}_{m_p\times n_q} \end{bmatrix}$$

其中 $\sum_{i=1}^{p} m_i = m$, $\sum_{j=1}^{q} n_j = n$。

分块矩阵的运算:

1. **转置**: 分块矩阵的转置，需先对子块整体转置，再对每个子块自身转置。

$$\boldsymbol{A}^T = \begin{bmatrix} (\boldsymbol{A}^{(1,1)})^T & (\boldsymbol{A}^{(2,1)})^T & \cdots & (\boldsymbol{A}^{(p,1)})^T \\ (\boldsymbol{A}^{(1,2)})^T & (\boldsymbol{A}^{(2,2)})^T & \cdots & (\boldsymbol{A}^{(p,2)})^T \\ \vdots & \vdots & \ddots & \vdots \\ (\boldsymbol{A}^{(1,q)})^T & (\boldsymbol{A}^{(2,q)})^T & \cdots & (\boldsymbol{A}^{(p,q)})^T \end{bmatrix}$$

2. **加法/减法**: 对应位置的子块相加减。

$$\boldsymbol{A} + \boldsymbol{B} = \begin{bmatrix} \boldsymbol{A}^{(1,1)} + \boldsymbol{B}^{(1,1)} & \boldsymbol{A}^{(1,2)} + \boldsymbol{B}^{(1,2)} & \cdots & \boldsymbol{A}^{(1,q)} + \boldsymbol{B}^{(1,q)} \\ \boldsymbol{A}^{(2,1)} + \boldsymbol{B}^{(2,1)} & \boldsymbol{A}^{(2,2)} + \boldsymbol{B}^{(2,2)} & \cdots & \boldsymbol{A}^{(2,q)} + \boldsymbol{B}^{(2,q)} \\ \vdots & \vdots & \ddots & \vdots \\ \boldsymbol{A}^{(p,1)} + \boldsymbol{B}^{(p,1)} & \boldsymbol{A}^{(p,2)} + \boldsymbol{B}^{(p,2)} & \cdots & \boldsymbol{A}^{(p,q)} + \boldsymbol{B}^{(p,q)} \end{bmatrix}$$

3. **数乘**: 标量乘以矩阵等于乘以每一个子块。

$$
a\boldsymbol{A} = \begin{bmatrix} a\boldsymbol{A}^{(1,1)} & a\boldsymbol{A}^{(1,2)} & \cdots & a\boldsymbol{A}^{(1,q)} \\ a\boldsymbol{A}^{(2,1)} & a\boldsymbol{A}^{(2,2)} & \cdots & a\boldsymbol{A}^{(2,q)} \\ \vdots & \vdots & \ddots & \vdots \\ a\boldsymbol{A}^{(p,1)} & a\boldsymbol{A}^{(p,2)} & \cdots & a\boldsymbol{A}^{(p,q)} \end{bmatrix}
$$

💡 *分块矩阵的乘法规则类似，但需确保子块的维度满足矩阵乘法要求。*

矩阵的迹

方阵 $\boldsymbol{A}$ 的**迹** (Trace) 定义为它的主对角线上的所有元素之和，记作 $\mathrm{tr}(\boldsymbol{A})$。

$$
\mathrm{tr}(\boldsymbol{A}) = \sum_{i=1}^{n} a_{ii}
$$

💡 *迹是一个标量，是矩阵的一个非常重要的数值特征。*

相似变换与相似不变量

线性变换是对整个空间的变换，其本质与描述它所采用的基向量（即变换矩阵的具体形式）无直接关联。线性变换与方阵的一一对应关系建立在空间**基向量确定**的前提下。

💡 *想象一个二维坐标系逆时针旋转45°的变换。在不同的基向量组下，描述这个同一旋转动作的矩阵是不同的，但这些矩阵之间是**相似**的。*

定义（相似矩阵）：
若存在可逆矩阵 $\boldsymbol{P}$，使得 n 阶方阵 $\boldsymbol{A}$ 与 $\boldsymbol{B}$ 满足如下关系：

$$
\boldsymbol{B} = \boldsymbol{P}\boldsymbol{A}\boldsymbol{P}^{-1}
$$

则称 $\boldsymbol{B}$ 是 $\boldsymbol{A}$ 的**相似矩阵**，或称 $\boldsymbol{A}$ 与 $\boldsymbol{B}$ **相似**。$\boldsymbol{P}\boldsymbol{A}\boldsymbol{P}^{-1}$ 的运算称为对 $\boldsymbol{A}$ 进行**相似变换**。

相似矩阵代表**同一个线性变换**，只是在不同基下的表达。因此，它们必然会留下一些共同的“痕迹”，即**相似不变量**。

重要性质:

1. **迹的相似不变性**: 若 $\boldsymbol{A} \sim \boldsymbol{B}$，则 $\mathrm{tr}(\boldsymbol{A}) = \mathrm{tr}(\boldsymbol{B})$。

💡 *相似矩阵的对角线元素和一定相等。*

2. **行列式的相似不变性**: 若 $\boldsymbol{A} \sim \boldsymbol{B}$，则 $\det(\boldsymbol{A}) = \det(\boldsymbol{B})$。

> 💡 *由行列式的几何意义（线性变换的伸缩比例）可知，该比例与坐标基的选择无关。*

多维缩放 (Multiple Dimensional Scaling, MDS) 是一种经典的**降维**方法，其核心思想是：在低维空间中重构数据点的坐标，使得这些点之间的**距离**尽可能接近原始高维空间中的距离。

问题设定与目标

假设有 m 个样本，并已知它们在高维空间中的**距离平方矩阵** $\boldsymbol{D}$，其元素为：

$$D_{i,j}^2 = \|\boldsymbol{x}_i - \boldsymbol{x}_j\|^2 = \sum_{k=1}^{d} (x_{i,k} - x_{j,k})^2$$

MDS的目标是找到一组低维表示 $\boldsymbol{z}_1, \boldsymbol{z}_2, ..., \boldsymbol{z}_m$，使得：

$$\|\boldsymbol{z}_i - \boldsymbol{z}_j\|^2 \approx D_{i,j}^2$$

从距离矩阵到内积矩阵

推导的关键在于将距离关系转化为内积关系。我们假设降维后的数据已经进行了**中心化**，即：

$$\sum_{i=1}^{m} \boldsymbol{z}_i = \boldsymbol{0}$$

定义降维后数据的**内积矩阵** (Gram Matrix) 为 $\boldsymbol{B}$，其元素为 $B_{i,j} = \boldsymbol{z}_i \boldsymbol{z}_j^T$。

距离的平方与内积存在以下关系：

$$\begin{aligned} D_{i,j}^2 &= \|\boldsymbol{z}_i - \boldsymbol{z}_j\|^2 \\ &= (\boldsymbol{z}_i - \boldsymbol{z}_j)(\boldsymbol{z}_i - \boldsymbol{z}_j)^T \\ &= \boldsymbol{z}_i \boldsymbol{z}_i^T + \boldsymbol{z}_j \boldsymbol{z}_j^T - 2\boldsymbol{z}_i \boldsymbol{z}_j^T \\ &= B_{i,i} + B_{j,j} - 2B_{i,j} \end{aligned}$$

为了求解 $\boldsymbol{B}$，需对上述关系式进行求和，并利用中心化条件($\sum_i B_{i,j} = 0, \sum_j B_{i,j} = 0$)进行推导：

$$\sum_{i=1}^{m} D_{i,j}^2 = \operatorname{tr}(\boldsymbol{B}) + mB_{j,j}$$

$$\sum_{j=1}^{m} D_{i,j}^2 = \operatorname{tr}(\boldsymbol{B}) + mB_{i,i}$$

$$\sum_{i=1}^{m}\sum_{j=1}^{m} D_{i,j}^2 = 2m \operatorname{tr}(\boldsymbol{B})$$

最终，可以通过以下公式从距离矩阵 $\boldsymbol{D}$ 还原出内积矩阵 $\boldsymbol{B}$：

$$B_{i,j} = -\frac{1}{2}\left(D_{i,j}^2 - \frac{1}{m}\sum_{i=1}^{m} D_{i,j}^2 - \frac{1}{m}\sum_{j=1}^{m} D_{i,j}^2 + \frac{1}{m^2}\sum_{i=1}^{m}\sum_{j=1}^{m} D_{i,j}^2\right)$$

特征分解与降维

内积矩阵 $\boldsymbol{B}$ 是一个实对称半正定矩阵，可以进行特征分解：

$$\boldsymbol{B} = \boldsymbol{V}\boldsymbol{\Lambda}\boldsymbol{V}^T$$

其中 $\boldsymbol{\Lambda} = \operatorname{diag}(\lambda_1, \lambda_2, ..., \lambda_m)$ $(\lambda_1 \geq \lambda_2 \geq ... \geq \lambda_m \geq 0)$，$\boldsymbol{V}$ 是正交矩阵。

取前 k 个最大的特征值及其对应的特征向量，即可得到降维后的数据矩阵：

$$\boldsymbol{Z} = \boldsymbol{V}_{:,1:k}\boldsymbol{\Lambda}_{1:k,1:k}^{1/2}$$

其中 $\boldsymbol{Z}$ 的每一行 $\boldsymbol{z}_i$ 就是一个样本在 k 维空间中的坐标。

💡 *MDS的魅力在于：我们无需知道原始高维数据X，仅凭样本间的距离矩D，就可以通过数学推导恢复出其内积矩阵B，并最终通过特征分解直接得到降维后的结果Z。*

LU分解的定义与条件

LU分解是将一个n阶方阵$\boldsymbol{A}$分解为一个**下三角矩阵 $\boldsymbol{L}$** (Lower triangular) 和一个**上三角矩阵 $\boldsymbol{U}$** (Upper triangular) 的乘积，即：

$$\boldsymbol{A} = \boldsymbol{L}\boldsymbol{U}$$

LU分解并非对所有矩阵都可行。一个充分条件是：矩阵$\boldsymbol{A}$的**各阶顺序主子式均不为零**。在此条件下，可以通过高斯消元法（仅使用第三种行初等变换：倍加行）得到分解：

$$\boldsymbol{L}^{(n-1)}\cdots\boldsymbol{L}^{(2)}\boldsymbol{L}^{(1)}\boldsymbol{A}=\boldsymbol{U}$$

其中$\boldsymbol{L}^{(k)}$是用于消元的初等矩阵。由此可得下三角矩阵$\boldsymbol{L}$为：

$$\boldsymbol{L}=(\boldsymbol{L}^{(1)})^{-1}(\boldsymbol{L}^{(2)})^{-1}\cdots(\boldsymbol{L}^{(n-1)})^{-1}$$

💡 *$\boldsymbol{L}$矩阵的构造非常巧妙，其主对角线以下的元素正好是高斯消元过程中所用的乘数。*

选主元与PLU分解

若放宽条件，只要求方阵$\boldsymbol{A}$为**满秩矩阵**，则可能在消元过程中遇到**0主元**。此时需要通过**行交换**（选主元）来继续分解。

这种情况下，LU分解可以推广为**PLU分解**（或称**带行置换的LU分解**）：

$$\boldsymbol{PA}=\boldsymbol{LU}$$

其中$\boldsymbol{P}$是一个**置换矩阵**（由单位矩阵行交换得到）。

💡 *置换矩阵是标准正交阵，其逆矩阵等于其转置，即 $\boldsymbol{P}^{-1} = \boldsymbol{P}^T$。*

LU分解的表示与优势

LU分解可以有多种等价的表示形式：

1. **外积展开式**: 若 $\boldsymbol{L}=[\boldsymbol{L}_{:,1},\boldsymbol{L}_{:,2},\ldots,\boldsymbol{L}_{:,n}]$，$\boldsymbol{U}=[\boldsymbol{U}_{1,:},\boldsymbol{U}_{2,:},\ldots,\boldsymbol{U}_{n,:}]^T$，则分解可写为：

$$\boldsymbol{PA}=\sum_{k=1}^{n}\boldsymbol{L}_{:,k}\boldsymbol{U}_{k,:}$$

2. **LDU分解**: 有时会将$\boldsymbol{U}$进一步分解为一个对角阵$\boldsymbol{D}$（由主元构成）和一个单位上三角矩阵的乘积，即 $\boldsymbol{A}=\boldsymbol{LDU}$。

LU分解的优势:

1. **节省存储**: 将一个矩阵分解为两个三角矩阵，并未引入新的存储需求。
2. **高效求解方程组**: 若将$\boldsymbol{A}$视作线性变换，给定变换后的向量$\boldsymbol{y}$，求解$\boldsymbol{Ax}=\boldsymbol{y}$时，关键变换信息已存储在$\boldsymbol{L}$和$\boldsymbol{U}$中，无需对$\boldsymbol{A}$重新计算。利用三角矩阵的特性，可通过**前向替换**和**回代**快速求解。
3. **批量求解**: 对于多个方程组 $\boldsymbol{Ax}_1=\boldsymbol{y}_1,\boldsymbol{Ax}_2=\boldsymbol{y}_2,\ldots$，只需进行一次LU分解即可高效求解。

矩阵的特征分解 (Eigen Decomposition)

特征值与特征向量

对于n阶方阵$\boldsymbol{A}$，若存在标量λ和非零向量$\boldsymbol{x}_i$，使得：

$$\boldsymbol{A}\boldsymbol{x}_i^T = \lambda_i \boldsymbol{x}_i^T$$

则称λ_i为矩阵$\boldsymbol{A}$的**特征值**，$\boldsymbol{x}_i$为对应的**左特征向量**（单位行向量）。

重要性质: 若方阵$\boldsymbol{A}$的特征值$\lambda_1, \lambda_2, \ldots, \lambda_n$**非零且互不相等**，则其对应的所有特征向量$\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_n$构成一个**线性无关组**。

💡 *该性质可通过数学归纳法证明。核心在于利用特征值互异的条件，证明任何一组特征向量都无法线性表出其他特征向量。*

对称矩阵的特征分解

对于**对称矩阵** $\boldsymbol{B}$ ($\boldsymbol{B}^T = \boldsymbol{B}$)，其性质更为优良：

1. **特征向量正交**: 对称矩阵的属于**不同特征值**的特征向量是彼此**正交**的。

$$\text{若 } \lambda_i \neq \lambda_j, \text{ 则 } \boldsymbol{x}_i \boldsymbol{x}_j^T = 0$$

2. **正交对角化**: 任意对称阵$\boldsymbol{B}$都可进行**特征分解**（或称**正交对角化**）：

$$\boldsymbol{B} = \boldsymbol{Q}\boldsymbol{\Lambda}\boldsymbol{Q}^T$$

其中：
- $\boldsymbol{\Lambda} = \text{diag}(\lambda_1, \lambda_2, \ldots, \lambda_n)$ 是由特征值构成的对角阵。
- $\boldsymbol{Q} = [\boldsymbol{x}_1^T, \boldsymbol{x}_2^T, \ldots, \boldsymbol{x}_n^T]$ 是由对应的**单位特征向量**（已标准化）构成的**正交矩阵**，满足 $\boldsymbol{Q}^T\boldsymbol{Q} = \boldsymbol{Q}\boldsymbol{Q}^T = \boldsymbol{E}$。

💡 *对称矩阵的特征分解是奇异值分解(SVD)和主成分分析(PCA)的理论基础。*

唯一性说明:

- 特征分解的结果可能不唯一（例如特征向量的符号或重特征值对应特征向量基的选取）。
- 通常约定将特征值按降序排列 $\lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_n$。在此约定下，若所有特征值**互不相等**，则特征分解是**唯一**的。

奇异值分解 (SVD)

奇异值分解 (Singular Value Decomposition, SVD) 是比特征分解更强大、更通用的矩阵分解工具，适用于任意形状的$m \times n$矩阵$\boldsymbol{A}$。

SVD的定义

任意矩阵$\boldsymbol{A}_{m\times n}$都可以被分解为以下三个矩阵的乘积：

$$\boldsymbol{A}_{m\times n} = \boldsymbol{U}_{m\times m}\boldsymbol{\Sigma}_{m\times n}\boldsymbol{V}^T_{n\times n}$$

其中：

- $\boldsymbol{U}$: $m \times m$ 的**左奇异向量矩阵**，是**正交矩阵**，满足 $\boldsymbol{U}\boldsymbol{U}^T = \boldsymbol{U}^T\boldsymbol{U} = \boldsymbol{E}_m$。其列向量称为左奇异向量。
- $\boldsymbol{\Sigma}$: $m \times n$ 的**奇异值矩阵**，是一个**对角矩阵**（但非方阵）。其对角线上的元素 $\sigma_1, \sigma_2, \ldots, \sigma_{\min(m,n)}$ 称为**奇异值**，通常按降序排列 $\sigma_1 \geq \sigma_2 \geq \ldots \geq 0$，其余元素均为0。
- $\boldsymbol{V}$: $n \times n$ 的**右奇异向量矩阵**，是**正交矩阵**，满足 $\boldsymbol{V}\boldsymbol{V}^T = \boldsymbol{V}^T\boldsymbol{V} = \boldsymbol{E}_n$。其列向量称为右奇异向量。

SVD的推导与性质

SVD可以通过对称矩阵$\boldsymbol{A}^T\boldsymbol{A}$和$\boldsymbol{A}\boldsymbol{A}^T$的特征分解来推导：

- $\boldsymbol{A}^T\boldsymbol{A}$的特征向量构成了SVD中的**右奇异向量** $\boldsymbol{V}$。
- $\boldsymbol{A}\boldsymbol{A}^T$的特征向量构成了SVD中的**左奇异向量** $\boldsymbol{U}$。
- 奇异值 σ_i 是 $\boldsymbol{A}^T\boldsymbol{A}$ (或 $\boldsymbol{A}\boldsymbol{A}^T$) 的特征值 λ_i 的平方根，即 $\sigma_i = \sqrt{\lambda_i}$。

紧奇异值分解 (Truncated SVD):
在实际应用中，特别是降维中，通常使用SVD的紧凑形式。只保留前r ($r = \text{rank}(\boldsymbol{A})$)个奇异值及其对应的左右奇异向量：

$$\boldsymbol{A}_{m\times n} \approx \boldsymbol{U}_{m\times r}\boldsymbol{\Sigma}_{r\times r}\boldsymbol{V}^T_{r\times n}$$

💡 *SVD是线性代数的顶峰之作，是现代数据科学和机器学习的基石，广泛应用于降维、推荐系统、自然语言处理、数据压缩等领域。*

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$\|\boldsymbol{x}\| = \sqrt{x_1^2 + x_2^2 + x_3^2 + ... + x_d^2}$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。 ↩