3. 线性代数

3.1 基础概念

3.1.1 标量、向量、矩阵与张量

标量（$Scalar$）：只有大小，没有方向的量，如整数，有理数，实数等。例如一个变量 x ，就是一个标量。

向量（$Vector$）：一维数组 R^n ，有大小和方向的量，例如向量 $\vec{x}=(x_1,x_2)$ 。

矩阵（$Matrix$）：二维数组 $A \in R^{m\times n}$ 。

张量（$Tensor$）：多维数组 $A \in R^{N\times C\times H\times W}$ 。

3.1.2 范数

向量的范数：只要满足非负、自反、三角不等式就可以称之为距离。范数是一种强化了的距离概念，它在定义上比距离多了一条数乘的运算法则。在数学上，范数包括向量范数和矩阵范数，向量范数表征向量空间中向量的大小，矩阵范数表征矩阵引起变化的大小。一种非严密的解释就是，对应向量范数，向量空间中的向量都是有大小的，这个大小如何度量，就是用范数来度量的，不同的范数都可以来度量这个大小，就好比米和尺都可以来度量远近一样；对于矩阵范数，学过线性代数，我们知道，通过运算 $AX=B$，可以将向量 X 变化为 B，矩阵范数就是来度量这个变化大小的。

向量范数需要满足的条件：

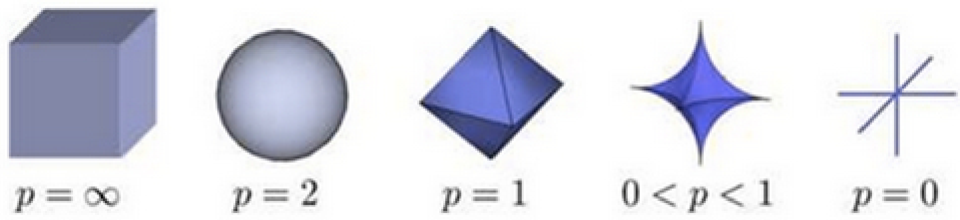
$$\begin{cases} f(x)\geq 0 \\ f(x)=0\Rightarrow x=0 \\ f(x+y)\leq f(x)+f(y) \\ \forall\alpha\in\mathbb{R}, f(\alpha x)=|\alpha|f(x) \end{cases}$$

L-P范数

闵可夫斯基距离的定义一样，**L-P** 范数不是一个范数，而是一组范数，其定义如下：

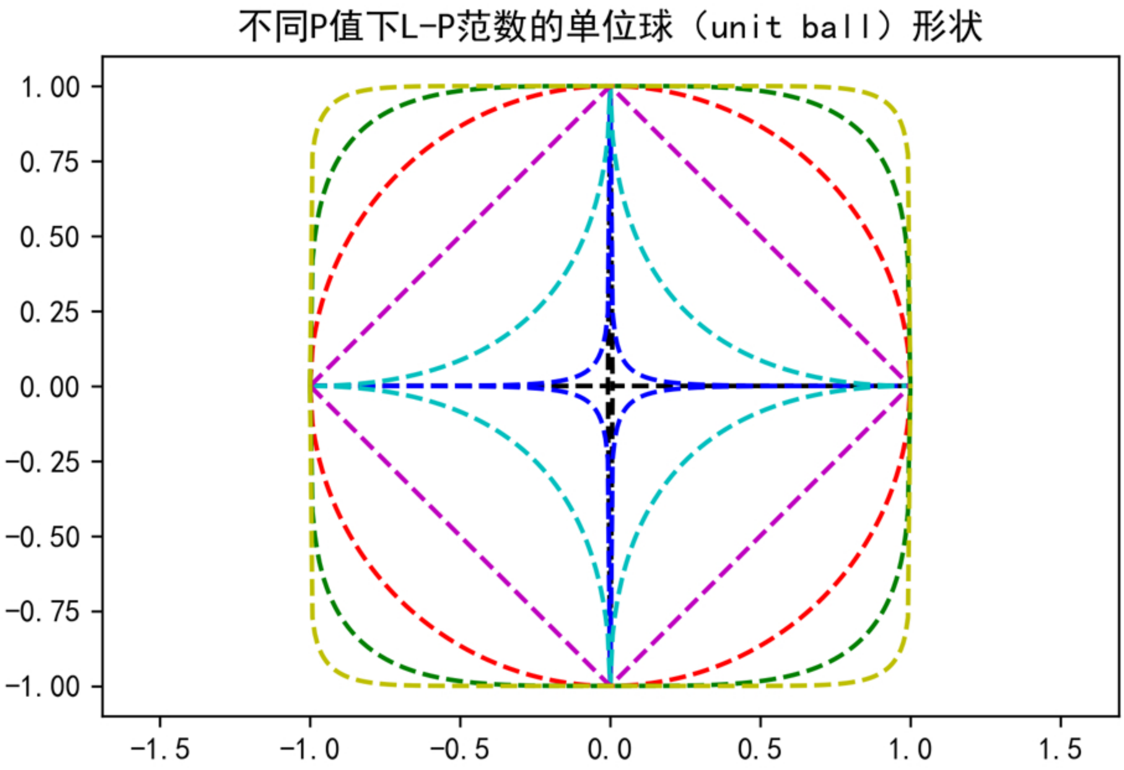
$$\mathrm{L}_p=\|X\|_p=\sqrt[p]{\sum_{i=1}^{n}x_i^p}, X=(x_1,x_2,\cdots,x_n)$$

根据 p 的变化，范数也有着不同的变化，一个经典的有关 p 范数的变化图如下：



上图表示了 p 从无穷到 0 变化时，三维空间中到原点的距离（范数）为1的点构成的图形的变化情况。以常见的L-2范数（$p=2$）为例，此时的范数也即欧氏距离，空间中到原点的欧氏距离为1的点构成了一个球面。图上是三维坐标轴，所以 $p=0$ 的三条线是三个坐标轴。

不同 p 值下 L_p 范数在二维坐标下的单位球（$unit\ ball$）形状如下图所示，从里到外 p 值分别为：$(0.01,0.25,0.5,1,2,4,10)$。



- （1）$L0$ 范数

$p=0$ 时，也就是 $L0$ 范数，$L0$ 范数并不满足三角不等式的性质，因此它并不是一个真正的范数，它主要被用来度量向量中非零元素的个数。用上面的 $L-P$ 定义可以得到的 L_0 的定义为：

$$\|X\|_0=\sqrt[0]{\sum_{i=1}^{n} x_i^0}$$

这里就有点问题了，我们知道非零元素的零次方为1，但零的零次方，非零数开零次方都是什么鬼，很不好说明 $L0$ 的意义，所以在通常情况下，大家都用的是：

$$\|X\|_0=\#(i|x_i\neq 0)$$

表示向量 X 中非零元素的个数。对于 $L0$ 范数，其优化问题为：

$$\begin{gathered}\min\|X\|_0\\ \text{s.t. } Ax=b\end{gathered}$$

在实际应用中，由于 $L0$ 范数本身不容易有一个好的数学表示形式，给出上面问题的形式化表示是一个很难的问题，故被人认为是一个 NP 难问题。所以在实际情况中，$L0$ 的最优问题会被放宽到 $L1$ 或 $L2$ 下的最优化。

- （2）$L1$ 范数

$L1$ 范数是我们经常见到的一种范数，它的定义如下：

$$\|X\|_1=\sum_{i=1}^{n}|x_i|$$

表示向量 X 中非零元素的绝对值之和。

$L1$ 范数有很多的名字，例如我们熟悉的曼哈顿距离、最小绝对误差等。使用 $L1$ 范数可以度量两个向量间的差异，如绝对误差和（Sum of Absolute Difference）：

$$SAD\left(x_1, x_2\right)=\sum_i^n\left|x_{1i}-x_{2i}\right|$$

对于 $L1$ 范数，它的优化问题如下：

$$\begin{array}{c}\min\|X\|_1 \\ \text{s.t. } Ax=b\end{array}$$

由于 $L1$ 范数的天然性质，对 $L1$ 优化的解是一个稀疏解，因此 $L1$ 范数也被叫做稀疏规则算子。通过 $L1$ 可以实现特征的稀疏，去掉一些没有信息的特征，例如在对用户的电影爱好做分类的时候，用户有100个特征，可能只有十几个特征是对分类有用的，大部分特征如身高体重等可能都是无用的，利用 $L1$ 范数就可以过滤掉。

- （3）$L2$ 范数

$L2$ 范数是我们最常见最常用的范数了，我们用的最多的度量距离欧氏距离就是一种 $L2$ 范数，它的定义如下：

$$\|X\|_2=\sqrt{\sum_{i=1}^n x_i^2}$$

表示向量元素的平方和再开平方。

像 $L1$ 范数一样， $L2$ 也可以度量两个向量间的差异，如平方差和（Sum of Squared Difference）：

$$SSD\left(x_1, x_2\right)=\sum_i^n\left(x_{1i}-x_{2i}\right)^2$$

对于 $L2$ 范数，它的优化问题如下：

$$\begin{array}{c}\min\|X\|_2 \\ \text{s.t. } Ax=b\end{array}$$

$L2$ 范数通常会被用来做优化目标函数的正则化项，防止模型为了迎合训练集而过于复杂造成过拟合的情况，从而提高模型的泛化能力。

- （4） L_∞ 范数

当 $p=\infty$ 时，也就是 L_∞ 范数，它主要被用来度量向量元素的最大值，与 $L0$ 一样，通常情况下表示为：

$$\|\mathrm{x}\|_\infty=\max\left(\left|x_i\right|\right)$$

可进一步阅读：

范数（norm） 几种范数的简单介绍（https://blog.csdn.net/a493823882/article/details/80569888）

3.2 行列式

3.2.1 行列式的基本定义

- 1阶方阵的行列式就是元素本身；

- n 级行列式 $\begin{vmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & \dots & a_{nn} \end{vmatrix}$ 等于所有取自不同行不同列的 n 个元素的乘积 $a_{1j_1}a_{2j_2}\dots a_{nj_n}$ 的代数和，这里 $j_1j_2\dots j_n$ 是 $1,2\dots n$ 的一个排列，当其为偶排列时，$a_{1j_1}a_{2j_2}\dots a_{nj_n}$ 带有正号，当其为奇排列时，$a_{1j_1}a_{2j_2}\dots a_{nj_n}$ 带有负号，因此上面定义可以改写成：

$$\begin{vmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & \dots & a_{mn} \end{vmatrix} = \sum_{j_1j_2\dots j_n} (-1)^{\tau(j_1j_2\dots j_n)} a_{1j_1}a_{j_2}\dots a_{nj_n}$$

- n 阶方阵的行列式等于任一行（或列）的各元素与其对应的代数余子式乘积的和：$|A| = \sum_{j=1}^{n} a_{ij} \cdot (-1)^{i+j} M_{ij}$ （行展开）。
- 1×1 的方阵，其行列式等于该元素本身。$A = (a_{11}) \quad |A| = a_{11}$ 。
- 2×2 的方阵，其行列式用主对角线元素乘积前去次对角线元素乘积：

$$A = \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix} \cdots |A| = a_{11}a_{22} - a_{12}a_{21}$$

- 3×3 的方阵：$A = \begin{pmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{pmatrix}$ ，根据“主对角线元素乘积前去次对角线元素的乘积”的原则得：

$$A = \left(\begin{array}{ccc|ccc|ccc} a_{11} & a_{12} & a_{13} & a_{11} & a_{12} & a_{13} & a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} & a_{21} & a_{22} & a_{23} & a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} & a_{31} & a_{32} & a_{33} & a_{31} & a_{32} & a_{33} \end{array}\right)$$

$$|A| = a_{11}a_{22}a_{33} + a_{12}a_{23}a_{31} + a_{13}a_{21}a_{32} - a_{11}a_{23}a_{32} - a_{12}a_{21}a_{33} - a_{13}a_{22}a_{31}$$

- 代数余子式：在一个 n 阶行列式 A 中，把 (i,j) 元素 a_{ij} 所在的第 i 行和第 j 列划去后，留下的 $n-1$ 阶方阵的行列式叫做元素 a_{ij} 的余子式，记做 M_{ij} 。代数余子式记为 A_{ij} ，$A_{ij} = (-1)^{i+j} M_{ij}$ 。

$$A = \begin{pmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{pmatrix}$$
$$\forall 1 \le j \le n, |A| = \sum_{i=1}^{n} a_{ij} \cdot (-1)^{i+j} M_{ij}$$
$$\forall 1 \le i \le n, |A| = \sum_{j=1}^{n} a_{ij} \cdot (-1)^{i+j} M_{ij}$$

$$|A| = a_{11}a_{22}a_{33} + a_{12}a_{23}a_{31} + a_{13}a_{21}a_{32} - a_{11}a_{23}a_{32} - a_{12}a_{21}a_{33} - a_{13}a_{22}a_{31}$$

任一行元素与其对应代数余子式乘积之和等于行列式，与其他行对应元素的代数余子式乘积之和为0 。

- 范德蒙行列式（Vandermonde）：

$$
D_n = \begin{vmatrix} 1 & 1 & 1 & \cdots & 1 \\ x_1 & x_2 & x_3 & \cdots & x_n \\ x_1^2 & x_2^2 & x_3^2 & \cdots & x_n^2 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ x_1^{n-1} & x_2^{n-1} & x_3^{n-1} & \cdots & x_n^{n-1} \end{vmatrix} = \prod_{i,j(n\geq i)>1)} (x_i - x_j)
$$

证明提示：数学归纳法，参考Lagrange/Newton插值法。

3.2.2 行列式的性质

1. 上三角和下三角行列式等于主对角线元素相乘：

$$
\begin{vmatrix} a_{11} & 0 & \dots & 0 \\ a_{21} & a_{22} & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & \dots & a_{n2} \end{vmatrix} = \begin{vmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ 0 & a_{22} & \dots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & a_n \end{vmatrix} = \begin{vmatrix} a_{11} & 0 & \dots & 0 \\ 0 & a_{22} & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & a_{nn} \end{vmatrix} = a_{11}a_{22}\dots a_{nn}
$$

2. 行列互换，行列式不变：

$$
\begin{vmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & \dots & a_m \end{vmatrix} = \begin{vmatrix} a_{11} & a_{21} & \dots & a_{n1} \\ a_{12} & a_{22} & \dots & a_{n2} \\ \vdots & \vdots & \ddots & \vdots \\ a_{1n} & a_{2n} & \dots & a_{mn} \end{vmatrix}
$$

3. 行列式提取公因子:

$$
\begin{vmatrix} ka_{11} & ka_{12} & \dots & ka_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & \dots & a_m \end{vmatrix} = k \begin{vmatrix} a_{11} & a_{21} & \dots & a_{n1} \\ a_{12} & a_{22} & \dots & a_{n2} \\ \vdots & \vdots & \ddots & \vdots \\ a_{1n} & a_{2n} & \dots & a_m \end{vmatrix}
$$

因此，当行列式某一行或某一列为零，那么行列式为零。（只需令 $k = 0$ 得证）

4. 某一行求和分解:

$$
\begin{vmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ & \vdots & \vdots & \vdots \\ b_1 + c_1 & b_2 + c_2 & \dots & b_n + c_n \\ \vdots & \vdots & & \vdots \\ a_{n1} & a_{n2} & \dots & a_{mn} \end{vmatrix} = \begin{vmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ \vdots & \vdots & & \vdots \\ b_1 & b_2 & \dots & b_n \\ \vdots & \vdots & & \vdots \\ a_{n1} & a_{n2} & \dots & a_{nn} \end{vmatrix} + \begin{vmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ \vdots & \vdots & & \vdots \\ c_1 & c_2 & \dots & c_n \\ \vdots & \vdots & & \vdots \\ a_{n1} & a_{n2} & \dots & a_{mn} \end{vmatrix}
$$

5. 行列式中有两行或两列完成相同，那么行列式为零。

6. 如果行列式中两行成比例，那么行列式为零:

$$
\begin{vmatrix} ka_{11} & ka_{12} & \dots & ka_{1n} \\ \vdots & \vdots & & \vdots \\ a_{i1} & a_{i2} & \dots & a_{in} \\ \vdots & \vdots & & \vdots \\ ka_{i1} & ka_{i2} & \dots & ka_{in} \\ \vdots & \vdots & \vdots & \vdots \\ a_{n1} & a_{n2} & \dots & a_m \end{vmatrix} = k \begin{vmatrix} ka_{11} & ka_{12} & \dots & ka_{1n} \\ \vdots & \vdots & & \\ a_{i1} & a_{i2} & \dots & a_{in} \\ \vdots & \vdots & & \vdots \\ a_{i1} & a_{i2} & \dots & a_{in} \\ \vdots & \vdots & \vdots & \\ a_{n1} & a_{n2} & \dots & a_m \end{vmatrix} = 0
$$

7. 把一行的倍数加到另一行，行列式不变:

$$
\begin{vmatrix}
a_{11} & a_{12} & \dots & a_{1n} \\
\vdots & \vdots & & \vdots \\
a_{i1} & a_{i2} & \dots & a_{in} \\
\vdots & \vdots & & \vdots \\
a_{j1}+ka_{i1} & a_{j2}+ka_{i2} & \dots & a_{jn}+ka_{in} \\
\vdots & \vdots & & \vdots \\
a_{n1} & a_{n2} & \dots & a_{nn}
\end{vmatrix}
$$

$$
=\begin{vmatrix}
a_{11} & a_{12} & \dots & a_{1n} \\
\vdots & \vdots & & \vdots \\
a_{i1} & a_{i2} & \dots & a_{in} \\
\vdots & \vdots & & \vdots \\
a_{j1} & a_{j2} & \dots & a_{jn} \\
\vdots & \vdots & \vdots & \vdots \\
a_{n1} & a_{n2} & \dots & a_{nn}
\end{vmatrix}
+k\begin{vmatrix}
a_{11} & a_{12} & \dots & a_{1n} \\
\vdots & \vdots & & \\
a_{i1} & a_{i2} & \dots & a_{in} \\
\vdots & \vdots & & \vdots \\
a_{i1} & a_{i2} & \dots & a_{in} \\
\vdots & \vdots & \vdots & \vdots \\
a_{n1} & a_{n2} & \dots & a_{nn}
\end{vmatrix}
=\begin{vmatrix}
a_{11} & a_{12} & \dots & a_{1n} \\
\vdots & \vdots & & \vdots \\
a_{i1} & a_{i2} & \dots & a_{in} \\
\vdots & \vdots & & \vdots \\
a_{j1} & a_{j2} & \dots & a_{jn} \\
\vdots & \vdots & \vdots & \vdots \\
a_{n1} & a_{n2} & \dots & a_{nn}
\end{vmatrix}
$$

8. 兑换行列式中两行的位置，行列式反号:

$$
\begin{vmatrix}
ka_{11} & ka_{12} & \dots & ka_{1n} \\
\vdots & \vdots & & \vdots \\
a_{i1} & a_{i2} & \dots & a_{in} \\
\vdots & \vdots & & \vdots \\
a_{j1} & a_{j2} & \dots & a_{jn} \\
\vdots & \vdots & \vdots & \vdots \\
a_{n1} & a_{n2} & \dots & a_{nn}
\end{vmatrix}
$$

$$
=\begin{vmatrix}
ka_{11} & ka_{12} & \dots & ka_{1n} \\
\vdots & \vdots & & \vdots \\
a_{i1}+a_{j1} & a_{i2}+a_{j2} & \dots & a_{in}+a_{jn} \\
\vdots & \vdots & & \vdots \\
a_{j1} & a_{j2} & \dots & a_{jn} \\
\vdots & \vdots & \vdots & \vdots \\
a_{n1} & a_{n2} & \dots & a_{m}
\end{vmatrix}
=\begin{vmatrix}
ka_{11} & ka_{12} & \dots & ka_{1n} \\
\vdots & \vdots & & \vdots \\
a_{i1}+a_{j1} & a_{i2}+a_{j2} & \dots & a_{in}+a_{jn} \\
\vdots & \vdots & & \vdots \\
-a_{i1} & -a_{i2} & \cdots & -a_{in} \\
\vdots & \vdots & \vdots & \vdots \\
a_{n1} & a_{n2} & \dots & a_{nn}
\end{vmatrix}
$$

$$
=\begin{vmatrix}
ka_{11} & ka_{12} & \dots & ka_{1n} \\
\vdots & \vdots & & \vdots \\
a_{j1} & a_{j2} & \dots & a_{jn} \\
\vdots & \vdots & & \vdots \\
-a_{i1} & -a_{i2} & \cdots & -a_{in} \\
\vdots & \vdots & \vdots & \vdots \\
a_{n1} & a_{n2} & \dots & a_{nn}
\end{vmatrix}
=-\begin{vmatrix}
ka_{11} & ka_{12} & \dots & ka_{1n} \\
\vdots & \vdots & & \vdots \\
a_{j1} & a_{j2} & \dots & a_{jn} \\
\vdots & \vdots & & \vdots \\
a_{i1} & a_{i2} & \dots & a_{in} \\
\vdots & \vdots & \vdots & \vdots \\
a_{n1} & a_{n2} & \dots & a_{nn}
\end{vmatrix}
$$

3.2.3 方程组及克拉默法则

1. 引言

设有三元线性方程组：

$$
\begin{cases}
a_{11}x_1 + a_{12}x_3 + a_{13}x_3 = b_1 \\
a_{21}x_1 + a_{22}x_3 + a_{23}x_3 = b_2 \\
a_{31}x_1 + a_{32}x_3 + a_{33}x_3 = b_3
\end{cases}
$$

则有三级行列式：

$$
d = \begin{vmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{vmatrix} = a_{11}a_{22}a_{33} + a_{12}a_{23}a_{31} + a_{13}a_{21}a_{32} - a_{11}a_{23}a_{32} - a_{12}a_{21}a_{33} - a_{13}a_{22}a_{31}
$$

其中当上述方程组的三级行列式 $d \neq 0$ 有唯一解，解为：

$$
x_1 = \frac{d_1}{d}, x_2 = \frac{d_2}{d}, x_3 = \frac{d_3}{d}
$$

其中，

$$
d_1 = \begin{vmatrix} b_1 & a_{12} & a_{13} \\ b_2 & a_{22} & a_{23} \\ b_3 & a_{32} & a_{33} \end{vmatrix}, d_2 = \begin{vmatrix} a_{11} & b_1 & a_{13} \\ a_{21} & b_2 & a_{23} \\ a_{31} & b_3 & a_{33} \end{vmatrix}, \quad d_3 = \begin{vmatrix} a_{11} & a_{12} & b_1 \\ a_{21} & a_{22} & b_2 \\ a_{31} & a_{32} & b_3 \end{vmatrix}
$$

2. 克拉默法则
   该结果可以推广到 n 元线性方程
   如果线性方程组：

$$
\begin{cases}
a_{11}x_1 + a_{12}x_2 + \ldots + a_{1n}x_n = b_1 \\
a_{21}x_1 + a_{22}x_2 + \ldots + a_{2n}x_n = b_2 \\
\cdots\cdots\cdots \\
a_{n1}x_1 + a_{n2}x_2 + \ldots + a_{nn}x_n = b_n
\end{cases}
$$

的系数矩阵：

$$
A = \begin{pmatrix} a_{11} & a_{12} & \ldots & a_{1n} \\ a_{21} & a_{22} & \ldots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & \ldots & a_{nn} \end{pmatrix}
$$

的行列式 $d = |A| \neq 0$ ，那么线性方程组有解，并且解是唯一解，解可以通过系数表为：

$$
x_1 = \frac{d_1}{d}, x_2 = \frac{d_2}{d}, \ldots, x_n = \frac{d_n}{d}
$$

其中 d_j 是把矩阵 A 中的第 j 列换成常数项的矩阵的行列式：

$$
d_j = \begin{vmatrix} a_{11} & \ldots & a_{1j-1} & b_1 & a_{1j+1} & \ldots & a_{1n} \\ a_{21} & \ldots & a_{2j-1} & b_2 & a_{2j+1} & \ldots & a_{2n} \\ \vdots & & \vdots & \vdots & \vdots & & \vdots \\ a_{n1} & \ldots & a_{nj-1} & b_n & a_{2j+1} & \ldots & a_{nn} \end{vmatrix}, j = 1, 2, \ldots, n
$$

根据克拉默法则，对于齐次线性方程组，即常数项 $b = 0$ ，因为零解一定是齐次线性方程组的解，因此齐次线性方程组有非零解的话，则对应的 $|A| = 0$ 。

在齐次线性方程组中，方程个数小于变量个数，则解不唯一。

实际上，如果齐次线性方程组的系数矩阵的行秩 r 小于未知数个数 n ，则有非零解。

3.2.4 线性方程组的解的结构

如果齐次线性方程组多个解的情况下：

$$\begin{cases} a_{11}x_1 + a_{12}x_2 + \ldots + a_{1n}x_n = 0 \\ a_{21}x_1 + a_{22}x_2 + \ldots + a_{2n}x_n = 0 \\ \ldots\ldots\ldots \\ a_{n1}x_1 + a_{n2}x_2 + \ldots + a_{nn}x_n = 0 \end{cases}$$

齐次线性方程组的一组解 $\eta_1, \eta_2, \ldots, \eta_n$ 称为一个基础解系，如果：

（1） 齐次线性方程组的任一个解 $\eta_1, \eta_2, \ldots, \eta_n$ 都能表成 的线性组合；

（2） $\eta_1, \eta_2, \ldots, \eta_n$ 线性无关；

注：如果 $\eta_1, \eta_2, \ldots, \eta_n$ 是线性相关，那么其中一个解就可以表示成剩下其他解的线性组合。

定理：在齐次线性方程组有非零解的情况下，它有基础解系，并且基础解系所含解的个数等于 $n-r$，这里 r 表示系数矩阵的秩，实际上，$n-r$ 也就是自由未知量的个数。

方程改写为:

$$\begin{cases} a_{11}x_1 + \ldots + a_{1r}x_r = -a_{1r+1}x_{r+1} - \ldots - a_{1n}x_n \\ a_{21}x_1 + \ldots + a_{2r}x_r = -a_{2r+1}x_{r+1} - \ldots - a_{2n}x_n \\ \vdots \\ a_{r1}x_1 + \ldots + a_{rr}x_r = -a_{r+1}x_{r+1} - \ldots - a_{rn}x_n \end{cases}$$

左边含有 r 项，右边（能够自由取值）的有 $n-r$ 项，右边依次代入 $n-r$ 组数：$(1,0,\ldots,0),(0,1,\ldots,0),\ldots,(0,0,\ldots,1)$， 就得到方程组的 $n-r$ 个解。

3.3 线性相关性

定义3.3.1： 向量 α 称为向量组 $\beta_1, \beta_2, \ldots, \beta_s$ 的一个线性组合，如果有数域 P 中的数 $k_1, k_2, \ldots, k_s$ 使得：$\alpha = k_1\beta_1 + k_2\beta_2 + \ldots k_s\beta_s$ 。

任一 n 维向量 $\alpha = (a_1, a_2, \ldots, a_n)$ 都是向量组 ε 的线性组合，其中 $\varepsilon = \begin{cases} \varepsilon_1 = (1,0,\ldots,0) \\ \varepsilon_2 = (0,1,\ldots,0) \\ \ldots\ldots \\ \varepsilon_n = (0,0,\ldots,1) \end{cases}$，因为$\alpha = a_1\varepsilon_1 + a_2\varepsilon_2 + \ldots + a_n\varepsilon_n$ ，其中 $\varepsilon_1, \varepsilon_2, \ldots \varepsilon_n$ 称为 n 维单位向量。

定义3.3.2： 如果向量组 $\alpha_1, \alpha_2, \ldots, \alpha_t$ 中每一个向量 $\alpha_i(i=1,2,\ldots,t)$ 都可以由向量组 $\beta_1, \beta_2, \ldots, \beta_s$ 线性表出，那么向量组 $\alpha_1, \alpha_2, \ldots, \alpha_t$ 就称为可以经向量组 $\beta_1, \beta_2, \ldots, \beta_s$ 线性标出。如果两个向量组互相可以线性表出，他们就称为等价。

定义3.3.3： 如果向量组 $\alpha_1, \alpha_2, \ldots, \alpha_s(s \geq 2)$ 中有一个向量可以由其余的向量线性表出，那么向量组 $\alpha_1, \alpha_2, \ldots, \alpha_s$ 称为线性相关的。

定义3.3.4： 向量组 $\alpha_1, \alpha_2, \ldots, \alpha_s(s \geq 2)$ 称为线性相关，如果有数据 P 中不全为零的数 $k_1, k_2, \ldots k_s$，使得 $k_1\alpha_1 + k_2\alpha_2 + \ldots k_s\alpha_s = 0$ 。

定义3.3.5： 向量组 $\alpha_1, \alpha_2, \ldots, \alpha_s(s \geq 1)$ 称为线性无关，如果由使得 $k_1\alpha_1 + k_2\alpha_2 + \ldots k_s\alpha_s = 0$ 可以推出 $k_1 = k_2 = \ldots = k_s = 0$ 。

定理3.3.1：

有向量组 $\alpha_1, \alpha_2, \ldots, \alpha_r$ 与 $\beta_1, \beta_2, \ldots, \beta_s$，如果：

（1）向量组 $\alpha_1, \alpha_2, \ldots, \alpha_r$ 可以经 $\beta_1, \beta_2, \ldots, \beta_s$ 线性表出；

（2） $r > s$；

那么向量组 $\alpha_1, \alpha_2, \ldots, \alpha_r$ 必定线性相关。

- 推论（1）如果向量组 $\alpha_1,\alpha_2,...,\alpha_r$ 可以经 $\beta_1,\beta_2,...,\beta_s$ 线性表出，且 $\alpha_1,\alpha_2,...,\alpha_r$ 线性无关，那么 $r \leq s$；
- 推论（2）任意 $n+1$ 个 n 维向量必线性相关；
- 推论（3）两个线性无关的等价向量组，他们含有相同个数的向量。

定义3.3.6： 一个向量组的一个部分组成为一个极大线性无关组，如果这个部分组本身是线性无关的，并且从这个向量组中任意添加一个向量，所得的部分向量组都线性相关。

- 推论得到：任意一个极大线性无关组都与向量组本身等价。

定理3.3.2： 一个向量组的极大线性无关组都含有相同个数的向量。我们把向量组中极大线性无关组中的向量个数称之为这个向量组的秩。因此一个向量组线性无关的充分必要条件是它的秩和它所含向量的个数相同。

一个n 阶行列式的值不等于零，它等价于矩阵的秩等于 n 。

3.4 矩阵

3.4.1 矩阵的基本运算

1. **矩阵的加法**

   A 和 B 是两个 $s \times n$ 的矩阵:

$$A = (a_{ij})_{sn} = \begin{pmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{s1} & a_{s2} & \dots & a_{sn} \end{pmatrix}$$

$$B = (b_{ij})_{sn} = \begin{pmatrix} b_{11} & b_{12} & \dots & b_{1n} \\ b_{21} & b_{22} & \dots & b_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ b_{s1} & b_{s2} & \dots & b_{sn} \end{pmatrix}$$

$$C = A + B = (c_{ij})_{sn} = (a_{ij} + b_{ij})_{sn} = \begin{pmatrix} a_{11}+b_{11} & a_{12}+b_{12} & \dots & a_{1n}+b_{1n} \\ a_{21}+b_{21} & a_{22}+b_{22} & \dots & a_{2n}+b_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{s1}+b_{s1} & a_{32}+b_{s2} & \dots & a_{sn}+b_{sn} \end{pmatrix}$$

矩阵运算满足加法结合律及加法交换律:

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

$A + B = B + A$

$秩（A+B） \leq 秩（A） + 秩（B）$

2. **矩阵的乘法**

   A 为 $m \times n$ 阶的矩阵，B 为 $s \times n$ 阶的矩阵，那么，$C = A \times B$ 是 $m \times n$ 阶的矩阵，其中，$c_{ij} = \sum_{k=1}^{s} a_{ik} b_{ij}$ 。
   矩阵的乘法适合结合律不适合交换律:

$(AB)C = A(BC)$
$AB \neq BA$

3. **矩阵的方幂运算**

   设 A 为方阵:

$$
\begin{aligned}
&A^1 = A\\
&A^{k+1} = A^k A\\
&A^k A^l = A^{(k+l)}\\
&\left(A^k\right)^l = A^{kl}
\end{aligned}
$$

4. **矩阵的转置**

$$
A = (a_{ij})_{sn} = \begin{pmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{s1} & a_{s2} & \dots & a_{sn} \end{pmatrix}
$$

$$
A' = (a_{ij})_{ns} = \begin{pmatrix} a_{11} & a_{22} & \dots & a_{s2} \\ \vdots & \vdots & \ddots & \vdots \\ a_{1n} & a_{2n} & \dots & a_{sn} \end{pmatrix}
$$

运算：

$$
\begin{aligned}
&(A')' = A\\
&(A+B)' = A' + B'\\
&(AB)' = B'A'\\
&(kA)' = kA'
\end{aligned}
$$

5. **矩阵乘积行列式与秩**

（1） $|AB| = |A||B|$

推论有：$|A^n| = |A|^n$

定义于数域 P 上的 $n \times n$ 矩阵 A 称为非退化的，如果 $|A| \neq 0$ ，矩阵 A 为非退化的充要条件是 A 的秩等于 n 。

（2） 秩（$A+B$） $\leq min($秩（A）,秩（B））

3.4.2 伴随矩阵和逆

1. 伴随矩阵

对于 $n \times n$ 矩阵任意元素 a_{ij} 都有各自的代数余子式 $A_{ij} = (-1)^{i+j} M_{ij}$ ，构造 $n \times n$ 的方阵 A^* ， 称 A^* 其为矩阵 A 的伴随矩阵。

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

注意，这里的第 i 行第 j 列元素 A_{ji} 是 a_{ji} 的代数余子式 。

2. 矩阵的逆

定义： n 级方阵 A 称为可逆的，如果有 n 级方程 B 使得 $AB = BA = E$，其中 E 是 n 级单位矩阵。

矩阵 A 可逆的充要条件是 A 是非退化的。

根据公式：$|A| = \sum_{j=1}^{n} a_{ij} \cdot (-1)^{i+j} M_{ij}$ 且 a_{ij} 与其他行对应元素的代数余子式乘积之和为0的性质，我们有：

$$A=\begin{pmatrix}a_{11}&a_{12}&\dots&a_{1n}\\a_{21}&a_{22}&\dots&a_{2n}\\\vdots&\vdots&\ddots&\vdots\\a_{n1}&a_{n2}&\dots&a_{nn}\end{pmatrix},A^*=\begin{pmatrix}A_{11}&A_{21}&\dots&A_{n1}\\A_{12}&A_{22}&\dots&A_{n2}\\\vdots&\vdots&\ddots&\vdots\\A_{1n}&A_{2n}&\dots&A_{nn}\end{pmatrix}$$

$$A\cdot A^*=\begin{pmatrix}|A|&0&\dots&0\\0&|A|&\dots&0\\\vdots&\vdots&\ddots&\vdots\\0&0&\dots&|A|\end{pmatrix}=|A|\cdot I$$

$$\Rightarrow A^{-1}=\frac{1}{|A|}\cdot A^*$$

逆的运算：

$$(A')^{-1}=\left(A^{-1}\right)'$$
$$(AB)^{-1}=A^{-1}B^{-1}$$

定理： A 是一个 $s\times n$ 矩阵，如果 P 是一个 $s\times s$ 可逆矩阵，Q 是一个 $n\times n$ 可逆矩阵，那么：

$$秩(A)=秩(PA)=秩(AQ)$$

证明：

令 $B=PA$，有 秩（B） $\leq$ 秩（A），
又因为 $A=P^{-1}B$，所以有 秩（A） $\leq$ 秩（B）

所以有 $秩(A)=秩(PA)=秩(AQ)$。

3. 矩阵的秩

定义： 矩阵的行秩就是矩阵行向量组的秩，矩阵的列秩就是矩阵的列向量组的秩。矩阵的行秩和列秩相同。

定理： $n\times n$ 矩阵 $A=\begin{pmatrix}a_{11}&a_{12}&\dots&a_{1n}\\a_{21}&a_{22}&\dots&a_{2n}\\\vdots&\vdots&\ddots&\vdots\\a_{n1}&a_{n2}&\dots&a_{nn}\end{pmatrix}$ 的行列式为零的充分必要条件是 A 的秩小于 n。

有线性方程组$\begin{cases}a_{11}x_1+a_{12}x_3+\ldots+a_{1n}x_n=b_1\\a_{21}x_1+a_{22}x_3+\ldots+a_{2n}x_n=b_2\\\ldots\ldots\ldots\\a_{s1}x_1+a_{s2}x_3+\ldots+a_{sn}x_n=b_s\end{cases}$，

引入向量 $\alpha_1=\begin{pmatrix}a_{11}\\a_{21}\\\vdots\\a_{51}\end{pmatrix},\alpha_2=\begin{pmatrix}a_{12}\\a_{22}\\\vdots\\a_{52}\end{pmatrix},\ldots\alpha_s=\begin{pmatrix}a_{1n}\\a_{2n}\\\vdots\\a_{sn}\end{pmatrix},\beta=\begin{pmatrix}b_1\\b_2\\\vdots\\b_s\end{pmatrix}$，

改写成：$x_1\alpha_1+x_2\alpha_2+\ldots+x_n\alpha_n=\beta$，

上述方程组有解的条件为它的系数矩阵与增广矩阵有相同的秩。

$$A=\begin{pmatrix}a_{11}&a_{12}&\dots&a_{1n}\\a_{21}&a_{22}&\dots&a_{2n}\\\vdots&\vdots&\ddots&\vdots\\a_{s1}&a_{s2}&\dots&a_{sn}\end{pmatrix},\overline{A}=\begin{pmatrix}a_{11}&a_{12}&\dots&a_{1n}&b_1\\a_{21}&a_{22}&\dots&a_{2n}&b_2\\\vdots&\vdots&&\vdots&\vdots\\a_{51}&a_{s2}&\dots&a_{sn}&b_s\end{pmatrix}$$

实际上就是 β 可以有 α 线性表出。

3.4.3 二次型和正定矩阵

正定阵：对于 n 阶方阵 A ，若任意 n 阶向量 X ，都有 $X^TAX>0$ ，则称 A 是正定阵。若条件变成 $X^TAX\geq 0$ ，则 A 称作半正定阵，类似的还有负定阵，半负定阵。
思考：给定任意 $m\times n$ 的矩阵 A ，证明 A^TA 一定是半正定方阵。（该结论在线性回归中将用到）

正定阵的判定：

（1）对称阵 A 为正定阵；

（2） A 的特征值都为正；

（3）A 的顺序主子式大于0；

以上三个命题等价。

$$(a_{11})\cdot\begin{pmatrix}a_{11}&a_{12}\\a_{21}&a_{22}\end{pmatrix}\cdot\begin{pmatrix}a_{11}&a_{12}&a_{13}\\a_{21}&a_{22}&a_{23}\\a_{31}&a_{32}&a_{33}\end{pmatrix}\cdot\begin{pmatrix}a_{11}&a_{12}&\cdots&a_{1n}\\a_{21}&a_{22}&\cdots&a_{2n}\\\vdots&\vdots&\ddots&\vdots\\a_{n1}&a_{n2}&\cdots&a_{m}\end{pmatrix}$$

合同：

数域 P 上的 $n\times n$ 矩阵 A ， B 称为合同的，如果有数域 P 上的可逆的 $n\times n$ 矩阵 C 使得：$B=C^{'}AC$ 。

合同关系的性质：

1. 反身性： $A=E'AE$ ；
2. 对称性：由 $B=C'AC$ 得到 $A=\left(C^{-1}\right)'BC^{-1}$ ；
3. 传递性：$A_1=C_1'AC_1,A_2=C_2'AC_2$ 得到 $A_1=(C_1C_2)'A(C_1C_2)$ 。

3.5 线性空间

3.5.1 映射

定义映射：设 M 与 M' 是两个集合，所谓集合 M 到集合 M' 的一个映射就是指一个法则，它使 M 中的每个元素 a 都有 M' 中一个确定的元素 a' 与之对应，如果映射 σ 使元素 $a'\in M'$ 与元素 $a\in M$ 对应，那么就记为：

$$\sigma(a)=a'$$

a' 就称为 a 在映射 σ 下的像，而 a 称为 a' 在映射 σ 下的一个原像。

需要注意的是，映射和函数还有一定的区别，从本质上讲，函数是一种特殊的映射，是从非空数集到非空数集的映射，而且只能是一对一映射或多对一映。

映射的不同分类是根据映射的结果进行的，从下面的三个角度进行：

1．根据结果的几何性质分类：满射（到上）与非满射（内的）

2．根据结果的分析性质分类：单射（一一的）与非单射

3．同时考虑几何与分析性质：满的单射（一一对应）。

定义如下：

满射：集合 M' 中的每个元素 a' 在集合 M 中具有原像；

单射：集合 M 中不同的元素 a 在集合 M' 中都有不同的象；

双射：既是单射又是满射的映射称为双射，亦称“一一映射”。

3.5.2 线性空间

定义：给定非空集合 V 和域 P，若：

1. 在 V 中定义了一种运算，称为加法，即对 V 中任意两个元素 α与 β 都按某一法则对应于 V 内惟一确定的一个元素 $\alpha+\beta$，称为 α 与 β 的和。
2. 在P与V的元素间还定义了一种运算，称为纯量乘法(亦称数量乘法)，即对V中任意元素α和P中任意元素k，都按某一法则对应V内惟一确定的一个元素$k\alpha$，称为k与α的积。
3. 加法与纯量乘法满足以下条件：

   1) $\alpha+\beta=\beta+\alpha$，对任意$\alpha$，$\beta\in V$.

   2) $\alpha+(\beta+\gamma)=(\alpha+\beta)+\gamma$，对任意$\alpha$，$\beta$，$\gamma\in V$.

   3) 存在一个元素$0\in V$，对一切$\alpha\in V$有$\alpha+0=\alpha$，元素 0 称为 V 的零元.

   4) 对任一$\alpha\in V$，都存在$\beta\in V$使$\alpha+\beta=0$，β 称为α的负元素，记为$\beta=-\alpha$.

   5) 对 P 中单位元1，有$1\alpha=\alpha(\alpha\in V)$.

   6) 对任意k，$l\in P$，$\alpha\in V$有$(kl)\alpha=k(l\alpha)$.

   7) 对任意k，$l\in P$，$\alpha\in V$有$(k+l)\alpha=k\alpha+l\alpha$.

   8) 对任意$k\in P$，α，$\beta\in V$有$k(\alpha+\beta)=k\alpha+k\beta$，

则称V为域P上的一个线性空间，或向量空间。

3.5.3 维数，基与坐标

定义维数：如果在线性空间中 V 有 n 个线性无关的向量，但是如果没有更多数目的线性无关的向量，那么 V 就被称为 n 维的；如果在 V 中可以找到任意多个线性无关的向量，那么 V 就称为无线维的。

定义基和坐标：在 n 维线性空间 V 中， n 个线性无关的向量 $\varepsilon_1,\varepsilon_2,\dots,\varepsilon_n$ 称为 V 的一组基，设 α 是 V 中任意向量，于是 $\varepsilon_1,\varepsilon_2,\dots,\varepsilon_n$ 与 α 线性相关，因此 α 可以被基 $\varepsilon_1,\varepsilon_2,\dots,\varepsilon_n$ 线性表出：$\alpha=a_1\varepsilon_1+a_2\varepsilon_2+\ldots+a_n\varepsilon_n$，其中系数 $a_1,a_2,\dots,a_n$ 是被向量 α 与基 $\varepsilon_1,\varepsilon_2,\dots,\varepsilon_n$ 唯一确定的，这组数就称为 α 在基 $\varepsilon_1,\varepsilon_2,\dots,\varepsilon_n$ 下的坐标，记为 :$(a_1,a_2,\dots,a_n)$

如果在线性空间 V 中有 n 个线性无关的向量 $\alpha_1,\alpha_2,\dots,\alpha_n$，且 V 中任一向量都可以用它们线性表出，那么 V 是 n 维的，并且 $\alpha_1,\alpha_2,\dots,\alpha_n$ 就是 V 的一组基。

3.5.4 基变换与坐标变换

1. 基变换

   设 $\varepsilon_1,\varepsilon_2,\dots,\varepsilon_n$ 与 $\varepsilon'_1,\varepsilon'_2,\dots,\varepsilon'_n$ 是 n 维线性空间 V 中的两组基，它们的关系是：

$$\begin{cases}\varepsilon'_1=a_{11}\varepsilon_1+a_{21}\varepsilon_2+\ldots+a_{n1}\varepsilon_1\\ \varepsilon'_2=a_{12}\varepsilon_1+a_{22}\varepsilon_2+\ldots+a_{n2}\varepsilon_2\\ \cdots\cdots\cdots\\ \varepsilon'_n=a_{1n}\varepsilon_1+a_{2n}\varepsilon_2+\ldots+a_{nn}\varepsilon_n\end{cases}$$

   的系数矩阵，

   设向量 ξ 在这两组基下的坐标分别为 $(x_1,x_2,\dots,x_n)$ 与 $(x'_1,x'_2,\dots,x'_n)$ 即有：

$$\xi=x_1\varepsilon_1+x_2\varepsilon_2+\ldots+x_n\varepsilon_n=x_1\varepsilon_1+x_2\varepsilon_2+\ldots+x_n\varepsilon_n$$

$$\xi=(\varepsilon_1,\varepsilon_2,\ldots,\varepsilon_n)\begin{pmatrix}x_1\\x_2\\\vdots\\x_n\end{pmatrix}$$

可以改写成：

$$(\varepsilon_1,\varepsilon_2',\ldots,\varepsilon_n')=(\varepsilon_1,\varepsilon_2,\ldots,\varepsilon_n)\begin{pmatrix}a_{11}&a_{12}&\ldots&a_{1n}\\a_{21}&a_{22}&\ldots&a_{2n}\\\vdots&\vdots&\ddots&\vdots\\a_{n1}&a_{n2}&\ldots&a_{nn}\end{pmatrix}$$

矩阵 $A=\begin{pmatrix}a_{11}&a_{12}&\ldots&a_{1n}\\a_{21}&a_{22}&\ldots&a_{2n}\\\vdots&\vdots&\ddots&\vdots\\a_{n1}&a_{n2}&\ldots&a_{nn}\end{pmatrix}$ 称为基 $\varepsilon_1,\varepsilon_2,\ldots,\varepsilon_n$ 到基 $\varepsilon_1',\varepsilon_2',\ldots,\varepsilon_n'$ 的过渡矩阵。

2. 坐标变换

由 $\xi=x_1\varepsilon_1+x_2\varepsilon_2+\ldots+x_n\varepsilon_n=x_1'\varepsilon_1'+x_2'\varepsilon_2'+\ldots+x_n'\varepsilon_n'$ 以及基变换公式，有

$$(\varepsilon_1',\varepsilon_2',\ldots,\varepsilon_n')\begin{pmatrix}{x_1}'\\{x_2}'\\\vdots\\{x_n}'\end{pmatrix}=\xi=(\varepsilon_1,\varepsilon_2,\ldots,\varepsilon_n)\begin{pmatrix}x_1\\x_2\\\vdots\\x_n\end{pmatrix}$$

得到：

$$(\varepsilon_1,\varepsilon_2,\ldots,\varepsilon_n)\begin{pmatrix}a_{11}&a_{12}&\ldots&a_{1n}\\a_{21}&a_{22}&\ldots&a_{2n}\\\vdots&\vdots&\ddots&\vdots\\a_{n1}&a_{n2}&\ldots&a_{nn}\end{pmatrix}\begin{pmatrix}{x_1}'\\{x_2}'\\\vdots\\{x_n}'\end{pmatrix}=(\varepsilon_1,\varepsilon_2,\ldots,\varepsilon_n)\begin{pmatrix}x_1\\x_2\\\vdots\\x_n\end{pmatrix}$$

即：

$$\begin{pmatrix}{x_1}'\\{x_2}'\\\vdots\\{x_n}'\end{pmatrix}=\begin{pmatrix}a_{11}&a_{12}&\ldots&a_{1n}\\a_{21}&a_{22}&\ldots&a_{2n}\\\vdots&\vdots&\ddots&\vdots\\a_{n1}&a_{n2}&\ldots&a_{nn}\end{pmatrix}^{-1}\begin{pmatrix}x_1\\x_2\\\vdots\\x_n\end{pmatrix}$$

或：

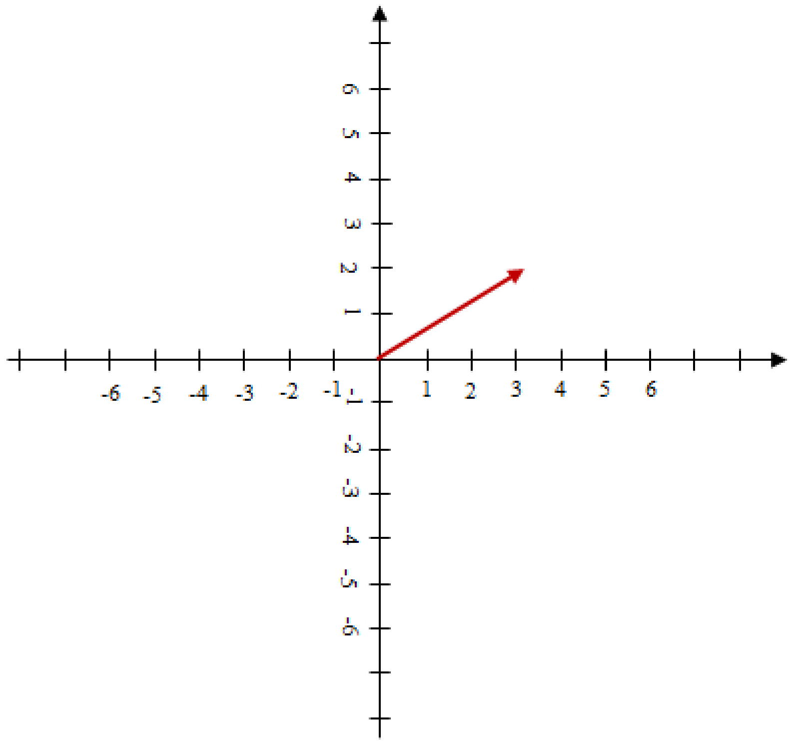
$$\begin{pmatrix}x_1\\x_2\\\vdots\\x_n\end{pmatrix}=\begin{pmatrix}a_{11}&a_{12}&\ldots&a_{1n}\\a_{21}&a_{22}&\ldots&a_{2n}\\\vdots&\vdots&\ddots&\vdots\\a_{n1}&a_{n2}&\ldots&a_{nn}\end{pmatrix}\begin{pmatrix}{x_1}'\\{x_2}'\\\vdots\\{x_n}'\end{pmatrix}$$

注意的是，当 $\varepsilon_1,\varepsilon_2,\ldots,\varepsilon_n$ 与 $\varepsilon_1',\varepsilon_2',\ldots,\varepsilon_n'$ 都是标准正交基，那么 A 是属于正交变换，有 $A'=A^T$ 。

应该是对的：假设投影变换后的新坐标系为 $w=\{w_1,w_2,w_m\}$（一组标准正交基），实际上 $w_1=(a_{11},a_{21},\ldots,a_{n1})$，所以变化后的新坐标有 $\begin{pmatrix}{x_1}'\\{x_2}'\\\vdots\\{x_n}'\end{pmatrix}=w^T\begin{pmatrix}x_1\\x_2\\\vdots\\x_n\end{pmatrix}$ 。

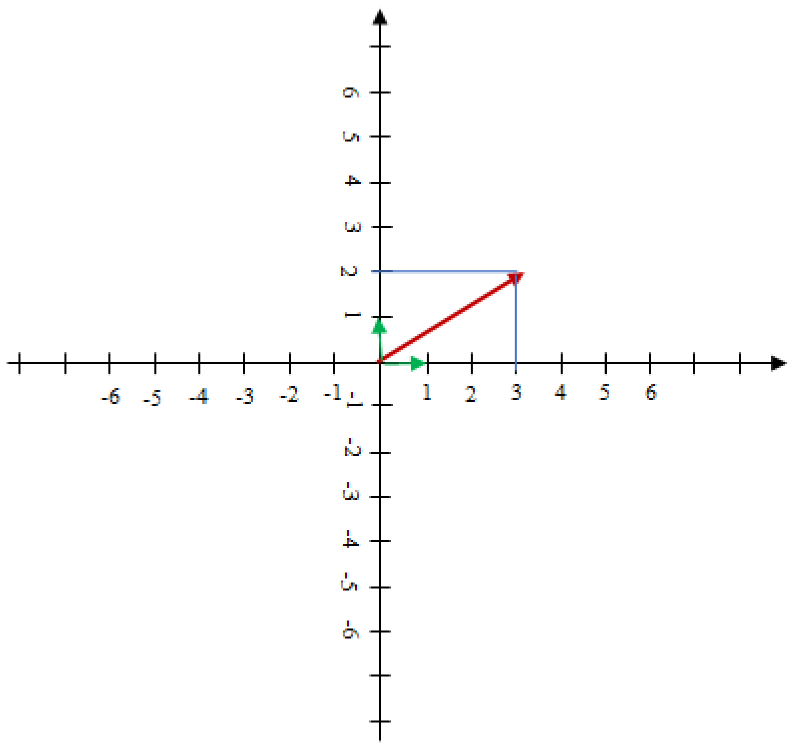
例子：

一个二维向量可以对应二维笛卡尔直角坐标系中从原点出发的一个有向线段。例如下面这个向量：



在代数表示方面，我们经常用线段终点的点坐标表示向量，例如上面的向量可以表示为(3,2)，这是我们再熟悉不过的向量表示。

不过我们常常忽略，只有一个(3,2)本身是不能够精确表示一个向量的。我们仔细看一下，这里的3实际表示的是向量在 x 轴上的投影值是3，在 y 轴上的投影值是2。也就是说我们其实隐式引入了一个定义：以 x 轴和 y 轴上正方向长度为1的向量为标准。那么一个向量(3,2)实际是说在 x 轴投影为3而 y 轴的投影为2。注意投影是一个矢量，所以可以为负。
更正式的说，向量 (x, y) 实际上表示线性组合：$x(1,0)^T + y(0,1)^T$ ，不难证明所有二维向量都可以表示为这样的线性组合。此处(1,0)和(0,1)叫做二维空间中的一组基。

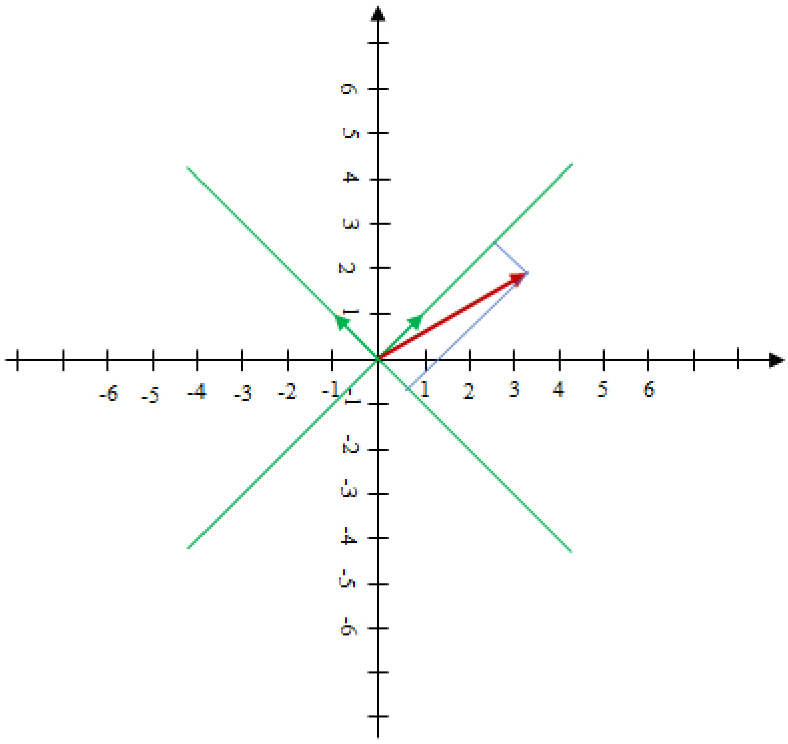


所以，要准确描述向量，首先要确定一组基，然后给出在基所在的各个直线上的投影值，就可以了。只不过我们经常省略第一步，而默认以(1,0)和(0,1)为基。 我们之所以默认选择(1,0)和(0,1)为基，当然是比较方便，因为它们分别是 x 和 y 轴正方向上的单位向量，因此就使得二维平面上点坐标和向量一一对应，非常方便。但实际上任何两个线性无关的二维向量都可以成为一组基，所谓线性无关在二维平面内可以直观认为是两个不在一条直线上的向量。

例如（1,1）以及（-1,1）也可以成为一组基（浩彬老师：二维空间中，2组线性无关向量）。当然因为这组基的模并不是1，所以我们要让这组基分别除以他们的模，

$\sqrt{1^2+1^2}=\sqrt{2}, \quad \sqrt{(-1)^2+1^2}=\sqrt{2}$，所以得到一组标准正交基 $\left(\frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}}\right), \left(-\frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}}\right)$

则新的坐标公式为 $\begin{pmatrix} \frac{1}{\sqrt{2}}, & \frac{1}{\sqrt{2}} \\ -\frac{1}{\sqrt{2}}, & \frac{1}{\sqrt{2}} \end{pmatrix} \begin{pmatrix} 3 \\ 2 \end{pmatrix} = \begin{pmatrix} \frac{5}{\sqrt{2}} \\ -\frac{1}{\sqrt{2}} \end{pmatrix}$



其中矩阵的两行分别为两个基，乘以原向量，其结果刚好为新基的坐标。可以稍微推广一下，如果我们有 m 个二维向量，只要将二维向量按列排成一个两行 m 列矩阵，然后用“基矩阵”乘以这个矩阵，就得到了所有这些向量在新基下的值。例如(1,1)，(2,2)，(3,3)，想变换到刚才那组基上，则可以这样表示：

$$\begin{pmatrix} 1/\sqrt{2} & 1/\sqrt{2} \\ -1/\sqrt{2} & 1/\sqrt{2} \end{pmatrix} \begin{pmatrix} 1 & 2 & 3 \\ 1 & 2 & 3 \end{pmatrix} = \begin{pmatrix} 2/\sqrt{2} & 4/\sqrt{2} & 6/\sqrt{2} \\ 0 & 0 & 0 \end{pmatrix}$$

于是一组向量的基变换被干净的表示为矩阵的相乘。

一般的，如果我们有M个 N 维向量，想将其变换为由 R 个 N 维向量表示的新空间中，那么首先将R个基按行组成矩阵A，然后将向量按列组成矩阵B，那么两矩阵的乘积AB就是变换结果，其中AB的第 m 列为A中第 m 列变换后的结果。数学表示为：

$$\begin{pmatrix} p_1 \\ p_2 \\ \vdots \\ p_R \end{pmatrix} \begin{pmatrix} a_1 & a_2 & \dots & a_M \end{pmatrix} = \begin{pmatrix} p_1a_1 & p_1a_2 & \cdots & p_1a_M \\ p_2a_1 & p_2a_2 & \cdots & p_2a_M \\ \vdots & \vdots & \ddots & \vdots \\ p_Ra_1 & p_Ra_2 & \cdots & p_Ra_M \end{pmatrix}$$

其中， P_i 是一个行向量，表示第 i 个基，a_j 是一个列向量，表示第 j 个原始数据记录。
其中生成结果中的算术 $(AB)_{ij}$ 表示的是第 j 个数据在第 i 维的坐标
设V是一个非空集合，P是一个数域。

3.6 线性变换

线性空间 V 的一个变换 $\mathcal{A}$ 称为线性变换，如果对于 V 中的任意的元素 α ， β 和数域P中任意数 k ，都有

$$\begin{aligned}
&\mathcal{A}(\alpha+\beta)=\mathcal{A}(\alpha)+\mathcal{A}(\beta)\\
&\mathcal{A}(k\alpha)=k\mathcal{A}(\alpha)
\end{aligned}$$

$\mathcal{A}(\alpha)$ 表示元素 α 在变换 $\mathcal{A}$ 下的像。简单来说，就是能够保持数乘和加法运算的变换。实际上，每个矩阵对应一个线性变换。

3.6.1 线性变换的矩阵

1. 线性变换的矩阵表示
   空间 V 中任一向量设 ξ 可以被基 $\varepsilon_1,\varepsilon_2,\ldots,\varepsilon_n$ 线性表出，即有关系是：

   $$\xi=x_1\varepsilon_1+x_2\varepsilon_2+\ldots+x_n\varepsilon_n$$

   由于线性变换保持线性关系不变，因而 ξ 的像 $\mathcal{A}(\xi)$ 与基的像 $\mathcal{A}\varepsilon_1,\mathcal{A}\varepsilon_2,\ldots,\mathcal{A}\varepsilon_n$ 也必然有相同的关系：

   $$\begin{aligned}
   \mathcal{A}\xi&=\mathcal{A}\left(x_1\varepsilon_1+x_2\varepsilon_2+\ldots+x_n\varepsilon_n\right)\\
   &=x_1\mathcal{A}\left(\varepsilon_1\right)+x_2\mathcal{A}\left(\varepsilon_2\right)+\ldots x_n\mathcal{A}\left(\varepsilon_n\right)
   \end{aligned}$$

   设 $\varepsilon_1,\varepsilon_2,\ldots,\varepsilon_n$ 是线性空间V中的一组基， $\mathcal{A}$ 是V中任意n个向量，存在唯一的线性变换 $\mathcal{A}$ 使：

   $$\mathcal{A}\varepsilon_i=a_i,i=1,2,\ldots n$$

   设 $\varepsilon_1,\varepsilon_2,\ldots,\varepsilon_n$ 是线性空间V中的一组基， $\mathcal{A}$ 是V中的一个线性变换，基向量的像可以被基线性表出：

   $$\left\{\begin{array}{l}
   \mathcal{A}\varepsilon_1=a_{11}\varepsilon_1+a_{21}\varepsilon_2+\ldots+a_{n1}\varepsilon_1\\
   \mathcal{A}\varepsilon_2=a_{12}\varepsilon_1+a_{22}\varepsilon_2+\ldots+a_{n2}\varepsilon_2\\
   \cdots\cdots\cdots\\
   \boldsymbol{A}\varepsilon_n=a_{1n}\varepsilon_1+a_{2n}\varepsilon_2+\ldots+a_{nn}\varepsilon_n
   \end{array}\right.$$

   用矩阵来表示就是：$\mathcal{A}\left(\varepsilon_1,\varepsilon_2,\ldots,\varepsilon_n\right)=\left(\mathcal{A}\varepsilon_1,\mathcal{A}\varepsilon_2,\ldots,\mathcal{A}\varepsilon_n\right)=\left(\varepsilon_1,\varepsilon_2,\ldots,\varepsilon_n\right)A$ ，其中：

   $$A=\left(\begin{array}{cccc}
   a_{11}&a_{12}&\ldots&a_{1n}\\
   a_{21}&a_{22}&\ldots&a_{2n}\\
   \vdots&\vdots&\ddots&\vdots\\
   a_{n1}&a_{n2}&\ldots&a_{nn}
   \end{array}\right)$$

   定理：设线性变换 $\mathcal{A}$ 在基 $\varepsilon_1,\varepsilon_2,\ldots,\varepsilon_n$ 下的矩阵是 A ，向量 ξ 在基 $\varepsilon_1,\varepsilon_2,\ldots,\varepsilon_n$ 下的坐标是 $(x_1,x_2,\ldots,x_n)$ ，则 $\mathcal{A}(\xi)$ 在基 $\varepsilon_1,\varepsilon_2,\ldots,\varepsilon_n$ 下的坐标$(y_1,y_2,\ldots,y_n)$，可以按公式：$\left(\begin{array}{c}y_1\\y_2\\\vdots\\y_n\end{array}\right)=A\left(\begin{array}{c}x_1\\x_2\\\vdots\\x_n\end{array}\right)$计算。

2. 相似
   定理：设线性空间 V 中线性变换 $\mathcal{A}$ 在两组基 $\varepsilon_1,\varepsilon_2,\ldots,\varepsilon_n$ 及 $\eta_1,\eta_2,\ldots,\eta_n$下的矩阵分别问$A$和$B$，从基 $\varepsilon_1,\varepsilon_2,\ldots,\varepsilon_n$ 到基 $\eta_1,\eta_2,\ldots,\eta_n$ 的矩阵是 X，有 $B=X^{-1}AX$ ，

$$\mathcal{A}\left(\varepsilon_1,\varepsilon_2,\ldots,\varepsilon_n\right)=\left(\varepsilon_1,\varepsilon_2,\ldots,\varepsilon_n\right)A$$

证明：已知，

$\mathcal{A}\left(\eta_1,\eta_2,\ldots,\eta_n\right)=\left(\eta_1,\eta_2,\ldots,\eta_n\right)B$，$\left(\eta_1,\eta_2,\ldots,\eta_n\right)=\left(\varepsilon_1,\varepsilon_2,\ldots,\varepsilon_n\right)X$

所以

$$\begin{aligned}
&\left(\mathcal{A}\eta_1,\mathcal{A}\eta_2,\ldots,\mathcal{A}\eta_n\right)=\mathcal{A}\left(\eta_1,\eta_2,\ldots,\eta_n\right)\\
&=\mathcal{A}\left[\left(\varepsilon_1,\varepsilon_2,\ldots,\varepsilon_n\right)X\right]=\left[\mathcal{A}\left(\varepsilon_1,\varepsilon_2,\ldots,\varepsilon_n\right)\right]X\\
&=\left(\mathcal{A}\varepsilon_1,\mathcal{A}\varepsilon_2,\ldots,\mathcal{A}\varepsilon_n\right)X=\left(\varepsilon_1,\varepsilon_2,\ldots,\varepsilon_n\right)AX\\
&=\left(\eta_1,\eta_2,\ldots,\eta_n\right)X^{-1}AX
\end{aligned}$$

得到 $B=X^{-1}AX$

定义：设A和B为数域P上两个n级矩阵，如果可以找到数据P上的n级可逆矩阵X，使得 $B=X^{-1}AX$ ，就说A相似B，记作： $A\sim B$，

相似的性质：

1. 反身性： $A\sim A$,因为 $A=E^{-1}AE$ 。
2. 对称性：如果 $A\sim B$，那么 $B\sim A$ 。
3. 传递性：如果$A\sim B$，$B\sim C$，那么 $A\sim C$ 。

定理：线性变换在不同基下所对应的矩阵是相似的，反过说，如果两个矩阵相似，那么他们可以看作同一个线性变换在两组基下所对应的矩阵；

矩阵相似的计算：
如果

$$\begin{aligned}
&B_1=X^{-1}A_1X,B_2=X^{-1}A_2X\\
&B_1+B_2=X^{-1}\left(A_1+A_2\right)X\\
&B_1B_2=X^{-1}\left(A_1A_2\right)X
\end{aligned}$$

如果 $B=X^{-1}AX$ ， $f(x)$ 是数域 P 上一多项式，那么 $f(B)=X^{-1}f(A)X$ 。

3.6.2 特征值与特征向量

1. 定义
   设 $\mathcal{A}$ 是数域 P上线性空间 V 的一个线性变换，如果对于数域P找那个一数 λ_0，存在一个非零向量 ξ，使得：$A\xi=\lambda_0\xi$,那么 λ_0 称为$\mathcal{A}$ 的一个特征值，而 ξ 称为称为 $\mathcal{A}$ 的一个特征向量。

   从几何上看，特征向量的方向经过线性变换后，保持在同一条直线上，这时方向不变（$\lambda_0>0$ ），或者方向相反（ $\lambda_0<0$ ），至于 $\lambda_0=0$ 时，特征向量就被线性变换为0。

   特征向量和特征值是一一对应的。
   如果 ξ 是线性变换 $\mathcal{A}$ 的属于特征值 λ_0 的特征向量，那么 ξ 的任意一个非零倍数 $k\xi$ 也是线性变换 $\mathcal{A}$ 的属于特征值 λ_0 的特征向量，因为$A(k\xi)=\lambda_0(k\xi)$ 。所以特征向量不是被特征值唯一所确定的，但是特征值却是被特征向量唯一所确定的。

2. 计算特征值和特征向量的方法
   设V是数域P上的 n 维线性空间， $\varepsilon_1,\varepsilon_2,\ldots,\varepsilon_n$ 是它的一组基，线性变换 $\mathcal{A}$ 在这组基下的矩阵为 A，设 λ_0 是特征值，它的一个特征向量 ξ 在 $\varepsilon_1,\varepsilon_2,\ldots,\varepsilon_n$ 下的坐标是 $x_{01},x_{02},\ldots,x_{0n}$ ，则有：

   $A\begin{pmatrix}x_{01}\\x_{02}\\\vdots\\x_{0n}\end{pmatrix}=\lambda_0\begin{pmatrix}x_{01}\\x_{02}\\\vdots\\x_{0n}\end{pmatrix}$ 或 $\left(\lambda_0E-A\right)\begin{pmatrix}x_{01}\\x_{02}\\\vdots\\x_{0n}\end{pmatrix}=0$,

   这说明特征向量 ξ 的坐标 $x_{01},x_{02},\ldots,x_{0n}$ 满足其次线性方程组

$$\begin{cases} a_{11}x_1 + a_{12}x_2 + \ldots + a_{1n}x_n = \lambda_0 x_1 \\ a_{21}x_1 + a_{22}x_2 + \ldots + a_{2n}x_n = \lambda_0 x_2 \\ \cdots\cdots\cdots\cdots \\ a_{n1}x_1 + a_{n2}x_2 + \ldots + a_{nn}x_n = \lambda_0 x_n \end{cases}$$ ，即：

$$\begin{cases} (\lambda_0 - a_{11}x_1) - a_{12}x_2 - \ldots - a_{1n}x_n = 0 \\ -a_{21}x_1 + (\lambda_0 - a_{22})x_2 - \ldots - a_{2n}x_n = 0 \\ \cdots\cdots\cdots \\ -a_{n1}x_1 - a_{n2}x_2 - \ldots + (\lambda_0 - a_{nn})x_n = 0 \end{cases}$$

因为特征向量 $\xi \neq 0$ ，所以它的坐标 $x_{01}, x_{02}, \ldots, x_{0n}$ 不全为零，即齐次线性方程组有非零解，所以它的系数行列式为零：

$$|\lambda_0 E - A| = \begin{pmatrix} \lambda_0 - a_{11} & -a_{12} & \ldots & -a_{1n} \\ -a_{21} & \lambda_0 - a_{22} & \ldots & -a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ -a_{n1} & -a_{n2} & \ldots & \lambda_0 - a_m \end{pmatrix}$$

我们把 $|\lambda_0 E - A|$ 称为 A 的特征多项式，那么特征值就是特征多项式的一个根，这时如果 $x_{01}, x_{02}, \ldots, x_{0n}$ 是方程组 $\begin{cases} (\lambda_0 - a_{11}x_1) - a_{12}x_2 - \ldots - a_{1n}x_n = 0 \\ -a_{21}x_1 + (\lambda_0 - a_{22})x_2 - \ldots - a_{2n}x_n = 0 \\ \cdots\cdots\cdots \\ -a_{n1}x_1 - a_{n2}x_2 - \ldots + (\lambda_0 - a_{nn})x_n = 0 \end{cases}$ 的一个非零解，那么非零向量 $\xi = x_{01}\varepsilon_1 + x_{02}\varepsilon_2 + \ldots + x_{0n}\varepsilon_n$ 就是特征值 λ_0 的特征向量。

例：设线性变换 $\mathcal{A}$ 在基 $\varepsilon_1, \varepsilon_2, \varepsilon_3$ 下的矩阵是

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

因为特征多项式为：

$$|\lambda E - A| = \begin{vmatrix} \lambda - 1 & -2 & -2 \\ -2 & \lambda - 1 & -2 \\ -2 & -2 & \lambda - 1 \end{vmatrix} = (\lambda + 1)^2(\lambda - 5)$$

所以特征值为-1（二重）和5

把特征值-1代入齐次方程组

$$\begin{cases} (\lambda - 1)x_1 - 2x_2 - 2x_3 = 0 \\ -2x_1 - (\lambda - 1)x_2 - 2x_3 = 0 \\ -2x_1 - 2x_2 - (\lambda - 1)x_3 = 0 \end{cases}$$

得到：

$\begin{cases} -2x_1 - 2x_2 - 2x_3 = 0 \\ -2x_1 - 2x_2 - 2x_3 = 0 \\ -2x_1 - 2x_2 - 2x_3 = 0 \end{cases}$ ，$x_3 = -x_1 - x_2$ 分别代入（1,0）和（0,1）得到基础解系：

$\begin{pmatrix} 1 \\ 0 \\ -1 \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \\ -1 \end{pmatrix}$ 因此得到属于-1的两个线性无关特征向量为 $\xi_1 = \varepsilon_1 - \varepsilon_3$ ，$\xi_2 = \varepsilon_2 - \varepsilon_3$

而属于-1的全部特征向量就是$k_1\xi_1 + k_2\xi_2$， k_1, k_2 取遍数域 P 中不全为零的全部数对。

同理代入特征值5，

$\begin{cases} 4x_1 - 2x_2 - 2x_3 = 0 \\ -2x_1 + 4x_2 - 2x_3 = 0, \\ -2x_1 - 2x_2 + 4x_3 = 0 \end{cases} \begin{cases} 2x_1 - x_2 = x_3 \\ -x_1 + 2x_2 = x_3 \end{cases}$ 得到基础解系 $\begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}$，

因此5的一个线性无关特征向量为 $\xi_1 = \varepsilon_1 + \varepsilon_2 + \varepsilon_3$ ，

而属于-1的全部特征向量就是 $k\xi$ ， k 取遍数域 P 中不全为零的全部数对。

3.6.3 对角矩阵

定理：设 $\mathcal{A}$ 是n维线性空间V中的一个线性变换， $\mathcal{A}$ 的矩阵在某一组基下为对角矩阵的充分必要条件是， $\mathcal{A}$ 有 n 个线性无关的特征向量。

证明：设 $\mathcal{A}$ 在基 $\varepsilon_1,\varepsilon_2,\ldots,\varepsilon_n$下具有具有对角矩阵 $\begin{pmatrix}\lambda_1 & & & \\ & \lambda_2 & & \\ & & \ddots & \\ & & & \lambda_n\end{pmatrix}$ ，即有 $\mathcal{A}\varepsilon_i=\lambda_i\varepsilon_i, i=1,2\ldots n$ ，所有 $\varepsilon_1,\varepsilon_2,\ldots,\varepsilon_n$ 就是 $\mathcal{A}$ 的n个线性无关的特征向量。

反过来，如果 $\mathcal{A}$ 有n个线性无关的特征向量 $\varepsilon_1,\varepsilon_2,\ldots,\varepsilon_n$ ，那么以 $\varepsilon_1,\varepsilon_2,\ldots,\varepsilon_n$ 为基，在这组基下 $\mathcal{A}$ 的矩阵就是对角矩阵。

定理：属于不同特征值的特征向量是线性无关的。

定理：如果 $\lambda_1,\lambda_2,\ldots,\lambda_k$ 是线性变换 $\mathcal{A}$ 的不同特征值，而 $a_{i1},\ldots a_{ir_i}$ 是属于特征值 λ_i 的线性无关特征向量，$i=1,2,\ldots,k$ ，那么向量组 $a_{11},\cdots a_{1r_1},\cdots,a_{k1},\cdots a_{1r_k}$ 也线性无关。

解读：不同特征值的特征向量固然不相同，有重根的情况下，n 重的基础解系得到的 n 个特征向量也是线性无关的也是没有问题。但要注意的是重根情况下，基础解系之上实际上是有无穷个特征向量， 因此为了严谨，就不能直接说特征向量都是无关的。

推论：如果在 n 维线性空间 V 中，线性变换 $\mathcal{A}$ 的特征多项式在数域P中有n个不同的根，即 $\mathcal{A}$ 有n个不同的特征值，那么 $\mathcal{A}$ 在某组基下的矩阵是对角形的。

在3.6.2的例子中，我们计算出特征值-1（二重）以及5的对应特征向量是：

$$\begin{aligned}\xi_1&=\varepsilon_1-\varepsilon_3\\ \xi_2&=\varepsilon_2-\varepsilon_3\\ \xi_3&=\varepsilon_1+\varepsilon_1+\varepsilon_3\end{aligned}$$

所以 $\mathcal{A}$ 在基 ξ_1,ξ_2,ξ_3 下的对角矩阵为：

$$B=\begin{pmatrix}-1 & 0 & 0\\ 0 & -1 & 0\\ 0 & 0 & 5\end{pmatrix}$$

而由 $\varepsilon_1,\varepsilon_2,\varepsilon_3$ 到 ξ_1,ξ_2,ξ_3 的对角矩阵为：

$$X=\begin{pmatrix}1 & 0 & 1\\ 0 & 1 & 1\\ -1 & -1 & 1\end{pmatrix}$$

于是有 $B=X^{-1}AX$

3.6.4 值域与核

定于：设 $\mathcal{A}$ 是线性空间 V 中的一个线性变换， $\mathcal{A}$ 的全体像组成的集合称为 $\mathcal{A}$ 的值域，用 $\mathcal{A}V$ 表示，所有被 $\mathcal{A}$ 编程零向量的向量组成的集合称为 $\mathcal{A}$ 的核，用 $\mathcal{A}^{-1}(0)$ 表示 。

特征值的性质：

设 n 阶矩阵 $A=(a_{ij})$ 的特征值为 $\lambda_1,\lambda_2,\ldots,\lambda_n$ ，则：

$\lambda_1+\lambda_2+\cdots+\lambda_n=a_{11}+a_{22}+\cdots a_{mn}$

$\lambda_1\lambda_2\cdots\lambda_n=|A|$

其中，矩阵 A 主行列式的元素和，称为矩阵 A 的迹。

思考：已知 λ 是方阵 A 的特征值，则：

（1） λ^2 是 A^2 的特征值；

（2） A 可逆时，λ^{-1} 是 A^{-1} 的特征值。

提示：可参考定义。

不同特征值对应的特征向量：

设 $\lambda_1, \lambda_2, \cdots \lambda_m$ 是方阵 A 的 m 个特征值， $p_1, p_2, \cdots p_m$ 是依次与之对应的特征向量，若 $\lambda_1, \lambda_2, \cdots \lambda_m$ 各不相等，则 $p_1, p_2, \cdots p_m$ 线性无关。

总结与思考：

（1）不同特征值对应的特征向量，线性无关。

（2）若方阵 A 是对称阵呢？结论是否加强？如：协方差矩阵、二次型矩阵、无向图邻接矩阵等：对称阵

引理：

1. **实数对称阵的特征值是实数**
   设负数 λ 为对称阵 A 的特征值，复向量 x 为对应的特征向量，即 $Ax = \lambda x (x \neq 0)$ ，用 $\overline{\lambda}$ 表示 λ 的共轭复数，$\overline{x}$ 表示 x 的共轭向量，而 A 是实矩阵，有 $\overline{A} = A$ 。

证明过程如下：

首先，$A\overline{x} = \overline{A}\overline{x} = \overline{Ax} = \overline{\lambda x} = \overline{\lambda}\overline{x}$

因为，$\overline{x}^T (Ax) = \overline{x}^T \lambda x = \lambda x^{-T} x$

$$\overline{x}^T (Ax) = \left(\overline{x}^T A^T\right) x = (A\overline{x})^T x = (\overline{\lambda}\overline{x})^T x = \overline{\lambda}\overline{x}^T x$$

从而，$\lambda \overline{x}^T x = \overline{\lambda}\overline{x}^T x \Rightarrow (\lambda - \overline{\lambda})\overline{x}^T x = 0$

而，$\overline{x}^T x = \sum_{i=1}^n \overline{x_i} x_i = \sum_{i=1}^n |x_i|^2 \neq 0$

所以 $\lambda - \overline{\lambda} = 0 \Rightarrow \lambda = \overline{\lambda}$

利用上述结论很快得到：

将实数 λ 带入方程组 $(A - \lambda \mathrm{I})x = 0$， 该方程组为实系数方程组，因此，实对称阵的特征向量可以取实向量。

2. **实对称阵不同特征值的特征向量正交：**
   令实对称矩阵为 A ，其两个不同的特征值 $\lambda_1 \lambda_2$ 对应的特征向量分别是 $\mu_1 \mu_2$ ； $\lambda_1 \lambda_2$ $\mu_1 \mu_2$ 都是实数或是实向量。

$$\begin{cases} A\mu_1 = \lambda_1 \mu_1 \\ A\mu_2 = \lambda_2 \mu_2 \Rightarrow \mu_1^T A \mu_2 = \mu_1^T \lambda_2 \mu_2 \end{cases}$$

$$\begin{aligned} &\Rightarrow \left(A^T \mu_1\right)^T \mu_2 = \lambda_2 \mu_1^T \mu_2 \\ &\Rightarrow (A\mu_1)^T \mu_2 = \lambda_2 \mu_1^T \mu_2 \\ &\Rightarrow (\lambda_1 \mu_1)^T \mu_2 = \lambda_2 \mu_1^T \mu_2 \\ &\Rightarrow \lambda_1 \mu_1^T \mu_2 = \lambda_2 \mu_1^T \mu_2 \\ &\frac{\lambda_1 \neq \lambda_2}{2} \mu_1^T \mu_2 = 0 \end{aligned}$$

3. **最终结论**
   设 A 为 n 阶对称阵，则必有正交阵 P ，使得 $P^{-1}AP = P^TAP = \Lambda$ ，其中， Λ 是以A 的 n 个

特征值为对角元的对角阵，该变换称为“合同变换”，A 和 Λ 互为合同矩阵。

注：在谱聚类、PCA等章节中将会继续讨论。

QR分解：

分解概念：对于 $m \times n$ 的列满秩矩阵 A ，必有：$A_{m \times n} = Q_{m \times n} \cdot R_{m \times n}$ 。其中，$Q^T Q = \mathrm{I}$ （即列正交矩阵），R 为非奇异上三角矩阵。当要求 R 的对角线元素为正时，该分解唯一，称该分解为 QR 分解。（可用于求解矩阵 A 的特征值、A 的逆等问题。）

QR 分解计算特征值：

计算 n 阶方阵 A 的特征值：

$$
\begin{aligned}
& A = Q \cdot R \Rightarrow A_1 = Q^T A Q = R \cdot Q \\
& \cdots \\
& A_k = Q_k \cdot R_k \Rightarrow A_{k+1} = R_k \cdot Q_k \\
& \cdots \\
& A_k \rightarrow \operatorname{diag}\{\lambda_1, \lambda_2, \cdots \lambda_n\}
\end{aligned}
$$

3.7 欧几里得空间

3.7.1 定义

定义：设 V 是实数域R上一线性空间，在V上定义了一个二元实函数，称为内积，记作 (α, β) ，它具有如下性质：

$(\alpha, \beta) = (\beta, \alpha)$

$(k\alpha, \beta) = k(\alpha, \beta)$

$(\alpha + \beta, \gamma) = (\alpha, \gamma) + (\beta, \gamma)$

$(\alpha, \alpha) \geq 0$ ，当且仅当 $\alpha = 0$ 时，$(\alpha, \alpha) = 0$

这里的 α，β, γ 是V中任意向量，k是任意实数，这样的线性空间称为欧几里得空间。

定义：非负实数 $\sqrt{(\alpha, \alpha)}$ 称为向量 α 的长度（模），记作 $|\alpha|$ 。

例：向量 $\alpha = (x, y)$ ，$|\alpha| = \sqrt{(x^2 + y^2)}$ ，

在解释几何中，向量 α, β 的夹角 $<\alpha, \beta>$ 的余弦可以通过内积来表示：

$$
cos < \alpha, \beta > = \frac{(\alpha, \beta)}{|\alpha||\beta|}
$$

定义：如果向量 α , β 的内积为零，即 $(\alpha, \beta) = 0$ ，那么 α , β 称为正交或互相垂直，记为 $\alpha \perp \beta$。

如果向量 $\alpha_1, \alpha_2, \ldots, \alpha_m$ 两两正交，那么有：

$$
|\alpha_1 + \alpha_2 + \ldots + \alpha_m|^2 = |\alpha_1|^2 + |\alpha_2|^2 + \ldots + |\alpha_m|^2
$$

接下来讨论有限维度的欧几里得空间，在 V 中取一组基 $\varepsilon_1, \varepsilon_2, \ldots, \varepsilon_n$ ，对V中任意两个向量：

$$
\begin{aligned}
\alpha = x_1 \varepsilon_1 + x_2 \varepsilon_2 + \ldots + x_n \varepsilon_n \\
\beta = y_1 \varepsilon_1 + y_2 \varepsilon_2 + \ldots + y_n \varepsilon_n
\end{aligned}
$$

由内积性质得知

$$
\begin{aligned}
& (\alpha, \beta) = (x_1 \varepsilon_1 + x_2 \varepsilon_2 + \ldots + x_n \varepsilon_n, y_1 \varepsilon_1 + y_2 \varepsilon_2 + \ldots + y_n \varepsilon_n) \\
& = \sum_{i=1}^n \sum_{j=1}^n (\varepsilon_i, \varepsilon_j) x_i y_j
\end{aligned}
$$

令 $a_{ij}=(\varepsilon_i,\varepsilon_j),(i,j=1,2,\ldots,n)$

显然 $a_{ij}=a_{ji}$，改写成矩阵形式 $(\alpha,\beta)=X'AY$.

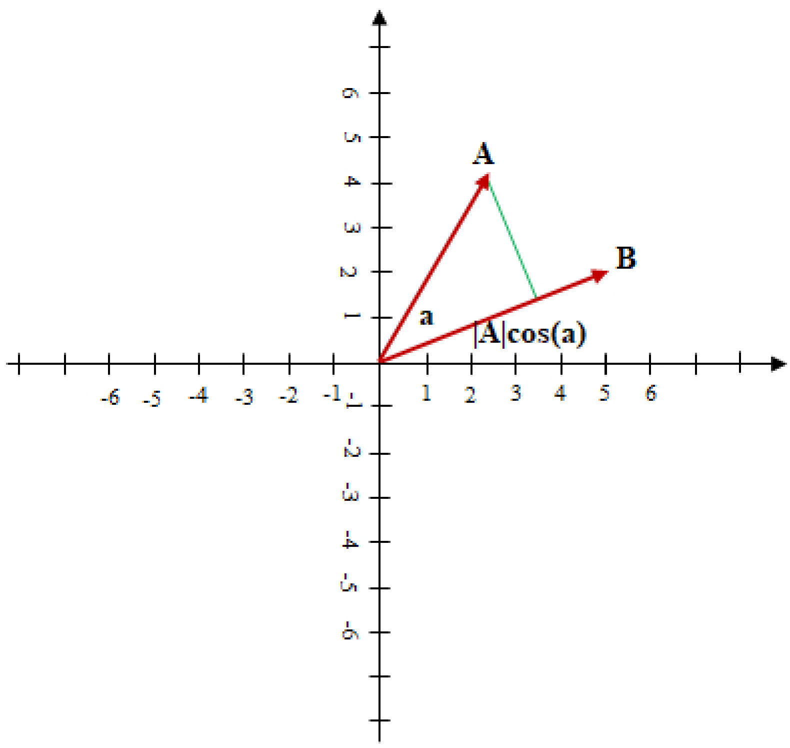
其中 $X=\begin{pmatrix}x_1\\x_2\\\vdots\\x_n\end{pmatrix},Y=\begin{pmatrix}y_1\\y_2\\\vdots\\y_n\end{pmatrix}$ 分别是 α , β 坐标

而矩阵 $A=(a_{ij})_{nn}$ 称为基 $\varepsilon_1,\varepsilon_2,\ldots,\varepsilon_n$ 的度量矩阵。

实际上 $A=\begin{pmatrix}\varepsilon_1\\\varepsilon_2\\\vdots\\\varepsilon_n\end{pmatrix}(\varepsilon_1,\varepsilon_2,\ldots,\varepsilon_n)$

上面标明，在知道一组基的度量矩阵后，任意两个向量的内积就可以通过坐标按上式计算。（可以看出度量矩阵是和两个向量无关）

例：设A和B均为二维向量，$A=（x_1,y_1）$,$B=(x_2,y_2)$, 在二维平面上A和B可以用两条自原点出发的有向线段表示，如下图：



现在我们从A点向 B 所在直线引一条垂线。我们知道垂线与B的交点叫做A在B上的投影，再设A与 B 的夹角是a，则投影的矢量长度为 $|A|cos(a)$，其中 $|A|=\sqrt{(x^2+y^2)}$ 是向量A的模，也就是A线段的标量长度。

A与 B 的内积等于A到 B 的投影长度乘以B的模。再进一步，如果我们假设 B 的模为1，即让$|B|=1$，那么就变成了：$A\cdot B=|A|cos(a)$。

设 $\eta_1,\eta_2,\ldots\eta_n$ 是空间 V 中的另一组基，而由 $\varepsilon_1,\varepsilon_2,\ldots,\varepsilon_n$ 到 $\eta_1,\eta_2,\ldots\eta_n$ 的过度矩阵为 C，即

$$
\begin{aligned}
&(\eta_1,\eta_2,\ldots\eta_n)=(\varepsilon_1,\varepsilon_2,\ldots,\varepsilon_n)\,C\\
&B=(b_{ij})=\begin{pmatrix}\eta_1\\ \eta_2\\ \vdots\\ \eta_n\end{pmatrix}(\eta_1,\eta_2,\ldots\eta_n)=((\varepsilon_1,\varepsilon_2,\ldots,\varepsilon_n)\,C)'\,((\varepsilon_1,\varepsilon_2,\ldots,\varepsilon_n)\,C)\\
&=C'AC
\end{aligned}
$$

这说明不同基的度量矩阵是合同的

同时，对于非零向量 α 因为 $X\neq\begin{pmatrix}0\\0\\ \vdots\\0\end{pmatrix}$ ，因此 $(\alpha,\alpha)=X'AX>0$ ，因此度量矩阵是正定的 。

3.7.2 标准正交基

定义：欧式空间 V 中有一组非零的向量，如果它们两两正交，就称为以正交向量组。

定义：在 n 维欧式空间中，由 n 个向量组成的正交向量组称为正交基；由单位向量组成的正交基称为标准正交基。

设 $\varepsilon_1,\varepsilon_2,\ldots,\varepsilon_n$ 是一组标准正交基，有：

$$(\varepsilon_i,\varepsilon_j)=\begin{cases}1,i=j\\0,i\neq j\end{cases}$$

因此一组基为标准正交基的充要条件是它的度量矩阵为单位矩阵。

在标准正交基下，向量的坐标可以通过内积简单表示：$\alpha=x_1\varepsilon_1+x_2\varepsilon_2+\ldots+x_n\varepsilon_n$ 用 ε_i 在等式两边作内积有：$(\alpha,\varepsilon_i)=x_i$ ，所以坐标公式有 $\alpha=(\varepsilon_1,\alpha)\varepsilon_1+(\varepsilon_2,\alpha)\varepsilon_2+\ldots+(\varepsilon_n,\alpha)\varepsilon_n$

因此在标准正交基下，内积也有特别简单的表示公式：

$$\begin{matrix}\alpha=x_1\varepsilon_1+x_2\varepsilon_2+\ldots+x_n\varepsilon_n\\ \beta=y_1\varepsilon_1+y_2\varepsilon_2+\ldots+y_n\varepsilon_n'\end{matrix},(\alpha,\beta)=x_1y_1+x_2y_2+\ldots+x_ny_n=X'Y$$

定理： n 维欧式空间中任一个正交向量组都能扩充成一组正交基

把一组线性无关的向量变成一单位正交向量组的方法称为施密特正交化。

例：把 $\alpha_1=(1,1,0,0),\alpha_2=(-1,0,0,1),\alpha_3=(1,0,1,0),\alpha_4=(1,-1,-1,1)$

先把向量组正交化，得到：

$$
\begin{aligned}
&\beta_1=\alpha_1=(1,1,0,0)\\
&\beta_2=\alpha_2-\tfrac{(\alpha_2,\beta_1)}{(\beta_1,\beta_1)}\beta_1=\left(\tfrac{1}{2},-\tfrac{1}{2},1,0\right)\\
&\beta_3=\alpha_3-\tfrac{(\alpha_3,\beta_1)}{(\beta_1,\beta_1)}\beta_1-\tfrac{(\alpha_3,\beta_2)}{(\beta_2,\beta_2)}\beta_2=\left(-\tfrac{1}{3},\tfrac{1}{3},\tfrac{1}{3},0\right)\\
&\beta_4=\alpha_4-\tfrac{(\alpha_4,\beta_1)}{(\beta_1,\beta_1)}\beta_1-\tfrac{(\alpha_4,\beta_2)}{(\beta_2,\beta_2)}\beta_2-\tfrac{(\alpha_4,\beta_3)}{(\beta_3,\beta_3)}\beta_3=(1,-1,-1,1)
\end{aligned}
$$

再将其单位化得：

$$
\begin{aligned}
&\eta_1=\left(\tfrac{1}{\sqrt{2}},\tfrac{1}{\sqrt{2}},0,0\right)\\
&\eta_2=\left(\tfrac{1}{\sqrt{6}},-\tfrac{1}{\sqrt{6}},\tfrac{2}{\sqrt{6}},0\right)\\
&\eta_3=\left(-\tfrac{1}{\sqrt{12}},\tfrac{1}{\sqrt{12}},\tfrac{1}{\sqrt{12}},\tfrac{3}{\sqrt{12}}\right)\\
&\eta_4=\left(\tfrac{1}{2},-\tfrac{1}{2},-\tfrac{1}{2},\tfrac{1}{2}\right)
\end{aligned}
$$

设 $\varepsilon_1,\varepsilon_2,\ldots,\varepsilon_n$ 与 $\eta_1,\eta_2,\ldots\eta_n$ 是欧式空间V中的两组标准正交基，他们的过渡矩阵 $A=(a_{ij})$ ，有

$$(\eta_1,\eta_2,\ldots,\eta_n)=(\varepsilon_1,\varepsilon_2,\ldots,\varepsilon_n)\begin{pmatrix} a_{11} & a_{12} & \ldots & a_{1n} \\ a_{21} & a_{22} & \ldots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & \ldots & a_{nn} \end{pmatrix}。$$

因为 $\eta_1,\eta_2,\ldots\eta_n$ 是标准正交基，所以 $(\eta_i,\eta_j)=\begin{cases}1, i=j\\ 0, i\neq j\end{cases}$ ，事实上矩阵 A 的各列就是 $\eta_1,\eta_2,\ldots\eta_n$ 在标准正交基 $\varepsilon_1,\varepsilon_2,\ldots,\varepsilon_n$ 下的坐标，因此 $(\eta_i,\eta_j)=\begin{cases}1, i=j\\ 0, i\neq j\end{cases}$ 根据内积的表达公式 $(\alpha,\beta)=x_1y_1+x_2y_2+\ldots+x_ny_n=X'Y$ ，可以表示为：
$a_{1i}a_{1j}+a_{2i}a_{2j}+\ldots+a_{ni}a_{nj}=\begin{cases}1, i=j\\ 0, i\neq j\end{cases}$ ，根据此式有 $A'A=E, A'=A^{-1}$ 。

定义：n 级实数矩阵 A 称为正交矩阵，如果 $A'A=E$

说明由标准正交基到标准正交基的过度矩阵是正交矩阵；反过来，如果第一组基是标准正交基，同时过渡矩阵是正交矩阵，那么第二组基也一定是标准正交基。

3.7.3 正交变换

欧式空间 V 的线性变换 $\mathcal{A}$ 称为正交变换，如果它保持向量的内积不变，即对于任意的 $\alpha,\beta\in V$，都有 $(\mathcal{A}\alpha,\mathcal{A}\beta)=(\alpha,\beta)$ 。

设 $\mathcal{A}$ 是n维欧式空间V中的一个线性变换，也是下面四个命题相互等价：

1. $\mathcal{A}$ 是正交变换
2. $\mathcal{A}$ 保持向量的长度不变，即对于 $\alpha\in V, |\mathcal{A}\alpha|=|\alpha|$
3. 如果 $\varepsilon_1,\varepsilon_2,\ldots,\varepsilon_n$ 是标准正交基，那么 $\mathcal{A}\varepsilon_1,\mathcal{A}\varepsilon_2,\ldots,\mathcal{A}\varepsilon_n$ 也是标准正交基
4. $\mathcal{A}$ 在任一组标准正交基下的矩阵是正交矩阵

3.7.4 实对称矩阵的标准型

设 A 是实对称矩阵，则 A 的特征值皆为实数；

设 A 是实对称矩阵，$\mathcal{A}$ 的定以如下：

对应与实对称矩阵 A ，在 n 维欧式空间 R^n 上定义一个线性变换如下：$\boldsymbol{\mathcal{A}}\begin{pmatrix}x_1\\x_2\\\vdots\\x_n\end{pmatrix}=A\begin{pmatrix}x_1\\x_2\\\vdots\\x_n\end{pmatrix}$ ，显然 $\mathcal{A}$ 在标准正交基下的矩阵就是 A

那么对任意 $\alpha,\beta\in R^n$ 有 $(\mathcal{A}\alpha,\beta)=(\alpha,\mathcal{A}\beta)$ 或 $\beta'(\mathcal{A}\alpha)=\alpha(\mathcal{A}\beta')$

定理：对于任意一个 n 级实对称矩阵 A，都存在一个 n 级正交矩阵 T，使 $T'AT=T^{-1}AT$ 成对角形

例：已知 $A=\begin{pmatrix}0&1&1&-1\\1&0&-1&1\\1&-1&0&1\\-1&1&1&0\end{pmatrix}$ 求一正交矩阵 T，使 $T'AT$ 成对角形

（1） 先求 A 的特征值

$$|\lambda E-A|=\begin{vmatrix}\lambda&-1&-1&1\\-1&\lambda&1&-1\\-1&1&\lambda&-1\\1&-1&-1&\lambda\end{vmatrix}=(\lambda-1)^3(\lambda+3)$$,所以 A 的特征值为1（三重），-3

（2） 求特征向量

把 $\lambda=1$ 代入，求得基础解系，分别对基础解系正交化及单位化

把$\lambda=-3$ 代入，把特征向量单位化

最后把每个正交化及单位化后特征向量按列组成正交矩阵：

$$T=\begin{pmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{6}} & -\frac{1}{\sqrt{12}} & \frac{1}{2} \\ \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{6}} & \frac{1}{\sqrt{12}} & -\frac{1}{2} \\ 0 & \frac{2}{\sqrt{6}} & \frac{1}{\sqrt{12}} & -\frac{1}{2} \\ 0 & 0 & \frac{3}{\sqrt{12}} & \frac{1}{2} \end{pmatrix} \text{ 有 } T'AT=\begin{pmatrix} 1 & & & \\ & 1 & & \\ & & 1 & \\ & & & -3 \end{pmatrix}$$

3.7.5 奇异值分解

奇异值分解($SingularValueDecomposition$，以下简称$SVD$)是在机器学习领域广泛应用的算法，它不光可以用于降维算法中的特征分解，还可以用于推荐系统，以及自然语言处理等领域。是很多机器学习算法的基石。

实际上，在3.6.2和3.6.3章节中，我们介绍了利用特征值分解将矩阵化为对角阵。求出特征值和特征向量有什么好处呢？ 就是我们可以将矩阵A特征分解。

如果我们求出了矩阵 A 的 n 个特征值 $\lambda_1\leq\lambda_2\leq\cdots\leq\lambda_n$ ，以及这 n 个特征值所对应的特征向量 $\omega_1,\omega_2,...,\omega_n$,那么矩阵A就可以用下式的特征分解表示：

$$A=WBW^{-1}$$

其中，W是这n个特征向量所张成的$n\times n$维矩阵，而B为这n个特征值为主对角线的$n\times n$维矩阵。一般我们会把W的这n个特征向量标准化，即满足 $||w_i||_2=1$ ，或者 $w_i^Tw_i=1$ ，此时W的n个特征向量为标准正交基，满足 $W^TW=I$ ，即 $W^T=W^{-1}$ ，也就是说W为酉矩阵。这样我们的特征分解表达式可以写成：

$$A=WBW^T$$

注意到要进行特征分解，矩阵A必须为方阵。

同样的，SVD也是对矩阵进行分解，但是和特征分解不同，SVD并不要求要分解的矩阵为方阵。

假设矩阵A是一个$m\times n$的矩阵，那么我们定义矩阵A的SVD 分解为：

$$A=UBV^T$$

其中U是一个$m\times m$的矩阵，B是一个$m\times n$的矩阵，除了主对角线上的元素以外全为0，主对角线上的每个元素都称为奇异值，V 是一个$n\times n$的矩阵。U 和 V 都是酉矩阵，即满足 $U^TU=I$，$V^TV=I$ 。

那么如何求出SVD分解后的 U,B,V 这三个矩阵呢？

首先，将 A 的转置和 A 做矩阵乘法，那么会得到一个 $n\times n$ 的方阵 A^TA 。然后对方阵 A^TA 进行特征分解，特征值和特征向量满足下述公式：

$$(A^TA)v_i=\lambda_iv_i$$

这样就得到矩阵 A^TA 的 n 个特征值和对应的 n 个特征向量 v。最后将 A^TA 的所有特征向量张成一个$n\times n$的矩阵V，就是公式里面的V矩阵了， V 中的每个特征向量称为 A 的右奇异向量。

同样的，我们也可以得到 U 矩阵。首先，将A和A的转置做矩阵乘法，那么会得到$m\times m$的一个方阵 AA^T 。然后对方阵 AA^T 进行特征分解，特征值和特征向量满足下式：

$$(AA^T)u_i = \sigma_i u_i$$

这样就得到矩阵 AA^T 的m个特征值和对应的m个特征向量u。最后将 AA^T 的所有特征向量张成一个$m\times m$的矩阵U，就是SVD公式里面的U矩阵了。一般我们将U中的每个特征向量叫做A的左奇异向量。

对于奇异值矩阵B 的求法，过程如下：

由于B除了对角线上是奇异值其他位置都是0，只需要求出每个奇异值 σ 就可以了，如下所示：

$$A = UBV^T \Rightarrow AV = UBV^TV \Rightarrow AV = UB \Rightarrow Av_i = \sigma_i u_i \Rightarrow \sigma_i = Av_i/u_i$$

说明：关于 A^TA 的特征向量组成的就是SVD中的V矩阵，而 AA^T 的特征向量组成的就是SVD中的U矩阵，根据证明如下，以 V 矩阵为例。

$$A = UBV^T \Rightarrow A^T = VBU^T \Rightarrow A^TA = VBU^TUBV^T = VB^2V^T$$

其中，$U^U = I, B^T = B$ 。可以看出 A^TA 的特征向量组成的的确就是 SVD 中的 V 矩阵。同样的也可以得到 U 矩阵的相关证明。

进一步还可以得到特征值矩阵等于奇异值矩阵的平方，也就是说特征值和奇异值满足如下关系：

$$\sigma_i = \sqrt{\lambda_i}$$

也就是说，我们也可以通过求出 A^TA 的特征值取平方根来求奇异值。

白化/漂泊（Whitening）：

$$x = \begin{bmatrix} x_{11} & x_{12} & \cdots & x_{1m} \\ x_{21} & x_{22} & \cdots & x_{2m} \\ \vdots & \vdots & \ddots & \vdots \\ x_{n1} & x_{n2} & \cdots & x_{nm} \end{bmatrix} \xlongequal{\Delta} \begin{pmatrix} x_1 & x_2 & \cdots & x_m \end{pmatrix}$$

计算观测数据 x 的 $n\times n$ 的对称阵 $x\cdot x^T$ 的特征值和特征向量，用特征值形成对角阵 D，特征向量形成正交阵 U，则：$x\cdot x^T = U^TDU$ 。

令：$\tilde{x} = U^TD^{-0.5}U\cdot x$

则：

$$\begin{aligned}\tilde{x}\cdot\tilde{x}^T &= \left(U^TD^{-0.5}U\cdot x\right)\left(U^TD^{-0.5}U\cdot x\right)^T \\ &= \left(U^TD^{-0.5}U\cdot x\right)\left(x^TU^TD^{-0.5}U\right) \\ &= U^TD^{-0.5}U\cdot\left(xx^T\right)\cdot U^TD^{-0.5}U \\ &= U^TD^{-0.5}U\cdot U^TDU\cdot U^TD^{-0.5}U \\ &= \mathrm{I}\end{aligned}$$

3.8 向量的导数（矩阵求导）

3.8.1 定义和约定

前言：

矩阵求导的本质是多元函数求导，可以理解为把求导的结果排列为矩阵形式，方便表达与计算。但是矩阵求导本身有一个混乱的地方，就是行、列向量的差异，所导致的结果的差异。而这个差异将导致最后结果出现转置问题。事实上，对于最后求导的结果，很多教材的处理是不一样的，本质上，这只是一个$Layout\ Convention$ 问题。本章采用非转置结果，也即是得到结果是雅克比矩阵，同时后面将解释非

转置结果。$Wiki$是目前找到关于矩阵求导最详细的结果，因此这里的结果将保持和$wiki$的结果一致，$wiki$相关内容可进一步阅读：https://en.wikipedia.org/wiki/Matrix_calculus#Layout_conventions。

约定：

- 标量用普通小写字母表示，如 x , y ,
- 向量用带箭头小写字母表示，如 $\vec{x}, \vec{y}$ ，且其中元素为：$\vec{x} = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix}$
- 矩阵用大写字母表示，如： $A = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{bmatrix}$
- 所有向量均默认为列向量，即 $\vec{x}$ 为列向量，而 $\vec{x}^T$ 为行向量
- 分子布局（$Numerator-layout$）：分子为 $\vec{y}$（列向量）或分母为 $\vec{x}^T$（行向量），或者说分子保持原始形式，分母为转置形式。
- 分母布局（$Denominator-layout$）：分子为 $\vec{y}^T$（行向量）或分母为 $\vec{x}$（列向量），或者说分子为转置形式，分母保持为原始形式。

需要注意的是，以下公式的原始计算式子，$wiki$上是不做转置的，但是为了统一，下文是使用了转置。例如分子布局的标量对向量，$wiki$ 上是 $\frac{\partial y}{\partial \vec{x}} = \left(\frac{\partial y}{\partial x_1}, \frac{\partial y}{\partial x_2}, \cdots, \frac{\partial y}{\partial x_n}\right)$ ，而本文是 $\frac{\partial y}{\partial (\vec{x})^T} = \left(\frac{\partial y}{\partial x_1}, \frac{\partial y}{\partial x_2}, \cdots, \frac{\partial y}{\partial x_n}\right)$ ，但无论是 $wiki$ 还是本文，原始式子的分母部分都是行向量。

3.8.2 分子布局

（1）标量对向量：分子 y 为标量，分母 $(\vec{x})^T$ 为行向量 $\frac{\partial y}{\partial (\vec{x})^T} = \left(\frac{\partial y}{\partial x_1}, \frac{\partial y}{\partial x_2}, \cdots, \frac{\partial y}{\partial x_n}\right)$

（2）向量对标量：分子 $\vec{y}$ 为列向量，分母 x 为标量 $\frac{\partial \vec{y}}{\partial x} = \begin{pmatrix} \frac{\partial y_1}{\partial x} \\ \frac{\partial y_2}{\partial x} \\ \cdots \\ \frac{\partial y_m}{\partial x} \end{pmatrix}$

（3）向量对向量：分子 $\vec{y}$ 为列向量，分母 $(\vec{x})^T$ 为行向量

例如A是 $m \times n$ 矩阵，$\vec{x}$ 是 $n \times 1$ 列向量，则 $A\vec{x}$ 是 $m \times 1$ 列向量，记 $\vec{y} = A \cdot \vec{x}$ ，那么 $\frac{\partial \vec{y}}{\partial (\vec{x})^T}$ ？

$$A = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{bmatrix}, \vec{x} = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix}, \quad A \cdot \vec{x} = \begin{pmatrix} a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n \\ a_{21}x_1 + a_{22}x_2 + \cdots + a_{2n}x_n \\ \vdots \\ a_{m1}x_1 + a_{m2}x_2 + \cdots + a_{mn}x_n \end{pmatrix}$$

$$\frac{\partial y}{\partial (\vec{x})^T} = \begin{pmatrix} \frac{\partial y_1}{\partial (\vec{x})^T} \\ \frac{\partial y_2}{\partial (\vec{x})^T} \\ \vdots \\ \frac{\partial y_m}{\partial (\vec{x})^T} \end{pmatrix} = \begin{pmatrix} \frac{\partial y_1}{\partial x_1} & \frac{\partial y_1}{\partial x_2} & \cdots & \frac{\partial y_1}{\partial x_n} \\ \frac{\partial y_2}{\partial x_1} & \frac{\partial y_2}{\partial x_2} & \cdots & \frac{\partial y_2}{\partial x_n} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial y_m}{\partial x_1} & \frac{\partial y_m}{\partial x_2} & \cdots & \frac{\partial y_m}{\partial x_n} \end{pmatrix} = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{bmatrix} = A$$

其中 A 又被称为雅克比矩阵。

（4）标量对矩阵。这里的书写与一些网上的博客不一样，但是计算结果是一样。网上一般是 $\frac{\partial y}{\partial X}$ ，但是为了避免混淆，这里的分母部分不再写为 X ，而是 X^T 。 X 是一个 $p \times q$ 矩阵。

$$\frac{\partial y}{\partial (X)^T}=\begin{pmatrix}\frac{\partial y}{\partial x_{11}} & \frac{\partial y}{\partial x_{21}} & \cdots & \frac{\partial y}{\partial x_{p1}}\\ \frac{\partial y}{\partial x_{12}} & \frac{\partial y}{\partial x_{22}} & \cdots & \frac{\partial y}{\partial x_{p2}}\\ \vdots & \vdots & & \vdots\\ \frac{\partial y}{\partial x_{1q}} & \frac{\partial y}{\partial x_{2q}} & \cdots & \frac{\partial y}{\partial x_{pq}}\end{pmatrix}$$

这里看到是转置的。

（5）矩阵对标量

$$\frac{\partial Y}{\partial x}=\begin{pmatrix}\frac{\partial y_{11}}{\partial x} & \frac{\partial y_{12}}{\partial x} & \cdots & \frac{\partial y_{1n}}{\partial x}\\ \frac{\partial y_{21}}{\partial x} & \frac{\partial y_{22}}{\partial x} & \cdots & \frac{\partial y_{2n}}{\partial x}\\ \vdots & \vdots & \ddots & \vdots\\ \frac{\partial y_{m1}}{\partial x} & \frac{\partial y_{m2}}{\partial x} & \cdots & \frac{\partial y_{mn}}{\partial x}\end{pmatrix}$$

3.8.3 分母布局

（1） 标量对向量：分子 y 为标量，分母 $\vec{x}$ 为列向量 $\frac{\partial y}{\partial \vec{x}}=\begin{pmatrix}\frac{\partial y}{\partial x_1}\\ \frac{\partial y}{\partial x_2}\\ \vdots\\ \frac{\partial y}{\partial x_n}\end{pmatrix}$

（2） 向量对标量：分子 $(\vec{y})^T$ 为行向量，分母 x 为标量 $\frac{\partial (\vec{y})^T}{\partial x}=\left(\frac{\partial y_1}{\partial x},\frac{\partial y_2}{\partial x},\cdots,\frac{\partial y_m}{\partial x}\right)$

（3） 向量对向量：分子 $(\vec{y})^T$ 为行向量，分母 $\vec{x}$ 为列向量

例如 A 是 $m\times n$ 矩阵，$\vec{x}$ 是 $n\times 1$ 列向量，则 $A\vec{x}$ 是 $m\times 1$ 列向量，记 $\vec{y}=A\cdot\vec{x}$，那么 $\frac{\partial (\vec{y})^T}{\partial \vec{x}}$？

$$A=\begin{bmatrix}a_{11} & a_{12} & \cdots & a_{1n}\\ a_{21} & a_{22} & \cdots & a_{2n}\\ \vdots & \vdots & \ddots & \vdots\\ a_{m1} & a_{m2} & \cdots & a_{mn}\end{bmatrix},\vec{x}=\begin{pmatrix}x_1\\ x_2\\ \vdots\\ x_n\end{pmatrix},\quad A\cdot\vec{x}=\begin{pmatrix}a_{11}x_1+a_{12}x_2+\cdots+a_{1n}x_n\\ a_{21}x_1+a_{22}x_2+\cdots+a_{2n}x_n\\ \vdots\\ a_{m1}x_1+a_{m2}x_2+\cdots+a_{mn}x_n\end{pmatrix}$$

$$\frac{\partial (\vec{y})^T}{\partial \vec{x}}=\begin{pmatrix}\frac{\partial y_1}{\partial \vec{x}} & \frac{\partial y_2}{\partial \vec{x}} & \cdots & \frac{\partial y_m}{\partial \vec{x}}\end{pmatrix}=\begin{pmatrix}\frac{\partial y_1}{\partial x_1} & \frac{\partial y_2}{\partial x_1} & \cdots & \frac{\partial y_m}{\partial x_1}\\ \frac{\partial y_1}{\partial x_2} & \frac{\partial y_2}{\partial x_2} & \cdots & \frac{\partial y_m}{\partial x_2}\\ \vdots & \vdots & \ddots & \vdots\\ \frac{\partial y_1}{\partial x_n} & \frac{\partial y_2}{\partial x_n} & \cdots & \frac{\partial y_m}{\partial x_n}\end{pmatrix}=\begin{bmatrix}a_{11} & a_{21} & \cdots & a_{m1}\\ a_{12} & a_{22} & \cdots & a_{m2}\\ \vdots & \vdots & \ddots & \vdots\\ a_{1n} & a_{2n} & \cdots & a_{mn}\end{bmatrix}=A^T$$

（4） 标量对矩阵。

$$\frac{\partial y}{\partial X}=\begin{pmatrix}\frac{\partial y}{\partial x_{11}} & \frac{\partial y}{\partial x_{12}} & \cdots & \frac{\partial y}{\partial x_{1q}}\\ \frac{\partial y}{\partial x_{21}} & \frac{\partial y}{\partial x_{22}} & \cdots & \frac{\partial y}{\partial x_{2q}}\\ \vdots & \vdots & \ddots & \vdots\\ \frac{\partial y}{\partial x_{p1}} & \frac{\partial y}{\partial x_{p2}} & \cdots & \frac{\partial y}{\partial x_{pq}}\end{pmatrix}$$

这里看到是正的。

3.8.4 其他写法

一个小众的写法（邹博写法）：向量对向量求偏导，可以看到分子分母都是列向量，这种情况其实不好推导。

A 是 $m \times n$ 矩阵，$\vec{x}$ 是 $n \times 1$ 列向量，则 $A\vec{x}$ 是 $m \times 1$ 列向量，记 $\vec{y} = A \cdot \vec{x}$ ，那么 $\frac{\partial \vec{y}}{\partial \vec{x}}$ 为?

$$A = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{bmatrix}, \vec{x} = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix}, \quad A \cdot \vec{x} = \begin{pmatrix} a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n \\ a_{21}x_1 + a_{22}x_2 + \cdots + a_{2n}x_n \\ \vdots \\ a_{m1}x_1 + a_{m2}x_2 + \cdots + a_{mn}x_n \end{pmatrix}$$

$$\frac{\partial \vec{y}}{\partial \vec{x}} = \frac{\partial A\vec{x}}{\partial \vec{x}} = \begin{bmatrix} a_{11} & a_{21} & \cdots & a_{m1} \\ a_{12} & a_{22} & \cdots & a_{m2} \\ \vdots & \vdots & \ddots & \vdots \\ a_{1n} & a_{2n} & \cdots & a_{mn} \end{bmatrix} = A^T$$

$A \cdot \vec{x}$ 第一个元素 $a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n$ 对第一个元素 x_1 求导，得到 a_{11} 。因为 $\vec{x}$ 是列向量，因此 $A \cdot \vec{x}$ 第一个元素 $a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n$ 对第二个元素 x_2 ，得到 a_{12} ，按照 $\vec{x}$ 形状，写在第二行第一列，以此类推，得到 A^T 。

推广得到:

$$\frac{\partial A\vec{x}}{\partial \vec{x}} = A^T$$
$$\frac{\partial A\vec{x}}{\partial \vec{x}^T} = A$$
$$\frac{\partial (\vec{x}A)}{\partial \vec{x}} = A$$

需要注意的是，所谓的布局，其实就是起始计算公式的分子和分母采用的行列设置。关于详细的各个矩阵求导的基础结果，可以进一步阅读$wiki$：https://en.wikipedia.org/wiki/Matrix_calculus#Scalar-by-vector_identities

3.8.5 常用的结果

一些常用的结果（原始向量不做说明，均为列向量）：

向量对向量：（向量对向量得矩阵，没严格验证）

$$\frac{\partial A\vec{x}}{\partial \vec{x}^T} = A$$
$$\frac{\partial A\vec{x}}{\partial \vec{x}} = A^T$$
$$\frac{\partial \left(\vec{x}^T A\right)}{\partial \vec{x}} = A$$
$$\frac{\partial \left(\vec{x}^T A\right)}{\partial \vec{x}^T} = A^T$$

标量对向量（标量对向量，向量是行得行，是列得列，没严格验证）

$$\frac{\partial \left(\vec{b}^T A\vec{x}\right)}{\partial \vec{x}} = A^T\vec{b}$$

$$\frac{\partial \left(\vec{b}^T A\vec{x}\right)}{\partial \vec{x}^T} = \vec{b}^T A$$

$\frac{\partial \left(\vec{x}^T A\vec{x}\right)}{\partial \vec{x}} = \left(A + A^T\right)\vec{x}$；若 A 是对称阵，$2A\vec{x}$ 。

$\frac{\partial \left(\vec{x}^T A\vec{x}\right)}{\partial \vec{x}^T} = \vec{x}^T \left(A + A^T\right)$；若 A 是对称阵，$2\vec{x}^T A$ 。

其实通过上面的式子可以发现，求导结果类似于我们函数求导的结果，唯一差异在于转置以及前后关系。而这一点，如果不想记忆，可以通过结果是行向量还是列向量"猜"到。

推导:

$$\frac{\partial\left(\vec{x}^T A \vec{x}\right)}{\partial \vec{x}}=\left(A+A^T\right) \vec{x}$$

$$\frac{\partial\left(\vec{x}^T A \vec{x}\right)}{\partial \vec{x}}=\frac{\partial\left(\sum_{i=1}^n \sum_{j=1}^n a_{ij} x_i x_j\right)}{\partial \vec{x}}=\begin{pmatrix}\sum_{j=1}^n a_{1j} x_j \\ \sum_{j=1}^n a_{2j} x_j \\ \vdots \\ \sum_{j=1}^n a_{nj} x_j\end{pmatrix}+\begin{pmatrix}\sum_{i=1}^n a_{i1} x_i \\ \sum_{i=1}^n a_{i2} x_i \\ \vdots \\ \sum_{i=1}^n a_{in} x_i\end{pmatrix}=\left(A+A^T\right) \vec{x}$$

接下来我们计算误差平方和 $\frac{\partial(\vec{y}-X\vec{w})^T(\vec{y}-X\vec{w})}{\partial \vec{w}}$ 的求导结果，需要注意的是，误差平方和是一个标量

所以分子肯定是行向量乘以列向量。而因为 $\vec{y}$ 是一个列向量，因此也一定是 $X\vec{w}$ 这种形式得到的列向量（而不是$\vec{w}X$）

$$\begin{aligned}\frac{\partial(\vec{y}-X\vec{w})^T(\vec{y}-X\vec{w})}{\partial \vec{w}}&=\frac{\partial(\vec{y}-X\vec{w})^T(\vec{y}-X\vec{w})}{\partial \vec{w}}\\&=\frac{\partial \vec{y}^T \vec{y}}{\partial \vec{w}}-\frac{\partial \vec{y}^T X \vec{w}}{\partial \vec{w}}-\frac{\partial \vec{w}^T X^T \vec{y}}{\partial \vec{w}}+\frac{\partial \vec{w}^T X^T X \vec{w}}{\partial \vec{w}}\end{aligned}$$

（1） $\frac{\partial \vec{y}^T \vec{y}}{\partial \vec{w}}=0$ （标量对向量，且标量和 $\vec{w}$ 无关）

（2） $\frac{\partial \vec{y}^T X \vec{w}}{\partial \vec{w}}=X^T \vec{y}$ （标量对向量，查看上面公式）

（3） $\frac{\partial \vec{w}^T X^T \vec{y}}{\partial \vec{w}}=\frac{\partial\left(\vec{w}^T X^T \vec{y}\right)^T}{\partial \vec{w}}=\frac{\partial \vec{y}^T X \vec{w}}{\partial \vec{w}}=X^T \vec{y}$（标量对向量，标量的转置等于本身，查看上面公式）

（4） $\frac{\partial \vec{w}^T X^T X \vec{w}}{\partial \vec{w}}=2\left(X^T X\right) \vec{w}$

上面式子整合，有:

$$\begin{aligned}\frac{\partial(\vec{y}-X\vec{w})^T(\vec{y}-X\vec{w})}{\partial \vec{w}}&=\frac{\partial \vec{y}^T \vec{y}}{\partial \vec{w}}-\frac{\partial \vec{y}^T X \vec{w}}{\partial \vec{w}}-\frac{\partial \vec{w}^T X^T \vec{y}}{\partial \vec{w}}+\frac{\partial \vec{w}^T X^T X \vec{w}}{\partial \vec{w}}\\&=0-X^T \vec{y}-X^T \vec{y}+2\left(X^T X\right) \vec{w}=-2X^T(\vec{y}-X\vec{w})\end{aligned}$$