空间解析几何

客观世界中有这样一类量，它们既有大小，又有方向，例如位移，速度，加速度，力，力矩等等，这一类量叫做向量（或矢量）。

由于一切向量的共性是他们都有大小和方向，因此在数学上我们只研究与起点无关的向量，并称这种向量为自由向量（以后简称向量），即只考虑向量的大小和方向，而不论它的起点在什么地方，当遇到与起点有关的向量时，可在一般原则下作特别处理。

由于我们只讨论自由向量，所以如果两个向量 a 和 b 的大小相等，且方向相同，我们就说向量 a 和 b 是相等的。记作 a=b，这就是说，经过平行移动后能完全重合的向量是相等的。

向量的大小叫做向量的模，$\overrightarrow{AB},\vec{a}$ 的模记作 $\left|\overrightarrow{AB}\right|,\left|\vec{a}\right|$，向量的模等于 1 的向量叫做单位向量，模等于零的向量叫做零向量。

1 点乘（内积、数量积）

1.1 点乘概念

向量是由 n 个实数组成的一个 n 行 1 列（n*1）或一个 1 行 n 列（1*n）的有序数组；向量的点乘,也叫向量的内积、数量积，对两个向量执行点乘运算，就是对这两个向量对应位一一相乘之后求和的操作，点乘的结果是一个标量。

点乘公式

对于向量 a 和向量 b：

$$\mathrm{a}=[a_1,a_2,\ldots,a_n]，\quad \mathrm{b}=[b_1,b_2,\ldots,b_n] \qquad （1）$$

a 和 b 的点积公式为：

$$\mathrm{a}\bullet b=a_1b_1+a_2b_2+\ldots+a_nb_n \qquad （2）$$

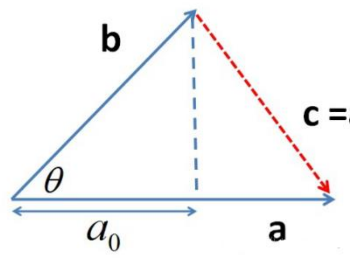
要求一维向量 a 和向量 b 的**行列数相同**。

1.2 点乘基本物理意义

点乘的几何意义是可以用来表征或计算两个向量之间的夹角，以及在 b 向量在 a 向量方向上的投影，有公式：

$$\mathrm{a}\bullet b=|a||b|\cos\theta \qquad （1）$$

推导过程如下，首先看一下向量组成：



定义向量：

$$c = a - b \tag{2}$$

根据三角形余弦定理有：

$$c^2 = a^2 + b^2 - 2|a||b|\cos\theta \tag{3}$$

根据关系 c=a-b（a、b、c 均为向量）有：

$$(a - b) \bullet (a - b) = a^2 + b^2 - 2a \bullet b = a^2 + b^2 - 2|a||b|\cos\theta \tag{4}$$

即：

$$a \bullet b = |a||b|\cos\theta \tag{5}$$

向量 a，b 的长度都是可以计算的已知量，从而有 a 和 b 间的夹角 θ：

$$\theta = \mathrm{acccos}(\frac{a \bullet b}{|a||b|}) \tag{6}$$

根据这个公式就可以计算向量 a 和向量 b 之间的夹角。从而就可以进一步判断这两个向量是否是同一方向，是否正交(也就是垂直)等方向关系，具体对应关系为：

a・b>0　方向基本相同，夹角在 0°到 90°之间

a・b=0　正交，相互垂直

a・b<0　方向基本相反，夹角在 90°到 180°之间

1.3 练习题

1.求平行于向量 a=(6,7,-6)的单位向量。

2.求点 M(4,-3,5)到各坐标轴的距离。

3.一向量的终点在点 B(2,-1,7)，它在 x 轴，y 轴和 z 轴上的投影依次为 4，-4，7，求这向量的起点 A 的坐标。

4.已知三点 M(1,1,1),A(2,2,1)和 B(2,1,2)，求∠AMB。

2 叉乘（外积、向量积）

2.1 叉乘概念

两个向量的叉乘，又叫向量积、外积、叉积，叉乘的运算结果是一个向量而不是一个标量。并且两个向量的叉积与这两个向量组成的坐标平面垂直。

对于向量 a 和向量 b：

$$\mathrm{a}=(x_1,y_1,z_1) \tag{1}$$

$$\mathrm{b}=(x_2,y_2,z_2) \tag{2}$$

a 和 b 的叉乘公式为：

$$\mathrm{a}\times b=\begin{vmatrix} i & j & k \\ x_1 & y_1 & z_1 \\ x_2 & y_2 & z_2 \end{vmatrix}=(y_1z_2-y_2z_1)\mathrm{i}-(x_1z_2-x_2z_1)j+(x_1y_2-x_2y_1)k \tag{3}$$

其中：$\mathrm{i}=(1,0,0)$，$\mathrm{j}=(0,1,0)$，$\mathrm{k}=(0,0,1)$，根据 i、j、k 间关系，有：

$$\mathrm{a}\times b=(y_1z_2-y_2z_1,-(x_1z_2-x_2z_1),x_1y_2-x_2y_1) \tag{4}$$

由向量积的定义可以推得：

（1）$\boldsymbol{a}\times\boldsymbol{a}=\boldsymbol{0}$.

这是因为夹角 $\theta=0$,所以 $|\boldsymbol{a}\times\boldsymbol{a}|=|\boldsymbol{a}|^2\sin 0=0$.

（2）对于两个非零向量 $\boldsymbol{a}$、$\boldsymbol{b}$,如果 $\boldsymbol{a}\times\boldsymbol{b}=\boldsymbol{0}$,那么 $\boldsymbol{a}/\!/\boldsymbol{b}$;反之,如果 $\boldsymbol{a}/\!/\boldsymbol{b}$,那么 $\boldsymbol{a}\times\boldsymbol{b}=\boldsymbol{0}$.

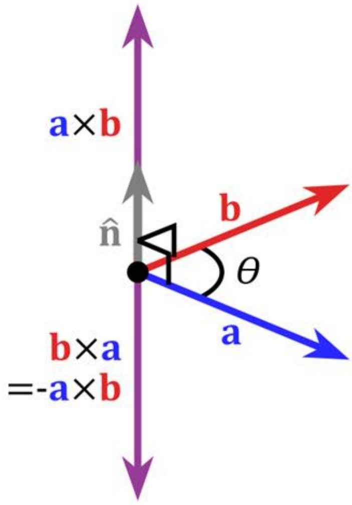
这是因为如果 $\boldsymbol{a}\times\boldsymbol{b}=\boldsymbol{0}$,由于 $|\boldsymbol{a}|\neq 0$, $|\boldsymbol{b}|\neq 0$,故必有 $\sin\theta=0$,于是 $\theta=0$ 或 π,即 $\boldsymbol{a}/\!/\boldsymbol{b}$;反之,如果 $\boldsymbol{a}/\!/\boldsymbol{b}$,那么 $\theta=0$ 或 π,于是 $\sin\theta=0$,从而 $|\boldsymbol{a}\times\boldsymbol{b}|=0$,即 $\boldsymbol{a}\times\boldsymbol{b}=\boldsymbol{0}$.

由于可以认为零向量与任何向量都平行,因此,上述结论可叙述为:向量 $\boldsymbol{a}/\!/\boldsymbol{b}$ 的充分必要条件是 $\boldsymbol{a}\times\boldsymbol{b}=\boldsymbol{0}$.

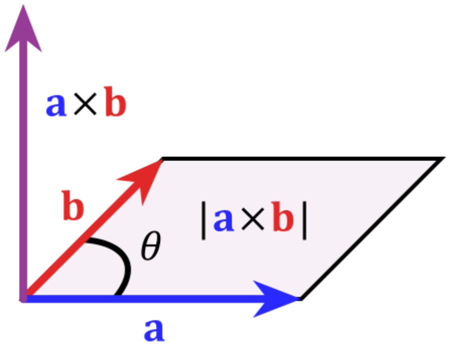
2.2 叉乘几何意义

在三维几何中，向量 a 和向量 b 的叉乘结果是一个向量，更为熟知的叫法是法向量，该向量垂直于 a 和 b 向量构成的平面。

在 3D 图像学中，叉乘的概念非常有用，可以通过两个向量的叉乘，生成第三个垂直于 a，b 的法向量，从而构建 X、Y、Z 坐标系。如下图所示：



在二维中，两个向量的向量积的模的绝对值等于由这两天向量组成的平行四边形的面积：|a×b|等于由向量 a 和向量 b 构成的平行四边形的面积。



2.3 练习题

1.已知三角形 ABC 的顶点分别是 A(1,2,3)、B(3,4,5)和(2,4,7)，求三角形 ABC 的面积。

2.已知 A(1,2,0)，B(2,3,1)，C(4,2,2)，M(x,y,z)四点共面，求点 M 的坐标 x、y、z 所满足的关系式。。

3 范数

3.1 向量范数

3.1.1 向量范数的定义

如果 V 是数据 K 上的线性空间，且对于 V 的任一向量 x，对应一个实值函数||x||,它满足三个条件：

1、非负性：当 x≠0 时，||x||>0;当 x=0 时，||x||=0；

2、齐次性：||ax||=|a| ||x|| (a∈K，x∈V)；

3、三角不等式：||x+y||≤||x||+||y|| (x,y∈V)。

则称||x||为 V 上向量 x 的范数，简称向量范数。

那么我们怎么理解什么是范数呢？

我们知道距离的定义是一个宽泛的概念，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。有时候为了便于理解，我们可以把范数当作距离来理解。

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。

这里简单地介绍以下几种向量范数的定义和含义：

3.1.1 p-范数

与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：

$$Lp=\sqrt[p]{\sum_{1}^{n}x_i^p},\ x=(x_1,x_2,\cdots,x_n)$$

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。

实际上，在 0≤p<1≤p<1 时，Lp 并不满足三角不等式的性质，也就不是严格意义下的范数。以 p=0.5，二维坐标(1,4)、(4,1)、(1,9)为例：

$$\sqrt[0.5]{(1+\sqrt{4})}+\sqrt[0.5]{(\sqrt{4}+1)}<\sqrt[0.5]{(1+\sqrt{9})}$$

因此这里的 L-P 范数只是一个概念上的宽泛说法。

3.1.2 1-范数

L1 范数是我们经常见到的一种范数，它的定义如下：

$$||x||_1 = \sum_i |x_i|$$

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）：

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

对于 L1 范数，它的优化问题如下：

$$min||x||_1$$
$$s.t. Ax = b$$

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。

3.1.3 2-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：

$$\|X\|_2 = \left(\sum_{i=1}^{n} x_i^{\ 2}\right)^{\frac{1}{2}} = \sqrt{\sum_{i=1}^{n} x_i^{\ 2}}$$

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）.

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

对于 L2 范数，它的优化问题如下：

$$min||x||_2$$
$$s.t. Ax = b$$

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。

3.1.4 ∞-范数

当 P=∞时，也就是 L-∞范数，它主要被用来度量向量元素的最大值。用上面的 L-P 定义可以得到的 L-∞的定义为：

$$||x||_\infty = \sqrt[\infty]{\sum_1^n x_i^\infty},\ x = (x_1, x_2, \cdots, x_n)$$

与 L0 一样，在通常情况下，大家都用的是：

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

来表示 L-∞范数。

3.1.5 加权范数

加权最小二乘法是对原模型进行加权，使之成为一个新的不存在异方差性的模型，然后采用普通最小二乘法估计其参数的一种数学优化技术。

一般最小二乘法将时间序列中的各项数据的重要性同等看待，而事实上时间序列各项数据对未来的影响作用应是不同的。一般来说，近期数据比起远期数据对未来的影响更大。因此比较合理的方法就是使用加权的方法，对近期数据赋以较大的权数，对远期数据则赋以较小的权数。加权最小二乘法采用指数权数 W^{n-i}，$0<W<1$，加权以后求得的参数估计值应满足：

$$S = \sum_{i=1}^{n} W^{n-i} (y_i - \hat{y}_i)^2 = min, (i = 1, 2, \cdots, n)$$

以直线模型 $\hat{y} = a + bt$ 为例，其加权的剩余平方和为：

$$S = \sum_{i=1}^{n} W^{n-i} (y_i - a - bt)^2$$

对上式分别求 a 和 b 的偏导数，得到标准方程组：

$$\begin{cases} \sum W^{n-1} y_i = a \sum W^{n-i} + b \sum W^{n-1} t \\ \sum W_{n-1} t y = a \sum W^{n-i} t + b \sum W^{n-i} t^2 \end{cases}$$

对上述方程解出 a 和 b，就得到加权最小二乘法直线模型。应用加权最小二乘法，W 的取值不同，解出的 a，b 也不同，因此 W 值取多少，需要经分析后确定。

3.1.6 例题

1. 向量 X=[2, 3, -5, -7]，求向量的 1-范数，2-范数和无穷范数。

向量的 1-范数：各个元素的绝对值之和：

$$\|x\|_1 = 2+3+5+7=17$$

向量的 2-范数：每个元素的平方和再开平方根：

$$\|x\|_2 = (2\times 2 + 3\times 3 + 5\times 5 + 7\times 7)^{\frac{1}{2}} = 9.3274$$

向量的无穷范数：

（1）正无穷范数：向量的所有元素的绝对值中最大的；即 X 的负无穷范数为：7；

（2）负无穷范数：向量的所有元素的绝对值中最小的；即 X 的负无穷范数为：2。

3.2 矩阵范数

3.2.1 矩阵范数的定义与性质

广义矩阵范数

定义 2.3 设 $A \in \mathbf{C}^{m\times n}$，定义一个实值函数 $\|A\|$，它满足以下三个条件

(1) 非负性：当 $A \neq O$ 时，$\|A\| > 0$；当 $A = O$ 时，$\|A\| = 0$；

(2) 齐次性：$\|\alpha A\| = |\alpha| \, \|A\| \quad (\alpha \in \mathbf{C})$；

(3) 三角不等式：$\|A+B\| \leqslant \|A\| + \|B\| \quad (B \in \mathbf{C}^{m\times n})$.

则称 $\|A\|$ 为 A 的**广义矩阵范数**[15].

矩阵范数

若对 $\mathbf{C}^{m\times n}$，$\mathbf{C}^{n\times l}$ 及 $\mathbf{C}^{m\times l}$ 上的同类广义矩阵范数 $\|\cdot\|$[①]，还应满足下面一个条件：

(4) 相容性：

$$\|AB\| \leqslant \|A\| \, \|B\| \quad (B \in \mathbf{C}^{n\times l}) \tag{2.2.1}$$

则称 $\|A\|$ 为 A 的**矩阵范数**.

F-范数

$$\|A\|_F = \left(\sum_{i=1}^{m}\sum_{j=1}^{n} |a_{ij}|^2\right)^{1/2} = (\mathrm{tr}(A^{\mathrm{H}}A))^{1/2} \tag{2.2.3}$$

范数(2.2.3)又称为 **Frobenius 范数**，或简称为 ***F*-范数**. 为了与例 2.7 的两种范数一致起见，这一范数亦可用 $\|A\|_{m_2}$ 表示.

3.2.2 几种常用的矩阵范数

从属范数

定理 2.4 已知 $\mathbf{C}^m$ 和 $\mathbf{C}^n$ 上的同类向量范数 $\|\cdot\|$，设 $A \in \mathbf{C}^{m\times n}$，则函数

$$\|A\| = \max_{\|x\|=1} \|Ax\| \tag{2.2.6}$$

是 $\mathbf{C}^{m\times n}$ 上的矩阵范数，且与已知的向量范数相容.

称由式(2.2.6)给出的矩阵范数为“**由向量范数导出的矩阵范数**”，简称为**从属范数**. 对于 $\mathbf{C}^{n\times n}$ 上的任何一种从属范数，有

$$\|I\| = \max_{\|x\|=1} \|Ix\| = 1$$

但对于一般的矩阵范数(设该矩阵范数与某向量范数相容，见例 2.9)，由于

$$\|x\| = \|Ix\| \leqslant \|I\| \, \|x\|$$

对任意的 $x \in \mathbf{C}^n$ 成立，所以 $\|I\| \geqslant 1$.

列和范数、谱范数、行和范数

定理 2.5 设 $A=(a_{ij})_{m\times n}\in \mathbf{C}^{m\times n}$，$x=(\xi_1,\xi_2,\cdots,\xi_n)^T\in \mathbf{C}^n$，则从属于向量 x 的三种范数 $\|x\|_1$，$\|x\|_2$，$\|x\|_\infty$ 的矩阵范数依次是：

(1) $\|A\|_1=\max\limits_{j}\sum\limits_{i=1}^{m}|a_{ij}|$； (2.2.7)

(2) $\|A\|_2=\sqrt{\lambda_1}$， λ_1 为 A^HA 的最大特征值； (2.2.8)

(3) $\|A\|_\infty=\max\limits_{i}\sum\limits_{j=1}^{n}|a_{ij}|$. (2.2.9)

通常称 $\|A\|_1$，$\|A\|_2$ 及 $\|A\|_\infty$ 依次为列和范数，谱范数及行和范数.

3.3 SVM 中数学建模中范数应用

求解这个"决策面"的过程，就是最优化。一个最优化问题通常有两个基本的因素：
1）目标函数，也就是你希望什么东西的什么指标达到最好；

2）优化对象，你期望通过改变哪些因素来使你的目标函数达到最优。在线性 SVM 算法中，目标函数显然就是那个"分类间隔"，而优化对象则是决策面。所以要对 SVM 问题进行数学建模，首先要对上述两个对象（"分类间隔"和"决策面"）进行数学描述。按照一般的思维习惯，我们先描述决策面。数学建模的时候，先在二维空间建模，然后再推广到多维。

3.3.1 "决策面"方程

我们都知道二维空间下一条直线的方式如下所示：

$$y = \mathrm{ax} + \mathrm{b}$$

现在我们做个小小的改变，让原来的 x 轴变成 x1，y 轴变成 x2。

$$\mathrm{x}_2 = \mathrm{ax}_1 + \mathrm{b}$$

移项得：

$$\mathrm{ax}_1 - \mathrm{x}_2 + \mathrm{b} = 0$$

将公式向量化得：

$$[\mathrm{a} \quad -1]\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} + \mathrm{b} = 0$$

进一步向量化，用 w 列向量和 x 列向量和标量 γ 进一步向量化：

$$\boldsymbol{\omega}^T \boldsymbol{x} + \gamma = 0$$

其中，向量 w 和 x 分别为：

$$\boldsymbol{\omega} = [\omega_1, \omega_2]^T, \boldsymbol{x} = [x_1, x_2]^T$$

这里 w1=a，w2=-1。我们都知道，最初的那个直线方程 a 和 b 的几何意义，a 表示直线的斜率，b 表示截距，a 决定了直线与 x 轴正方向的夹角，b 决定了直线与 y 轴交点位置。那么向量化后的直线的 w 和 r 的几何意义是什么呢？

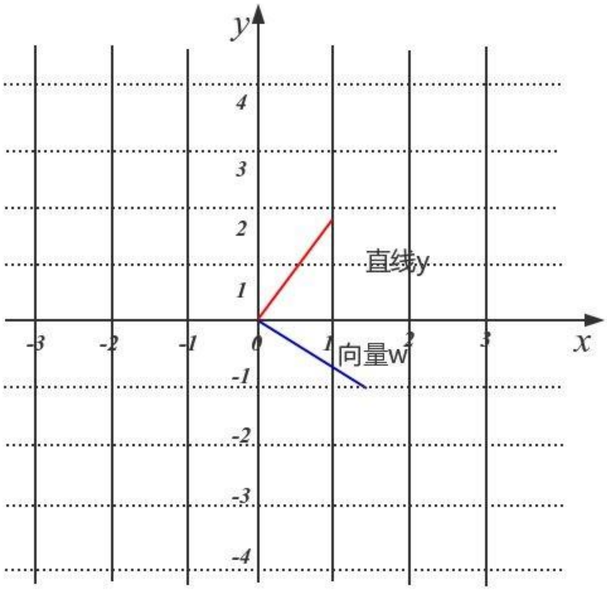
现在假设：

$$a = \sqrt{3}, b = 0$$

可得：

$$\boldsymbol{\omega} = [\sqrt{3}, -1]^T$$

在坐标轴上画出直线和向量 w：



蓝色的线代表向量 w，红色的先代表直线 y。我们可以看到向量 w 和直线的关系为垂直关系。这说明了向量 w 也控制这直线的方向，只不过是与这个直线的方向是垂直的。标量 γ 的作用也没有变，依然决定了直线的截距。此时，我们称 w 为直线的法向量。

二维空间的直线方程已经推导完成，将其推广到 n 为空间，就变成了超平面方程。(一个超平面，在二维空间的例子就是一个直线)但是它的公式没变，依然是：

$$\boldsymbol{\omega}^T \boldsymbol{x} + \gamma = 0$$

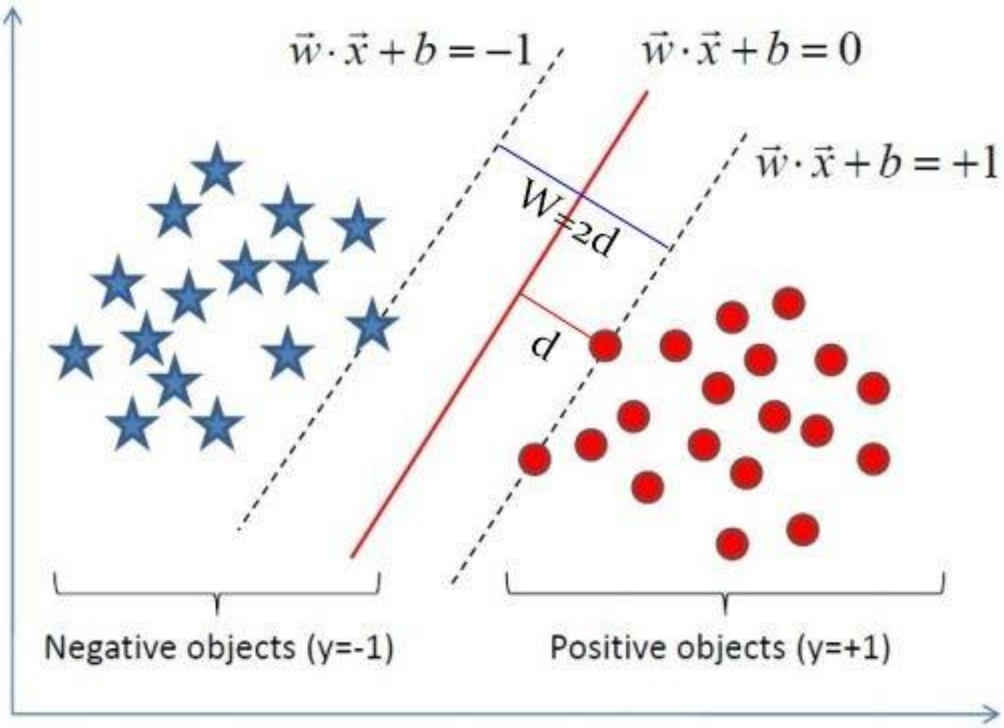
不同之处在于：

$$\boldsymbol{\omega} = [\omega_1, \omega_2, \cdots, \omega_n]^T$$
$$\boldsymbol{x} = [x_1, x_2, \cdots, x_n]^T$$

我们已经顺利推导出了"决策面"方程，它就是我们的超平面方程，之后，我们统称其为超平面方程。

3.3.2 分类间隔"方程

现在，我们依然对于一个二维平面的简单例子进行推导。



我们已经知道间隔的大小实际上就是支持向量对应的样本点到决策面的距离的二倍。那么图中的距离 d 我们怎么求？我们高中都学过，点到直线的距离距离公式如下：

$$d=\left|\frac{Ax_0+By_0+C}{\sqrt{A^2+B^2}}\right|$$

公式中的直线方程为 Ax0+By0+C=0， 点 P 的坐标为(x0,y0)。

现在，将直线方程扩展到多维，求得我们现在的超平面方程，对公式进行如下变形：

$$d=\frac{|\boldsymbol{\omega}^T\boldsymbol{x}+\gamma|}{||\boldsymbol{\omega}||}$$

这个 d 就是"分类间隔"。其中||w||表示 w 的二范数，求所有元素的平方和，然后再开方。比如对于二维平面：

$$\boldsymbol{\omega}=[\omega_1,\omega_2]^T$$

那么，

$$||\boldsymbol{\omega}||=\sqrt[2]{{\omega_1}^2+{\omega_2}^2}$$

我们目的是为了找出一个分类效果好的超平面作为分类器。分类器的好坏的评定依据是分类间隔 W=2d 的大小，即分类间隔 W 越大，我们认为这个超平面的分类效果越好。此

时，求解超平面的问题就变成了求解分类间隔 W 最大化的为题。W 的最大化也就是 d 最大化的。

3.3.3 约束条件

看起来，我们已经顺利获得了目标函数的数学形式。但是为了求解 w 的最大值。我们

- 不得不面对如下问题：我们如何判断超平面是否将样本点正确分类？
- 我们知道要求距离 d 的最大值，我们首先需要找到支持向量上的点，怎么在众多的点中选出支持向量上的点呢？

上述我们需要面对的问题就是约束条件，也就是说我们优化的变量 d 的取值范围受到了限制和约束。事实上约束条件一直是最优化问题里最让人头疼的东西。但既然我们已经知道了这些约束条件确实存在，就不得不用数学语言对他们进行描述。但 SVM 算法通过一些巧妙的小技巧，将这些约束条件融合到一个不等式里面。

这个二维平面上有两种点，我们分别对它们

- 进行标记：红颜色的圆点标记为 1，我们

  人为规定其为正样本；
- 蓝颜色的五角星标记为-1，我们人为规定其为负样本。

对每个样本点 xi 加上一个类别标签 yi：

$$y_i = \begin{cases} +1 & \text{红色点} \\ -1 & \text{蓝色点} \end{cases}$$

如果我们的超平面方程能够完全正确地对上图的样本点进行分类，就会满足下面的方程：

$$\begin{cases} \boldsymbol{\omega}^T x_i + \gamma > 0 & y_i = 1 \\ \boldsymbol{\omega}^T x_i + \gamma < 0 & y_i = -1 \end{cases}$$

如果我们要求再高一点，假设决策面正好处于间隔区域的中轴线上，并且相应的支持向量对应的样本点到决策面的距离为 d，那么公式进一步写成：

$$\begin{cases} \dfrac{\boldsymbol{\omega}^T x_i + \gamma}{||\boldsymbol{\omega}||} \geq \mathrm{d} & \forall y_i = 1 \\ \dfrac{\boldsymbol{\omega}^T x_i + \gamma}{||\boldsymbol{\omega}||} \leq -\mathrm{d} & \forall y_i = -1 \end{cases}$$

上述公式的解释就是，对于所有分类标签为 1 和-1 样本点，它们到直线的距离都大于等于 d(支持向量上的样本点到超平面的距离)。公式两边都除以 d，就可以得到：

$$\begin{cases} \boldsymbol{\omega_d}^T x_i + \gamma_d \geq 1 & \forall y_i = 1 \\ \boldsymbol{\omega_d}^T x_i + \gamma_d \leq -1 & \forall y_i = -1 \end{cases}$$

其中，

$$\boldsymbol{\omega}_d = \frac{\boldsymbol{\omega}}{||\boldsymbol{\omega}||d},\ \ \gamma_d = \frac{\gamma}{||\boldsymbol{\omega}||d}$$

因为||w||和 d 都是标量。所上述公式的两个矢量，依然描述一条直线的法向量和截距。

$$\boldsymbol{\omega}_d^T \boldsymbol{x} + \gamma_d = 0$$
$$\boldsymbol{\omega}^T \boldsymbol{x} + \gamma = 0$$

上述两个公式，都是描述一条直线，数学模型代表的意义是一样的。现在，让我们对 wd 和 γd 重新起个名字，就叫它们 w 和 γ。因此，我们就可以说："对于存在分类间隔的两类样本点，我们一定可以找到一些超平面面，使其对于所有的样本点均满足下面的条件："

$$\begin{cases} \boldsymbol{\omega}^T x_i + \gamma \geq 1 & \forall y_i = 1 \\ \boldsymbol{\omega}^T x_i + \gamma \leq -1 & \forall y_i = -1 \end{cases}$$

上述方程即给出了 SVM 最优化问题的约束条件。这时候，可能有人会问了，为什么标记为 1 和-1 呢？因为这样标记方便我们将上述方程变成如下形式：

$$y_i(\boldsymbol{\omega}^T x_i + \gamma) \geq 1 \ \ \forall x_i$$

正是因为标签为 1 和-1，才方便我们将约束条件变成一个约束方程，从而方便我们的计算。

3.3.4 线性 SVM 优化问题基本描述

现在整合一下思路，我们已经得到我们的目标函数：

$$d = \frac{|\boldsymbol{\omega}^T \boldsymbol{x} + \gamma|}{||\boldsymbol{\omega}||}$$

我们的优化目标是是 d 最大化。我们已经说过，我们是用支持向量上的样本点求解 d 的最大化的问题的。那么支持向量上的样本点有什么特点呢？

$$|\boldsymbol{\omega}^T x_i + \gamma| = 1 \quad \forall \text{支持向量上的样本点} x_i$$

你赞同这个观点吗？所有支持向量上的样本点，都满足如上公式。如果不赞同，请重看"分类间隔"方程推导过程。现在我们就可以将我们的目标函数进一步化简：

$$\mathrm{d} = \frac{1}{||\boldsymbol{\omega}||}$$

因为，我们只关心支持向量上的点。随后我们求解 d 的最大化问题变成了||w||的最小化问题。进而||w||的最小化问题等效于

$$\min \frac{1}{2} ||\omega||^2$$

为什么要做这样的等效呢？这是为了在进行最优化的过程中对目标函数求导时比较方便，但这绝对不影响最优化问题最后的求解。我们将最终的目标函数和约束条件放在一起进行描述：

$$\min \frac{1}{2} ||\omega||^2$$
$$\text{s.t.} \quad y_i(\boldsymbol{\omega}^T x_i + b) \geq 1, i = 1,2, \ldots, n$$

这里 n 是样本点的总个数，缩写 s.t.表示"Subject to"，是"服从某某条件"的意思。上述公式描述的是一个典型的不等式约束条件下的二次型函数优化问题，同时也是支持向量机的基本数学模型。

4 欧氏空间

4.1 欧氏空间的定义

定义 1.22 设 V 是实数域 $\mathbf{R}$ 上的线性空间，对于 V 中任二向量 $\boldsymbol{x}$ 与 $\boldsymbol{y}$，按某规则定义一个实数，用 $(\boldsymbol{x}, \boldsymbol{y})$ 表示，且它满足下列四个条件：

(1) 交换律：$(\boldsymbol{x}, \boldsymbol{y}) = (\boldsymbol{y}, \boldsymbol{x})$；

(2) 分配律：$(\boldsymbol{x}, \boldsymbol{y}+\boldsymbol{z}) = (\boldsymbol{x}, \boldsymbol{y}) + (\boldsymbol{x}, \boldsymbol{z})$；

(3) 齐次性：$(k\boldsymbol{x}, \boldsymbol{y}) = k(\boldsymbol{x}, \boldsymbol{y})$ $(\forall k \in \mathbf{R})$；

(4) 非负性：$(\boldsymbol{x}, \boldsymbol{x}) \geqslant 0$，当且仅当 $\boldsymbol{x} = \mathbf{0}$ 时，$(\boldsymbol{x}, \boldsymbol{x}) = 0$.

则称 V 为 **Euclid** 空间，简称欧氏空间或实内积空间 .

4.2 度量矩阵

假定 $x_1, x_2, \cdots, x_n$ 是 n 维欧氏空间 V^n 的基,对于 V^n 的任意两个向量

$$x = \xi_1 x_1 + \xi_2 x_2 + \cdots + \xi_n x_n,\ y = \eta_1 x_1 + \eta_2 x_2 + \cdots + \eta_n x_n$$

由性质 3 可得

$$(x, y) = \sum_{i,j=1}^{n} \xi_i \eta_j (x_i, x_j) = \sum_{i,j=1}^{n} a_{ij} \xi_i \eta_j \tag{1.3.4}$$

其中 $a_{ij} = (x_i, x_j)\ (i, j = 1, 2, \cdots, n)$,用矩阵表示就有

$$(x, y) = (\xi_1, \xi_2, \cdots, \xi_n) A \begin{bmatrix} \eta_1 \\ \eta_2 \\ \vdots \\ \eta_n \end{bmatrix} \tag{1.3.5}$$

这里

$$A = (a_{ij}) = \begin{bmatrix} (x_1, x_1) & \cdots & (x_1, x_n) \\ (x_2, x_1) & \cdots & (x_2, x_n) \\ \vdots & & \vdots \\ (x_n, x_1) & \cdots & (x_n, x_n) \end{bmatrix} \tag{1.3.6}$$

由式(1.3.6)可以看出,只要知道其中任意两个基向量的内积,也就知道了矩阵 A,从而也就知道了任意两个向量的内积. 因此,称式(1.3.6)中的 A 为 V^n 对于基 $x_1, x_2, \cdots, x_n$ 的**度量矩阵**(或 **Gram 矩阵**). 因为 $(x_i, x_j) = (x_j, x_i)\ (i, j = 1, 2, \cdots, n)$,所以有 $a_{ij} = a_{ji}\ (i, j = 1, 2, \cdots, n)$,即 A 是对称矩阵;又因为对任意向量 $x \neq 0$,由式(1.3.4)或式(1.3.5)知 $(x, x) > 0$,故 A 是正定矩阵.

显然,若矩阵 A 非奇异,则 $A^T A$ 必对称正定. 这是因为 A 的列向量可视为 R^n 的一个基,而 R^n 对于该基的度量矩阵正是 $A^T A$.

4.3 向量的长度(模)

定义 1.23 在欧氏空间 V 中,非负实数 $\sqrt{(x, x)}$ 称为 V 中向量 x 的**长度**(或**模**,**范数**),记为 $|x|$(或 $\|x\|_2$),即

$$|x| = \sqrt{(x, x)} \tag{1.3.7}$$

由此定义可知,非零向量的长度是正数,只有零向量的长度才是零. 这样定义的长度符合熟知的如下性质

$$|kx| = |k|\,|x| \tag{1.3.8}$$

$$|x + y| \leqslant |x| + |y| \tag{1.3.9}$$

4.4 单位向量

单位向量是指模等于 1 的向量。由于是非零向量，单位向量具有确定的方向。

一个非零向量除以它的模，可得所需单位向量。

4.5 规范化

长度为 1 的向量称为**单位向量**．如果 $x \neq 0$，由式(1.3.8)知

$$\frac{1}{|x|}x \qquad (1.3.10)$$

是一个单位向量．用向量 x 的长度 $|x|$ 去除 x，得到一个与 x 同方向的单位向量，通常称此过程为把 x **单位化**或**规范化**．

5 向量的距离

数学上距离通常是指用两个向量的距离。向量距离的大小是很多算法中的重要参考数据。

5.1 Euclidean Distance(欧几里德距离)

欧几里德距离是最常见的距离。在二维平面上，就是连接两个点的线段的长度。

对于给定的两个点：p = (p1, p2,..., pn)和 q = (q1, q2,..., qn)这间的距离可以计算如下:

D(p,q)=D(q,p) = sqrt(∑(p-q)^2)

5.2 Manhattan Distance(曼哈顿距离)

英文也叫做 Taxicab Geometry。

Taxicab 是出租车的意思。想想出租车在格子状（垂直相交）的城市道路行驶的距离，就是曼哈顿距离的典型例子。车子从 A 点到 B 点，途中有好几个十字路口，则车子无所谓在哪个路口拐弯或不拐弯，只要是朝着目标的方向走，距离总是一样的。

数学可以表示为:

D(p,q)=D(q,p) = ∑abs(p-q)

5.3 Chebyshev Distance(切比雪夫距离)

切比雪夫距离描述了向量在各个维度上的距离的最大值。

先看数学表示：

D(p,q)=D(q,p) = max(abs(p-q))

国际象棋中的王或后可以横向移动，也可以斜向移动，但都表示一步，或者说移动距离是 1 个单位。

5.4 K-近邻算法示例

k-近邻算法采用测量不同特征值之间的距离来进行分类：

电影可以按照题材分类，每个题材又是如何定义的呢？那么假如两种类型的电影，动作片和爱情片。动作片有哪些公共的特征？那么爱情片又存在哪些明显的差别呢？我们发现动作片中打斗镜头的次数较多，而爱情片中接吻镜头相对更多。当然动作片中也有一些接吻镜头，爱情片中也会有一些打斗镜头。所以不能单纯通过是否存在打斗镜头或者接吻镜头来判断影片的类别。那么现在我们有 6 部影片已经明确了类别，也有打斗镜头和接吻镜头的次数，还有一部电影类型未知。

电影名称	打斗镜头	接吻镜头	电影类型
California Man	3	104	爱情片
He's not Really into dues	2	100	爱情片
Beautiful Woman	1	81	爱情片
Kevin Longblade	101	10	动作片
Robo Slayer 3000	99	5	动作片
Amped II	98	2	动作片
?	18	90	未知

那么我们使用 K-近邻算法来分类爱情片和动作片：存在一个样本数据集合，也叫训练样本集，样本个数 M 个，知道每一个数据特征与类别对应关系，然后存在未知类型数据集合 1 个，那么我们要选择一个测试样本数据中与训练样本中 M 个的距离，排序过后选出最近的 K 个，这个取值一般不大于 20 个。选择 K 个最相近数据中次数最多的分类。那么我们根据这个原则去判断未知电影的分类

电影名称	与未知电影的距离
California Man	20.5
He's not Really into dues	18.7
Beautiful Woman	19.2
Kevin Longblade	115.3
Robo Slayer 3000	117.4
Amped II	118.9

我们假设 K 为 3，那么排名前三个电影的类型都是爱情片，所以我们判定这个未知电影也是一个爱情片。那么计算距离是怎样计算的呢？

6 向量夹角和正交向量

在解析几何中，两个向量 $\boldsymbol{x}$ 与 $\boldsymbol{y}$ 的夹角 $\langle \boldsymbol{x}, \boldsymbol{y} \rangle$ 的余弦，可以通过数量积，即内积表示为

$$\cos\langle \boldsymbol{x}, \boldsymbol{y} \rangle = \frac{(\boldsymbol{x}, \boldsymbol{y})}{|\boldsymbol{x}|\,|\boldsymbol{y}|} \tag{1.3.11}$$

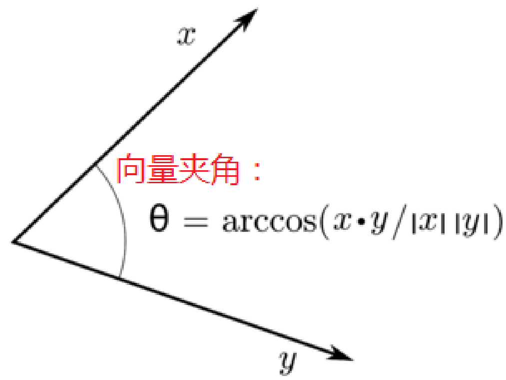
为了在欧氏空间中利用式(1.3.11)引入向量夹角的概念，须要证明不等式

$$\left| \frac{(\boldsymbol{x}, \boldsymbol{y})}{|\boldsymbol{x}|\,|\boldsymbol{y}|} \right| \leqslant 1$$

或

$$|(\boldsymbol{x}, \boldsymbol{y})| \leqslant |\boldsymbol{x}|\,|\boldsymbol{y}| \tag{1.3.12}$$

对任意两个向量 $\boldsymbol{x}$，$\boldsymbol{y}$ 均成立，其中当且仅当 $\boldsymbol{x}$，$\boldsymbol{y}$ 线性相关时，等号才成立．式(1.3.12)就是所谓的 Cauchy-Буняковский 不等式．



6.1 正交向量

通常，两个向量垂直的充分必要条件是它们夹角的余弦为零，亦即它们的数量积为零．在一般的欧氏空间中，仍以内积定义二向量夹角的余弦．因而很自然地引入下面的定义．

定义 1.25　如果对于欧氏空间中的两个向量 $\boldsymbol{x}$ 与 $\boldsymbol{y}$，有 $(\boldsymbol{x}, \boldsymbol{y}) = 0$，则称 $\boldsymbol{x}$ 与 $\boldsymbol{y}$ **正交或垂直**，记为 $\boldsymbol{x} \perp \boldsymbol{y}$.

由定义 1.25 可知，当 $\boldsymbol{x}$ 与 $\boldsymbol{y}$ 正交时，$\boldsymbol{y}$ 与 $\boldsymbol{x}$ 也正交；零向量与任意向量均正交；如果 $\boldsymbol{x} \perp \boldsymbol{y}$，且 $\boldsymbol{x}$ 与 $\boldsymbol{y}$ 线性相关，则此二向量中至少有一个是零向量．

6.2 正交向量组

定义 1.26　如果欧氏空间中一组非零向量两两正交，则称为**正交向量组**．

7 标准正交基

7.1 正交基、标准正交基

定义 1.27　在欧氏空间 V^n 中，由 n 个非零向量组成的正交向量组称为 V^n 的**正交基**；由单位向量组成的正交基称为**标准正交基**或**法正交基**．

7.2 正交规范化

定理 1.33　对于欧氏空间 V^n 的任一基 $\boldsymbol{x}_1, \boldsymbol{x}_2, \cdots, \boldsymbol{x}_n$，都可找到一个标准正交基 $\boldsymbol{y}_1, \boldsymbol{y}_2, \cdots, \boldsymbol{y}_n$. 换言之，任一非零欧氏空间都有正交基和标准正交基.

证　应用下面论述的关于向量组的 Schmidt 正交化方法（或过程），给出定理的构造性证明. 为此取 $\boldsymbol{y}_1' = \boldsymbol{x}_1$，作为所求正交基中的第一个向量. 再令

$$\boldsymbol{y}_2' = \boldsymbol{x}_2 + k\boldsymbol{y}_1'$$

由正交条件$(\boldsymbol{y}_2', \boldsymbol{y}_1') = 0$来决定待定常数$k$. 由

$$(\boldsymbol{x}_2 + k\boldsymbol{y}_1', \boldsymbol{y}_1') = (\boldsymbol{x}_2, \boldsymbol{y}_1') + k(\boldsymbol{y}_1', \boldsymbol{y}_1') = 0$$

得
$$k = -\frac{(\boldsymbol{x}_2, \boldsymbol{y}_1')}{(\boldsymbol{y}_1', \boldsymbol{y}_1')}$$

这样就得到两个正交的向量$\boldsymbol{y}_1'$, $\boldsymbol{y}_2'$,且$\boldsymbol{y}_2' \neq \boldsymbol{0}$. 又令

$$\boldsymbol{y}_3' = \boldsymbol{x}_3 + k_2\boldsymbol{y}_2' + k_1\boldsymbol{y}_1'$$

再由正交条件$(\boldsymbol{y}_3', \boldsymbol{y}_2') = 0$及$(\boldsymbol{y}_3', \boldsymbol{y}_1') = 0$来决定出$k_1$和$k_2$为

$$k_2 = -\frac{(\boldsymbol{x}_3, \boldsymbol{y}_2')}{(\boldsymbol{y}_2', \boldsymbol{y}_2')}, \quad k_1 = -\frac{(\boldsymbol{x}_3, \boldsymbol{y}_1')}{(\boldsymbol{y}_1', \boldsymbol{y}_1')}$$

到此,已经做出三个两两正交的向量$\boldsymbol{y}_1'$, $\boldsymbol{y}_2'$, $\boldsymbol{y}_3'$,且$\boldsymbol{y}_3' \neq \boldsymbol{0}$. 继续这样进行下去,设已做出$m$个两两正交且不为零的向量$\boldsymbol{y}_1'$, $\boldsymbol{y}_2'$, $\cdots$, $\boldsymbol{y}_m'$,为求出第$m+1$个与之正交的向量,令

$$\boldsymbol{y}_{m+1}' = \boldsymbol{x}_{m+1} + l_m\boldsymbol{y}_m' + l_{m-1}\boldsymbol{y}_{m-1}' + \cdots + l_2\boldsymbol{y}_2' + l_1\boldsymbol{y}_1'$$

使用m个正交条件

$$(\boldsymbol{y}_{m+1}', \boldsymbol{y}_i') = 0 \quad (i = 1, 2, \cdots, m)$$

来决定l_m, l_{m-1}, $\cdots$, l_2, l_1. 根据$\boldsymbol{y}_1'$, $\boldsymbol{y}_2'$, $\cdots$, $\boldsymbol{y}_m'$为两两正交的假定,可得

$$(\boldsymbol{x}_{m+1}, \boldsymbol{y}_i') + l_i(\boldsymbol{y}_i', \boldsymbol{y}_i') = 0$$

故

$$l_i = -\frac{(\boldsymbol{x}_{m+1}, \boldsymbol{y}_i')}{(\boldsymbol{y}_i', \boldsymbol{y}_i')} \quad (i = 1, 2, \cdots, m) \tag{1.3.18}$$

于是$\boldsymbol{y}_{m+1}'$就被确定出来了.

采用上述 Schmidt 正交化方法,可由已知基构造出n个两两正交的非零向量$\boldsymbol{y}_1'$, $\boldsymbol{y}_2'$, $\cdots$, $\boldsymbol{y}_n'$. 根据定理 1.32,知$\boldsymbol{y}_1'$, $\boldsymbol{y}_2'$, $\cdots$, $\boldsymbol{y}_n'$线性无关,从而它们形成V^n的一个正交基. 再以$|\boldsymbol{y}_i'|$除$\boldsymbol{y}_i'$ $(i = 1, 2, \cdots, n)$,就得到定理所要求的标准正交基

$$\boldsymbol{y}_i = \frac{1}{|\boldsymbol{y}_i'|}\boldsymbol{y}_i' \quad (i = 1, 2, \cdots, n)$$ 证毕

上面是由基 $\boldsymbol{x}_1, \boldsymbol{x}_2, \cdots, \boldsymbol{x}_n$ 求标准正交基 $\boldsymbol{y}_1, \boldsymbol{y}_2, \cdots, \boldsymbol{y}_n$ 的过程，有时也称为把基 $\boldsymbol{x}_1, \boldsymbol{x}_2, \cdots, \boldsymbol{x}_n$ **正交单位化或正交规范化**.