

# 算法组小萌新线性代数快速入门指南

MicDZ

2023 年 10 月



# 前言

大一新生在没有学习线性代数的情况下入门算法组存在障碍。从头到尾系统学习线性代数需要较长时间。本文期望解决这一痛点。本文主要参考 Introduction to Linear Algebra[3] 和 Linear Algebra and Its Applications[2]。推荐参考 MIT 网课 [1]。

本文项目仓库在 <https://github.com/MicDZ/linear-algebra-quick-start>。

- 本文面向的对象有：大一新生、对线性代数实质不太理解的高年级同学。
- 本文讲解的重点内容有：
  1. 线性代数的基本概念内涵
  2. 线性代数的实质
  3. 线性代数的一般应用
- 本文不会讲解的内容有：
  1. 线性代数的概念的具体定义
  2. 线性代数性质定理的证明
  3. 线性代数的例题
  4. 矩阵论的高级内容

注意：本文将不会是一篇指导完成课业内容的文章。本文更注重立足于实践。旨在帮助你真正“会用”线性代数。

MicDZ  
micdz@hust.edu.cn



# 目录

前言	iii
<b>1 向量及向量的线性组合</b>	<b>1</b>
1.1 向量 . . . . .	1
1.2 向量的线性组合 . . . . .	1
<b>2 矩阵</b>	<b>5</b>
2.1 矩阵的基本概念 . . . . .	5
2.2 线性独立与线性相关 . . . . .	6
2.3 向量空间 . . . . .	7
2.3.1 零空间 . . . . .	8
2.3.2 转置后的空间 . . . . .	9
2.3.3 正交性 . . . . .	10
2.4 行列式 . . . . .	10
<b>3 线性变换</b>	<b>13</b>
3.1 平移 . . . . .	13
3.2 旋转 . . . . .	13
3.3 投影 . . . . .	14
3.3.1 投影到一条直线上 . . . . .	14
3.3.2 投影到子空间中 . . . . .	15
<b>4 线性代数的应用</b>	<b>17</b>
4.1 线性回归 . . . . .	17

4.2	坐标变换 . . . . .	18
4.3	概率论 . . . . .	18

# 第 1 章 向量及向量的线性组合

## 1.1 向量

高中我们都学过向量。我们用向量来表示一个有大小有方向的量。过去，我们用一个“数对”表示向量。

$$\begin{aligned}\vec{a} &= (x_1, y_1) \quad \vec{b} = (x_2, y_2) \\ |\vec{a}| &= \sqrt{x_1^2 + y_1^2} \\ \vec{a} \cdot \vec{b} &= x_1 x_2 + y_1 y_2 \\ \cos \theta &= \frac{\vec{a} \cdot \vec{b}}{||\vec{a}|| \, ||\vec{b}||}\end{aligned}$$

与高中不同，现在我们通常用“列”来表示向量（也有行向量，为方便后文均采用列向量）。例如，在线性代数中，我们将二维的列向量理解为一个两行一列的“矩阵”（矩阵的概念在下节介绍）。向量的模的表示也有所区别，你需要习惯这种表示方法。

$$\begin{aligned}\vec{a} &= \begin{bmatrix} x_1 \\ y_1 \end{bmatrix} \quad \vec{b} = \begin{bmatrix} x_2 \\ y_2 \end{bmatrix} \\ ||\vec{a}|| &= \sqrt{x_1^2 + y_1^2}\end{aligned}$$

## 1.2 向量的线性组合

从这里开始算是正式进入了线性代数的学习。很明显，“线性”是线性代数研究的关键。我们首先介绍一下什么是“线性”。线性的最直观理解就是一

条直线。

$$y = kx + b$$

我们引入线性组合的概念。设有两条直线：

$$y = kx + b \dots (1)$$

$$y = mx + n \dots (2)$$

我们将两条直线进行任意的“线性组合”，很容易发现，所得结果仍为一条线性直线。

$$(1) \times c + (2) \times d$$

$$\begin{aligned} y &= c(kx + b) + d(mx + n) \\ &= (ck + dm)x + (cb + dn) \end{aligned}$$

其实这种特性可以用“线性运算封闭性”高度概括。换言之，将直线方程乘一个数，或者将两条直线方程相加，所得结果仍为一条直线。这是理解后续“线性空间”的重要基础。

如果将上述过程用我们刚才所学的向量表示，那么可以写成（暂时理解为一种记号，但其实有其几何含义）：

$$\begin{aligned} \vec{u} &= \begin{bmatrix} k \\ b \end{bmatrix} & \vec{v} &= \begin{bmatrix} m \\ n \end{bmatrix} \\ c\vec{u} + d\vec{v} &= \begin{bmatrix} ck + dm \\ cb + dn \end{bmatrix} \end{aligned}$$

我们就将  $c\vec{u} + d\vec{v}$  称为向量之间的**线性组合**。这里，我们也可以发现二维向量之间的线性组合所得结果仍然可以是一个二维向量。

如果我们将情况拓展到三维。看下面一个例子：

$$\vec{u} = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} \quad \vec{v} = \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix} \quad \vec{w} = \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}$$



通过这三个向量的线性组合，我们可以得到整个三维空间内的所有向量。因为对于任意一个三维向量我们都可以拆分到这三个分量上。

$$\begin{bmatrix} a \\ b \\ c \end{bmatrix} = a\vec{u} + b\vec{v} + c\vec{w}$$

但是，是否任意给定三个向量都可实现通过三个向量之间的线性组合得到三维空间内的所有向量呢？请你先思考一下，这将会在后面介绍。

$$\vec{u} = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} \quad \vec{v} = \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix} \quad \vec{w} = \begin{bmatrix} 0 \\ 2 \\ 0 \end{bmatrix}$$



## 第2章 矩阵

### 2.1 矩阵的基本概念

矩阵是线性代数最基础的概念。后面我们所有的讨论也围绕矩阵展开，但你不用太担心，现在你需要做的是将矩阵看做一个“记号”，其本质和 xyz 这种变量没有区别。

假定我现在有三个向量，我需要对他们进行线性组合。

$$\vec{u} = \begin{bmatrix} 1 \\ -1 \\ 0 \end{bmatrix} \quad \vec{v} = \begin{bmatrix} 0 \\ 1 \\ -1 \end{bmatrix} \quad \vec{w} = \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}$$

$$x_1\vec{u} + x_2\vec{v} + x_3\vec{w} = \begin{bmatrix} x_1 \\ x_2 - x_1 \\ x_3 - x_2 \end{bmatrix}$$

$x_1\vec{u} + x_2\vec{v} + x_3\vec{w}$  这个写法未免太复杂，我们用“矩阵”将这个式子改写一下：

$$A\mathbf{x} = \begin{bmatrix} 1 & 0 & 0 \\ -1 & 1 & 0 \\ 0 & -1 & 1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} x_1 \\ x_2 - x_1 \\ x_3 - x_2 \end{bmatrix} = \begin{bmatrix} b_1 \\ b_2 \\ b_3 \end{bmatrix} = \mathbf{b}$$

这里用了**矩阵的乘法**，其本质是“行列（点）积”，也就是将前一个矩阵中的每一个行向量与后一个矩阵中的列向量求点积，然后将点积所得结果放在对应的位置。

在这里因为后面的矩阵只有一个列向量，那么就可以将矩阵乘法这么

理解

$$A\mathbf{x} = \begin{bmatrix} 1 & 0 & 0 \\ -1 & 1 & 0 \\ 0 & -1 & 1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} (1, 0, 0) \cdot (x_1, x_2, x_3) \\ (-1, 1, 0) \cdot (x_1, x_2, x_3) \\ (0, -1, 1) \cdot (x_1, x_2, x_3) \end{bmatrix}$$

那么就可以得到一个非常简单的写法，我们将它称为**线性方程组**：

$$A\mathbf{x} = \mathbf{b}$$

如果我们已知  $A$  和  $\mathbf{b}$ ，要求得  $\mathbf{x}$  怎么做呢？看上去很简单，只需要

$$\mathbf{x} = A^{-1}\mathbf{b}$$

这里  $A^{-1}$  就被称为**矩阵的逆**。这里也只是一个记号，并不代表实际就是求“负一次幂”，实际计算参考教材。

但实际情况并不会这么简单， $\mathbf{x}$  是否有解必然是与  $A$  和  $\mathbf{b}$  有关的。下面一节我们讨论这种相关性。

## 2.2 线性独立与线性相关

还是考虑刚才的例子，我们再添加一个向量，记为  $\vec{w}^*$ 。

$$\vec{u} = \begin{bmatrix} 1 \\ -1 \\ 0 \end{bmatrix} \quad \vec{v} = \begin{bmatrix} 0 \\ 1 \\ -1 \end{bmatrix} \quad \vec{w} = \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} \quad \vec{w}^* = \begin{bmatrix} -1 \\ 0 \\ 1 \end{bmatrix}$$

我们可以发现， $\vec{w}^* = -\vec{u} - \vec{v}$ ，也就是说  $\vec{w}^*$  可以由  $\vec{u}$  和  $\vec{v}$  线性组合得来。那么我们称  $\vec{w}^*, \vec{u}, \vec{v}$  **线性相关**。相反地  $\vec{w}$  不可以由  $\vec{u}$  和  $\vec{v}$  线性组合得来，那么我们称  $\vec{w}, \vec{u}, \vec{v}$  **线性独立**。

我们在这里尝试在空间中理解这个事实。

$\vec{u}, \vec{v}, \vec{w}$  线性独立，在空间中的具体体现就是  $\vec{w}, \vec{u}, \vec{v}$  不会在同一个平面内。如图2.1。

而  $\vec{w}^*, \vec{u}, \vec{v}$  线性相关，那么这三个向量将会在一个平面上，如图2.2。现在我们考虑这个方程的解

$$A\mathbf{x} = \mathbf{b}$$

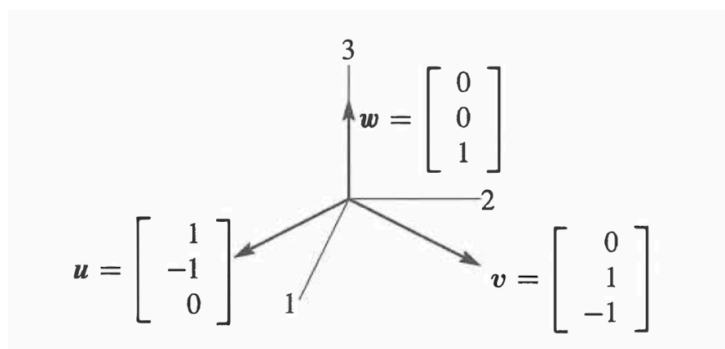


图 2.1: 不共面的三个向量

$A$  表示的就是  $\vec{u}, \vec{v}, \vec{w}$  这一组向量。 $A\mathbf{x}$  就是将这一组向量进行线性组合。

显然，对于  $\vec{u}, \vec{v}, \vec{w}$  这一组向量，无论  $\mathbf{b}$  落在空间内的哪一个点，都可以解出对应的  $\mathbf{x}$ ；而对于  $\vec{u}, \vec{v}, \vec{w}^*$  这一组向量，只有当  $\mathbf{b}$  落在他们所张成的平面上时， $\mathbf{x}$  才会有解。

## 2.3 向量空间

还是以刚才四个向量为例

$$\vec{u} = \begin{bmatrix} 1 \\ -1 \\ 0 \end{bmatrix} \quad \vec{v} = \begin{bmatrix} 0 \\ 1 \\ -1 \end{bmatrix} \quad \vec{w} = \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} \quad \vec{w}^* = \begin{bmatrix} -1 \\ 0 \\ 1 \end{bmatrix}$$

$\vec{u}, \vec{v}, \vec{w}$  这一组向量，张成了一个三维的空间，那么我们就称这个三维空间为三维的**向量空间**，记为  $C(A) = \mathbf{R}^3$ 。同理  $\vec{u}, \vec{v}, \vec{w}^*$  这一组向量，张成了一个二维的空间，我们称为二维的向量空间，记为  $\mathbf{R}^2$ 。

那么我们将刚才的可解性问题转化一下术语，就可以这样表述：

$A\mathbf{x} = \mathbf{b}$  当且仅当  $\mathbf{b}$  属于  $A$  的向量空间内时有解。（需要自己理解一下什么叫做“属于”）

这里我们就可以将**矩阵的秩**有一个直观的理解了，矩阵的秩就是组成矩阵的向量组所张成的向量空间的维数。

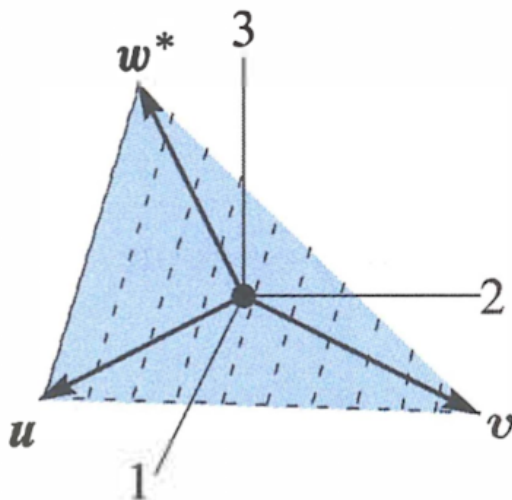


图 2.2: 共面的三个向量

### 2.3.1 零空间

上一节，我们能够直观理解线性方程组有解和无解的条件，但是究竟如何解这个方程呢，我们从这一小节开始讨论。

为了解  $A\mathbf{x} = \mathbf{b}$  这一方程，我们引入了“零空间”的概念。简单来说，**零空间**是  $A\mathbf{x} = \mathbf{0}$  的解空间。为什么叫做零“空间”呢？假设有一个解  $\mathbf{x}$  满足  $A\mathbf{x} = \mathbf{0}$ ，那么对  $\mathbf{x}$  点乘任何数有  $A(c\mathbf{x}) = cA\mathbf{x} = c\mathbf{0} = \mathbf{0}$ ，仍为  $\mathbf{0}$ 。也就是说，任何满足  $A\mathbf{x} = \mathbf{0}$  的解经过线性组合后，仍为这个方程的解。我们设其中一个解为  $\mathbf{x}_n$ 。记这个空间为  $N(A)$ 。

同样的对于  $A\mathbf{x} = \mathbf{b}$  这个方程，如果有某一个解  $\mathbf{x}_p$ ，那么  $\mathbf{x}_p + \mathbf{x}_n$  也一定是其解。

那么我们就可以用  $\mathbf{x}_p + \mathbf{x}_n$  表示  $A\mathbf{x} = \mathbf{b}$  的所有解。可以证明所有的解都可用这个式子表示。

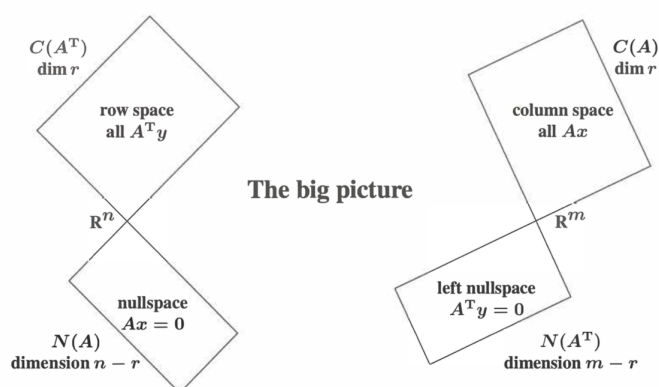


图 2.3: 四个空间之间的相互关系

### 2.3.2 转置后的空间

把  $m \times n$  矩阵  $A$  的行换成同序数的列得到一个  $n \times m$  矩阵，此矩阵叫做  $A$  的转置矩阵，记为  $A^T$ 。

例如

$$A = \begin{bmatrix} 1 & 2 & 0 \\ 3 & -1 & 4 \end{bmatrix} \quad A^T = \begin{bmatrix} 1 & 3 \\ 2 & -1 \\ 0 & 4 \end{bmatrix}$$

我们就有了四个空间：

1.  $C(A)$ : 行空间
2.  $C(A^T)$ : 列空间
3.  $N(A)$ : 零空间
4.  $N(A^T)$ : 左零空间

这四个空间之间显然存在一些相关关系，期望读者自行学习一下。

### 2.3.3 正交性

对于两个向量而言

$$\mathbf{u} = \begin{bmatrix} 1 \\ -1 \end{bmatrix} \quad \mathbf{v} = \begin{bmatrix} -1 \\ 1 \end{bmatrix}$$

$$\mathbf{u}^\top \mathbf{v} = 0$$

利用矩阵乘法的性质，我们可以用  $\mathbf{u}^\top \mathbf{v}$  表示两个向量的点乘，点乘的值为 0 那么两个向量就正交。

如果是两个矩阵正交：可以采用同样的定义  $\mathbf{A}^\top \mathbf{B} = \mathbf{0}$ 。

线性代数的理论部分大致就到此结束。

## 2.4 行列式

很多线性代数开篇就是介绍行列式的计算，但行列式本身的意义其实必须要在能够理解向量空间后能够理解。所以本文将行列式的介绍放在了一个较后的位置。

对于一个  $n \times n$  的矩阵，行列式的几何含义是：

定义在  $n$  维空间下一组单位正交基所构成的体积为 1。

一个矩阵的行列式就等于，这个  $n$  维矩阵的列（或行）向量组所张成的体积。

我们看一个非常简单的例子，如果是一个  $3 \times 3$  的矩阵：

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

这一组列向量显然是将  $y$  轴拉伸了 2 倍，张成的体积就是标准单位正交基张成空间的两倍。自然其行列式的值就是 2。

现实中的向量组不会像这里的都是正交的，遇到非正交的向量组很难再通过求解体积的方式来计算行列式，所以你还是需要掌握行列式的求解方法。



通过这个定义，你就可以明白，为什么必须是方阵才能计算行列式，为什么不满秩的方阵其行列式一定为 0（两个问题都请读者思考一下）。



## 第3章 线性变换

线性变换是线性代数最常见的应用，本文将线性代数单列为一章，希望可以引起读者的重视。

在正式开始讲具体的变换之前，希望读者能从宏观上对“变换”有一个理解。我们往往将变换称为“映射”，对应的是将一个集合映射到另一个集合。函数就是一种最简单的变换。后文我们要讨论的变换也与此类似。

我们暂时先用  $T$  来表示这种变换，其实矩阵乘法就是一种映射（想想为什么）。线性变换对变换本身有要求，与此前的线性运算封闭非常类似

$$T(\mathbf{u} + \mathbf{v}) = T(\mathbf{u}) + T(\mathbf{v})$$

$$T(c\mathbf{u}) = cT(\mathbf{u})$$

### 3.1 平移

平移明显不是一个线性变换（想想为什么）。平移可以直接利用向量的加法实现

$$T(\mathbf{w}) = \mathbf{w} + \mathbf{u}_0$$

### 3.2 旋转

在这里不讨论旋转矩阵的具体细节，感兴趣的同学可以参考??。

两个向量的点积在被旋转后保持不变。记旋转矩阵为  $\mathbf{M}$ 。

$$\mathbf{a}^\top \cdot \mathbf{b} = (\mathbf{M}\mathbf{a})^\top \cdot \mathbf{M}\mathbf{b} = \mathbf{a}^\top (\mathbf{M}^\top \mathbf{M}) \mathbf{b}$$

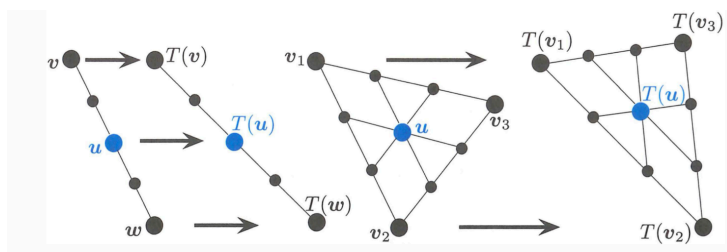


图 3.1: 旋转矩阵的作用效果

所以可得出

$$\mathbf{M}\mathbf{M}^T = \mathbf{I}$$

### 3.3 投影

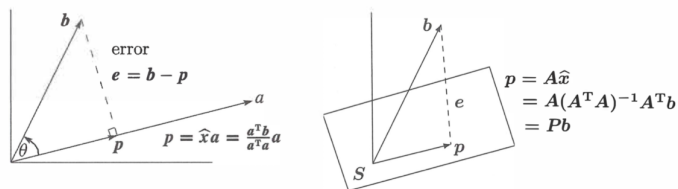
投影是线性代数中很重要的一个概念，巧妙利用投影可以实现很多有意思的应用。

#### 3.3.1 投影到一条直线上

我们先考虑将向量投影到一条直线上的情况。 $\mathbf{b}$  是原向量， $\mathbf{a}$  是投影的方向。 $\mathbf{p}$  是投影向量。

投影的定义可以简单概括为，在  $\mathbf{a}$  方向上找到一个向量  $\mathbf{p}$  使其距离  $\mathbf{b}$  最近。

因为  $\mathbf{p}$  在  $\mathbf{a}$  方向上，所以  $\mathbf{p}$  可以表示为  $\mathbf{a}\hat{x}$ 。 $\mathbf{p}$  和  $\mathbf{b}$  的“距离”可以用  $\mathbf{e} = \mathbf{b} - \mathbf{p}$  表示。

图 3.2: 将  $\mathbf{b}$  投影到  $\mathbf{a}$  方向上

当  $\mathbf{e}$  和  $\mathbf{a}$  正交时这个“距离”最小。

$$\begin{aligned}
\mathbf{e} \cdot \mathbf{a} &= 0 \\
(\mathbf{b} - \mathbf{p}) \cdot \mathbf{a} &= 0 \\
(\mathbf{b} - \mathbf{a}\hat{x}) \cdot \mathbf{a} &= 0 \\
\hat{x} &= \frac{\mathbf{a} \cdot \mathbf{b}}{\mathbf{a} \cdot \mathbf{a}} = \frac{\mathbf{a}^\top \mathbf{b}}{\mathbf{a}^\top \mathbf{a}}
\end{aligned}$$

因而  $\mathbf{p} = \mathbf{a}\hat{x} = \mathbf{a} \frac{\mathbf{a}^\top \mathbf{b}}{\mathbf{a}^\top \mathbf{a}}$  ,

我们设  $P$  为投影矩阵, 可以实现将一个向量  $\mathbf{b}$  投影到我们所需要的空间中, 记为  $P\mathbf{b}$ 。

$$P\mathbf{b} = \mathbf{p} = \mathbf{a} \frac{\mathbf{a}^\top \mathbf{b}}{\mathbf{a}^\top \mathbf{a}}$$

故将向量投影到  $\mathbf{a}$  方向的投影矩阵为  $P = \frac{\mathbf{a}\mathbf{a}^\top}{\mathbf{a}^\top \mathbf{a}}$  。

### 3.3.2 投影到子空间中

问题可以转换为

找到一个向量组合  $\mathbf{p} = \hat{x}_1 \mathbf{a}_1 + \cdots + \hat{x}_n \mathbf{a}_n$  , 使得距离  $\mathbf{b}$  最近。

与刚才的操作类似, 不同的是, 所得的  $\mathbf{e}$  需要与所有的  $\mathbf{a}_i$  垂直。

$$\begin{aligned}
\mathbf{a}_1^\top (\mathbf{b} - A\hat{\mathbf{x}}) &= 0 \\
&\dots \\
\mathbf{a}_n^\top (\mathbf{b} - A\hat{\mathbf{x}}) &= 0
\end{aligned}$$

写成一个矩阵的形式:

$$\begin{aligned}
A^\top (\mathbf{b} - A\hat{\mathbf{x}}) &= \mathbf{0} \quad \text{or} \quad A^\top \mathbf{b} = A^\top A\hat{\mathbf{x}} \\
\mathbf{p} = A\hat{\mathbf{x}} &= A(A^\top A)^{-1} A^\top \mathbf{b}
\end{aligned}$$

就可以得到投影矩阵:

$$P = A(A^\top A)^{-1} A^\top$$

这样的投影矩阵在工程中应用非常广泛。



## 第 4 章 线性代数中的应用

### 4.1 线性回归

在前文的讨论中，我们经常面临  $A\mathbf{x} = \mathbf{b}$  无解的问题。一个常见的例子是，一个平面内给定三个点，若这三个点不共线的话，无法解出一条直线经过这三点。但我们有“最小二乘法”，可以利用最小化损失的方式“拟合”出一条最合适的“解”。

线性代数工具可以将平面上的回归拓展到  $n$  维空间。

我们将  $\mathbf{b}$  分解，一部分为  $\mathbf{b}$  向  $A$  的列空间的投影  $\mathbf{p}$ ，另一部分为  $\mathbf{e}$ 。显然有  $\mathbf{e} = \mathbf{b} - \mathbf{p}$ 。 $A\mathbf{x} = \mathbf{b}$  无解，但是  $A\hat{\mathbf{x}} = \mathbf{p}$  有解。

将向量的模长平方可以得到

$$\|A\mathbf{x} - \mathbf{b}\|^2 = \|A\mathbf{x} - \mathbf{p}\|^2 + \|\mathbf{e}\|^2$$

由图4.1，可以看出  $\mathbf{b}$  不在空间  $A$  中，这也是为什么  $A\mathbf{x} = \mathbf{b}$  无解。但

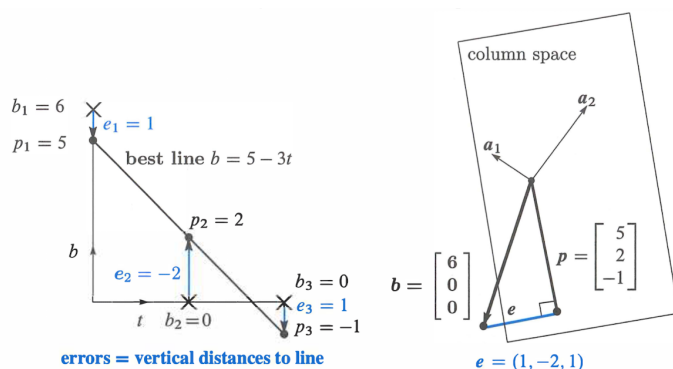


图 4.1: 线性回归问题的两种表述方式

我们可以通过投影实现最小化  $\mathbf{e} = \mathbf{b} - A\hat{\mathbf{x}}$ 。

我们也可以对  $E = \|A\mathbf{x} - \mathbf{b}\|$  求偏导，使偏导值为零，可以得到当  $A^T A \hat{\mathbf{x}} = A^T \mathbf{b}$  偏导值为  $\mathbf{0}$ 。我们发现，这个结果与投影形式是一致的。

至此， $\hat{\mathbf{x}}$  就是我们想要得到的结果

$$\hat{\mathbf{x}} = (A^T A)^{-1} A^T \mathbf{b}$$

读者可以尝试在二维条件下验证一下这个公式是否与最小二乘法公式吻合。

## 4.2 坐标变换

坐标变换是工程中非常常用的操作，对于两个坐标系之间的变换，我们可以拆分为旋转和平移两个操作。

如下操作，我们实现了将坐标  $\mathbf{v}$  转换到  $\mathbf{u}$

$$\mathbf{u} = R\mathbf{v} + \mathbf{t}$$

其中， $R$  是前文所提到过的旋转矩阵， $\mathbf{t}$  为前文提到的平移向量。

通常我们用增广矩阵的工具，使这个式子更加简结

$$\begin{bmatrix} \mathbf{u} \\ 1 \end{bmatrix} = \begin{bmatrix} R & \mathbf{t} \end{bmatrix} \begin{bmatrix} \mathbf{v} \\ 1 \end{bmatrix}$$

## 4.3 概率论

首先介绍一下概率论中我们的几个基本概念。

- 期望： $\mathbf{E}[\mathbf{x}] = \sum_{i=1}^n p_i x_i$
- 样本方差： $\mathbf{S}[\mathbf{x}] = \frac{1}{n-1} \sum_{i=1}^n (x_i - m)^2$ ，其中  $m$  是样本的平均值。
- 方差： $\sigma^2 = \mathbf{E}[(\mathbf{x} - m)^2] = \sum_{i=1}^n p_i (x_i - m)^2$
- 协方差： $\mathbf{E}[(X - \mathbf{E}X)(Y - \mathbf{E}Y)]$



协方差可能会是一个较陌生的概念，从直观上来看，协方差表示的是两个变量总体误差的期望。我们可以用协方差来表示两个变量之间的相关性。协方差为 0 的两个随机变量称为是不相关的。

当变量多于两个时，我们计算变量间的协方差可以借助线性代数进行。

$$\mathbf{S} = \mathbf{E}[(\mathbf{X} - \mathbf{EX})(\mathbf{X} - \mathbf{EX})^{\top}]$$

我们就称  $\mathbf{S}$  为协方差矩阵。

还是以二维的两个随机变量  $X, Y$  为例

$$\begin{aligned} \mathbf{S} &= \mathbf{E}\left[\begin{bmatrix} X - \mathbf{EX} \\ Y - \mathbf{EY} \end{bmatrix} \begin{bmatrix} X - \mathbf{EX} & Y - \mathbf{EY} \end{bmatrix}\right] \\ &= \begin{bmatrix} \mathbf{E}(X - \mathbf{EX})(X - \mathbf{EX}) & \mathbf{E}(X - \mathbf{EX})(Y - \mathbf{EY}) \\ \mathbf{E}(Y - \mathbf{EY})(X - \mathbf{EX}) & \mathbf{E}(Y - \mathbf{EY})(Y - \mathbf{EY}) \end{bmatrix} \end{aligned}$$

可以看出，协方差矩阵是一个对称矩阵，对角线上是所有变量的方差，其余位置为两两的协方差。通过协方差矩阵，我们可以很轻松地显示出变量之间的相关性。



## 参考文献

- [1] 麻省理工学院 - mit - 线性代数 - bilibili. <https://www.bilibili.com/video/BV16Z4y1U7oU/>.
- [2] David C. Lay. *Linear Algebra and Its Applications*. 2003.
- [3] Gilbert Strang. *Introduction to Linear Algebra*. Wellesley, MA, 2009.