前言

这篇序言的一个目标可以马上实现。您需要了解麻省理工学院(MIT)线性代数课程 Math 18.06 的视频讲座。这些视频与本书配套，并且是麻省理工学院开放课程的一部分。关于线性代数的直接链接是

https://ocw.mit.ed/curse/8-6-linear-algebra-pring-2010/
https://ocw.mit.edu/courses/18-06sc-linea-algebra-fall-2011/

在 YouTube 上，这些讲座的链接是 https://ocw.mit.edu/1806videos 和 /1806scvideos 。

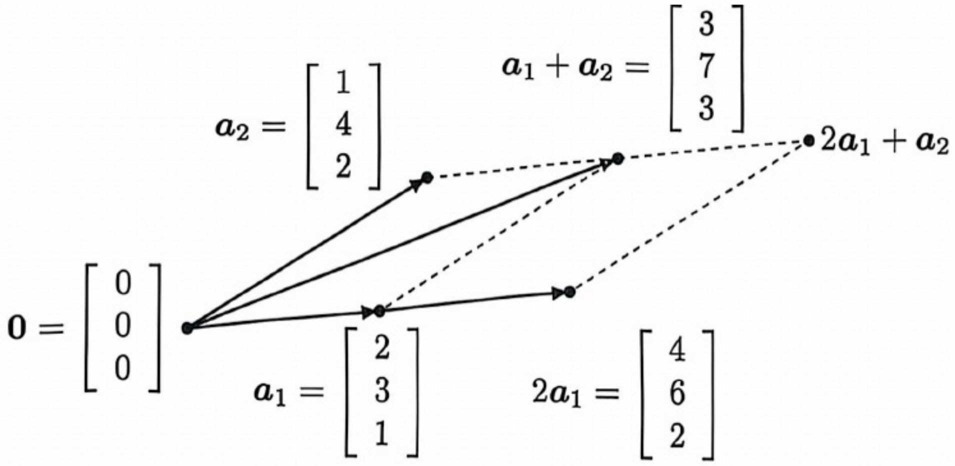
第一个链接带来了开放课程资源诞生之初的原版讲座。在 2011 年的新讲座中添加了研究生的问题解决方案(非常出色)，还有对线性代数的简短介绍。如今这门课程有了新的开端——线性无关性的关键思想以及矩阵的列空间已移至前面。

我想在这篇序言中跟您讲讲那些想法。

从两个列向量 a1 和 a2 开始。它们各自可以有三个分量，因此它们对应于三维空间中的点。该图形需要一个中心点，该中心点位于零向量处：

$$\boldsymbol{a}_1 = \begin{bmatrix} 2 \\ 3 \\ 1 \end{bmatrix} \qquad \boldsymbol{a}_2 = \begin{bmatrix} 1 \\ 4 \\ 2 \end{bmatrix} \qquad \textbf{zero vector} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}.$$

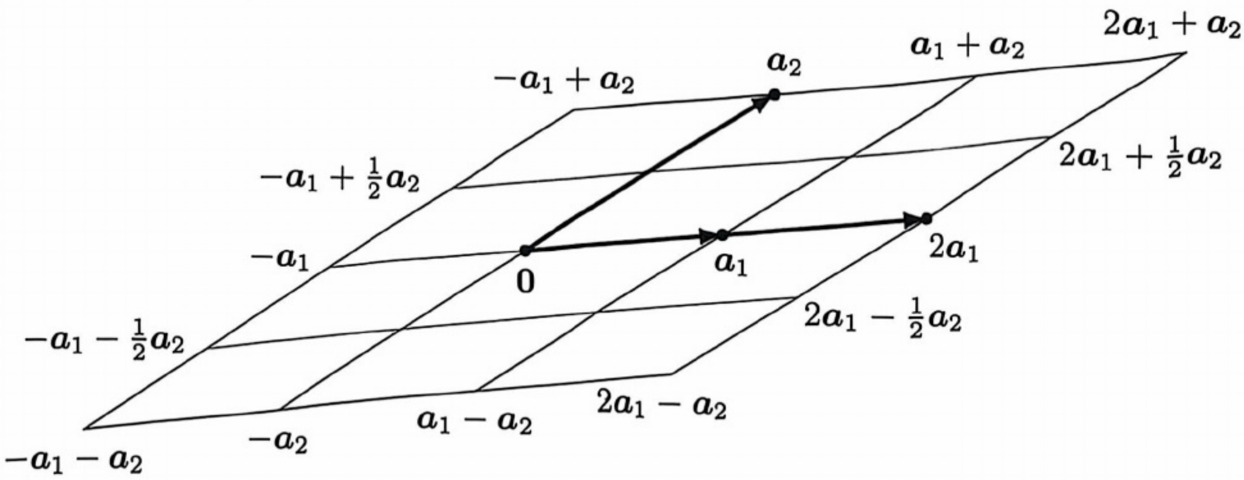
这些向量绘制在这个二维页面上。但我们都可以想象出三维图像。这里有 a1、a2、2a1 以及向量之和 a1 + a2 。



这幅图展示了两个基本运算——向量的加法 a1 + a2 和将向量乘以2 。将这些运算结合起来就得到了一个“线性组合” 2a1 + a2：

$$\boxed{\textbf{Linear combination} = ca_1 + da_2 \;\; \textbf{for any numbers } c \textbf{ and } d}$$

那些数字 c 和 d 可以是负数。在这种情况下，cai 和 daz 的方向会反转：它们从右向左移动。同样重要的是，c 和 d 可以包含分数。这里有一幅展示更多线性组合的图片。最终，我们想要所有的向量 ca1 + daz 。



这就是关键！组合 ca1 + da2 填满了一个整个平面。这是三维空间中的一个无限平面。通过使用越来越多的分数和小数 c 和 d，我们填满了一个完整的平面。平面上的每一点都是 a1 和 a2 的组合。

现在，线性代数中有一个基本概念：矩阵。矩阵 A 包含 n 个列向量 a1、a2……an。在这个例子中，我们的矩阵有两个列向量 a1 和 a2，它们都是三维空间中的向量。所以这个矩阵有三行两列。

3 by 2 matrix
$m = $ **3 rows**
$n = $ **2 columns**

$$A = \begin{bmatrix} & & \\ a_1 & & a_2 \\ & & \end{bmatrix} = \begin{bmatrix} 2 & 1 \\ 3 & 4 \\ 1 & 2 \end{bmatrix}$$

这两列向量的组合在三维空间中构成了一个平面。这个平面有一个自然的名称，矩阵的列空间。对于任何 A 矩阵，A 的列空间包含了所有列的组合。

这是到目前为止介绍的四个观点，您将在第一章中全面的看到它们。

1. 三维中的列向量 a1 和 a2

2. 这些向量的线性组合 ca1 + da2

3. 包含列向量 a1 和 a2的矩阵A 。

4. 矩阵的列空间 = 该列的所有线性组合 = 平面

现在我们在 A 中又增加了两列。
这四个列向量处于三维空间中。

$$A = \begin{bmatrix} 2 & 1 & 3 & 0 \\ 3 & 4 & 7 & 0 \\ 1 & 2 & 3 & -1 \end{bmatrix}$$

线性代数旨在理解每一个列空间。让我试试这个。

第 1 列和第 2 列产生相同的平面(相同的 a1 和 a2)

第三列没有提供任何新的信息，因为 a3在该平面上:a3=a1 + a2

第 4 列不在平面上：加入 c4a4 会使平面升高或降低

这个矩阵 A 的列空间是整个三维空间：所有的点!

您可以看到我们是如何一次处理一列，从左到右的。每一列可以独立于前一列，也可以是对前几列的组合。要生成三维空间中的每一个点，您需要三个独立的列。

矩阵乘法 A = CR

使用“线性组合”和“独立列”这两个词，可以很好地描述那个 3 行 4 列的矩阵 A。第 3 列是第 1 列和第 2 列的线性组合。第 1 列、第 2 列、第 4 列是独立的。要通过独立列 1、2、4 的组合得到零向量，唯一的办法是将所有这些列都乘以零。

我们离第一章的一个关键思想已经很近了，我必须继续下去。矩阵乘法是我们所知道可以解决问题的完美方法。从 A 的 4 列中，我们选取独立列 a1、a2、a4组成矩阵C。R 的每一列都告诉我们 C 中 a1、a2、a4 的组合如何产生 A 的一列。A 等于 C 乘以 R：

$$A = \begin{bmatrix} 2 & 1 & \mathbf{3} & 0 \\ 3 & 4 & \mathbf{7} & 0 \\ 1 & 2 & \mathbf{3} & -1 \end{bmatrix} = \begin{bmatrix} \mathbf{2} & \mathbf{1} & 0 \\ \mathbf{3} & \mathbf{4} & 0 \\ \mathbf{1} & \mathbf{2} & -1 \end{bmatrix} \begin{bmatrix} 1 & 0 & \mathbf{1} & 0 \\ 0 & 1 & \mathbf{1} & 0 \\ 0 & 0 & \mathbf{0} & 1 \end{bmatrix} = CR$$

A 的第 3 列由第 1 列和第 2 列决定，R 的第 3 列展示了其方式。将 C 的独立列 1 和2 相加，得到 A 的第 3 列 a3 = a1 + a2 = (3, 7, 3) 。

矩阵乘法：CR 的每一列 j 都是 C 乘以 R 的第 j 列

本书的第 1.3 节将介绍矩阵与向量相乘(两种方法)。然后第 1.4 节将介绍矩阵与矩阵相乘。这是线性代数中的关键运算。重要的是，这种乘法运算不止一种好的方法。

我就说到这里。前言的通常目的是向您介绍全貌。接下来的几页将为您介绍两种规划学习该学科的方式——尤其是前七章，它们的内容足以充实大多数线性代数课程。然后是可选章节，接着是当今应用中最活跃的主题：深度学习。

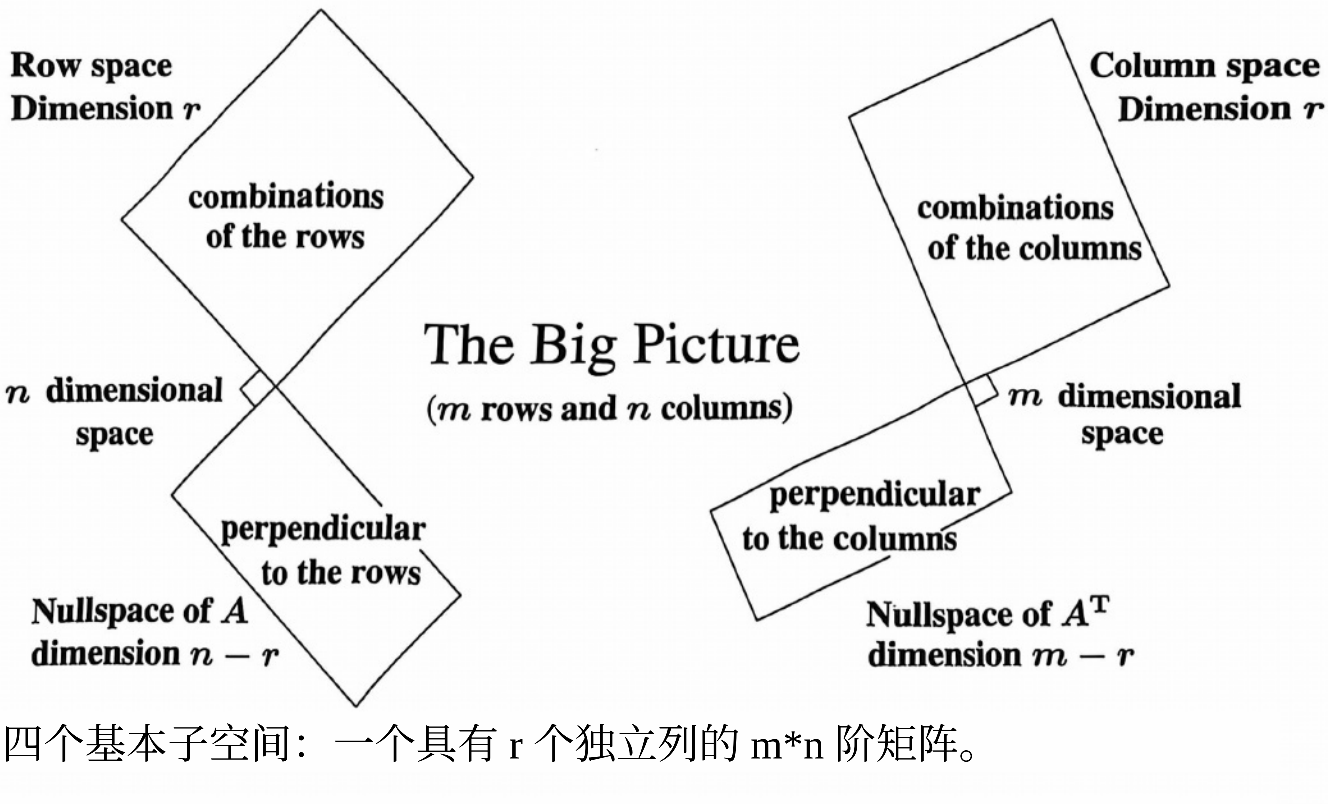
四个基本子空间

您刚刚看到了本课程是如何从矩阵 A 的列开始展开的。有两个关键步骤。其中一个步骤是取列的所有组合 cai + daz + eas + fan 。这得到了 A 的列空间。另一个步骤是将矩阵分解为 C 乘以 R 。矩阵 C 包含了一组完整的独立列。

我充分意识到这只是这本书的序言。你对矩阵的列空间还毫无经验(对 C 和 R 的经验更少)。但好消息是：这些是正确的开始方向。最终，每个矩阵都会引出四个基本空间。与 A 的列空间相伴的是行空间——所有行的组合。当我们对 n 个列向量的所有组合和 m 个行向量的所有组合进行组合时——这些组合会填满向量的“空间”。

另外两个子空间完成了这幅图景。假设行空间是三维中的一个平面。那么在三维图景中存在一个特殊方向——这个方向垂直于行空间。那条垂直线就是矩阵的零空间。我们会看到零空间中的向量(垂直于所有行)满足 Ax = 0：这是线性方程中最基本的情况。

如果垂直于所有行的向量很重要，那么垂直于所有列的向量也同样重要。以下是四个基本子空间的图示。



四个基本子空间：一个具有 r 个独立列的 m*n 阶矩阵。

这张关于四个子空间的图出现在第三章。垂直空间的概念在第四章中得到发展。而在第七章中发现了所有四个子空间的特殊“基向量”。这是线性代数基本定理的最后一部分。该定理包含了关于任何矩阵(方阵或矩形矩阵)的一个惊人事实：

独立列的数量等于独立行的数量。

矩阵的五种分解

以下是线性代数的组织原则。当我们的矩阵具有特殊性质时，这些分解将体现出来。一章又一章之后，它们以直接且有用的方式表达了关键思想。

沿着列表往下看，其用处会逐渐增加。正交矩阵最终胜出，因为它们的列是垂直的单位向量。这就是完美。

以下是来自第 1 ,2 , 4 , 6 , 7 章的五个分解:

1 A = CR = R 将 C 中的独立列组合起来，以给出 A 的所有列
2 A = LU = 下三角 L 乘上三角 U
4 A = QR = 正交矩阵 Q 乘以上三角矩阵 R
6 S = QAQT = (正交 Q)(A 的特征值)(正交 QT)
7 A = UΣVT = (正交矩阵 U)(Σ 中的奇异值)(正交矩阵 VT)

我想提醒您注意最后一个——奇异值分解(SVD)。它适用于任何矩阵 A。那些因子 U 和 V 的列是垂直的——且长度均为 1。用 U 或 V 乘以任何向量都会得到一个长度相同的向量——因此计算不会增大或减小。而 Σ 是一个由“奇异值”组成的正对角矩阵。如果您在第六章中学习了特征值和特征向量，请翻到几页后的第七章第一节继续了解奇异值。

深度学习

要真实地展现线性代数，就必须将其应用包括在内。完整性是完全不可能的。此刻，应用数学的主导方向有一个特殊的要求：它不能完全是线性的！

这个方向的一个名称是“深度学习”。它是一种极其成功的解决一个基本科学问题的方法：从数据中学习。在很多情况下，数据以矩阵的形式出现。我们的目标是观察矩阵内部变量之间的联系。我们不是求解矩阵方程或微分方程来表示已知的输入输出规则，而是必须找到这些规则。深度学习的成功在于构建一个函数 F(x，v)，其输入为两种类型的 x 和 v:

向量 v 描述了训练数据的特征。
矩阵 z 为这些特征分配权重。
对于该训练数据，函数 F(x，v)接近正确的输出。
当 v 变为没见过的测试数据时，F(x, u)仍接近正确值。

这种成功部分源于学习函数 F 的形式，这种形式使其能够包含大量的数据。最终，线性函数 F 会完全不够用。对于 F 而言，分段线性函数是最佳选择。它将简单性与通用性结合起来。

书中的应用及网站上的应用

我希望这本书在线性代数课程结束后仍能对你有所帮助。正是线性代数在各个方面的应用使这成为可能。矩阵承载数据，而其他矩阵对数据进行运算。目标是“透视矩阵”，通过理解其特征值和特征向量以及奇异值和奇异向量来实现。而且每种应用都有特殊的矩阵——这里列举四个例子：

Markov matrices M	Each column is a set of probabilities adding to 1.
Incidence matrices A	Graphs and networks start with a set of nodes. The matrix A tells the *connections* (edges) between those nodes.
Transform matrices F	The Fourier matrix uncovers the *frequencies* in the data.
Covariance matrices C	The **variance** is key information about a random variable. *The covariance explains dependence between variables.*

在本第六版中，我们纳入了上述应用以及更多内容。对于深度学习中矩阵权重关键计算，第 9 章介绍了优化的思想。这是线性代数与微积分的交汇之处：由于 F(z) 存在众多变量，在最小值点处，导数等于零会变成一个矩阵方程。

第五版中的几个主题虽已放弃其位置，但并未失去其重要性。这些部分只是转移到了网络上。第六版新版的网站是

math.mit.edu/linearalgebra

该网站包含了这一新版的示例章节以及所有问题集的答案。这些章节(及更多内容)是从第五版中在线保存的：

Fourier Series	**Norms and Condition Numbers**
Iterative Methods and Preconditioners	**Linear Algebra for Cryptography**

这里有一个关于线性代数的小小提示——在这门课程深入展开之前的三个小问题：

1. 假设你在这张纸上画出三条长度分别为 r、s 和 t 的直线段。要使这三个线段能构成一个三角形，这三个长度应满足什么条件？在这个问题中，你可以选择这三条线的方向。

2. 现在假设三条直线 u、v、w 的方向是固定且不同的。但你可以将这三条直线拉伸为 au、bv、cw，其中 a、b、c 是任意数。你能总是用这三个向量 au、bv、cw 构成一个封闭三角形吗？

3. 线性代数并非局限于平面！假设在三维空间中有四条不同方向的直线 u、v、w、z。能否总是选择出不全为零的数 a、b、c、d 使得 au + bv + cw + dz = 0？