目录

线性方程组

1.概念

包含未知数$x_1, x_2, \dots, x_n$的一个**线性方程**是形如

$$a_1x_1 + a_2x_2 + \cdots + a_nx_n = b$$

的方程，其中b与系数$a_1, a_2, \dots, a_n$是实数或复数，通常是已知数。

而**线性方程组**是由一个或多个包含相同变量$x_1, x_2, \dots, x_n$的线性方程组成的。例如

$$\begin{cases} 2x_1 - x_2 + 1.5x_3 = 8 \\ \quad x_1 - \ 4x_3 = -7 \end{cases}$$

线性方程组的一组解是一组数$(s_1, \dots, s_n)$，用这组数分别代替$x_1, x_2, \dots, x_n$时所有方程的两边相等。例如(5,6.5,3)代入方程组变成等式$8 = 8$和$7 = -7$.

方程组所有可能的解的集合称为线性方程组的**解集**。若是两个线性方程组有相同的解集，则它们等价。

求包含两个未知数的两个方程组成的方程组的解，等价于求两条直线的交点，例如

$$\begin{cases} x_1 - 2x_2 = -1 \\ -x_1 + 3x_2 = 2 \end{cases}$$

这两个方程的图形都是直线，两条直线可能交于一点、不相交(平行)、重合，分别对应于一

个解、无解、无穷多解的情况。

2.矩阵

2.1 矩阵引入

一个线性方程组包含的主要信息可以用**矩阵**表示，给出方程组

$$\begin{cases} x_1 - 2x_2 + x_3 = 0 \\ 2x_2 - 8x_3 = 8 \\ -4x_1 + 5x_2 + 9x_3 = -9 \end{cases}$$

把每一个变量的系数写在对齐的一列中，矩阵

$$\begin{bmatrix} 1 & -2 & 1 \\ 0 & 2 & -8 \\ -4 & 5 & 9 \end{bmatrix}$$

称之为该方程组的**系数矩阵**，而

$$\begin{bmatrix} 1 & -2 & 1 & 0 \\ 0 & 2 & -8 & 8 \\ -4 & 5 & 9 & -9 \end{bmatrix}$$

称之为**增广矩阵**。

2.2 线性方程组的求解

基本的思路是把方程组用一个更容易解的等价方程组代替。

粗略地说，我们用第一个方程组中第一个方程中含x_1的项消去其他方程中的含x_1的项，然后用第二个方程中含x_2的项消去其他方程中含x_2的项，以此类推，最后得到一个很简单的等价方程组(阶梯状的方程组)。

例如

$$\begin{cases} x_1 - 2x_2 + x_3 = 0 \\ 2x_2 - 8x_3 = 8 \\ -4x_1 + 5x_2 + 9x_3 = -9 \end{cases} \qquad \begin{bmatrix} 1 & -2 & 1 & 0 \\ 0 & 2 & -8 & 8 \\ -4 & 5 & 9 & -9 \end{bmatrix}$$

可以将第一个方程*4 加上第 3 个方程，从而得到新的方程 3.如下所示

$$\begin{array}{r} 4\cdot[\text{方程 1}]:\quad 4x_1 - 8x_2 + 4x_3 = 0 \\ +[\text{方程 3}]: -4x_1 + 5x_2 + 9x_3 = -9 \\ \hline [\text{新方程 3}]: -3x_2 + 13x_3 = -9 \end{array}$$

就可以得到一个新的方程组和矩阵。以此类推最终得到阶梯状的方程组为

$$\begin{cases} x_1 - 2x_2 + x_3 = 0 \\ x_2 - 4x_3 = 4 \\ x_3 = 3 \end{cases} \qquad \begin{bmatrix} 1 & -2 & 1 & 0 \\ 0 & 1 & -4 & 4 \\ 0 & 0 & 1 & 3 \end{bmatrix}$$

就这样，我们可以很容易地知道原方程组的解为(29,16,3)。

下面我们引入行初等变换：

1. (倍加变换) 把某一行换成它本身与另一行的倍数的和。
2. (对数变换) 把两行对换。
3. (倍乘变换) 把某一行的所有元素乘以同一个非零数。

若其中一个矩阵经过若干行初等变换变换为另一个矩阵，我们称这两个矩阵为行等价的。

2.3 主元

对于一个阶梯型矩阵，我们定义**主元元素**与**主元列**。下面是几个概念：

先导元素：(非零行中)该行最左边的非零元素。

主元位置：阶梯型矩阵中先导元素的位置

主元列：含有先导元素的列

我们举个例子，如下所示

$$\begin{bmatrix} 1 & 4 & 5 & -9 & -7 \\ 0 & 2 & 4 & -6 & -6 \\ 0 & 0 & 0 & -5 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

可以知道，主元列是第 1、2、4 列，主元为 1、2、-5.

通过先前的知识可以知道，矩阵可以对应于一个线性方程组，例如：

$$\begin{bmatrix} 1 & 0 & -5 & 1 \\ 0 & 1 & 1 & 4 \\ 0 & 0 & 0 & 0 \end{bmatrix} \rightarrow \begin{cases} x_1 - 5x_3 = 1 \\ x_2 + x_3 = 4 \\ 0 = 0 \end{cases}$$

对应于主元列的变量x_1和x_2称为**基本变量**，其他变量(x_3)称为**自由变量**。

3.向量方程

3.1 向量

仅含一列的矩阵称为**列向量**，或简称**向量**。

若n是正整数，$\mathbb{R}^n$表示所有n个实数数列(或有序n元组)的集合，通常写成$n \times 1$列矩阵的形式，如

$$\vec{u} = \begin{bmatrix} u_1 \\ u_2 \\ \vdots \\ u_n \end{bmatrix}$$

所有元素都是零的向量称为零向量，用$\vec{0}$表示。

下面举个例子：

$$\vec{u} = \begin{bmatrix} 3 \\ -1 \end{bmatrix}, \vec{w} = \begin{bmatrix} w_1 \\ w_2 \end{bmatrix}$$

这里w_1和w_2是任意实数。所有的两个元素的向量的集记为$\mathbb{R}^{\mathbf{2}}$，$\mathbb{R}$表示向量中的元素是实数，

而指数 2 表示每个向量包含两个元素。

向量之间是可以做加减的，就是对应元素做加减得到新的向量；也支持和实数做标量乘法，就是实数与每个元素做乘法得到新的向量。

有时为了方便，我们会将列向量写成括号的形式，例如

$$\begin{bmatrix} 3 \\ -1 \end{bmatrix} \to (3, -1)$$

可以理解为是一个集合点对，$\mathbb{R}^2$看作是平面上所有点的集合，而向量的几何表示是一条由原点(0,0)指向点(3,-1)的有向线段，如图。

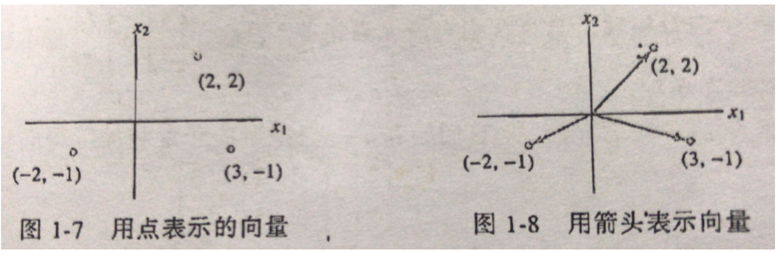


图 1-7 用点表示的向量 图 1-8 用箭头表示向量

其他维度的也同理。

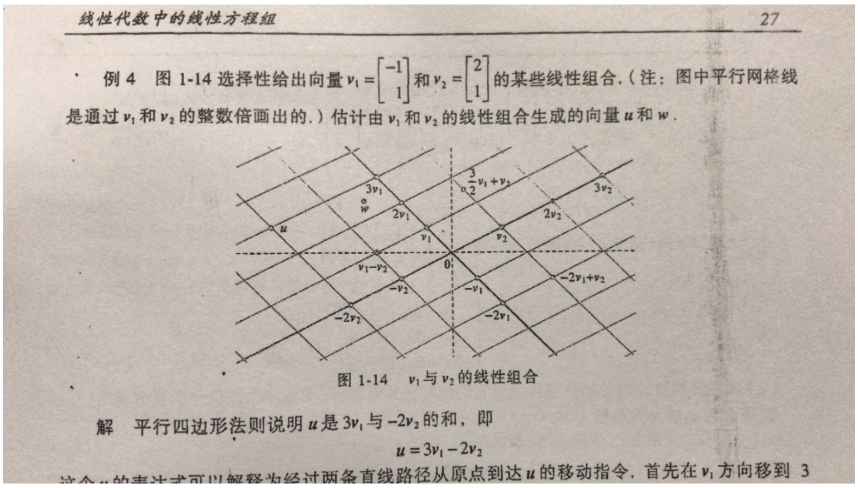
3.2 线性组合

给定$\mathbb{R}^n$中向量$\overrightarrow{v_1}, \overrightarrow{v_2}, \dots, \overrightarrow{v_p}$和标量$c_1, c_2, \dots, c_p$，向量

$$\vec{y} = c_1\overrightarrow{v_1} + \cdots + c_p\overrightarrow{v_p}$$

称为向量$\overrightarrow{v_1}, \overrightarrow{v_2}, \dots, \overrightarrow{v_p}$以$c_1, c_2, \dots, c_p$为**权**的**线性组合**。

线性组合可以在几何上去理解，线性组合后的向量与其他向量一样，均是空间中的一个点。它的形成方式如下所示


线性代数中的线性方程组 27

例 4 图 1-14 选择性给出向量 $v_1 = \begin{bmatrix} -1 \\ 1 \end{bmatrix}$ 和 $v_2 = \begin{bmatrix} 2 \\ 1 \end{bmatrix}$ 的某些线性组合.（注：图中平行网格线是通过 v_1 和 v_2 的整数倍画出的.）估计由 v_1 和 v_2 的线性组合生成的向量 u 和 w.

图 1-14 v_1 与 v_2 的线性组合

解 平行四边形法则说明 u 是 $3v_1$ 与 $-2v_2$ 的和，即

$$u = 3v_1 - 2v_2$$

也就是例如

$$\vec{u} = 3\overrightarrow{v_1} - 2\overrightarrow{v_2}$$

可以当成是从原点开始沿着$\overrightarrow{v_1}$的方向走了三个单位，再沿着$\overrightarrow{v_2}$方向走了-2 个单位。

线性代数的一个主要思想是研究可以表示为某一固定向量集合$\{\overrightarrow{v_1}, \overrightarrow{v_2}, \dots, \overrightarrow{v_p}\}$的线性组合的所有向量。若$\overrightarrow{v_1}, \overrightarrow{v_2}, \dots, \overrightarrow{v_p}$是$\mathbb{R}^n$中的向量，则$\overrightarrow{v_1}, \overrightarrow{v_2}, \dots, \overrightarrow{v_p}$的所有线性组合所称的集合用记号$\mathbf{Span}\{\overrightarrow{v_1}, \overrightarrow{v_2}, \dots, \overrightarrow{v_p}\}$表示称为由$\overrightarrow{v_1}, \overrightarrow{v_2}, \dots, \overrightarrow{v_p}$所生成(或张成)的$\mathbb{R}^n$的子集，也就是说，

$\text{Span}\{\overrightarrow{v_1},\overrightarrow{v_2},\ldots,\overrightarrow{v_p}\}$是所有形如

$$c_1\overrightarrow{v_1}+\cdots+c_p\overrightarrow{v_p}$$

的向量的集合，其中$c_1,\ldots,c_p$为标量。

要判断向量$\vec{b}$是否属于$\text{Span}\{\overrightarrow{v_1},\overrightarrow{v_2},\ldots,\overrightarrow{v_p}\}$，就是判断向量方程

$$x_1\overrightarrow{v_1}+\cdots+x_p\overrightarrow{v_p}=\vec{b}$$

是否有解，或等价地，判断增广矩阵$[\overrightarrow{v_1}\ \ldots\ \overrightarrow{v_p}\ \ \vec{b}]$的线性方程组是否有解。注意，一定包含零向量。

下面是$\text{Span}\{\vec{u}\}$与$\text{Span}\{\vec{u},\vec{v}\}$的几何解释。

设$\vec{u}$是$\mathbb{R}^3$中的向量，那么$\text{Span}\{\vec{u}\}$就是$\vec{u}$的所有数量倍数的集合，也就是通过$\vec{u}$和原点的直线上所有点的集合。

若$\vec{u}$和$\vec{v}$是$\mathbb{R}^3$的非零向量，$\vec{v}$不是$\vec{u}$的倍数，则$\text{Span}\{\vec{u},\vec{v}\}$是$\mathbb{R}^3$中通过$\vec{u}$、$\vec{v}$和原点的平面，可以结合我们之前的线性组合的几何理解上去理解。

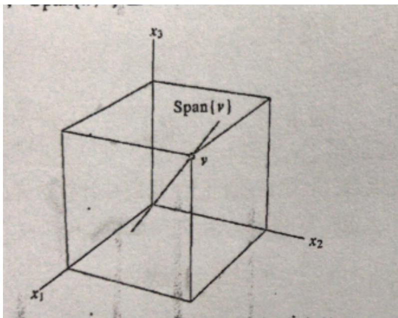


图 1-16 Span{v} 是通过原点的直线

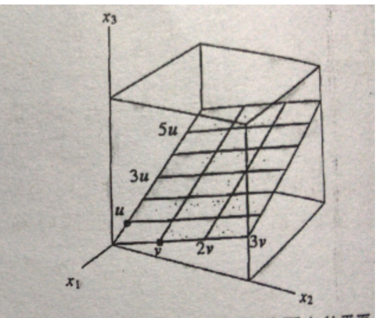


图 1-17 Span{u, v} 是通过原点的平面

4.矩阵方程 $A\vec{x}=\vec{b}$

线性代数的一个基本思想是把向量的线性组合看作矩阵与向量的积。

若A是$n\times m$矩阵，它的各列为$a_1,\ldots,a_n$。若$\vec{x}$是$\mathbb{R}^n$中向量，则A与$\vec{x}$的积，记为$A\vec{x}$，就是A的各列**以$\vec{x}$中对应元素为权的线性组合**，即

$$A\vec{x}=[\overrightarrow{a_1}\ \overrightarrow{a_2}\ \ldots\ \overrightarrow{a_n}]\begin{bmatrix}x_1\\x_2\\\vdots\\x_n\end{bmatrix}=x_1\overrightarrow{a_1}+x_2\overrightarrow{a_2}+\cdots+x_n\overrightarrow{a_n}$$

注意维度要求，A的列数必须要等于$\vec{x}$中的元素个数。例如

$$\begin{bmatrix}1&2&-1\\0&-5&3\end{bmatrix}\begin{bmatrix}4\\3\\7\end{bmatrix}=4\begin{bmatrix}1\\0\end{bmatrix}+3\begin{bmatrix}2\\-5\end{bmatrix}+7\begin{bmatrix}-1\\3\end{bmatrix}=\begin{bmatrix}3\\6\end{bmatrix}$$

通过这个例子，相信能更好的理解矩阵相乘(即为什么要第一个矩阵的列数要等于第二个矩阵的行数)，还可以结合神经网络的前向传播来理解。

定理 设A是$n\times m$矩阵，则下列命题是逻辑上等价的，也就是说，对某个A，它们都成立或者都不成立。

a. 对$\mathbb{R}^m$中的每个$\vec{b}$，方程$A\vec{x}=\vec{b}$有解

b. $\mathbb{R}^m$中的每个$\vec{b}$都是A的列的一个线性组合

c. A的各列生成$\mathbb{R}^m$

d. A在每一列都有一个主元位置

其实要理解不难，(a)和(d)是等价的，每一列均有一个主元，即阶梯型的矩阵每一列均不会出现$0=\vec{b}$的情况，可以从最后一行开始逐步往上求解，所以(a)也是成立的。

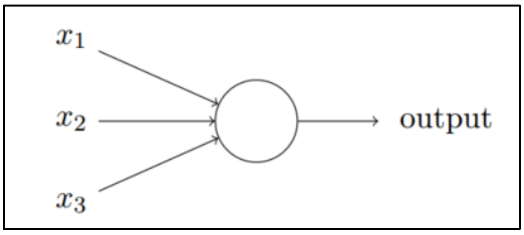
对于(b)，因为A是一个可以到达最后一行的阶梯型的矩阵，所以自然是可以表示出任何m行向量。

4.1 矩阵相乘在神经网络的应用

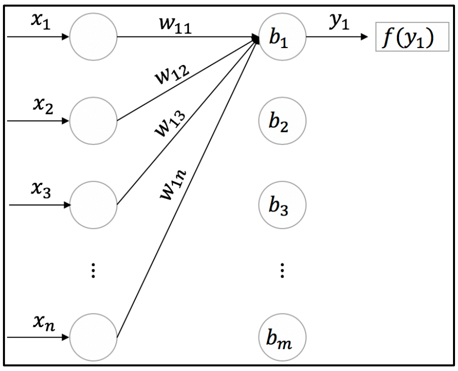
这一小节介绍一下矩阵计算在神经网络中的应用，有助于加深理解。

神经网络中最基本的成分就是神经元模型。和生物的神经元类似，神经元和其他神经元相连，当它"兴奋"时就会向其他神经元发送物质，当其他神经元接收到了物质超过了一定的阈值，就会接着往其他神经元发送物质。

神经网络中的神经元也是如此，可以将其抽象为如下图所示。$x_1+x_2+x_3$为输入，如果超过了阈值就触发激活函数，并产生输出传给下一个神经元。这一过程，我们称之为前向传播。



下面就是介绍如何利用矩阵相乘来快速计算前向传播。我们给出一个神经网络其中两层的传播过程，如下所示



其中，我们描述下前向传播的过程：

- 前一层包含n个结点，后一层包含m个点，两层之间的结点都是全连接的。
- 每一条连接边都带有一个权重w，每一个连接边的输入都会乘以权重。
- 后一层的每个结点都会接收来自前一层的n个输入，接收的输入累加，再加上一个结点所带的偏置b。
- 加上偏置后得到y，y会再经过一个激活函数成为下一层的输入$f(y)$，特别说明，为

了简化接下来的计算，我们将不考虑激活函数。

根据我们的描述，可以得到如下方程组：

$$\begin{cases} w_{11}x_1 + w_{12}x_2 + \cdots + w_{1n}x_n + b_1 = y_1 \\ w_{21}x_1 + w_{22}x_2 + \cdots + w_{2n}x_n + b_2 = y_2 \\ \vdots \\ w_{m1}x_1 + w_{m2}x_2 + \cdots + w_{mn}x_n + b_m = y_m \end{cases}$$

是不是很熟悉，这就是一个线性方程组，我们可以将其写为矩阵相乘的形式，如下所示

$$\begin{bmatrix} w_{11} & w_{12} & \cdots & w_{1n} \\ w_{21} & w_{22} & \cdots & w_{2n} \\ \vdots & \vdots & \vdots & \vdots \\ w_{m1} & w_{m2} & \cdots & w_{mn} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} + \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_m \end{bmatrix} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_m \end{bmatrix}$$

由此可看出，一个神经网络的前向传播过程是可以用矩阵运算替代的。

众所周知，目前用 GPU 训练神经网络比 CPU 要快好几个数量级，其原因就在于 GPU 是成千上万个小处理器核心构成的，CPU 最多有几十个大处理器核心，因此 GPU 处理矩阵运算具有巨大优势（GPU 负责图形处理，例如游戏效果)。虽然 CPU 每个核心的综合能力强很多，但面对神经网络大量的并行加减乘除运算，就不如 GPU 并行处理速度快了。

4.2 齐次线性方程与线性相/无关

若它可以写成$A\vec{x} = \vec{0}$的形式，则称该线性方程组为**齐次线性方程组**。这样的方程组至少有一个解，即$\vec{x} = \vec{0}$，这个解称为它的**平凡解**。我们往往关注的是它的非平凡解。

我们将矩阵方程转换为向量方程，例如考虑方程

$$x_1 \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix} + x_2 \begin{bmatrix} 4 \\ 5 \\ 6 \end{bmatrix} + x_3 \begin{bmatrix} 2 \\ 1 \\ 0 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}$$

此方程当然有平凡解，即$x_1 = x_2 = x_3 = 0$，不过我们更关心的是平凡解是否是唯一解。

我们定义，若向量方程仅有平凡解，$\mathbb{R}^n$中的一组向量$\{v_1, \ldots, v_p\}$称为**线性无关**的，如下所示

$$x_1\overrightarrow{v_1} + \cdots + x_p\overrightarrow{v_p} = \vec{0}$$

若存在不全为零的权$c_1, \ldots, c_p$，使下式成立，则称向量组(集) $\{v_1, \ldots, v_p\}$称为**线性相关**的

$$c_1\overrightarrow{v_1} + \cdots + c_p\overrightarrow{v_p} = \vec{0}$$