1. 线性代数基础

1.学习目标

1. 学习准备知识: 线性代数中的概念和基本运算.
2. 学习在Python语言中实现线性代数中的运算.

2.线性代数基本概念

2.1 矢量

数学上看，矢量空间的元素，或矢量空间中的点，就是**矢量**．更具体地看，有序的值的序列就构成了一个矢量．一般用小写黑体字母表示，如

$$\mathbf{x}=\begin{pmatrix}2\\3\end{pmatrix}=(2,3)^T$$

是二维空间中的矢量．我们默认都采用笛卡尔坐标系．又如

$$\mathbf{y}=\begin{pmatrix}1.0\\1.0\\1.0\end{pmatrix}$$

是三维空间中的矢量．矢量的元素用下标标识，

$$x_1=2,x_2=3.$$

注意:

1. 举例: 我们所在的三维空间可以近似看作是一个三维矢量空间．取定某点为原点，那么空间中每一点就是该空间中的一个矢量．所有的矢量就构成了这个三维空间本身．
2. 在学习Python时，区分列表[2,3], 元组(2,3)和矢量

$$\mathbf{x}=\begin{pmatrix}2\\3\end{pmatrix}.$$

2.2 矢量的基本操作

2.2.1 矢量的加法:

$$\mathbf{x}+\mathbf{z}=\begin{pmatrix}2\\3\end{pmatrix}+\begin{pmatrix}1\\-2\end{pmatrix}=\begin{pmatrix}3\\1\end{pmatrix}$$

一般地，对于两个二维矢量求和，结果为对应坐标相加

$$\mathbf{a}+\mathbf{b}=\begin{pmatrix}a_1+b_1\\a_2+b_2\end{pmatrix}$$

注意：

1. 只有维度相同的两个矢量才可以相加；
2. 对于高维空间中的矢量的求和，可以类推.

2.2.2 矢量的数乘

已知常数\lambda,和n维矢量{\bf x}．数乘定义为

$$\lambda\mathbf{x}=\lambda\begin{pmatrix}x_1\\x_2\\\vdots\\x_n\end{pmatrix}=\begin{pmatrix}\lambda x_1\\\lambda x_2\\\vdots\\\lambda x_n\end{pmatrix}.$$

例如，有矢量

$$\mathbf{x}=\begin{pmatrix}1\\0\\0\end{pmatrix},$$

数乘为

$$3\mathbf{x}=3\begin{pmatrix}1\\0\\0\end{pmatrix}=\begin{pmatrix}3\\0\\0\end{pmatrix};$$

矢量为

$$\mathbf{y}=\begin{pmatrix}1\\1\end{pmatrix},$$

$$0.707\mathbf{x}=0.707\begin{pmatrix}1\\1\end{pmatrix}=\begin{pmatrix}0.707\\0.707\end{pmatrix};$$

说明：

1. 当常数为绝对值大于 1 的数时，数乘将矢量**伸长**，当常数为绝对值小于 1 的数时，数乘将矢量缩短.
2. 常数可以是**实数**，也可以是复数．(本课程中，大家可以不要考虑常数为复数的情形)

2.2.3 矢量的长度

矢量的长度，又称模. 现在举例说明之．例如,对于

$$\mathbf{x}=\begin{pmatrix}3\\4\end{pmatrix},$$

$$|\mathbf{x}| = \sqrt{3^2 + 4^2} = 5$$

通常，**矢量长度**取为各元素的平方和之平方根.

2.2.4 矢量的点积

两个矢量的点积，又叫标量积，定义为两个矢量的对应元素的乘积. 例如，对于二维矢量x和y,其点积

$$\mathbf{x} \cdot \mathbf{y} = x_1 y_1 + x_2 y_2.$$

举例，

$$x = \begin{pmatrix} 4 \\ 5 \end{pmatrix}; y = \begin{pmatrix} 1 \\ 0 \end{pmatrix}; \mathbf{x} \cdot \mathbf{y} = 4 \times 1 + 5 \times 0 = 4$$

注意：

1. 矢量的点积的结果为一个数，而不是矢量.
2. 矢量的点积满足**交换律**：

$$\mathbf{x} \cdot \mathbf{y} = \mathbf{y} \cdot \mathbf{x}.$$

3. 在**书写**矢量点积的表达式中，点积符号(圆点)不可省略:

$$\mathbf{x} \cdot \mathbf{y} \neq \mathbf{xy}$$

4. 举例：一个人有年龄，性别，身高，体重，职业，收入，婚否，母语，国籍，宗教信仰. 这些特征中的每一个都可以看成是一个维度. 每个维度上的信息，可以用数字来编码. 最后，对于这个人我们可以得到十维空间中的一个矢量！（注意：严格来说这些量具有不同的单位，这样组成的量不是矢量. 但如果我们选择适当的单位，合适地编码，我们可以把它们看成矢量.）

$$V_{\text{person}} = \begin{pmatrix} 24 \\ F \\ 170 \\ 60 \\ OL \\ 5.5 \\ N \\ Chinese \\ CN \\ Tao \end{pmatrix} = \begin{pmatrix} 24 \\ 0 \\ 170 \\ 60 \\ 3 \\ 5.5 \\ 0 \\ 1 \\ 3 \\ 0 \end{pmatrix}$$

4. 同理，对于网页上的一篇文章，交易数据，甚至任何一个事件，我们都可以将它表示成一个矢量. 例如，利用bag-of_words技术，我们可以把两篇文章分别表示成矢量，通过计算这两个矢量之差，或者两个矢量的点积，我们就可以判断两篇文章的相似度，从而可以有助于我们进行文章的归类，鉴别剽窃与否等等.

习题1：(30分)

2.3 矩阵

由特定行数和列数的数构成的数学对象，称之为矩阵．形如

$$X=\begin{pmatrix} X_{11} & X_{12} & X_{13} \\ X_{21} & X_{22} & X_{23} \end{pmatrix}$$

者，就是一个2×3矩阵..

例如，像

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

由2行2列的数排列在一起构成的数学对象，就称为**2×2矩阵**．　一般地，一个M×N的矩阵就是M×N个数排成M行N列的一个数学对象. 这里M,N都是大于或等于１的整数．　一般，矩阵用大写字母表示．在此，X, Z都是2行２列的矩阵．举例来说，一个矩阵描述的是对一个对象的操作．这里所说的对象，往往可以用一个矢量来描述．　矩阵作用在矢量上，会得到一个数，一个新矩阵，或者**另一个矢量**．矩阵本身也可以描述一个对象／状态，为之建模．一个M维矢量可看是一个M×1矩阵．

矩阵的分量用两个下标来标注，其中，第一个下标为行数，第二个下标表示分量所在的列数．在程序中，矩阵A的第i行第j列的分量，往往写作

$$A[i][j]$$

2.4 矩阵的基本操作

2.4.1 矩阵的和

只需记住一点：若已经矩阵A和B，它们具有相同的形状(size)，则A与B的和由下面式子定义

$$(A+B)_{ij}=A_{ij}+B_{ij}$$

分量得出，则得到A+B.

2.4.2 矩阵的数乘

矩阵的数乘，也叫矩阵的标量积．计算公式为：

$$(\lambda A)_{ij}=\lambda A_{ij}$$

例如

$$5\begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}=\begin{pmatrix} 5\times 1 & 5\times 0 \\ 5\times 0 & 5\times(-1) \end{pmatrix}=\begin{pmatrix} 5 & 0 \\ 0 & -5 \end{pmatrix}$$

2.4.3 矩阵的乘法

不是任意两个矩阵都可以定义矩阵乘法．只有当第一个矩阵的列数等于第二个矩阵的行数时才能定义．

举一例说明. 例如

$$AB=\begin{pmatrix}0&1&2\\3&4&5\end{pmatrix}\begin{pmatrix}0&1\\2&3\\4&5\end{pmatrix}=\begin{pmatrix}0\times0+1\times2+2\times4&0\times1+1\times3+2\times5\\3\times0+4\times2+5\times4&3\times1+4\times3+5\times5\end{pmatrix}=\begin{pmatrix}9&13\\17&40\end{pmatrix}$$

2×3 矩阵乘 3×2 矩阵得到一个 2×2 矩阵.

再举个例:

$$X\mathbf{a}=\begin{pmatrix}0&1\\1&0\end{pmatrix}\begin{pmatrix}1\\0\end{pmatrix}=\begin{pmatrix}0\times1+1\times0\\1\times1+0\times0\end{pmatrix}=\begin{pmatrix}0\\1\end{pmatrix}$$

2×2 矩阵乘 2×1 矩阵得到一个 2×1 矩阵，即结果为一个新的矢量.

再举两个例:

$$Z\mathbf{a}=\begin{pmatrix}1&0\\0&-1\end{pmatrix}\begin{pmatrix}1\\0\end{pmatrix}=\begin{pmatrix}1\times1+0\times0\\0\times1+(-1)\times0\end{pmatrix}=\begin{pmatrix}1\\0\end{pmatrix}$$

$$Z\mathbf{b}=\begin{pmatrix}1&0\\0&-1\end{pmatrix}\begin{pmatrix}1\\1\end{pmatrix}=\begin{pmatrix}1\times1+0\times1\\0\times1+(-1)\times1\end{pmatrix}=\begin{pmatrix}1\\-1\end{pmatrix}$$

一个 2×2 矩阵(Z)乘 2×1 矩阵**a**得到一个 2×1 矩阵，即结果为*原来*的**矢量**！从(21)式可以看出，当Z与**b**相乘时得到的结果是一个*新的*矢量.

矩阵乘法的性质:

1. 对于一般矩阵，矩阵的乘法**不满足交换律**．对于矩阵A, B,

   $$AB\neq BA$$

   举例子．假设你在赤道上一点**x**．　你先向东走1000km,再向北走1000km到达**y**；　若从该点先向北走1000km后，再先向东走1000km，到达**z**．　最终你会发现：**y，z**两点并不重合----两种走法到达了不同的地方！　又如，两次转动魔方的顺序交换一下，得到的魔方的形态是可能不一样的.

2. 矩阵乘法**满足结合律**．对于矩阵A,B,C,

   $$(AB)C=A(BC)$$

2.4.4 转置矩阵

矩阵A的**转置矩阵**定义为

$$(A^T)_{ij}=A_{ji}$$

例如

$$A=\begin{pmatrix}1&2\\3&4\end{pmatrix};\Rightarrow A^T=\begin{pmatrix}1&3\\2&4\end{pmatrix}$$

二维矢量是 2×1 矩阵，因此也可以计算其转置矩阵．例如

$$\mathbf{x} = \begin{pmatrix} 4 \\ 5 \end{pmatrix}; \Rightarrow \mathbf{x}^T = \begin{pmatrix} 4 & 5 \end{pmatrix}$$

应用：利用矩阵的**转置矩阵**，矢量的点积可以矩阵相乘形式来表示.

$$\mathbf{x} \cdot \mathbf{y} = \begin{pmatrix} 4 & 5 \end{pmatrix} \begin{pmatrix} 1 \\ 0 \end{pmatrix} = 4 \times 1 + 5 \times 0 = 4$$

2.4.5 逆矩阵

单位矩阵I的定义：元素满足条件

$$I_{ij} = \begin{cases} 1 & \text{for } i = j \\ 0 & \text{for } i \neq j \end{cases}$$

的N×N矩阵(方阵). 单位矩阵的性质就是：任何矩阵乘以I都仍然得该矩阵本身:

$$AI = IA = A$$

例如:

$$\begin{pmatrix} 1 & 2 \\ 3 & 4 \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} = \begin{pmatrix} 1 & 2 \\ 3 & 4 \end{pmatrix}$$

对于一个方阵A，它的逆矩阵A^{-1}定义为

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

对于奇异矩阵，其逆矩阵不存在.

2.4.6 解线性方程

矩阵可用来解决线性方程组．什么是线性方程组呢?

看一个例子．鸡兔同笼是中国的数学名题之一. 参考1500年前《孙子算经》：今有雉兔同笼，上有三十五头，下有九十四足，问雉兔各几何?

解法1:

设有鸡兔分别有x,y只，可列方程组:

$$\begin{aligned} x + y = 35 \\ 2x + 4y = 94 \end{aligned}$$

你可以用代入消元法解这个方程组.

解法2:

请小鸡和小兔都举起自己的两只脚．那么它们将有70只脚离开地面．于是，地面将只剩下94-70=24只脚．现在地上已经没有鸡脚了．所以只有小兔的脚，所以小兔的数目为24/2 = 12.　故小鸡的数目为35-12 = 23.

解法3:

我们也可以把这个方程组写成矩阵形式：

$$\begin{pmatrix} 1 & 1 \\ 2 & 4 \end{pmatrix}\begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} 35 \\ 94 \end{pmatrix}$$

也就是这样的形式：

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

我们只需要求出矩阵A的逆矩阵，然后把它在上式两边同时乘上就行了！

要想解出x, 可以在上式两边从左边乘上A矩阵的逆矩阵：

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

也就是

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

这是很**通用**的解法！

又如，线性方程组

$$\begin{aligned} 2x + 2y - z = 4 \\ -x + 2y - z = -6 \\ -3x + y + 2z = -3 \end{aligned}$$

可以写成如下形式

$$\begin{pmatrix} 2 & 2 & -1 \\ -1 & 2 & -1 \\ -3 & 1 & 2 \end{pmatrix}\begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} 4 \\ -6 \\ -3 \end{pmatrix}$$

即

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

要想解出x, 可以在上式两边从左边乘上A矩阵的逆矩阵：

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

也就是

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

多元一次方程组当然也可以这样来求解！现在，我们就找到了一种求解线性方程组的一般方法. 我们可以把求解线性方程组问题用矩阵的语言重新表示出来. **解线性方程组的关键步骤就是: 求其系数组成的矩阵的逆矩阵**！

3. 总结

人工智能(AI)是一个宏大的学科，因为它包含很多方面．包括其组成的复杂性，语言和划分等． AI也是一个广阔的学科，对人类自身及社会有极其重大的影响．

AI系统不一定要模拟人类或自然之机制．例如，如鸟一样，飞机是受到鸟的飞行的启发．飞机可以飞行，但机制完全不一样．又如，人类如何思考也远在我们当前科学能理解的范围之外．大家记住，我们AI课程的核心方法就是：**能做正确的事情的智能体**，而不是一味地模拟人类或自然之机制.

尽管完全的理性是不能做到的，因为环境太复杂．我们甚至不知道什么是**正确的事情**．但我们要尽量做到正确，这里，确定性丧失了，复杂因素变多了，**概率和统计**的作用非常关键，因而我们有必要了解有关概率和统计的基本知识．(几天以后课程再介绍)

在媒体和公众那里经常被问起的一个问题就是：是否AI是人类的威胁？   的确，AI可能让我们产生一些对未来不确定的感觉，并且人们会想是否能保住自己的工作，AI系统是否会接管地球导致人类灭绝．(包括Stephen Hawkins，他告诉BBC，发展完全AI导致人类的灭绝. 他对人工智能思考得比较多，因为他亲身体验过它们在用于交流时是多么强大.)   相比这些问题挑战，我们的社会有更多机遇和进步的可能．我相信人们有信心能解决．未来社会，每一个人都会和AI息息相关．AI是激动人心又繁荣的学科领域，它将改变我们的生活方式．正如计算机和网络改变了我们与父辈祖辈的生活方式一样．人人都可认识AI,利用AI,**为AI做贡献**.

希望大家要将不同的学科的知识融汇贯通，做一个跨学科的有创造力的人．AI领域，就可以让你有机会把自己培养成一个有创造性的跨学科工作者----人才． 举例，时间 - 空间，引力 - 加速度，物质 - 能量, 人员管理系统 / 金融 - 图论，游戏 - 树 / 栈 / 队列 / 堆，设计 / 规划 - 局部搜索，...

4. 习题

习题地址：   https://github.com/hg08/ai_lecture/blob/master/week1/ex_week1_1.pdf

0 : (**10分**)写出**4维**矢量**a**,**b**的和

$$\mathbf{a}+\mathbf{b}$$

的表达式，用它们的分量表示.

1：(**30分**)两矢量点积可以用两矢量的长度，以及夹角的余弦表示为

$$\mathbf{x}\cdot\mathbf{y}=|\mathbf{x}||\mathbf{y}|\cos\theta.$$

已知

$$\mathbf{a}=\begin{pmatrix}4\\0\end{pmatrix},\mathbf{b}=\begin{pmatrix}-4\\0\end{pmatrix},\mathbf{c}=\begin{pmatrix}4\\3\end{pmatrix}.$$

求矢量**a**与矢量**c**的夹角；求矢量**b**与矢量**c**的夹角．用python代码实现这个功能．程序名取为"**dot_product.py**". 使得运行命令

python dot_product.py

能够正常运行．要求输出结果到名为"**output_**.txt**"的文件中(**为你的姓名，可以用拼音). 文件中包含两行内容，格式如下：

a与c夹角的余弦: ***

a与c的夹角(rad): ***

a与c的夹角(°)： ***

说明：

1. 有模板可用，下载地址: https://github.com/hg08/ai_lecture/blob/master/week1/dot_product.py
2. 模板不是必需的，你可以完全自己写．(以后省略此说明)

2: (**20分**)

(1)计算下面各式,

$$-1\begin{pmatrix}1\\1\\0\end{pmatrix};-2\begin{pmatrix}1\\0\\1\end{pmatrix}$$

(2) 并对计算结果再求出其模．

(3) 由对矢量做数乘后所得的结果,试解释上面的数乘(常数为**-1**, -2)分别表示对原矢量做了什么操作?

3: **(10分**)矢量和矩阵有什么区别? 请列出来．矢量和矩阵在Python里的实现方式有什么共同点?

4: (**30分**) 假设我们有一本词典， 里面只有五个单词 [a, b, c, d, e]. 这里有三个文档：

文档A: [a,b, b, d,d,e]

文档B: [b,b,b,e,e,e,d,a]

文档C: [d,b,b,e]

使用bag-of-words模型将每个文档表示成五维向量,每个维度上的分量分别代表a,b,c,d,e这五个单词在文档中出现的次数. 例如文档A可表示为

$$\mathbf{a}=\begin{pmatrix}1\\2\\0\\2\\1\end{pmatrix}.$$

(1)将文档B和文档C按照上述模型表示成矢量.

(2) 若定义两矢量夹角的余弦值为这两个矢量所对应的文档的**相似性**．计算文档A与文档B的相似性，计算文档A与文档C的相似性.

5. 参考文献

[1] A. C. Mueller, S. Guido, Python机器学习入门,O' Reilly, 2016

6.扩展

[1]