线性代数

-17 课　正交矩阵和 Gram-Schmidt 正交化

一、知识概要

这一节从上一节结尾介绍的标准正交向量谈起，主要介绍标准正交向量组的性质与优点，以及将一组向量化为标准正交向量组的方法:Gram-Schmidt 正交化。

二．标准正交向量

2.1 回顾标准正交向量

上节介绍过标准正交向量，我们通过一个式子进行回顾。

设 q 是标准正交向量组中的任意向量，则

$$q_i^T q_j \begin{cases} 0 & (i \neq j) \\ 1 & (i = j) \end{cases}$$

这很好地表现了标准正交向量组内各向量的性质。“标准”→ 长度为 1。

机器学习算法与自然语言处理公众号
丁坤博

2.2 标准正交矩阵 Q

另外这节介绍另一个概念：标准正交矩阵 Q。

所谓标准正交矩阵 Q，就是将标准正交向量组中的q_1，q_2 ⋯ q_n列在同一个矩阵中：

$$Q = \begin{bmatrix} q_1 & \cdots\cdots & q_n \end{bmatrix}$$

这样的标准正交矩阵有一个很好的性质：

$$Q^T Q = \begin{bmatrix} q_1 \\ q_2 \\ \cdots \\ q_n \end{bmatrix} \begin{bmatrix} q_1 & q_2 & \cdots & q_n \end{bmatrix} = \begin{bmatrix} 1 & 0 & \dots & 0 & 0 \\ 0 & 1 & \dots & 0 & 0 \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & \dots & 1 & 0 \\ 0 & 0 & \dots & 0 & 1 \end{bmatrix} = I$$

特别地，当 Q 是方阵时，我们将这样的矩阵 Q 称为：**正交矩阵**。

为什么我们将方阵单独拿出来呢？因为方阵有逆矩阵，根据上面对标准正交矩阵 Q 的讨论，很明显，由$Q^T Q = I$，可以得到 Q 的逆矩阵。

此时：$$Q^T = Q^{-1}$$

例如：$Q = \begin{bmatrix} 0 & 0 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix}$， $Q^T = Q^{-1} = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \end{bmatrix}$。

此时 $\begin{bmatrix} 0 & 0 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix}\begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \end{bmatrix} = \mathrm{I}$

另一个例子是根据我们之前介绍的正交向量组：$\begin{bmatrix} cos\theta \\ sin\theta \end{bmatrix}$，$\begin{bmatrix} -sin\theta \\ cos\theta \end{bmatrix}$。写成矩阵形式即为：

$$Q = \begin{bmatrix} cos\theta & -sin\theta \\ sin\theta & cos\theta \end{bmatrix}$$

此外，正交矩阵不要忘了单位化。例如：$\begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix}$，这个矩阵各列是正交的，但并不是正交矩阵，因为没有单位化，所以，**正交矩阵不要忘了单位化向量。**正确的正交矩阵是：$\frac{1}{\sqrt{2}}\begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix}$。由这个矩阵可以延伸出阿达马矩阵，这里不做详细介绍。

2.3 标准正交矩阵的作用

上面介绍了标准正交矩阵 Q 的各种性质，很显然这是一种新的性质优良的矩阵，接下来主要介绍它的具体应用之一：投影矩阵。

记得上面介绍的投影矩阵 $P = A(A^TA)^{-1}A^T$。那么当取的 A 矩阵是标准正交矩阵 Q 时，很明显：

$$Q(Q^TQ)^{-1}Q^T = QQ^T$$

特别的，当 Q 时方阵（正交阵）时，由于此时$Q^T = Q^{-1}$。所以投影矩阵即为 I。

化为 $P = QQ^T$后，很容易验证投影矩阵的两条性质：投影矩阵为对称矩阵，且$P^2 = P$。带进去计算就好了。

还有之前介绍的拟合方程：

$$A^Tb = A^TA\hat{x}$$

当使用标准正交基，即 A = Q 为标准正交矩阵时，原式可化简为：

$$Q^TQ\hat{x} = Q^Tb$$

$$\leftrightarrow \quad I\hat{x} = Q^Tb$$

$$\leftrightarrow \quad \hat{x} = Q^Tb$$

这样化简之后，很明显$\hat{x}$的每个分量都是 Q 中对应列向量与 b 的点乘结果。即：

$$\hat{x_i} = {q_i}^T \mathrm{b}$$

这个式子的意义就是，如果我们已知标准正交基，那么 b 在第 i 个基上的投影就是对应基向量${q_i}^T$b。

很明显，当我们选择标准正交向量作为基时，投影矩阵相关公式中的 A 都可以代换为 Q，这样很多公式都可被化简。

三、Gram-Schmidt 正交化

这部分介绍一个方法，从线性无关向量组入手，将其矩阵标准正交化。

【例】有两个线性无关的向量 a，b。我们想从中得到标准正交向量q_1，q_2。

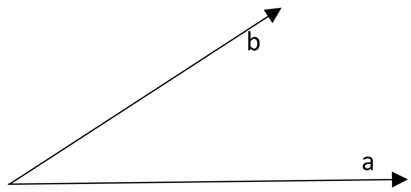
也可以理解为：空间原来的两个基 a，b 无法满足标准正交，现在我们通过 Gram-Schmidt 正交化并单位化将其变为两个标准正交基q_1，q_2。

解：

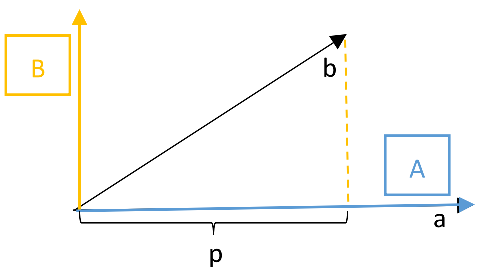
首先假设已知一组基 A,B 是正交的。那么$q_1 = \frac{A}{|A|}$，$q_2 = \frac{B}{|B|}$。

接下来问题的关键就在于：找到两个正交的基。

就以这两个 a，b 为例，线性无关即不共线，如下图：



怎样将它们化为互相正交的两个基呢？联想我们之前学习的投影，可以这样取：



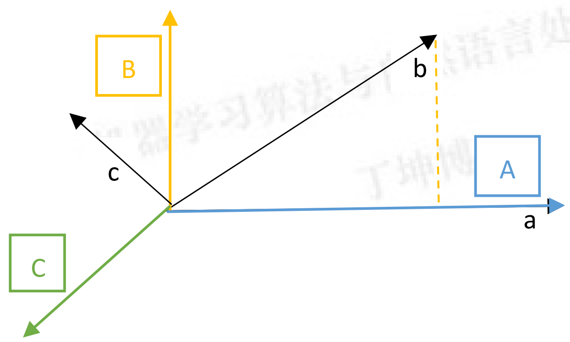
将 a 向量定为 A 向量，然后将 b 向量投影到 a 向量上。取投影垂线所在直线的正方向为另一个基向量 B 的方向，取投影垂线长度为 B 的长度。这时 A 与 B 明显正交。

上一节中学习过，这个 B 即为 (b-p)，联系之前 15 课学习的投影，不难得到：

$$B = b - \frac{A^T b}{A^T A}A$$

带进去检验一下，$A^T B = A^T(b - \frac{A^T b}{A^T A}A) = 0$。说明 A 与 B 是正交的，我们的方法正确。接下来通过$q_1 = \frac{A}{|A|}$，$q_2 = \frac{B}{|B|}$ 单位化各个向量，就得到了同 a，b 空间的标准正交基q_1，q_2。

同样的道理，推广到三维：



寻找三个正交的向量 A,B,C 的话，其中 A，B 方法不变：

$$A = a$$

$$B = b - \frac{A^T b}{A^T A}A$$

而 C 通过将 c 减去在 A,B 上的投影就可以得到：

$$C = c - \frac{A^T c}{A^T A}A - \frac{B^T c}{B^T B}B$$

得到三个正交的向量 A,B,C，再进行单位化即可。

接下来通过一个例子熟悉下这个方法：

【例】a = $\begin{bmatrix}1\\1\\1\end{bmatrix}$，b = $\begin{bmatrix}1\\0\\2\end{bmatrix}$，求标准正交矩阵 Q。

根据之前的步骤，A = a = $\begin{bmatrix}1\\1\\1\end{bmatrix}$。

$$B = b - \frac{A^T b}{A^T A}A = \begin{bmatrix}1\\0\\2\end{bmatrix} - \frac{3}{3}\begin{bmatrix}1\\1\\1\end{bmatrix} = \begin{bmatrix}0\\-1\\1\end{bmatrix}$$

再进行单位化，得到标准正交矩阵 Q：

$$Q = [q_1 \quad q_2] = \begin{bmatrix}1/\sqrt{3} & 0\\1/\sqrt{3} & -1/\sqrt{2}\\1/\sqrt{3} & 1/\sqrt{2}\end{bmatrix}$$

这就是 Gram-Schmidt 正交化方法。

机器学习算法与自然语言处理公众号
丁坤博

再看看这个例子，原始的矩阵 A = $\begin{bmatrix}1 & 1\\1 & 0\\1 & 2\end{bmatrix}$，而我们求得的 Q = $\begin{bmatrix}1/\sqrt{3} & 0\\1/\sqrt{3} & -1/\sqrt{2}\\1/\sqrt{3} & 1/\sqrt{2}\end{bmatrix}$。观察两个矩阵的列空间，它们是相同的，也就是说我们的正交化过程都是在同一个空间中进行的，只是最后得到了一个更好的标准正交基而已。

从矩阵的角度来看，类似于 A 的 LU 分解，在 Gram-Schmidt 正交化中，A 可分解为 Q 与 R。其中 R 是上三角矩阵：

$$A = QR$$

A = $[a_1 \quad a_2]$(原始向量，a，b，c⋯)

Q = $[q_1 \quad q_2]$(标准正交基，q_1，q_2⋯)

$$R = \begin{bmatrix}{a_1}^T q_1 & {a_2}^T q_1\\{a_1}^T q_2 & {a_2}^T q_2\end{bmatrix}$$

△其中 R 中的${a_1}^T q_2$为 0，类似于上面例子中，a_1=A，q_2是 B 方向上的，所以其内积为 0。其他情况类似，这也是 Gram-Schmidt 正交化的一个性质。

四、学习感悟

这一节主要内容围绕 Gram-Schmidt 正交化，即将一个空间的基化为互相标准正交的一组基，这样会方便我们的很多计算。这部分内容重在步骤，掌握其过程方法为主要，需要一定的习题量来熟练。

机器学习算法与自然语言处理公众号
丁坤博