降维（一）

矩阵降维后可以用更小规模的矩阵表示原矩阵，且不改变原矩阵的特征。在主成分分析（PCA）中，分析特征值和特征值的使用。本文介绍奇异值分解（SVD），一种比UV分解更强大的分解。然后在SVD的基础上介绍它的变种，CUR分解，CUR分解可以在原矩阵是稀疏时，保持分解后矩阵的稀疏性的分解方式。

基础概念

欧氏空间

定义了内积的实数域上的向量空间，称为欧几里得空间。

奇异矩阵

方阵A的秩不是满秩，则A是奇异矩阵。

1. 可逆矩阵是非奇异矩阵，非奇异矩阵也是可逆矩阵。
2. 如果A为奇异矩阵，则$Ax=0$有无穷解，$Ax=b$有无穷解或者无解。
3. 如果A为非奇异矩阵，则$Ax=0$有且只有唯一0解，$Ax=b$有唯一解。

正定矩阵

1. 一个矩阵半正定当且仅当它的每个特征值大于或者等于0。
2. 一个矩阵正定当且仅当它的每个特征值都大于0。

向量范数

1. 非负性：$||x|| \geq 0$
2. 齐次性：$||ax|| = |a| \cdot ||x||$
3. 三角不等式：$||x|| + ||y|| \geq ||x+y||$
4. 1-范数：$||x||_1 = |x_1| + |x_2| + \ldots + |x_n| = \sum_{i=1}^{n} |x_i|$
5. 2-范数（Euclidean范数）：$||x||_2 = \sqrt{x_1^2 + x_2^2 + \ldots + x_n^2} = (\sum_{i=1}^{2} x_i^2)^{\frac{1}{2}}$
6. 无穷范数：$||x||_\infty = max(|x_1|, |x_2|, \ldots, |x_n|)$

矩阵范数

如果满足范数三条件，同时满足矩阵乘法相容性（次乘性）：$||A|| \cdot ||B|| \geq ||A \cdot B||$

1. 列和范数（1-范数）：$||A||_1 = max_{1 \leq j \leq n}(\sum_{i=1}^{m} |a_{ij}|)$
2. 行和范数（∞-范数）：$||A||_\infty = max_{1 \leq m}(\sum_{j=1}^{n} |a_{ij}|)$
3. Frobenius范数（F-范数）：$||A||_F = (\sum_{i=1}^{m} \sum_{j=1}^{n} |a_{ij}|^2)^{\frac{1}{2}}$
4. 2-范数（谱模）：$||A||_2 = \sqrt{\lambda_{max}(A^T A)}$，矩阵$A^T A$的最大特征值开平方根。

矩阵的乘法表达

设矩阵乘法$C_{m \times k} = A_{m \times n} B_{n \times k}$

行列相乘

$$c_{ij} = a_i \cdot b_j \tag{1}$$

列行相乘

$$AB = \begin{pmatrix} a_1 & a_2 & \dots & a_n \end{pmatrix} \cdot \begin{pmatrix} b_1 \\ b_2 \\ \dots \\ b_n \end{pmatrix} = a_1 b_1 + a_2 b_2 + \dots + a_n b_n \tag{2}$$

$$= \begin{pmatrix} a_{11}b_{11} & a_{11}b_{12} & \dots & a_{11}b_{1k} \\ a_{21}b_{11} & a_{21}b_{12} & \dots & a_{21}b_{1k} \\ \dots & \dots & \dots & \dots \\ a_{m1}b_{11} & a_{m1}b_{12} & \dots & a_{m1}b_{1k} \end{pmatrix} + \dots + \begin{pmatrix} a_{1n}b_{n1} & a_{1n}b_{n2} & \dots & a_{1n}b_{nk} \\ a_{2n}b_{n1} & a_{2n}b_{n2} & \dots & a_{2n}b_{nk} \\ \dots & \dots & \dots & \dots \\ a_{mn}b_{n1} & a_{mn}b_{n2} & \dots & a_{mn}b_{nk} \end{pmatrix}$$

行行相乘

$$C = AB = \begin{pmatrix} a_1 \\ a_2 \\ \dots \\ a_m \end{pmatrix} \cdot \begin{pmatrix} b_1 \\ b_2 \\ \dots \\ b_n \end{pmatrix} \tag{3}$$

考虑C的每一个行向量

$$c_i = a_i \cdot \begin{pmatrix} b_1 \\ b_2 \\ \dots \\ b_n \end{pmatrix} = a_{i1} b_1 + a_{i2} b_2 + \dots + a_{in} b_n \tag{4}$$

矩阵C的每一个行向量，是B的行向量的一个线性组合，该线性组合中的系数是a_i的各个元素。从这个角度说C的每一个行向量都存在于B的行向量空间内。

列列相乘

$$C = AB = \begin{pmatrix} a_1 & a_2 & \dots & a_n \end{pmatrix} \cdot \begin{pmatrix} b_1 & b_2 & \dots & b_k \end{pmatrix} \tag{5}$$

考虑C的每一个列向量

$$c_i = \begin{pmatrix} a_1 & a_2 & \dots & a_n \end{pmatrix} \cdot b_i = b_{i1} a_1 + b_{i2} a_2 + \dots + b_{in} a_n \tag{6}$$

矩阵C的每一个列向量，是A的列向量的一个线性组合，该线性组合中的系数是b_i的各个元素。从这个角度说C的每一列都存在于A的列向量空间内。

特征值和特征向量

定义

设定M为一个方阵，设定λ为一个常数，e是一个非零列向量，和M的行数相同，那么称λ为M的特征值，e为M的特征向量。如果e是M的特征向量，c是任意常数，那么ce也是一个M的特征向量，且有相同的特征值。向量乘以一个常量只改变向量的长度，不改变向量的方向，因此为了避免长度的不同，规定单位向量，意味着**向量每个元素的平方和为1**.为了防止出现乘以-1的情况，我们要求特征向量的第一个非零分量为正。

Example 11.1 : Let M be the matrix

$$\begin{bmatrix} 3 & 2 \\ 2 & 6 \end{bmatrix}$$

One of the eigenvectors of M is

$$\begin{bmatrix} 1/\sqrt{5} \\ 2/\sqrt{5} \end{bmatrix}$$

and its corresponding eigenvalue is 7. The equation

$$\begin{bmatrix} 3 & 2 \\ 2 & 6 \end{bmatrix} \begin{bmatrix} 1/\sqrt{5} \\ 2/\sqrt{5} \end{bmatrix} = 7 \begin{bmatrix} 1/\sqrt{5} \\ 2/\sqrt{5} \end{bmatrix}$$

demonstrates the truth of this claim. Note that both sides multiply to

$$\begin{bmatrix} 7/\sqrt{5} \\ 14/\sqrt{5} \end{bmatrix}$$

Also observe that the eigenvector is a unit vector, because $(1/\sqrt{5})^2 + (2/\sqrt{5})^2 = 1/5 + 4/5 = 1$. □

计算特征值和特征向量

1. 幂迭代法

   等待更新

2. 行列式为0

   由$Me = \lambda e$，可得$(M - \lambda I)e = 0$

   即$M - \lambda I = 0$

   该行列式为n次，因此有n个解，从这n个解中，可以求出特征向量$Me = ce$，并将e调整为单位向量。

通过幂迭代找到特征对（特征值 + 特征向量）

待完善

矩阵特征向量

对于方阵M的列特征向量$e_1, e_2, \ldots, e_n$，构造矩阵E，其中E的第i列是e_i，那么有$EE^T = E^TE = I$。因为矩阵的特征向量是正交的，这些是正交的单位向量。

主成分分析（PCA）

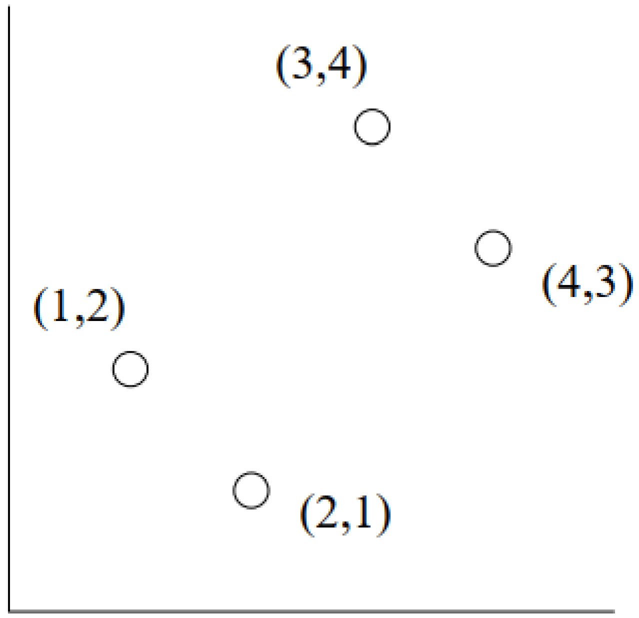
简单来说，主成分分析是为了找到高维矩阵的主特征值，通过该特征向量来表示高维矩阵的特征的方法。

参考链接：如何通俗易懂地讲解什么是 PCA 主成分分析？ - 论智的回答 - 知乎 https://www.zhihu.com/question/41120789/answer/474222214

如果不想清楚原理，下面的文章给出了应用计算的例子：

https://blog.csdn.net/zhongkelee/article/details/44064401

书中同样给出一个简单例子，如下：



上图是一个二维点集，该二维点集对应的矩阵为（第一列为x坐标，第二列为y坐标）

$$M = \begin{bmatrix} 1 & 2 \\ 2 & 1 \\ 3 & 4 \\ 4 & 3 \end{bmatrix}$$

计算M^TM，得到

$$M^TM = \begin{bmatrix} 1 & 2 & 3 & 4 \\ 2 & 1 & 4 & 3 \end{bmatrix} \begin{bmatrix} 1 & 2 \\ 2 & 1 \\ 3 & 4 \\ 4 & 3 \end{bmatrix} = \begin{bmatrix} 30 & 28 \\ 28 & 30 \end{bmatrix}$$

计算特征值为$\lambda = 58, \lambda = 2$，则特征向量为$\begin{bmatrix} \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \end{bmatrix}$和$\begin{bmatrix} -\frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \end{bmatrix}$，这里第一个分量选择负值是为了后续的坐标变换更简单。

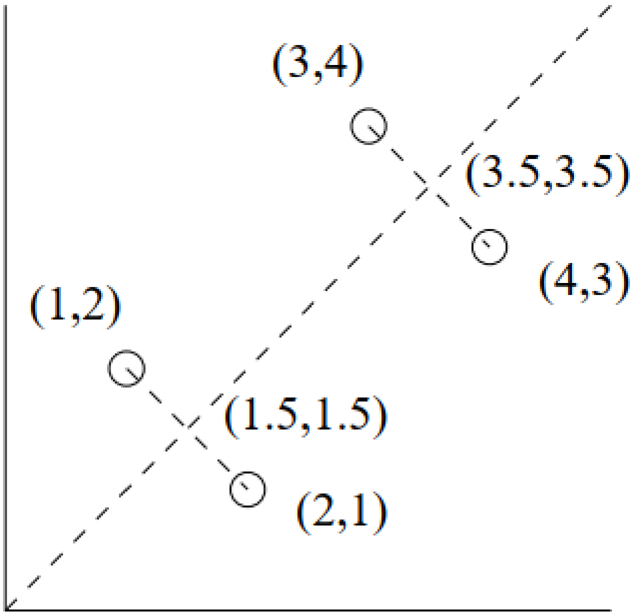
矩阵M^TM的特征向量矩阵为

$$E = \begin{bmatrix} 1\sqrt{2} & -1\sqrt{2} \\ 1\sqrt{2} & 1\sqrt{2} \end{bmatrix}$$

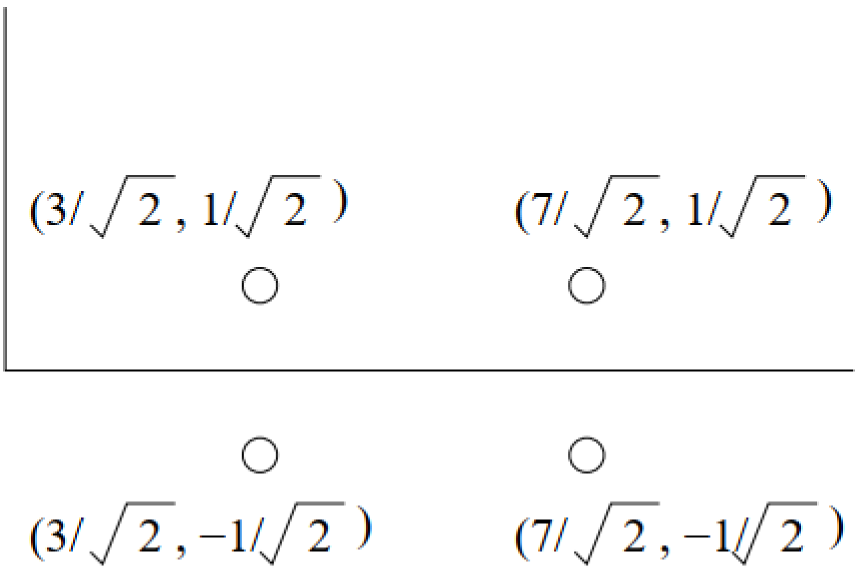
正交向量组成的矩阵可以看成是欧几里得空间的沿轴的旋转，上图的矩阵可以看作是45°逆时针旋转。例如处理矩阵M后，

$$ME = \begin{bmatrix} 1 & 2 \\ 2 & 1 \\ 3 & 4 \\ 4 & 3 \end{bmatrix} \begin{bmatrix} 1\sqrt{2} & -1\sqrt{2} \\ 1\sqrt{2} & 1\sqrt{2} \end{bmatrix} = \begin{bmatrix} 3/\sqrt{2} & 1/\sqrt{2} \\ 3/\sqrt{2} & -1/\sqrt{2} \\ 7/\sqrt{2} & 1/\sqrt{2} \\ 7/\sqrt{2} & -1/\sqrt{2} \end{bmatrix}$$

对应的第一个特征向量如下图虚线所示



此时点$[1, 2]$转化为$[3/\sqrt{2}, 1/\sqrt{2}]$，新坐标系的y轴和虚线垂直，如图



使用特征向量降维

1. 通过上例可知，如果矩阵M的每一行代表欧几里得空间得一个点，则可以计算$M^T M$然后计算特征向量和特征值，定义矩阵E的每一列为对应的特征向量，按照大的特征值对应的特征向量排列。定义矩阵L是矩阵$M^T M$的特征值在主对角线按照从大到小的顺序排列，其他位置为0的矩阵，由于对每一个特征向量e和特征值λ有$M^T Me = \lambda e = e\lambda$，可以得出$M^T ME = EL$。
2. 矩阵ME表示矩阵M中的点转化到一个新的坐标系空间中，该空间中最重要的轴对应特征值最大的轴，第二个轴对应的则是特征值次大的轴，以此类推。如果想要将矩阵M转化到空间维度更小的空

间中，保留相关性最大的特征值（即最大的特征值）对应的轴，并依次选择，忽视特征值很小的轴（维度）。
3. 假如E_k表示矩阵E的前k列，那么ME_k则表示矩阵M的一个k维表达。

例：

考虑上图中的例子，由于原坐标系只有2维，因此只能降为1维，因此计算ME_1如下，

$$\begin{bmatrix} 1 & 2 \\ 2 & 1 \\ 3 & 4 \\ 4 & 3 \end{bmatrix} \begin{bmatrix} 1/\sqrt{2} \\ 1/\sqrt{2} \end{bmatrix} = \begin{bmatrix} 3/\sqrt{2} \\ 3/\sqrt{2} \\ 7/\sqrt{2} \\ 7/\sqrt{2} \end{bmatrix}$$

新矩阵代替了原矩阵，此时所有的点都被虚线上的投影代替，尽管前两个点（后两个点）的投影相同，这样的投影表达了在一维情况下对M中点的最好的分割。

矩阵的距离

考虑MM^T的特征值，因为M的行数多于列数，因此MM^T的大小要比M^TM大，但是如果M的列数多于行数，反而M^TM的大小会更大。如图，

$$MM^T = \begin{bmatrix} 1 & 2 \\ 2 & 1 \\ 3 & 4 \\ 4 & 3 \end{bmatrix} \begin{bmatrix} 1 & 2 & 3 & 4 \\ 2 & 1 & 4 & 3 \end{bmatrix} = \begin{bmatrix} 5 & 4 & 11 & 10 \\ 4 & 5 & 10 & 11 \\ 11 & 10 & 25 & 24 \\ 10 & 11 & 24 & 25 \end{bmatrix}$$

和M^TM一样，MM^T也是一个对称矩阵，在M^TM和MM^T的特征值之间有较强的联系，设e是M^TM的特征向量，则有

$$M^TMe = \lambda e \tag{7}$$

两边乘以M，有

$$MM^T(Me) = M\lambda e = \lambda(Me) \tag{8}$$

因此，只要Me不是0向量，那么它是MM^T的特征向量，λ既是M^TM也是MM^T的特征值。该结论对于MM^T也适用（不再详细说明）。

当$M^Te = 0$时，$MM^Te = 0$，由$\lambda e = 0$可知，$\lambda = 0$。

结论：MM^T的特征值（包括0）也是M^TM的特征值，如果MM^T的维数小于M^TM，那么结论反过来成立。

例：

$$\begin{bmatrix} 3/\sqrt{116} & 1/2 & 7/\sqrt{116} & 1/2 \\ 3/\sqrt{116} & -1/2 & 7/\sqrt{116} & -1/2 \\ 7/\sqrt{116} & 1/2 & -3/\sqrt{116} & -1/2 \\ 7/\sqrt{116} & -1/2 & -3/\sqrt{116} & 1/2 \end{bmatrix}$$

上图为MM^T的特征向量矩阵，MM^T的特征向量中包含M^TM的特征值58和2，但是因为有4个值，因此剩余的为0。

附注

定理1

实对称矩阵的不同的单位特征向量是正交的。

证明

设矩阵A的特征向量p,q满足$Ap=mp, Aq=nq$，其中m,n为特征值，则有

$$p^T(Aq)=p^T(nq)=np^Tq \tag{9}$$

$$(p^TA)q=(p^TA^T)q=(Ap)^Tq=(mp)^Tq=mp^Tq \tag{10}$$

又$p^T(Aq)=(p^TA)q$，则有

$$(m-n)p^Tq=0 \tag{11}$$

因此$p^Tq=0$，因此p,q正交。