PCA计算题

利用PCA方法将这组二维数据降到一维。

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

PCA计算题答案

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

为例，我们用PCA方法将这组二维数据其降到一维。

因为这个矩阵的每行已经是零均值，这里我们直接求协方差矩阵：

$$C = \frac{1}{5}\begin{pmatrix} -1 & -1 & 0 & 2 & 0 \\ -2 & 0 & 0 & 1 & 1 \end{pmatrix}\begin{pmatrix} -1 & -2 \\ -1 & 0 \\ 0 & 0 \\ 2 & 1 \\ 0 & 1 \end{pmatrix} = \begin{pmatrix} \frac{6}{5} & \frac{4}{5} \\ \frac{4}{5} & \frac{6}{5} \end{pmatrix}$$

然后求其特征值和特征向量，具体求解方法不再详述，可以参考相关资料。求解后特征值为：

$$\lambda_1 = 2, \lambda_2 = 2/5$$

其对应的特征向量分别是：

$$c_1\begin{pmatrix} 1 \\ 1 \end{pmatrix}, c_2\begin{pmatrix} -1 \\ 1 \end{pmatrix}$$

其中对应的特征向量分别是一个通解，c1和c2可取任意实数。那么标准化后的特征向量为：

$$\begin{pmatrix} 1/\sqrt{2} \\ 1/\sqrt{2} \end{pmatrix}, \begin{pmatrix} -1/\sqrt{2} \\ 1/\sqrt{2} \end{pmatrix}$$

PCA计算题答案

因此我们的矩阵P是：

$$P = \begin{pmatrix} 1/\sqrt{2} & 1/\sqrt{2} \\ -1/\sqrt{2} & 1/\sqrt{2} \end{pmatrix}$$

可以验证协方差矩阵C的对角化：

$$PCP^{\mathsf{T}} = \begin{pmatrix} 1/\sqrt{2} & 1/\sqrt{2} \\ -1/\sqrt{2} & 1/\sqrt{2} \end{pmatrix} \begin{pmatrix} 6/5 & 4/5 \\ 4/5 & 6/5 \end{pmatrix} \begin{pmatrix} 1/\sqrt{2} & -1/\sqrt{2} \\ 1/\sqrt{2} & 1/\sqrt{2} \end{pmatrix} = \begin{pmatrix} 2 & 0 \\ 0 & 2/5 \end{pmatrix}$$

最后我们用P的第一行乘以数据矩阵，就得到了降维后的表示：

$$Y = \begin{pmatrix} 1/\sqrt{2} & 1/\sqrt{2} \end{pmatrix} \begin{pmatrix} -1 & -1 & 0 & 2 & 0 \\ -2 & 0 & 0 & 1 & 1 \end{pmatrix} = \begin{pmatrix} -3/\sqrt{2} & -1/\sqrt{2} & 0 & 3/\sqrt{2} & -1/\sqrt{2} \end{pmatrix}$$

稀疏表示手工计算题

) 设 $A=\begin{pmatrix} 1 & 1 & 1 \\ 0 & 0 & 0 \\ -1 & -1 & -1 \end{pmatrix}$，求 $\|A\|_1, \|A\|_2, \|A\|_\infty, \|A\|_F$；

计算向量 $x=(1,2,-3)^T$的向量范数 $\|x\|_1$, $\|x\|_2$, $\|x\|_\infty$

稀疏表示基本知识

在向量空间 $R^n(C^n)$中，设$x=(x_1,x_2,\cdots,x_n)^T$

常用的向量 x的范数有

$$\|x\|_2=(|x_1|^2+|x_2|^2+\cdots+|x_n|^2)^{1/2}$$ x的2－范数或欧氏范数

$$\|x\|_1=|x_1|+|x_2|+\cdots+|x_n|$$ x的1－范数

$$\|x\|_\infty=\max_{1\le i\le n}|x_i|$$ x的∞－范数或最大范数

$$\|x\|_p=(|x_1|^p+|x_2|^p+\cdots+|x_n|^p)^{1/p}$$ x的p－范数，$p\ge 1$

稀疏表示基本知识

常用的矩阵范数

(1) $\|A\|_1 = \max_{1\le j\le n} \sum_{i=1}^{n} |a_{ij}|$

A的每列绝对值之和的最大值， 称A的列范数

(2) $\|A\|_\infty = \max_{1\le i\le n} \sum_{j=1}^{n} |a_{ij}|$

A的每行绝对值之和的最大值， 称A的行范数

(3) $\|A\|_2 = \sqrt{\lambda_{\max}(A^T A)}$ 称 A 的 2 － 范数

其中$\lambda_{max}(A^T A)$为$A^T A$的特征值的绝对值的最 大值

LLE手工推导作业

- 给定如下数据

$$X = [x_1, x_2, ..., x_N], D \times N$$

$$w = [w_1, w_2, ..., w_N], k \times N$$

求解约束优化问题

$$\arg\min_{w} \sum_{i=1}^{N} ||x_i - \sum_{j=1}^{k} w_{ji} x_{ji}||^2, s.t. \sum_{j=1}^{k} w_{ji} = 1$$

CCA推导作业1

8.6 对140名学生进行了阅读速度 、阅读能力 、运算速度和运算能力 的四种测验，所得成绩的相关系数阵为：

$$R=\begin{bmatrix} 1 & 0.03 & 0.24 & 0.59 \\ 0.03 & 1 & 0.06 & 0.07 \\ 0.24 & 0.06 & 1 & 0.24 \\ 0.59 & 0.07 & 0.24 & 1 \end{bmatrix}$$

试对阅读本领与运算本领之间进行典型相关分析。

CCA推导作业1答案

解：根据已知可得

$$R_{11}=\begin{pmatrix}1 & 0.03\\ 0.03 & 1\end{pmatrix},\quad R_{12}=R_{21}'=\begin{pmatrix}0.24 & 0.59\\ 0.06 & 0.07\end{pmatrix},\quad R_{22}=\begin{pmatrix}1 & 0.24\\ 0.24 & 1\end{pmatrix}.$$

$$R_{11}^{-1}=\frac{1}{1-0.0009}\begin{pmatrix}1.0009 & -0.030027\\ -0.030027 & 1.0009\end{pmatrix},\quad R_{22}^{-1}=\begin{pmatrix}1.06112 & -0.254669\\ -0.254669 & 1.06112\end{pmatrix}.$$

$$M_{1z}=R_{11}^{-1}R_{12}R_{22}^{-1}R_{21}$$

$$=\begin{pmatrix}1.0009 & -0.030027\\ -0.030027 & 1.0009\end{pmatrix}\begin{pmatrix}0.24 & 0.59\\ 0.06 & 0.07\end{pmatrix}\begin{pmatrix}1.06112 & -0.254669\\ -0.254669 & 1.06112\end{pmatrix}\begin{pmatrix}0.24 & 0.06\\ 0.59 & 0.07\end{pmatrix}$$

$$=\begin{pmatrix}0.357322 & 0.0456454\\ 0.035091 & 0.00551094\end{pmatrix}.$$

$$M_{2z}=R_{22}^{-1}R_{21}R_{11}^{-1}R_{12}$$

$$=\begin{pmatrix}1.06112 & -0.254669\\ -0.254669 & 1.06112\end{pmatrix}\begin{pmatrix}0.24 & 0.06\\ 0.59 & 0.07\end{pmatrix}\begin{pmatrix}1.0009 & -0.030027\\ -0.030027 & 1.0009\end{pmatrix}\begin{pmatrix}0.24 & 0.59\\ 0.06 & 0.07\end{pmatrix}$$

$$=\begin{pmatrix}0.0273146 & 0.0638401\\ 0.137808 & 0.335516\end{pmatrix}.$$

CCA推导作业1答案

计算得 M_{1z}, M_{2z} 的特征值为：λ_1^2=0.361817，λ_2^2=0.00101553.

M_{1z} 对应特征值 λ_1^2=0.361817 的单位特征向量为：$\vec{a} = \begin{pmatrix} 0.995185 \\ 0.0980115 \end{pmatrix}$

M_{2z} 对应特征值 λ_1^2=0.361817 的单位特征向量为：$\vec{b} = \begin{pmatrix} 0.187468 \\ 0.982271 \end{pmatrix}$

所以提取第一对典型变量为：$U_1 = 0.995185 Z_1^{(1)} + 0.0980115 Z_2^{(1)}$

$$V_1 = 0.187468 Z_1^{(2)} + 0.982271 Z_2^{(2)}$$

典型相关系数为：$\lambda_1 = \sqrt{0.361817} = 0.601512$.

其中 $Z_i^{(1)}, Z_j^{(2)}$ 分别为原始变量 X_i, Y_j 标准化后的结果。

ISOMAP作业

- ***ISOMAP*与*MDS*计算距离时的不同点是?**

答：*MDS*高维空间中计算两个点的距离:

$$d_{ij}^{2} = \| z_i - z_j \|^2 = \| z_i \|^2 + \| z_j \|^2 - 2 z_i^T z_j$$

*ISOMAP*高维空间中计算两个点的距离：
两点的最短路径，然后采用内积形式进行推导。

ISOMAP作业

请写出$ISOMAP$算法的推导过程

ISOMAP作业答案

1、计算平方距离矩阵D_X， $D_{X(ij)} = \left(d_{kl}^X\right)^2$；

2、 $H = I - \frac{1}{N}11^T$, 1为全1向量，$T = -\frac{1}{2}HD_XH$；

3、对T进行特征分解$T = U\Lambda U^T$，确定T的前d个特征向量，记为U'， d为

希望的维数;

4、并使$Y = \Lambda'^{\frac{1}{2}}U^T$，$\Lambda'$为$T$的d个特征值组成的对角阵。