典型相关分析

耿修瑞

中国科学院空天信息创新研究院

gengxr@sina.com.cn

2025.4

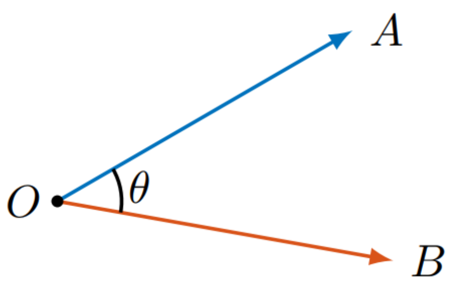
提纲

- 问题背景
- 互相关分析
- 典型相关分析
- 几何解释
- 相关进展

问题背景

两个向量之间的相似性：可以用距离、内积、角度、相关系数等度量来衡量

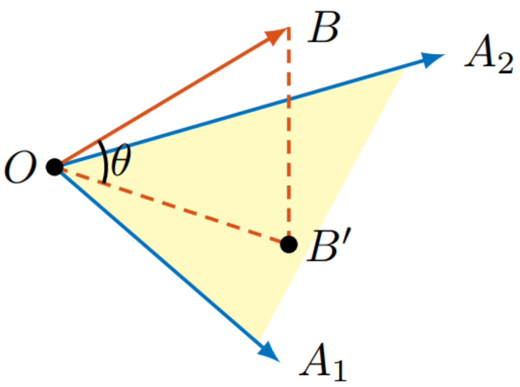
$$\theta = \arccos\left(\frac{\mathbf{a}^{\mathrm{T}}\mathbf{b}}{\|\mathbf{a}\|\|\mathbf{b}\|}\right)$$



角度越大代表两个向量之间的相似度越小；而角度越小则代表两个向量之间的相似度越大

问题背景

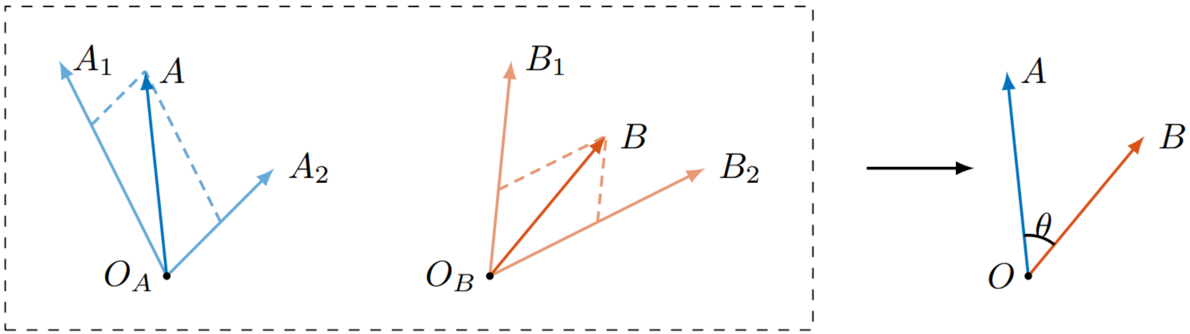
一个向量与一组向量之间的相似性：最小二乘法



通过最小二乘法，可以在一组向量所张成的超平面中找到和一个向量相似性最大的代表向量。

问题背景

两组向量之间的相似性：如何衡量？



衡量两组向量之间的相似性，需要在两组向量所张成的超平面各找到一个代表向量，使得两个代表向量之间的相似度最大。那么该如何寻找这两个代表向量呢？

互相关分析

给定 n 维空间的两组变量，$\{\mathbf{x}_1,\mathbf{x}_2,\cdots,\mathbf{x}_p\},\{\mathbf{y}_1,\mathbf{y}_2,\cdots,\mathbf{y}_q\}$ 接下来利用互相关分析来分析这两组变量的相关性，不妨记 $\mathbf{X}=\begin{bmatrix}\mathbf{x}_1 & \mathbf{x}_2 & \cdots & \mathbf{x}_p\end{bmatrix},\mathbf{Y}=\begin{bmatrix}\mathbf{y}_1 & \mathbf{y}_2 & \cdots & \mathbf{y}_q\end{bmatrix}$

首先对两组变量进行线性组合分别得到两个向量 $\mathbf{z}=\mathbf{X}\mathbf{u}$，$\mathbf{w}=\mathbf{Y}\mathbf{v}$，然后计算二者内积为 $\mathbf{z}^{\mathrm{T}}\mathbf{w}=\mathbf{u}^{\mathrm{T}}\mathbf{X}^{\mathrm{T}}\mathbf{Y}\mathbf{v}=n\mathbf{u}^{\mathrm{T}}\Sigma_{12}\mathbf{v}$，其中 $\boldsymbol{\Sigma}_{12}=\frac{1}{n}\mathbf{X}^{\mathrm{T}}\mathbf{Y}$，互相关分析模型如下

$$\begin{cases}\max\limits_{\mathbf{u}}\mathbf{u}^{\mathrm{T}}\Sigma_{12}\mathbf{v}\\ s.t.\ \mathbf{u}^{\mathrm{T}}\mathbf{u}=1,\mathbf{v}^{\mathrm{T}}\mathbf{v}=1\end{cases}$$

通过拉格朗日乘子法，可以得到上述优化问题的解可以归结为如下矩阵的特征值与特征向量问题（不妨设 $p\le q$）：

$$\begin{cases}\boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{21}\mathbf{u}=\lambda^2\mathbf{u}\\ \boldsymbol{\Sigma}_{21}\boldsymbol{\Sigma}_{12}\mathbf{v}=\lambda^2\mathbf{v}\end{cases}\text{或}\begin{cases}\boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{21}\mathbf{U}=\boldsymbol{\Lambda}_1\mathbf{U}\\ \boldsymbol{\Sigma}_{21}\boldsymbol{\Sigma}_{12}\mathbf{V}=\boldsymbol{\Lambda}_2\mathbf{V}\end{cases}$$

其中 $\boldsymbol{\Sigma}_{21}=\boldsymbol{\Sigma}_{12}^{\mathrm{T}}=\frac{1}{n}\mathbf{Y}^{\mathrm{T}}\mathbf{X}$

$$\mathbf{U}=\begin{bmatrix}\mathbf{u}_1 & \mathbf{u}_2 & \cdots & \mathbf{u}_p\end{bmatrix}$$

$$\mathbf{V}=\begin{bmatrix}\mathbf{v}_1 & \mathbf{v}_2 & \cdots & \mathbf{v}_q\end{bmatrix}$$

$$\boldsymbol{\Lambda}_1=\begin{bmatrix}\lambda_1^2 & & \\ & \ddots & \\ & & \lambda_p^2\end{bmatrix}$$

$$\boldsymbol{\Lambda}_2=\begin{bmatrix}\lambda_1^2 & & & & & \\ & \ddots & & & & \\ & & \lambda_p^2 & & & \\ & & & 0 & & \\ & & & & \ddots & \\ & & & & & 0\end{bmatrix}$$

互相关分析

矩阵$\mathbf{\Sigma}_{12}\mathbf{\Sigma}_{21},\mathbf{\Sigma}_{21}\mathbf{\Sigma}_{12}$的特征分解等价于$\mathbf{\Sigma}_{12}$的奇异值分解，即

$$\mathbf{\Sigma}_{12}=\mathbf{U}\mathbf{\Lambda}\mathbf{V}^{\mathrm{T}}$$

其中

$$\mathbf{\Lambda}=\mathbf{U}^{\mathrm{T}}\mathbf{\Sigma}_{12}\mathbf{V}=\begin{bmatrix}\mathbf{u}_1^{\mathrm{T}}\mathbf{\Sigma}_{12}\mathbf{v}_1 & & & 0 & \cdots & 0\\ & \ddots & & \vdots & \ddots & \vdots\\ & & \mathbf{u}_p^{\mathrm{T}}\mathbf{\Sigma}_{12}\mathbf{v}_p & 0 & \cdots & 0\end{bmatrix}=\begin{bmatrix}\lambda_1 & & & 0 & \cdots & 0\\ & \ddots & & \vdots & \ddots & \vdots\\ & & \lambda_p & 0 & \cdots & 0\end{bmatrix}$$

不妨设$\lambda_1^2\geq\lambda_2^2\cdots\geq\lambda_p^2>0$，则分别用$\mathbf{U},\mathbf{V}$对两组数据进行变换，即可得到$p$组互相关成分，

$$\begin{cases}\mathbf{Z}=\mathbf{X}\mathbf{U}\\ \mathbf{W}=\mathbf{Y}\mathbf{V}\end{cases}$$

其中与λ_1对应的成分具有最大相似性，而与λ_p对应的成分具有最小相似性。

讨论：互相关分析是否为正交变换？同一组变量的各个成分之间是否正交？

互相关分析

例1 试用互相关分析评估$\{\mathbf{x}_1,\mathbf{x}_2\},\{\mathbf{y}_1,\mathbf{y}_2\}$ 两组向量的相关性，其中

$$\mathbf{x}_1=\begin{bmatrix}\frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} & 0\end{bmatrix}^{\mathrm{T}},\mathbf{x}_2=\begin{bmatrix}\frac{1}{\sqrt{6}} & \frac{1}{\sqrt{6}} & -\frac{2}{\sqrt{6}}\end{bmatrix}^{\mathrm{T}},\mathbf{y}_1=\begin{bmatrix}0 & 1 & 0\end{bmatrix}^{\mathrm{T}},\mathbf{y}_2=\begin{bmatrix}0 & 0 & 1\end{bmatrix}^{\mathrm{T}}$$

记 $\mathbf{X}=[\mathbf{x}_1 \quad \mathbf{x}_2],\mathbf{Y}=[\mathbf{y}_1 \quad \mathbf{y}_2]$，计算可得$\mathbf{X}$和$\mathbf{Y}$ 的互相关矩阵为

$$\boldsymbol{\Sigma}_{12}=\frac{1}{3}\mathbf{X}^{\mathrm{T}}\mathbf{Y}=\frac{1}{3}\begin{bmatrix}-\frac{1}{\sqrt{2}} & 0\\ \frac{1}{\sqrt{6}} & -\frac{2}{\sqrt{6}}\end{bmatrix}$$

对其进行奇异值分解，有

$$\boldsymbol{\Sigma}_{12}=\mathbf{U}\boldsymbol{\Lambda}\mathbf{V}^{\mathrm{T}}=\begin{bmatrix}-\frac{1}{2} & \frac{\sqrt{3}}{2}\\ \frac{\sqrt{3}}{2} & \frac{1}{2}\end{bmatrix}\begin{bmatrix}\frac{1}{3} & 0\\ 0 & \frac{1}{3\sqrt{3}}\end{bmatrix}\begin{bmatrix}\frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}}\\ -\frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}}\end{bmatrix}$$

可以得到互相关分析结果为

$$\mathbf{Z}=\mathbf{X}\mathbf{U}=\begin{bmatrix}0 & \frac{2}{\sqrt{6}}\\ \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{6}}\\ -\frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{6}}\end{bmatrix}=[\mathbf{z}_1 \quad \mathbf{z}_2],\ \mathbf{W}=\mathbf{Y}\mathbf{V}=\begin{bmatrix}0 & 0\\ \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}}\\ -\frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}}\end{bmatrix}=[\mathbf{w}_1 \quad \mathbf{w}_2]$$

互相关分析

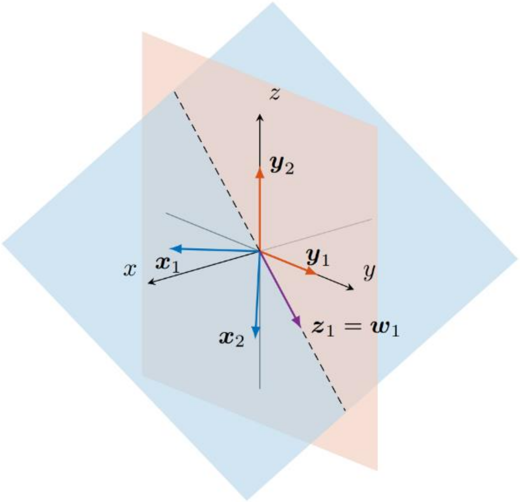
与 $\lambda_1=\frac{1}{3}$ 对应的两组向量的线性组合为

$$\begin{cases}\mathbf{z}_1=\begin{bmatrix}0 & \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}}\end{bmatrix}^{\mathrm{T}}\\ \mathbf{w}_1=\begin{bmatrix}0 & \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}}\end{bmatrix}^{\mathrm{T}}\end{cases}$$

显然二者重合，并且正好位于 $span(\mathbf{x}_1,\mathbf{x}_2)$与$span(\mathbf{y}_1,\mathbf{y}_2)$ 的交线上，这说明$\mathbf{\Sigma}_{12}$的最大奇异值对应的两个奇异向量组合出来的向量 $\mathbf{z}_1$和$\mathbf{w}_1$具有最大的相关性。这表明互相关分析有能力找到使得两组向量相似度最大的线性组合。

此外，奇异值 $\lambda_2=\frac{1}{3\sqrt{3}}$ 对应的两组向量的线性组合为

$$\begin{cases}\mathbf{z}_2=\begin{bmatrix}\frac{2}{\sqrt{6}} & -\frac{1}{\sqrt{6}} & -\frac{1}{\sqrt{6}}\end{bmatrix}^{\mathrm{T}}\\ \mathbf{w}_2=\begin{bmatrix}0 & -\frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}}\end{bmatrix}^{\mathrm{T}}\end{cases}$$



互相关分析

存在的问题：从互相关分析的目标函数可以看出，它最大化的是线性组合向量 $\mathbf{z}=\mathbf{X}\mathbf{u}$和$\mathbf{w}=\mathbf{Y}\mathbf{v}$ 的内积，因此结果必然会都到$\mathbf{z}$和$\mathbf{w}$的长度的影响。接下来我们对例1中的数据稍加改造，然后对改造后的数据进行互相关分析。

例2 试用互相关分析评估$\{\mathbf{x}_1,\mathbf{x}_2\},\{\mathbf{y}_1,\tilde{\mathbf{y}}_2\}$ 两组向量的相关性，其中

$$\mathbf{x}_1=\begin{bmatrix}\frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} & 0\end{bmatrix}^{\mathrm{T}},\mathbf{x}_2=\begin{bmatrix}\frac{1}{\sqrt{6}} & \frac{1}{\sqrt{6}} & -\frac{2}{\sqrt{6}}\end{bmatrix}^{\mathrm{T}},\mathbf{y}_1=\begin{bmatrix}0 & 1 & 0\end{bmatrix}^{\mathrm{T}},\tilde{\mathbf{y}}_2=\begin{bmatrix}0 & 1 & \frac{3}{2}\end{bmatrix}^{\mathrm{T}}$$

记 $\tilde{\mathbf{X}}=\begin{bmatrix}\mathbf{x}_1 & \mathbf{x}_2\end{bmatrix},\tilde{\mathbf{Y}}=\begin{bmatrix}\mathbf{y}_1 & \tilde{\mathbf{y}}_2\end{bmatrix}$，计算可得 $\tilde{\mathbf{X}}$和$\tilde{\mathbf{Y}}$ 的互相关矩阵为

$$\tilde{\mathbf{\Sigma}}_{12}=\frac{1}{3}\tilde{\mathbf{X}}^{\mathrm{T}}\tilde{\mathbf{Y}}=\frac{1}{3}\begin{bmatrix}-0.7071 & -0.7071\\ -0.4082 & -0.8165\end{bmatrix}$$

对其进行奇异值分解，有

$$\tilde{\mathbf{\Sigma}}_{12}=\tilde{\mathbf{U}}\tilde{\mathbf{\Lambda}}\tilde{\mathbf{V}}^{\mathrm{T}}=\begin{bmatrix}-0.7992 & 0.6011\\ -0.6011 & -0.7992\end{bmatrix}\begin{bmatrix}0.3667 & 0\\ 0 & 0.2617\end{bmatrix}\left[\begin{bmatrix}0.2898 & 0.9571\\ -0.9571 & 0.2898\end{bmatrix}\right]$$

互相关分析

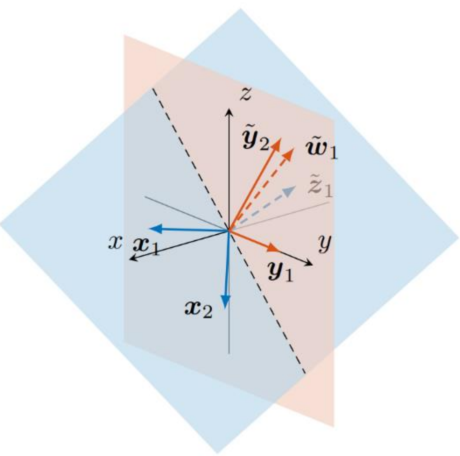
可以得到互相关分析结果为

$$\tilde{\mathbf{Z}}=\tilde{\mathbf{X}}\tilde{\mathbf{U}}=\begin{bmatrix}-0.8105 & 0.0988\\ 0.3197 & -0.7513\\ 0.4908 & 0.6525\end{bmatrix}=\begin{bmatrix}\tilde{\mathbf{z}}_1 & \tilde{\mathbf{z}}_2\end{bmatrix},\ \tilde{\mathbf{W}}=\tilde{\mathbf{Y}}\tilde{\mathbf{V}}=\begin{bmatrix}0 & 0\\ 1.2469 & -0.6673\\ 1.4356 & 0.4347\end{bmatrix}=\begin{bmatrix}\tilde{\mathbf{w}}_1 & \tilde{\mathbf{w}}_2\end{bmatrix}$$

与 $\lambda_1=0.3667$ 对应的两组向量的线性组合为

$$\begin{cases}\tilde{\mathbf{z}}_1=\begin{bmatrix}-0.8105 & 0.3197 & 0.4908\end{bmatrix}^{\mathrm{T}}\\ \tilde{\mathbf{w}}_1=\begin{bmatrix}0 & 1.2469 & 1.4356\end{bmatrix}^{\mathrm{T}}\end{cases}$$

可以发现 $\tilde{\mathbf{z}}_1$和$\tilde{\mathbf{w}}_1$ 不再具有相同的方向。这意味着改变任意一组向量中任意一个向量的长度，会导致不同的互相关分析结果。而我们的初衷是为了分析两组向量张成的平面之间的相似性，理论上这种相似性不应该受到各自平面所选的代表向量的影响。



值得注意的是，尽管 $\tilde{\mathbf{U}}$是正交矩阵，但

$$\tilde{\mathbf{Z}}^{\mathrm{T}}\tilde{\mathbf{Z}}=\begin{bmatrix}3.6158 & -0.2080\\ -0.2080 & 0.6342\end{bmatrix}$$

并不是一个对角矩阵。这意味着每一组向量的互相关分析的各个成分之间未必正交。

讨论：如何克服互相关分析的上述问题？

典型相关分析

上一节我们看到，在互相关分析中，由于采用了内积来衡量向量之间相似性，所以最终的结果会受到向量的模的影响，这导致互相关分析的结果不能反映两组向量所张成的平面之间的相似性。为了克服这一弊端，我们有必要将互相关分析中的用来衡量向量相似性的目标函数更换成一个与向量的模无关的量，而角度就是满足此条件的自然而然的选择。

仍记 $\mathbf{X}=\begin{bmatrix}\mathbf{x}_1 & \mathbf{x}_2 & \cdots & \mathbf{x}_p\end{bmatrix}^{\mathrm{T}}, \mathbf{Y}=\begin{bmatrix}\mathbf{y}_1 & \mathbf{y}_2 & \cdots & \mathbf{y}_q\end{bmatrix}^{\mathrm{T}}$，对两组变量进行线性组合分别得到两个向量 $\mathbf{z}=\mathbf{X}\mathbf{u}$，$\mathbf{w}=\mathbf{Y}\mathbf{v}$，这两个向量的夹角的余弦为

$$\cos(\theta)=\frac{\mathbf{z}^{\mathrm{T}}\mathbf{w}}{\sqrt{\mathbf{z}^{\mathrm{T}}\mathbf{z}}\sqrt{\mathbf{w}^{\mathrm{T}}\mathbf{w}}}=\frac{\mathbf{u}^{\mathrm{T}}\Sigma_{12}\,\mathbf{v}}{\sqrt{\mathbf{u}^{\mathrm{T}}\Sigma_{11}\,\mathbf{u}}\sqrt{\mathbf{v}^{\mathrm{T}}\Sigma_{22}\,\mathbf{v}}}$$

使得两个组合后的向量具有最小夹角的优化模型为

$$\max_{\mathbf{u},\mathbf{v}}\cos(\theta)=\frac{\mathbf{u}^{\mathrm{T}}\Sigma_{12}\,\mathbf{v}}{\sqrt{\mathbf{u}^{\mathrm{T}}\Sigma_{11}\,\mathbf{u}}\sqrt{\mathbf{v}^{\mathrm{T}}\Sigma_{22}\,\mathbf{v}}}$$

典型相关分析

上述无约束优化模型可以转化为如下等式约束优化模型：

$$\begin{cases} \max\limits_{\mathbf{u}} \mathbf{u}^{\mathrm{T}}\boldsymbol{\Sigma}_{12}\mathbf{v} \\ s.t.\ \mathbf{u}^{\mathrm{T}}\boldsymbol{\Sigma}_{11}\mathbf{u}=1, \mathbf{v}^{\mathrm{T}}\boldsymbol{\Sigma}_{22}\mathbf{v}=1 \end{cases}$$

利用拉格朗日乘子法，模型的解可以转化为如下矩阵的特征值特征向量问题

$$\begin{cases} \boldsymbol{\Sigma}_{11}^{-1}\boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}\boldsymbol{\Sigma}_{21}\mathbf{u}=\lambda^2\mathbf{u} \\ \boldsymbol{\Sigma}_{22}^{-1}\boldsymbol{\Sigma}_{21}\boldsymbol{\Sigma}_{11}^{-1}\boldsymbol{\Sigma}_{12}\mathbf{v}=\lambda^2\mathbf{v} \end{cases}$$

仍假设 $p \le q$，且 $\boldsymbol{\Sigma}_{\mathrm{I}}=\boldsymbol{\Sigma}_{11}^{-1}\boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}\boldsymbol{\Sigma}_{21}$ 可逆，则它必有 p 个正特征值 $\lambda_1^2 \ge \lambda_2^2 \cdots \ge \lambda_p^2$ 相应的特征向量记为 $\mathbf{U}=\begin{bmatrix}\mathbf{u}_1 & \mathbf{u}_2 & \cdots & \mathbf{u}_p\end{bmatrix}$，$\boldsymbol{\Sigma}_{\mathrm{II}}=\boldsymbol{\Sigma}_{22}^{-1}\boldsymbol{\Sigma}_{21}\boldsymbol{\Sigma}_{11}^{-1}\boldsymbol{\Sigma}_{12}$ 的特征向量记为 $\mathbf{V}=\begin{bmatrix}\mathbf{v}_1 & \mathbf{v}_2 & \cdots & \mathbf{v}_q\end{bmatrix}$，则最终的典型相关变换表达式为

$$\begin{cases} \mathbf{Z}=\mathbf{XU} \\ \mathbf{W}=\mathbf{YV} \end{cases}$$

讨论：典型相关分析是否为正交变换？
同一组变量的各个成分之间是否正交？

典型相关分析

$$\begin{cases} \mathbf{\Sigma}_{11}^{-1}\mathbf{\Sigma}_{12}\mathbf{\Sigma}_{22}^{-1}\mathbf{\Sigma}_{21}\mathbf{u} = \lambda^2\mathbf{u} \\ \mathbf{\Sigma}_{22}^{-1}\mathbf{\Sigma}_{21}\mathbf{\Sigma}_{11}^{-1}\mathbf{\Sigma}_{12}\mathbf{v} = \lambda^2\mathbf{v} \end{cases}$$

等价于如下广义特征问题（验证一下）：

$$\begin{bmatrix} 0 & \mathbf{\Sigma}_{12} \\ \mathbf{\Sigma}_{21} & 0 \end{bmatrix} \begin{bmatrix} \mathbf{u} \\ \mathbf{v} \end{bmatrix} = \lambda \begin{bmatrix} \mathbf{\Sigma}_{11} & 0 \\ 0 & \mathbf{\Sigma}_{22} \end{bmatrix} \begin{bmatrix} \mathbf{u} \\ \mathbf{v} \end{bmatrix}$$

典型相关分析

例3 试用典型相关分析评估 $\{\mathbf{x}_1,\mathbf{x}_2\},\{\mathbf{y}_1,\mathbf{y}_2\}$ 两组向量的相关性，其中

$$\mathbf{x}_1=\begin{bmatrix}\frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} & 0\end{bmatrix}^{\mathrm{T}},\mathbf{x}_2=\begin{bmatrix}\frac{1}{\sqrt{6}} & \frac{1}{\sqrt{6}} & -\frac{2}{\sqrt{6}}\end{bmatrix}^{\mathrm{T}},\mathbf{y}_1=\begin{bmatrix}0 & 1 & 0\end{bmatrix}^{\mathrm{T}},\mathbf{y}_2=\begin{bmatrix}0 & 0 & 1\end{bmatrix}^{\mathrm{T}}$$

记 $\mathbf{X}=[\mathbf{x}_1 \quad \mathbf{x}_2],\mathbf{Y}=[\mathbf{y}_1 \quad \mathbf{y}_2]$，计算可得

$$\mathbf{\Sigma}_{\mathrm{I}}=\begin{bmatrix}\frac{1}{2} & -\frac{1}{2\sqrt{3}}\\ -\frac{1}{2\sqrt{3}} & \frac{5}{6}\end{bmatrix},\ \mathbf{\Sigma}_{\mathrm{II}}=\begin{bmatrix}\frac{2}{3} & -\frac{1}{3}\\ -\frac{1}{3} & \frac{2}{3}\end{bmatrix}$$

这两个矩阵的特征值和特征向量矩阵分别为

$$\mathbf{\Lambda}_1=\begin{bmatrix}1 & 0\\ 0 & \frac{1}{3}\end{bmatrix},\mathbf{U}=\begin{bmatrix}-\frac{1}{2} & \frac{\sqrt{3}}{2}\\ \frac{\sqrt{3}}{2} & \frac{1}{2}\end{bmatrix};\mathbf{\Lambda}_2=\begin{bmatrix}1 & 0\\ 0 & \frac{1}{3}\end{bmatrix},\mathbf{V}=\begin{bmatrix}\frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}}\\ -\frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}}\end{bmatrix}$$

可以得到典型相关分析结果为

$$\mathbf{Z}=\mathbf{XU}=\begin{bmatrix}0 & \frac{2}{\sqrt{6}}\\ \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{6}}\\ -\frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{6}}\end{bmatrix}=[\mathbf{z}_1 \quad \mathbf{z}_2],\ \mathbf{W}=\mathbf{YV}=\begin{bmatrix}0 & 0\\ \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}}\\ -\frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}}\end{bmatrix}=[\mathbf{w}_1 \quad \mathbf{w}_2]$$

典型相关分析

从上面的例子可以看出，典型相关分析和互相关分析对于这个数据取得相同的结果。接下来我们将数据调整为例2中的数据

例4 试用典型相关分析评估 $\{\mathbf{x}_1,\mathbf{x}_2\},\{\mathbf{y}_1,\tilde{\mathbf{y}}_2\}$ 两组向量的相关性，其中

$$\mathbf{x}_1=\begin{bmatrix}\frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} & 0\end{bmatrix}^{\mathrm{T}},\mathbf{x}_2=\begin{bmatrix}\frac{1}{\sqrt{6}} & \frac{1}{\sqrt{6}} & -\frac{2}{\sqrt{6}}\end{bmatrix}^{\mathrm{T}},\mathbf{y}_1=\begin{bmatrix}0 & 1 & 0\end{bmatrix}^{\mathrm{T}},\tilde{\mathbf{y}}_2=\begin{bmatrix}0 & 1 & \frac{3}{2}\end{bmatrix}^{\mathrm{T}}$$

记 $\tilde{\mathbf{X}}=[\mathbf{x}_1 \quad \mathbf{x}_2],\tilde{\mathbf{Y}}=[\mathbf{y}_1 \quad \tilde{\mathbf{y}}_2]$，计算可得

$$\boldsymbol{\Sigma}_{\mathrm{I}}=\begin{bmatrix}\frac{1}{2} & -\frac{1}{2\sqrt{3}}\\ -\frac{1}{2\sqrt{3}} & \frac{5}{6}\end{bmatrix},\ \boldsymbol{\Sigma}_{\mathrm{II}}=\begin{bmatrix}\frac{8}{9} & -\frac{5}{18}\\ -\frac{2}{9} & \frac{4}{9}\end{bmatrix}$$

其特征值和特征向量矩阵分别为

$$\tilde{\boldsymbol{\Lambda}}_1=\begin{bmatrix}1 & 0\\ 0 & \frac{1}{3}\end{bmatrix},\ \tilde{\mathbf{U}}=\begin{bmatrix}-\frac{1}{2} & \frac{\sqrt{3}}{2}\\ \frac{\sqrt{3}}{2} & \frac{1}{2}\end{bmatrix},\tilde{\boldsymbol{\Lambda}}_2=\begin{bmatrix}1 & 0\\ 0 & \frac{1}{3}\end{bmatrix},\ \tilde{\mathbf{V}}=\begin{bmatrix}\frac{5}{\sqrt{29}} & -\frac{1}{\sqrt{5}}\\ -\frac{2}{\sqrt{29}} & -\frac{2}{\sqrt{5}}\end{bmatrix},$$

典型相关分析结果为

$$\tilde{\mathbf{Z}}=\tilde{\mathbf{X}}\tilde{\mathbf{U}}=\begin{bmatrix}0 & \frac{4}{\sqrt{6}}\\ \frac{1}{\sqrt{2}} & -\frac{2}{\sqrt{6}}\\ -\frac{1}{\sqrt{2}} & -\frac{2}{\sqrt{6}}\end{bmatrix}=[\tilde{\mathbf{z}}_1 \quad \tilde{\mathbf{z}}_2],\ \tilde{\mathbf{W}}=\tilde{\mathbf{Y}}\tilde{\mathbf{V}}=\begin{bmatrix}0 & 0\\ \frac{3}{\sqrt{29}} & -\frac{3}{\sqrt{5}}\\ -\frac{3}{\sqrt{29}} & -\frac{3}{\sqrt{5}}\end{bmatrix}=[\tilde{\mathbf{w}}_1 \quad \tilde{\mathbf{w}}_2]$$

典型相关分析

例4 试用典型相关分析评估 $\{\mathbf{x}_1,\mathbf{x}_2\},\{\mathbf{y}_1,\tilde{\mathbf{y}}_2\}$两组向量的相关性，其中

$$\mathbf{x}_1=\begin{bmatrix}\frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} & 0\end{bmatrix}^{\mathrm{T}},\mathbf{x}_2=\begin{bmatrix}\frac{1}{\sqrt{6}} & \frac{1}{\sqrt{6}} & -\frac{2}{\sqrt{6}}\end{bmatrix}^{\mathrm{T}},\mathbf{y}_1=\begin{bmatrix}0 & 1 & 0\end{bmatrix}^{\mathrm{T}},\tilde{\mathbf{y}}_2=\begin{bmatrix}0 & 1 & \frac{3}{2}\end{bmatrix}^{\mathrm{T}}$$

简单验证可知，$\tilde{\mathbf{z}}_1$和$\tilde{\mathbf{w}}_1$的夹角余弦为1，因此典型相关分析能够克服互相关分析的不足，从而真正反映两组向量张成的平面的相关关系。

此外，容易验证，典型相关分析不是正交变换，但对于同一组向量得到的各个成分之间却是正交的。

典型相关分析

典型相关分析去除高光谱图像中的条带噪声

高光谱图像的简化噪声模型

$$\widetilde{\mathcal{X}} = \mathcal{X} + \mathcal{S} + \mathcal{N}_x$$

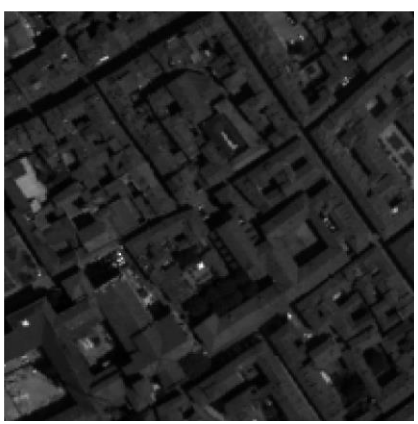
$$\widetilde{\mathcal{Y}} = \mathcal{Y} + \mathcal{S} + \mathcal{N}_y$$

$\widetilde{\mathcal{X}}$ $\widetilde{\mathcal{Y}}$ ：两幅相近的时间内采集的高光谱图像

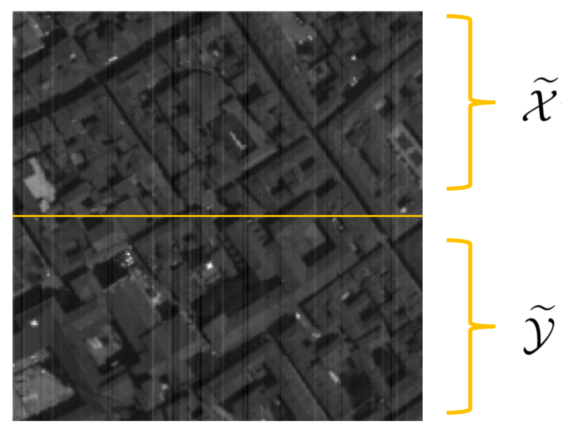
$\mathcal{N}_x, \mathcal{N}_y$ 随机噪声

$\mathcal{S}$ 是两幅图像中条带噪声。

示例：向80波段的高光谱图像的任意4个波段添加模拟条带噪声。



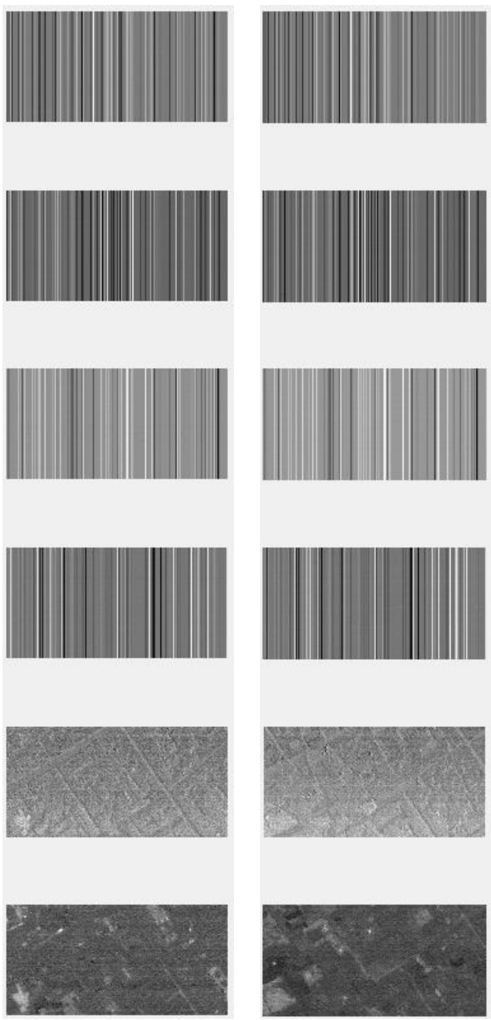
无条带波段



模拟条带波段

典型相关分析

两幅图像相关投影结果及相关系数：



0.9992

0.9989

0.9979

0.9976

条带噪声集中，真实信号较少。

0.3508

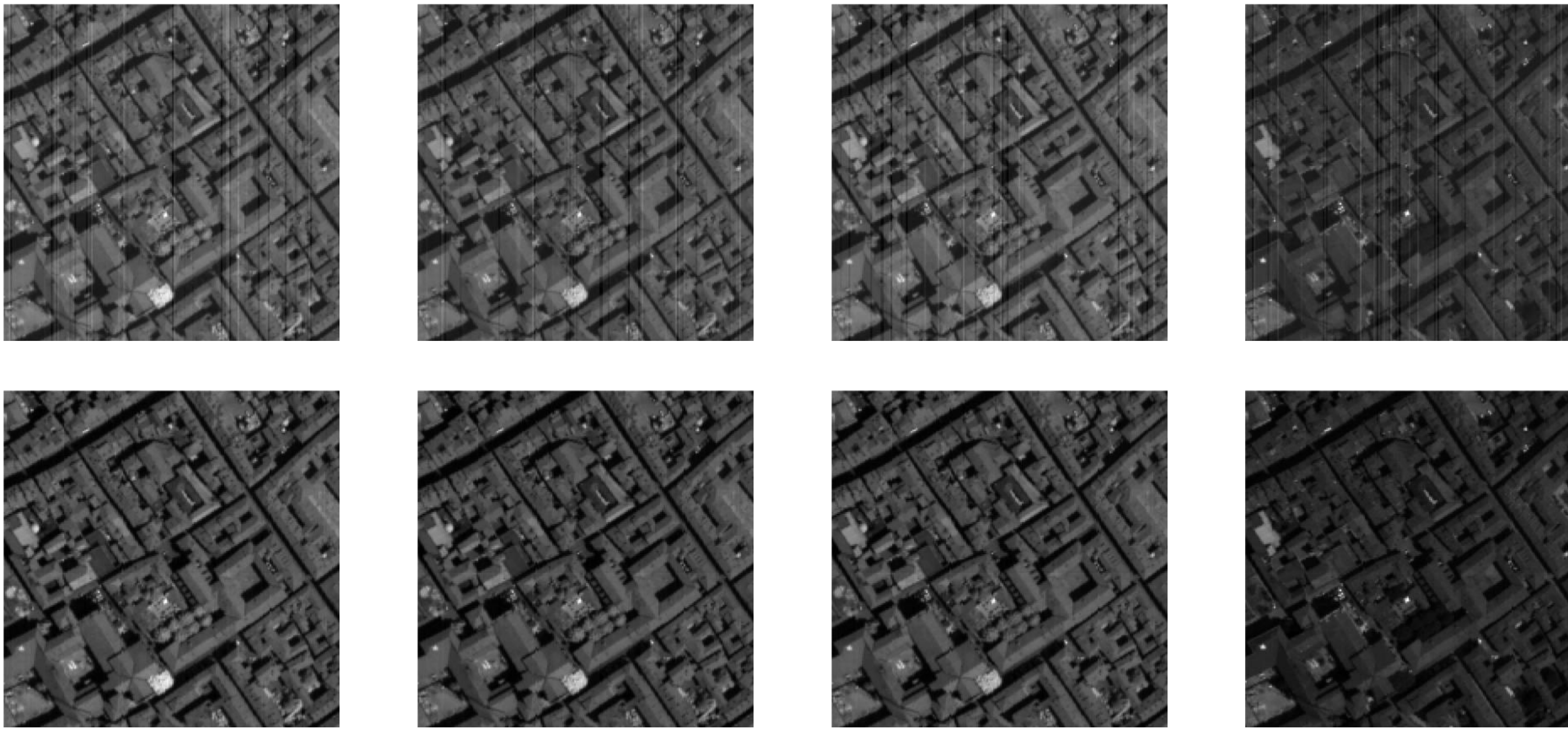
0.2118

⋮

真实信号集中，条带噪声较少。

典型相关分析

去噪后的投影，反投影回原空间得到最终去噪结果：



第一行为噪声波段，第二行为对应的去噪结果

典型相关分析与互相关分析

定理1 典型相关分析等价于白化后的互相关分析

尝试证明？

定理2 典型相关分析的每一组向量的各个成分必然相互正交

尝试证明？

典型相关分析的几何解释

幂法

假定 $n\times n$ 方阵 $\mathbf{A}$ 有 n 个线性无关的特征向量 $\mathbf{u}_1,\mathbf{u}_2,\cdots,\mathbf{u}_n$， 且相应的特征值满足 $\lambda_1>|\lambda_2|\geq\cdots\geq|\lambda_n|$ ， 对于任意的 n 维向量 $\mathbf{x}$，均可由 $\mathbf{A}$ 的特征向量线性表出，即存在 $c_1,c_2,\cdots,c_n$， 使得

$$\mathbf{x}=c_1\mathbf{u}_1+c_2\mathbf{u}_2+\cdots+c_n\mathbf{u}_n$$

故

$$\mathbf{Ax}=c_1\lambda_1\mathbf{u}_1+c_2\lambda_2\mathbf{u}_2+\cdots+c_n\lambda_n\mathbf{u}_n$$

从而

$$\mathbf{A}^k\mathbf{x}=c_1\lambda_1^k\mathbf{u}_1+c_2\lambda_2^k\mathbf{u}_2+\cdots+c_n\lambda_n^k\mathbf{u}_n=\lambda_1^k\left(c_1\mathbf{u}_1+c_2\frac{\lambda_2^k}{\lambda_1^k}\mathbf{u}_2+\cdots+c_n\frac{\lambda_n^k}{\lambda_1^k}\mathbf{u}_n\right)$$

因而有

$$\frac{\mathbf{A}^k\mathbf{x}}{\lambda_1^k}\xrightarrow{k\to\infty}c_1\mathbf{u}_1$$

当 k 足够大时

$$\lambda_1=\frac{\left\|\mathbf{A}^k\mathbf{x}\right\|}{\left\|\mathbf{A}^{k-1}\mathbf{x}\right\|},\ \mathbf{u}_1=\mathbf{A}^k\mathbf{x}\text{或者}\mathbf{u}_1=\frac{\mathbf{A}^k\mathbf{x}}{\left\|\mathbf{A}^k\mathbf{x}\right\|}$$

典型相关分析的几何解释

几何解释

根据前面的分析，典型相关分析可以归结为 $\mathbf{\Sigma}_{\mathrm{I}}$，$\mathbf{\Sigma}_{\mathrm{II}}$ 两个矩阵的特征值与特征向量问题，以 $\mathbf{\Sigma}_{\mathrm{I}}$ 为例，接下来我们用幂法计算该矩阵的最大的特征值所对应的特征向量。任意给一个初始向量 $\mathbf{u}^{(0)}$，幂法的本质就是给一个足够大的 k，使得 $\mathbf{u}^{(k)}=\mathbf{\Sigma}_{\mathrm{I}}^{k}\mathbf{u}^{(0)}$ 逐渐趋向于 $\mathbf{\Sigma}_{\mathrm{I}}$ 的第一个特征向量 $\mathbf{u}_1$。当用这个特征向量对数据 $\mathbf{X}$ 进行线性组合，则可近似得到 $\mathbf{X}$ 的第一个典型成分

$$\mathbf{X}\mathbf{u}^{(k)}=\mathbf{X}\mathbf{\Sigma}_{\mathrm{I}}^{k}\mathbf{u}^{(0)}$$

鉴于

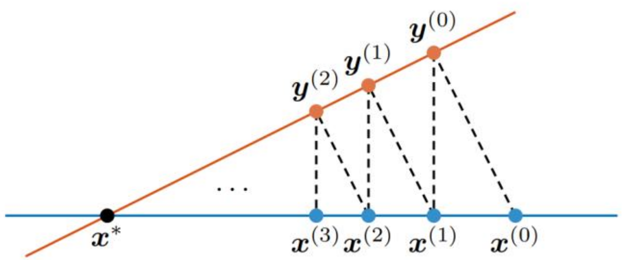
$$\begin{aligned}\mathbf{X}\mathbf{\Sigma}_{\mathrm{I}}^{k}\mathbf{u}^{(0)}&=\mathbf{X}\left(\mathbf{\Sigma}_{11}^{-1}\mathbf{\Sigma}_{12}\mathbf{\Sigma}_{22}^{-1}\mathbf{\Sigma}_{21}\right)^{k}\mathbf{u}^{(0)}=\mathbf{X}\left(\left(\mathbf{X}^{\mathrm{T}}\mathbf{X}\right)^{-1}\mathbf{X}^{\mathrm{T}}\mathbf{Y}\left(\mathbf{Y}^{\mathrm{T}}\mathbf{Y}\right)^{-1}\mathbf{Y}^{\mathrm{T}}\mathbf{X}\right)^{k}\mathbf{u}^{(0)}\\&=\mathbf{X}\left(\mathbf{X}^{\#}\mathbf{Y}\mathbf{Y}^{\#}\mathbf{X}\right)^{k}\mathbf{u}^{(0)}=\mathbf{X}\left(\mathbf{X}^{\#}\mathbf{P}_{\mathrm{Y}}\left(\mathbf{P}_{\mathrm{X}}\mathbf{P}_{\mathrm{Y}}\right)^{k-1}\mathbf{X}\right)\mathbf{u}^{(0)}=\left(\mathbf{P}_{\mathrm{X}}\mathbf{P}_{\mathrm{Y}}\right)^{k}\mathbf{X}\mathbf{u}^{(0)}\end{aligned}$$

其中 $\mathbf{X}^{\#}=\left(\mathbf{X}^{\mathrm{T}}\mathbf{X}\right)^{-1}\mathbf{X}^{\mathrm{T}},\mathbf{Y}^{\#}=\left(\mathbf{Y}^{\mathrm{T}}\mathbf{Y}\right)^{-1}\mathbf{Y}^{\mathrm{T}}$ 分别为 $\mathbf{X},\mathbf{Y}$ 的广义逆，$\mathbf{P}_{\mathrm{X}}=\mathbf{X}\mathbf{X}^{\#},\mathbf{P}_{\mathrm{Y}}=\mathbf{Y}\mathbf{Y}^{\#}$ 为由 $\mathbf{X},\mathbf{Y}$ 构建的正交投影矩阵。

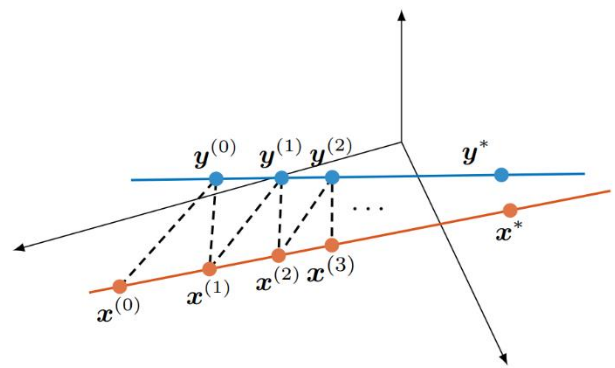
典型相关分析的几何解释

几何解释

因此，典型相关分析在几何上就是在两个超平面交替投影寻找最优解的过程。



超平面相交情形



超平面不相交情形

定理3 假设 Ω_1,Ω_2 是 $\mathbb{R}^n$ 中的两个相交的闭凸集，对于任意的 $\mathbf{x}^{(0)} \in \Omega_1$，通过交替投影法得到的两个序列 $\mathbf{x}^{(k)} \in \Omega_1, \mathbf{y}^{(k)} \in \Omega_2$ 必然收敛于 $\Omega_1 \cap \Omega_2$ 中的点。

定理4 假设 Ω_1,Ω_2 是 $\mathbb{R}^n$ 中的两个不交的闭凸集，对于任意的 $\mathbf{x}^{(0)} \in \Omega_1$，通过交替投影法得到的两个序列 $\mathbf{x}^{(k)} \in \Omega_1, \mathbf{y}^{(k)} \in \Omega_2$ 必然收敛于两个集合中距离最近的点。

典型相关分析的变形

配对典型相关分析

记 $\mathbf{X}_1, \mathbf{X}_2, \cdots, \mathbf{X}_m$ 是 n 维空间的 m 组向量，其中 $\mathbf{X}_i = \begin{bmatrix} \mathbf{x}_{1i} & \cdots & \mathbf{x}_{p_i i} \end{bmatrix}$ 为第 i 组向量。分别对他们进行线性组合得到 m 个成分

$$\mathbf{z}_1 = \mathbf{X}_1\mathbf{u}_1, \mathbf{z}_2 = \mathbf{X}_2\mathbf{u}_2, \cdots, \mathbf{z}_m = \mathbf{X}_m\mathbf{u}_m$$

为了使得这 m 个成分具有最大的相关性，配对典型相关分析采用如下优化模型来获取相应的系数

$$\max_{\mathbf{u}_1,\cdots,\mathbf{u}_m} \sum_{i,j=1}^{m} \frac{\mathbf{u}_i^{\mathrm{T}} \Sigma_{ij} \mathbf{u}_j}{\sqrt{\mathbf{u}_i^{\mathrm{T}} \Sigma_{ii} \mathbf{u}_i} \sqrt{\mathbf{u}_j^{\mathrm{T}} \Sigma_{jj} \mathbf{u}_j}}$$

该模型可以转化为

$$\begin{cases} \max\limits_{\mathbf{u}_1,\cdots,\mathbf{u}_m} \sum\limits_{i,j=1}^{m} \mathbf{u}_i^{\mathrm{T}} \Sigma_{ij} \mathbf{u}_j \\ s.t.\ \ \mathbf{u}_i^{\mathrm{T}} \Sigma_{ii} \mathbf{u}_i = 1, i = 1, 2, \cdots, m \end{cases}$$

典型相关分析的变形

配对典型相关分析

为了便于求解，将上述模型的约束弱化，转化为

$$\begin{cases} \max\limits_{\mathbf{u}_1,\cdots,\mathbf{u}_m} \sum\limits_{i,j=1}^{m} \mathbf{u}_i^{\mathrm{T}} \Sigma_{ij} \mathbf{u}_j \\ s.t. \ \sum\limits_{i=1}^{m} \mathbf{u}_i^{\mathrm{T}} \Sigma_{ii} \mathbf{u}_i = 1 \end{cases}$$

利用拉格朗日乘子法，该模型的解可以归结为如下广义特征值与特征向量问题

$$\begin{bmatrix} \Sigma_{11} & \cdots & \Sigma_{1m} \\ \vdots & \ddots & \vdots \\ \Sigma_{m1} & \cdots & \Sigma_{mm} \end{bmatrix} \begin{bmatrix} \mathbf{u}_1 \\ \vdots \\ \mathbf{u}_m \end{bmatrix} = \lambda \begin{bmatrix} \Sigma_{11} & & \\ & \ddots & \\ & & \Sigma_{mm} \end{bmatrix} \begin{bmatrix} \mathbf{u}_1 \\ \vdots \\ \mathbf{u}_m \end{bmatrix}$$

尝试推导一下？

典型相关分析的变形

张量典型相关分析

仍记 $\mathbf{x}_1,\mathbf{x}_2,\cdots,\mathbf{x}_m$ 是 n 维空间的 m 组向量，其中 $\mathbf{X}_i=\begin{bmatrix}\mathbf{x}_{1i} & \cdots & \mathbf{x}_{p_i i}\end{bmatrix}$ 为第 i 组向量。分别对他们进行线性组合得到 m 个成分

$$\mathbf{z}_1=\mathbf{X}_1\mathbf{u}_1,\ \mathbf{z}_2=\mathbf{X}_2\mathbf{u}_2,\cdots,\mathbf{z}_m=\mathbf{X}_m\mathbf{u}_m$$

为了使得这 m 个成分具有最大的相关性，张量典型相关分析引入了高阶相关性的概念

$$corr\left(\mathbf{z}_1,\mathbf{z}_2,\cdots,\mathbf{z}_m\right)=\frac{\left(\mathbf{z}_1\odot\mathbf{z}_2\odot\cdots\odot\mathbf{z}_m\right)^{\mathrm{T}}\mathbf{1}}{\sqrt{\mathbf{z}_1^{\mathrm{T}}\mathbf{z}_1}\sqrt{\mathbf{z}_2^{\mathrm{T}}\mathbf{z}_2}\cdots\sqrt{\mathbf{z}_m^{\mathrm{T}}\mathbf{z}_m}}$$

其中

$$\left(\mathbf{z}_1\odot\mathbf{z}_2\odot\cdots\odot\mathbf{z}_m\right)^{\mathrm{T}}\mathbf{1}=C\times_1\mathbf{u}_1\times_2\mathbf{u}_2\cdots\times_m\mathbf{u}_m$$

其中 C 是大小为 $p_1\times p_2\times\cdots\times p_m$ 的 m 阶统计张量。

典型相关分析的变形

张量典型相关分析

张量典型相关分析的优化模型为

$$\begin{cases} \max\limits_{\mathbf{u}_1,\cdots,\mathbf{u}_m} C \times_1 \mathbf{u}_1 \times_2 \mathbf{u}_2 \cdots \times_m \mathbf{u}_m \\ s.t.\ \ \mathbf{u}_i^{\mathrm{T}} \Sigma_{ii} \mathbf{u}_i = 1, i = 1, 2, \cdots, m \end{cases}$$

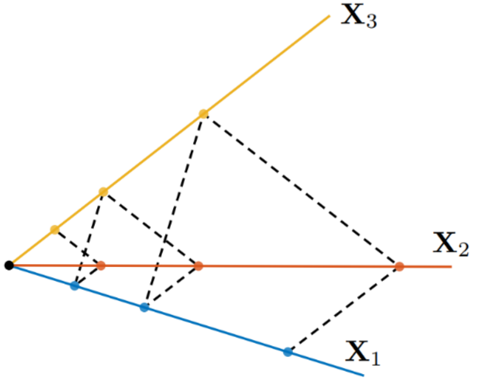
该模型的解归结为对张量 C 的各种分解。

典型相关分析的变形

循环投影法

前面，我们给出了典型相关分析的几何解释，即每一个典型相关成分的求解都相当于在两组向量张成的超平面的交替投影。自然而然地，对于n 维空间的m组向量 $\mathbf{x}_1,\mathbf{x}_2,\cdots,\mathbf{x}_m$，我们也可以把典型相关成分的求解归结为在这$m$ 个超平面上的循环投影。相应的特征值与特征向量问题为

$$\begin{cases} \boldsymbol{\Sigma}_{11}^{-1}\boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}\boldsymbol{\Sigma}_{23}\cdots\boldsymbol{\Sigma}_{m-1,m-1}^{-1}\boldsymbol{\Sigma}_{m-1,m}\boldsymbol{\Sigma}_{mm}^{-1}\boldsymbol{\Sigma}_{m1}\mathbf{u}_1 = \lambda\mathbf{u}_1 \\ \boldsymbol{\Sigma}_{22}^{-1}\boldsymbol{\Sigma}_{23}\boldsymbol{\Sigma}_{33}^{-1}\boldsymbol{\Sigma}_{34}\cdots\boldsymbol{\Sigma}_{mm}^{-1}\boldsymbol{\Sigma}_{m1}\boldsymbol{\Sigma}_{11}^{-1}\boldsymbol{\Sigma}_{12}\mathbf{u}_2 = \lambda\mathbf{u}_2 \\ \vdots \\ \boldsymbol{\Sigma}_{mm}^{-1}\boldsymbol{\Sigma}_{m1}\boldsymbol{\Sigma}_{11}^{-1}\boldsymbol{\Sigma}_{12}\cdots\boldsymbol{\Sigma}_{m-2,m-2}^{-1}\boldsymbol{\Sigma}_{m-2,m-1}\boldsymbol{\Sigma}_{m-1,m-1}^{-1}\boldsymbol{\Sigma}_{m-1,m}\mathbf{u}_m = \lambda\mathbf{u}_m \end{cases}$$



典型相关分析是循环投影法的特例，但典型相关分析的特征值必然为实数，而上述特征分析问题的特征值却未必均为实数。

小结

1. 互相关分析以内积为目标函数研究两组向量之间的相关关系。

2. 典型相关分析以角度（相关系数）为目标函数研究两组向量之间的相关关系。

3.互相关分析的特征向量构成正交矩阵，但同一组向量的不同的互相关成分之间未必正交；典型相关分析的特征向量一般情况下不构成正交矩阵，但同一组向量的不同的典型相关成分必然正交。

4. 典型相关分析等价于自化后的互相关分析。

5. 在几何上，典型相关分析的求解过程可以用交替投影法来解释。

谢　谢

耿修瑞

中国科学院空天信息创新研究院

gengxr@sina.com.cn