第十八讲 典则相关分析CCA

2024.5.13

PCA和CCA都是投影极大化、对角化

PCA: $\mathrm{var}(\mathbf{v}^\top \mathbf{x}) = \mathbf{v}^\top \Sigma_{\mathbf{xx}} \mathbf{v} = \max!$

$$\Sigma_{\mathbf{xx}} = V \Lambda V^\top \Rightarrow V^\top \Sigma_{\mathbf{xx}} V = \Lambda$$

CCA: $\mathrm{cov}(\mathbf{u}^\top \mathbf{x}, \mathbf{v}^\top \mathbf{y}) = \mathbf{u}^\top \Sigma_{\mathbf{xy}} \mathbf{v} = \max!$

$$\Sigma_{\mathbf{xy}} = U D V^\top \Rightarrow U^\top \Sigma_{\mathbf{xy}} V = D$$

Recap

引理1. 假设矩阵$A_{q\times p}$的秩为r,其奇异值分解$A=UDV^{\mathsf T}$,
其中$D_{r\times r}=\mathrm{diag}(\sqrt{\lambda_1},\ldots,\sqrt{\lambda_r})$, $\lambda_1\ge\ldots\ge\lambda_r>0$为$A^{\mathsf T}A$的非0特征根,
$U^{\mathsf T}U=V^{\mathsf T}V=I_r, U_{q\times r}=(\mathbf{u}_1,\ldots,\mathbf{u}_r), V_{p\times r}=(\mathbf{v}_1,\ldots,\mathbf{v}_r)$。则

$$\max_{\substack{\mathbf{u}\in R^q,\mathbf{v}\in R^p\\ ||\mathbf{u}||=1,||\mathbf{v}||=1}} \mathbf{u}^{\mathsf T}A\mathbf{v}=\sqrt{\lambda_1},$$

其中极大值在$\mathbf{u}=\mathbf{u}_1,\mathbf{v}=\mathbf{v}_1$时达到。进一步，对任何$k\le r$,
限制$\mathbf{u}\perp\mathbf{u}_1,\ldots,\mathbf{u}_{k-1}$， $\mathbf{v}\perp\mathbf{v}_1,\ldots,\mathbf{v}_{k-1}$,则(*)在$\mathbf{u}=\mathbf{u}_k,\mathbf{v}=\mathbf{v}_k$时
达到最大,最大值等于第k大奇异值，即 $\max\limits_{\mathbf{u}\perp\mathbf{u}_1,\ldots,\mathbf{u}_{k-1},\ \mathbf{v}\perp\mathbf{v}_1,\ldots,\mathbf{v}_{k-1}} \mathbf{u}^{\mathsf T}A\mathbf{v}=\sqrt{\lambda_k}$。

引理1的证明1：由Cauchy-Schwarz不等式，

$$\mathbf{u}^{\mathsf T}A\mathbf{v}\le\sqrt{\mathbf{u}^{\mathsf T}\mathbf{u}}\sqrt{\mathbf{v}^{\mathsf T}A^{\mathsf T}A\mathbf{v}}=\sqrt{\mathbf{v}^{\mathsf T}A^{\mathsf T}A\mathbf{v}}$$

等号成立当且仅当 $\mathbf{u}\propto A\mathbf{v}$。由第9讲定理1，对任何$\mathbf{v}\in R^p$ $(\mathbf{v}^{\mathsf T}\mathbf{v}=1)$

$$\sqrt{\mathbf{v}^{\mathsf T}A^{\mathsf T}A\mathbf{v}}\le\sqrt{\lambda_1(A^{\mathsf T}A)},$$

等号成立当且仅当$\mathbf{v}=\mathbf{v}_1$（$A^{\mathsf T}A$的最大特征根λ_1对应的特征向量),
此时，$\mathbf{u}\propto A\mathbf{v}_1=\mathbf{u}_1\sqrt{\lambda_1}$,所以$\mathbf{u}=\mathbf{u}_1$。依次类推，其它证明类似。

典则相关分析（CCA）

典则相关分析（Canonical correlation analysis,CCA）研究**两个随机向量**之间的相关性，它试图在两个向量内部分别发现最能代表两组之间相关性的线性组合。CCA由H. Hotelling提出。

例1. 140 个七年级学生进行了4项测试，

$$x_1 = \text{reading speed},\ x_2 = \text{reading power};$$
$$y_1 = \text{arithmetic speed},\ y_2 = \text{arithmetic power};$$

相关系数矩阵如下：

$$\Sigma = \begin{array}{c c} & \begin{array}{cccc} x_1 & x_2 & y_1 & y_2 \end{array} \\ \begin{array}{c} x_1 \\ x_2 \\ y_1 \\ y_2 \end{array} & \begin{pmatrix} 1.00 & 0.63 & 0.24 & 0.06 \\ & 1.00 & -0.06 & 0.07 \\ & & 1.00 & 0.42 \\ & & & 1.00 \end{pmatrix} \end{array} = \begin{pmatrix} \Sigma_{\mathbf{xx}} & \Sigma_{\mathbf{xy}} \\ \Sigma_{\mathbf{yx}} & \Sigma_{\mathbf{yy}} \end{pmatrix},$$

- ❑ 阅读$\mathbf{x} = (x_1, x_2)^\top$与数学$\mathbf{y} = (y_1, y_2)^\top$的协方差$\Sigma_{\mathbf{xy}}$是一个矩阵，如何用一个实数度量它们之间的相关性？
- ❑ 如果$\mathbf{x}, \mathbf{y}$高度相关，原因是什么 （是因为x_2与y_1强相关?）

问题的提出

假设 $\begin{pmatrix}\mathbf{x}_{p\times 1}\\ \mathbf{y}_{q\times 1}\end{pmatrix}=\begin{pmatrix}\Sigma_{\mathbf{xx}} & \Sigma_{\mathbf{xy}}\\ \Sigma_{\mathbf{yx}} & \Sigma_{\mathbf{yy}}\end{pmatrix}$，$\mathbf{x},\mathbf{y}$ 在这哪些方向上相关性最大?

第1步:

求解方向$\mathbf{a}\in R^p$和$\mathbf{b}\in R^q$，使得投影坐标$\mathbf{a}^\top\mathbf{x}$和$\mathbf{b}^\top\mathbf{y}$之间的相关性最大：

$$corr(\mathbf{a}^\top\mathbf{x},\mathbf{b}^\top\mathbf{y})=\mathbf{a}^\top\Sigma_{\mathbf{xy}}\mathbf{b}=\max!$$

为了解唯一，约束
$\mathrm{var}(\mathbf{a}^\top\mathbf{x})=\mathbf{a}^\top\Sigma_{\mathbf{xx}}\mathbf{a}=1$,
$\mathrm{var}(\mathbf{b}^\top\mathbf{y})=\mathbf{b}^\top\Sigma_{\mathbf{yy}}\mathbf{b}=1$

记最优方向为$\mathbf{a}_1$和$\mathbf{b}_1$，我们称$\xi_1=\mathbf{a}_1^\top\mathbf{x},\eta_1=\mathbf{b}_1^\top\mathbf{y}$ 为第一对典则变量，最大值为第一典则相关系数。

第k步：$2\le k\le r=rank(\Sigma_{\mathbf{xy}})$

在$\mathbf{a}^\top\mathbf{x}$与$\xi_1,\dots,\xi_{k-1}$不相关、$\mathbf{b}^\top\mathbf{y}$与$\eta_1,\dots,\eta_{k-1}$不相关以及单位方差的约束下, 求解

$$corr(\mathbf{a}^\top\mathbf{x},\mathbf{b}^\top\mathbf{y})=\max!$$

约束：$\mathbf{a}^\top\Sigma_{\mathbf{xx}}\mathbf{a}=\mathbf{b}^\top\Sigma_{\mathbf{yy}}\mathbf{b}=1$
$\mathrm{cov}(\mathbf{a}^\top\mathbf{x},\mathbf{a}_i^\top\mathbf{x})=\mathbf{a}^\top\Sigma_{\mathbf{xx}}\mathbf{a}_i=0$
$\mathrm{cov}(\mathbf{b}^\top\mathbf{y},\mathbf{b}_i^\top\mathbf{y})=\mathbf{b}^\top\Sigma_{\mathbf{yy}}\mathbf{b}_i=0$

记最优解为$\mathbf{a}_k$和$\mathbf{b}_k$，我们称$\xi_k=\mathbf{a}_k^\top\mathbf{x}$，$\eta_k=\mathbf{b}_k^\top\mathbf{y}$为第$k$对典则变量，最大值为第$k$典则相关系数。

典则相关系数和典则变量

命题1（课本Result10.1）假设$\Sigma = \mathrm{cov}\begin{pmatrix} \mathbf{x}_{p\times 1} \\ \mathbf{y}_{q\times 1} \end{pmatrix} = \begin{pmatrix} \Sigma_{\mathbf{xx}} & \Sigma_{\mathbf{xy}} \\ \Sigma_{\mathbf{yx}} & \Sigma_{\mathbf{yy}} \end{pmatrix}$,

假设$\Sigma_{\mathbf{xy}}$的标准化矩阵的奇异值分解为

$$A \overset{\Delta}{=} \Sigma_{\mathbf{xx}}^{-1/2}\Sigma_{\mathbf{xy}}\Sigma_{\mathbf{yy}}^{-1/2} = UDV^{\mathsf{T}}, \quad r = rank(A),$$

其中$D = diag(\sqrt{\lambda_1},\ldots,\sqrt{\lambda_r}), \lambda_1 \ge \lambda_2 \ge \ldots \ge \lambda_r > 0$为$\Phi = AA^{\mathsf{T}}$或$\Psi = A^{\mathsf{T}}A$的共同非零特征根，$U = (\mathbf{u}_1,\ldots,\mathbf{u}_r)$，$V = (\mathbf{v}_1,\ldots,\mathbf{v}_r)$，$U^{\mathsf{T}}U = V^{\mathsf{T}}V = I_r$,

则第k典则相关系数为$\sqrt{\lambda_k}$，第k对典则变量

$$(\xi_k, \eta_k) = (\mathbf{u}_k^{\mathsf{T}}\Sigma_{\mathbf{xx}}^{-1/2}\mathbf{x}, \mathbf{v}_k^{\mathsf{T}}\Sigma_{\mathbf{yy}}^{-1/2}\mathbf{y}), \quad k = 1,\ldots,r,$$

所有典则变量分别为

$$\mathbf{x}_{cca} = (\xi_1,\ldots,\xi_r)^{\mathsf{T}} = U^{\mathsf{T}}\Sigma_{\mathbf{xx}}^{-1/2}\mathbf{x},$$

$$\mathbf{y}_{cca} = (\eta_1,\ldots,\eta_r)^{\mathsf{T}} = V^{\mathsf{T}}\Sigma_{\mathbf{yy}}^{-1/2}\mathbf{y},$$

且$\mathrm{cov}\begin{pmatrix} \mathbf{x}_{cca} \\ \mathbf{y}_{cca} \end{pmatrix} = \begin{pmatrix} I_r & D \\ D & I_r \end{pmatrix}$。

证明：在$\mathbf{a}^{\top}\Sigma_{\mathbf{xx}}\mathbf{a}=\mathbf{b}^{\top}\Sigma_{\mathbf{yy}}\mathbf{b}=1$约束下，令$\mathbf{u}=\Sigma_{\mathbf{xx}}^{1/2}\mathbf{a},\mathbf{v}=\Sigma_{\mathbf{yy}}^{1/2}\mathbf{b}$，则$\|\mathbf{u}\|=\|\mathbf{v}\|=1$,

$$\rho(\mathbf{a},\mathbf{b})=\mathbf{a}^{\top}\Sigma_{\mathbf{xy}}\mathbf{b}=\mathbf{u}^{\top}\Sigma_{\mathbf{xx}}^{-1/2}\Sigma_{\mathbf{xy}}\Sigma_{\mathbf{yy}}^{-1/2}\mathbf{v}\overset{\Delta}{=}\mathbf{u}^{\top}A\mathbf{v}$$

由引理1，最大值为$A=\Sigma_{\mathbf{xx}}^{-1/2}\Sigma_{\mathbf{xy}}\Sigma_{\mathbf{yy}}^{-1/2}$的最大奇异值

$$\max\rho(\mathbf{a},\mathbf{b})=\sqrt{\lambda_1(A^{\top}A)}=\sqrt{\lambda_1}\qquad\text{(第一典则相关系数)}$$

上述最大值在$\mathbf{u}=\mathbf{u}_1,\mathbf{v}=\mathbf{v}_1$达到，即$\mathbf{a}=\Sigma_{\mathbf{xx}}^{-1/2}\mathbf{u}_1\overset{\Delta}{=}\mathbf{a}_1,\mathbf{b}=\Sigma_{\mathbf{yy}}^{-1/2}\mathbf{v}_1\overset{\Delta}{=}\mathbf{b}_1$时达到，

第一对典则变量：$\xi_1=\mathbf{a}_1^{\top}\mathbf{x}=\mathbf{u}_1^{\top}\Sigma_{\mathbf{xx}}^{-1/2}\mathbf{x}$，$\eta_1=\mathbf{b}_1^{\top}\mathbf{x}=\mathbf{v}_1^{\top}\Sigma_{\mathbf{yy}}^{-1/2}\mathbf{y}$，且

$$\mathrm{cov}\begin{pmatrix}\xi_1\\ \eta_1\end{pmatrix}=\begin{pmatrix}1&\sqrt{\lambda_1}\\ \sqrt{\lambda_1}&1\end{pmatrix}.$$

现假设$\xi=\mathbf{a}^{\top}\mathbf{x}=\mathbf{u}^{\top}\Sigma_{\mathbf{xx}}^{-1/2}\mathbf{x}$与$\xi_1=\mathbf{a}_1^{\top}\mathbf{x}=\mathbf{u}_1^{\top}\Sigma_{\mathbf{xx}}^{-1/2}\mathbf{x}$不相关，

$\eta=\mathbf{b}^{\top}\mathbf{y}=\mathbf{v}^{\top}\Sigma_{\mathbf{yy}}^{-1/2}\mathbf{y}$与$\eta_1=\mathbf{b}_1^{\top}\mathbf{y}=\mathbf{v}_1^{\top}\Sigma_{\mathbf{yy}}^{-1/2}\mathbf{y}$不相关，即

$$\mathbf{a}^{\top}\Sigma_{\mathbf{xx}}\mathbf{a}_1=\mathbf{v}^{\top}\mathbf{v}_1=0,\quad\mathbf{b}^{\top}\Sigma_{\mathbf{yy}}\mathbf{b}_1=\mathbf{u}^{\top}\mathbf{u}_1=0$$

且$\|\mathbf{u}\|=\|\mathbf{v}\|=1$,

由引理1，

$$\max_{\substack{\|\mathbf{u}\|=\|\mathbf{v}\|=1 \\ \mathbf{u}\perp\mathbf{u}_1,\mathbf{v}\perp\mathbf{v}_1}} \rho(\mathbf{a},\mathbf{b}) = \max_{\substack{\|\mathbf{u}\|=\|\mathbf{v}\|=1 \\ \mathbf{u}\perp\mathbf{u}_1,\mathbf{v}\perp\mathbf{v}_1}} \mathbf{u}^\top A\mathbf{v} = \sqrt{\lambda_2(A^\top A)} \quad \text{(第二典则相关系数)}$$

最大值在$\mathbf{u}=\mathbf{u}_2$，$\mathbf{v}=\mathbf{v}_2$达到，即在$\mathbf{a}_2=\Sigma_{\mathbf{xx}}^{-1/2}\mathbf{u}_2$，$\mathbf{b}_2=\Sigma_{\mathbf{yy}}^{-1/2}\mathbf{v}_2$达到最大。

另外，$\mathrm{cov}(\xi_1,\eta_2)=\mathrm{cov}(\mathbf{u}_1^\top\Sigma_{\mathbf{xx}}^{-1/2}\mathbf{x},\mathbf{v}_2^\top\Sigma_{\mathbf{yy}}^{-1/2}\mathbf{y})=\mathbf{u}_1^\top A\mathbf{v}_2=0$，$\mathrm{cov}(\xi_2,\eta_1)=0$，

$$\Rightarrow \quad \mathrm{cov}\begin{pmatrix}\xi_1\\ \xi_2\\ \eta_1\\ \eta_2\end{pmatrix}=\begin{pmatrix}1 & 0 & \sqrt{\lambda_1} & 0\\ 0 & 1 & 0 & \sqrt{\lambda_2}\\ \sqrt{\lambda_1} & 0 & 1 & 0\\ 0 & \sqrt{\lambda_2} & 0 & 1\end{pmatrix}.$$

依此类推共得到r对典则变量：(ξ_k,η_k), $k=1,\dots,r=\mathrm{rank}(\Sigma_{\mathbf{xy}})$，记

$$\mathbf{x}_{cca}=(\xi_1,\dots,\xi_r)^\top=U^\top\Sigma_{\mathbf{xx}}^{-1/2}\mathbf{x},$$

$$\mathbf{y}_{cca}=(\eta_1,\dots,\eta_r)^\top=V^\top\Sigma_{\mathbf{yy}}^{-1/2}\mathbf{y},$$

则$\mathrm{var}(\mathbf{x}_{cca})=U^\top\Sigma_{\mathbf{xx}}^{-1/2}\Sigma_{\mathbf{xx}}\Sigma_{\mathbf{xx}}^{-1/2}U=U^\top U=I_r$, $\mathrm{var}(\mathbf{y}_{cca})=I_r$,

且$\mathrm{cov}(\mathbf{x}_{cca},\mathbf{y}_{cca})=U^\top\Sigma_{\mathbf{xx}}^{-1/2}\Sigma_{\mathbf{xy}}\Sigma_{\mathbf{yy}}^{-1/2}V=D \Rightarrow \mathrm{cov}\begin{pmatrix}\mathbf{x}_{cca}\\ \mathbf{y}_{cca}\end{pmatrix}=\begin{pmatrix}I_r & D\\ D & I_r\end{pmatrix}$。证毕。

随机向量$\mathbf{y}_{q\times1}$，$\mathbf{x}_{p\times1}$，$\mathrm{cov}\begin{pmatrix}\mathbf{y}\\ \mathbf{x}\end{pmatrix}=\begin{pmatrix}\Sigma_{\mathbf{yy}} & \Sigma_{\mathbf{yx}}\\ \Sigma_{\mathbf{xy}} & \Sigma_{\mathbf{xx}}\end{pmatrix}$.

$\mathbf{y},\mathbf{x}$的标准化：$\Sigma_{\mathbf{yy}}^{-1/2}\mathbf{y}$，$\Sigma_{\mathbf{xx}}^{-1/2}\mathbf{x}$

上述典则求解过程主要包含两部分，总结如下：

- ❖ $\Sigma_{\mathbf{yx}}$的标准化（约束投影为1）：$A=\Sigma_{\mathbf{xx}}^{-1/2}\Sigma_{\mathbf{xy}}\Sigma_{\mathbf{yy}}^{-1/2}$
- ❖ SVD对角化：$A=UDV^\top \Rightarrow U^\top AV=D$

具体过程如下

$$\mathrm{cov}(\mathbf{x},\ \mathbf{y})=\Sigma_{\mathbf{xy}}$$

$$\Rightarrow\ \mathrm{cov}(\Sigma_{\mathbf{xx}}^{-1/2}\mathbf{x},\ \Sigma_{\mathbf{yy}}^{-1/2}\mathbf{y})=\Sigma_{\mathbf{xx}}^{-1/2}\Sigma_{\mathbf{xy}}\Sigma_{\mathbf{yy}}^{-1/2}$$

$$\Rightarrow \text{奇异值分解：}\ \Sigma_{\mathbf{xx}}^{-1/2}\Sigma_{\mathbf{xy}}\Sigma_{\mathbf{yy}}^{-1/2}=UDV^\top,$$

左乘$U^\top$，右乘左V得对角化

$$\begin{aligned}D &= U^\top\Sigma_{\mathbf{xx}}^{-1/2}\Sigma_{\mathbf{xy}}\Sigma_{\mathbf{yy}}^{-1/2}V\\ &= \mathrm{cov}\left(U^\top\Sigma_{\mathbf{xx}}^{-1/2}\mathbf{x},\ V^\top\Sigma_{\mathbf{yy}}^{-1/2}\mathbf{y}\right)\\ &\triangleq \mathrm{cov}(\mathbf{x}_{cca},\mathbf{y}_{cca})\end{aligned}$$

红蓝两个随机向量分别是典则向量

$$\mathbf{x}_{cca} = U^\top \Sigma_{\mathbf{xx}}^{-1/2}\mathbf{x}, \quad \mathbf{y}_{cca} = V^\top \Sigma_{\mathbf{yy}}^{-1/2}\mathbf{y}$$

它们的方差都是单位矩阵,它们的协方差为对角阵D。$\mathbf{x}_{cca}$和$\mathbf{y}_{cca}$的第k个分量构成第k对典则变量

$$\xi_k = \mathbf{u}_k^\top \Sigma_{\mathbf{xx}}^{-1/2}\mathbf{x}, \quad \eta_k = \mathbf{v}_k^\top \Sigma_{\mathbf{yy}}^{-1/2}\mathbf{y}$$

是否可以直接对协方差矩阵$\Sigma_{\mathbf{xy}}$做奇异值分解、实现对角化,得到$\mathbf{x}$，$\mathbf{y}$的线性组合的最大相关性？

可以：

$$\Sigma_{\mathbf{xy}} = \tilde{U}\tilde{D}\tilde{V}^\top \overset{\text{对角化}}{\Rightarrow} \tilde{D} = \tilde{U}^\top \Sigma_{\mathbf{xy}} V = \mathrm{cov}(\tilde{U}^\top \mathbf{x}, \tilde{V}^\top \mathbf{y})$$

其中$\mathrm{cov}(\mathbf{u}_1^\top \mathbf{x}, \mathbf{v}_1^\top \mathbf{y}) = \max\limits_{\|\mathbf{u}\|=\|\mathbf{v}\|=1} \mathrm{cov}(\mathbf{u}^\top \mathbf{x}, \mathbf{v}^\top \mathbf{y})$，等等。

但$\tilde{U}^\top \mathbf{x}, \tilde{V}^\top \mathbf{y}$不是标准/典则的，其方差不是单位阵。

CCA的回归解释

我们知道，相关分析与线性回归联系密切，下面我们从回归的角度解释CCA。

典则相关系数:
多元线性回归的“决定系数”

$A=\Sigma_{\mathbf{xx}}^{-1/2}\Sigma_{\mathbf{xy}}\Sigma_{\mathbf{yy}}^{-1/2}$，$\Phi=A^{\mathsf T}A,\Psi=AA^{\mathsf T}$的含义？为什么其特征根小于1?

$\Phi=A^{\mathsf T}A=\Sigma_{\mathbf{yy}}^{-1/2}\Sigma_{\mathbf{yx}}\Sigma_{\mathbf{xx}}^{-1}\Sigma_{\mathbf{xy}}\Sigma_{\mathbf{yy}}^{-1/2}\le I_q$为$\mathbf{y}$的方差中$\mathbf{x}$所能解释的"比例"。

$\Psi=AA^{\mathsf T}=\Sigma_{\mathbf{xx}}^{-1/2}\Sigma_{\mathbf{xy}}\Sigma_{\mathbf{yy}}^{-1}\Sigma_{\mathbf{yx}}\Sigma_{\mathbf{xx}}^{-1/2}\le I_p$为$\mathbf{x}$的方差中$\mathbf{y}$所能解释的"比例"。

解释如下：将$\mathbf{y}_{q\times 1}$表示成与$\mathbf{x}_{p\times 1}$相关的部分和不相关的部分：

$$\mathbf{y}=B^{\mathsf T}\mathbf{x}+\boldsymbol{\varepsilon}，\boldsymbol{\varepsilon}与\mathbf{x}不相关$$

即多元线性回归模型/去相关化/正交化或正交分解

其中B是$p\times q$是回归系数矩阵。$\boldsymbol{\varepsilon}$与$\mathbf{x}$不相关$\Rightarrow B^{\mathsf T}=\Sigma_{\mathbf{yx}}\Sigma_{\mathbf{xx}}^{-1}$，模型即：

$$\mathbf{y}=B^{\mathsf T}\mathbf{x}+\boldsymbol{\varepsilon}=\Sigma_{\mathbf{yx}}\Sigma_{\mathbf{xx}}^{-1}\mathbf{x}+(\mathbf{y}-\Sigma_{\mathbf{yx}}\Sigma_{\mathbf{xx}}^{-1}\mathbf{x})$$

两边同时取方差，

$$\Sigma_{\mathbf{yy}}=\mathrm{var}(\mathbf{y})=\mathrm{var}(B^{\mathsf T}\mathbf{x})+\mathrm{var}(\boldsymbol{\varepsilon})=\Sigma_{\mathbf{yx}}\Sigma_{\mathbf{xx}}^{-1}\Sigma_{\mathbf{xy}}+\Sigma_{\mathbf{yy}\bullet\mathbf{x}},$$

其中$\Omega=\Sigma_{\mathbf{yx}}\Sigma_{\mathbf{xx}}^{-1}\Sigma_{\mathbf{xy}}$代表$\mathbf{y}$中$\mathbf{x}$能解释的部分，它在$\mathbf{y}$的方差$\Sigma_{\mathbf{yy}}$中所占"比例"

$$\Sigma_{\mathbf{yy}}^{-1/2}\Omega\Sigma_{\mathbf{yy}}^{-1/2}=\Sigma_{\mathbf{yy}}^{-1/2}\left(\Sigma_{\mathbf{yx}}\Sigma_{\mathbf{xx}}^{-1}\Sigma_{\mathbf{xy}}\right)\Sigma_{\mathbf{yy}}^{-1/2}=\Phi$$

互换$\mathbf{x}$，$\mathbf{y}$位置,将$\mathbf{x}$分解为与$\mathbf{y}$相关的和不相关的两部分

$$\mathbf{x} = C^{\mathsf{T}}\mathbf{y} + \boldsymbol{\delta}，\boldsymbol{\delta}与\mathbf{y}不相关$$

则 $\mathrm{var}(\mathbf{x}) = \mathrm{var}(C^{\mathsf{T}}\mathbf{y}) + \mathrm{var}(\boldsymbol{\delta}) = \Sigma_{\mathbf{xy}}\Sigma_{\mathbf{yy}}^{-1}\Sigma_{\mathbf{yx}} + (\Sigma_{\mathbf{xx}} - \Sigma_{\mathbf{xy}}\Sigma_{\mathbf{yy}}^{-1}\Sigma_{\mathbf{yx}})$

其中$\tilde{\Omega} = \Sigma_{\mathbf{xy}}\Sigma_{\mathbf{yy}}^{-1}\Sigma_{\mathbf{yx}}$在$\mathbf{x}$的方差$\Sigma_{\mathbf{xx}}$中所占"比例"

$$\Sigma_{\mathbf{xx}}^{-1/2}\tilde{\Omega}\Sigma_{\mathbf{xx}}^{-1/2} = \Sigma_{\mathbf{xx}}^{-1/2}\left(\Sigma_{\mathbf{xy}}\Sigma_{\mathbf{yy}}^{-1}\Sigma_{\mathbf{yx}}\right)\Sigma_{\mathbf{xx}}^{-1/2} = \Psi,$$

其特征根与Φ相同，都是典则相关系数。

总之，典则相关系数和典则变量关于$\mathbf{x}, \mathbf{y}$对称。

特别地，当$q = 1$时，Φ是实数，Ψ秩1，唯一非0特征根$\lambda_1 = \Phi$，在回归分析中称为决定系数（参见后面“线性回归解释”）。
同样，$p = 1$时, Ψ是实数，Φ秩1，唯一非0特征根$\lambda_1 = \Psi$

2. CCA: 多元回归的消元法/对角化

CCA可看作是通过对多元线性模型（多个方程）的两端进行标准化变换和奇异值分解，将模型重整为若干个简单线性模型的过程，也就是解线性方程组的消元法/对角化过程：

$$q\text{个复杂模型}\,\mathbf{y}=B^{\top}\mathbf{x}+\boldsymbol{\varepsilon}\ \overset{\text{cca}}{\Rightarrow}\ r\text{个简单模型}\,\mathbf{y}_{\text{cca}}=D\,\mathbf{x}_{\text{cca}}+\boldsymbol{\delta}$$

线性模型：$\mathbf{y}=B^{\top}\mathbf{x}+\boldsymbol{\varepsilon}$

标准化: $\tilde{\mathbf{y}}=\Sigma_{\mathbf{yy}}^{-1/2}\,\mathbf{y},\ \tilde{\mathbf{x}}=\Sigma_{\mathbf{xx}}^{-1/2}\,\mathbf{x}$

$$\tilde{\mathbf{y}}=\Sigma_{\mathbf{yy}}^{-1/2}B^{\top}\Sigma_{\mathbf{xx}}^{1/2}\tilde{\mathbf{x}}+\Sigma_{\mathbf{yy}}^{-1/2}\boldsymbol{\varepsilon}=A^{\top}\tilde{\mathbf{x}}+\tilde{\boldsymbol{\varepsilon}}$$

SVD：$A=\Sigma_{\mathbf{xx}}^{-1/2}\Sigma_{\mathbf{xy}}\Sigma_{\mathbf{yy}}^{-1/2}=UDV^{\top}$

$$\tilde{\mathbf{y}}=VDU^{\top}\tilde{\mathbf{x}}+\tilde{\boldsymbol{\varepsilon}}\Rightarrow V^{\top}\tilde{\mathbf{y}}=DU^{\top}\tilde{\mathbf{x}}+V^{\top}\tilde{\boldsymbol{\varepsilon}}$$

记$\mathbf{y}_{cca}=V^{\top}\tilde{\mathbf{y}},\ \mathbf{x}_{cca}=U^{\top}\tilde{\mathbf{x}}$

$$\mathbf{y}_{cca}=D\mathbf{x}_{cca}+\boldsymbol{\delta}\Leftrightarrow\begin{cases}\eta_1=\sqrt{\lambda_1}\xi_1+\delta_1\\ \vdots\\ \eta_r=\sqrt{\lambda_r}\xi_r+\delta_r\end{cases}$$

原始的q元线性模型等价于r个一元简单回归模型：

$$\mathbf{y} = B^\top \mathbf{x} + \boldsymbol{\varepsilon} \Leftrightarrow \mathbf{y}_{cca} = D\mathbf{x}_{cca} + \boldsymbol{\delta} \Leftrightarrow \begin{cases} \eta_1 = \sqrt{\lambda_1}\xi_1 + \delta_1 \\ \vdots \\ \eta_r = \sqrt{\lambda_r}\xi_r + \delta_r \end{cases} \Big\Downarrow$$

重要性下降

- 一元(一个响应)简单(一个自变量)回归模型

$$\eta_i = \sqrt{\lambda_i}\xi_i + \delta_i$$

中η_i，ξ_i都是标准化的（均值为0，方差为1）.

- 回归系数 $\sqrt{\lambda_i} = \mathrm{corr}(\eta_i, \xi_i)$.
- r个方程的决定系数从大到小排列：$\lambda_1 \geq \cdots \geq \lambda_r$.
- 可视化: 如果用一个模型替代原始的q个模型，则取

$$\eta_1 = \sqrt{\lambda_1}\xi_1 + \delta_1,\ \text{即}\ \mathbf{u}_1^\top \Sigma_{\mathbf{yy}}^{-1/2}\mathbf{y} = \sqrt{\lambda_1}\mathbf{v}_1^\top \Sigma_{\mathbf{xx}}^{-1/2}\mathbf{x} + \delta_1$$

3. CCA 与分块对角化

$$\mathrm{var}\begin{pmatrix}\mathbf{x}_{cca}\\ \mathbf{y}_{cca}\end{pmatrix}=\begin{pmatrix}I_r & D\\ D & I_r\end{pmatrix}\Leftrightarrow\begin{pmatrix}U^{\mathsf T}\Sigma_{\mathbf{xx}}^{-1/2} & \\ & V^{\mathsf T}\Sigma_{\mathbf{yy}}^{-1/2}\end{pmatrix}\begin{pmatrix}\Sigma_{\mathbf{xx}} & \Sigma_{\mathbf{xy}}\\ \Sigma_{\mathbf{yx}} & \Sigma_{\mathbf{yy}}\end{pmatrix}\begin{pmatrix}\Sigma_{\mathbf{xx}}^{-1/2}U & \\ & \Sigma_{\mathbf{yy}}^{-1/2}V\end{pmatrix}=\begin{pmatrix}I_r & D\\ D & I_r\end{pmatrix}$$

CCA作为分块对角化代数问题的描述：

给定矩阵$\Sigma=\begin{pmatrix}\Sigma_{\mathbf{xx}} & \Sigma_{\mathbf{xy}}\\ \Sigma_{\mathbf{yx}} & \Sigma_{\mathbf{yy}}\end{pmatrix}$，求解$A,B$使得四块都是对角矩阵：

$$\begin{pmatrix}A & \\ & B\end{pmatrix}\begin{pmatrix}\Sigma_{\mathbf{xx}} & \Sigma_{\mathbf{xy}}\\ \Sigma_{\mathbf{yx}} & \Sigma_{\mathbf{yy}}\end{pmatrix}\begin{pmatrix}A^{\mathsf T} & \\ & B^{\mathsf T}\end{pmatrix}=\begin{pmatrix}A\Sigma_{\mathbf{xx}}A^{\mathsf T} & A\Sigma_{\mathbf{xy}}B^{\mathsf T}\\ B\Sigma_{\mathbf{yx}}A^{\mathsf T} & B\Sigma_{\mathbf{yy}}B^{\mathsf T}\end{pmatrix}$$

解：

- 先将主对角阵单位化：

$$\begin{pmatrix}\Sigma_{\mathbf{xx}}^{-1/2} & \mathbf{0}\\ \mathbf{0} & \Sigma_{\mathbf{yy}}^{-1/2}\end{pmatrix}\begin{pmatrix}\Sigma_{\mathbf{xx}} & \Sigma_{\mathbf{xy}}\\ \Sigma_{\mathbf{yx}} & \Sigma_{\mathbf{yy}}\end{pmatrix}\begin{pmatrix}\Sigma_{\mathbf{xx}}^{-1/2} & \mathbf{0}\\ \mathbf{0} & \Sigma_{\mathbf{yy}}^{-1/2}\end{pmatrix}=\begin{pmatrix}I_p & \Sigma_{\mathbf{xx}}^{-1/2}\Sigma_{\mathbf{xy}}\Sigma_{\mathbf{yy}}^{-1/2}\\ \Sigma_{\mathbf{yy}}^{-1/2}\Sigma_{\mathbf{yx}}\Sigma_{\mathbf{xx}}^{-1/2} & I_q\end{pmatrix}$$

- 应用SVD：$\Sigma_{\mathbf{xx}}^{-1/2}\Sigma_{\mathbf{xy}}\Sigma_{\mathbf{yy}}^{-1/2}=U_{p\times r}D_{r\times r}V^{\mathsf T}{}_{r\times q},D=diag(\sqrt{\lambda_1},\dots,\sqrt{\lambda_r})$

  得分块对角化：$\begin{pmatrix}U^{\mathsf T}\Sigma_{\mathbf{xx}}^{-1/2} & \\ & V^{\mathsf T}\Sigma_{\mathbf{yy}}^{-1/2}\end{pmatrix}\begin{pmatrix}\Sigma_{\mathbf{xx}} & \Sigma_{\mathbf{xy}}\\ \Sigma_{\mathbf{yx}} & \Sigma_{\mathbf{yy}}\end{pmatrix}\begin{pmatrix}\Sigma_{\mathbf{xx}}^{-1/2}U & \\ & \Sigma_{\mathbf{yy}}^{-1/2}V\end{pmatrix}=\begin{pmatrix}I_r & D\\ D & I_r\end{pmatrix}$,

- 但上述对角阵的大小与原矩阵不同，将U,V分别补全成正交阵$\tilde{U}_{p\times p},\tilde{V}_{q\times q}$即可。

样本典则相关分析

样本CCA

数据： $X_{n\times p}=\begin{pmatrix}\mathbf{x}_1{}^{\top}\\ \\ \mathbf{x}_n{}^{\top}\end{pmatrix}$, $Y_{n\times q}=\begin{pmatrix}\mathbf{y}_1{}^{\top}\\ \\ \mathbf{y}_n{}^{\top}\end{pmatrix}$, 假设都已经中心化。

样本协方差矩阵： $S=\begin{pmatrix}S_{\mathbf{xx}} & S_{\mathbf{xy}}\\ S_{\mathbf{yx}} & S_{\mathbf{yy}}\end{pmatrix}$,

令$A=S_{\mathbf{xx}}^{-1/2}S_{\mathbf{xy}}S_{\mathbf{yy}}^{-1/2}\overset{\text{svd}}{=}UDV^{\top}$

奇异值分解 : $A=S_{\mathbf{yy}}^{-1/2}S_{\mathbf{yx}}S_{\mathbf{xx}}^{-1/2}=UDV^{\top}$,

其中$U=(\mathbf{u}_1,\ldots,\mathbf{u}_r),V=(\mathbf{v}_1,\ldots,\mathbf{v}_r),U^{\top}U=V^{\top}V=I_r$,

$D=diag(\lambda_1,\ldots,\lambda),\lambda_1\geq\ldots\geq\lambda_r>0.$

- 对$i=1,\ldots,n$, 第k对典则变量 $\xi_{ik}=\mathbf{u}_{\mathrm{k}}{}^{\top}S_{\mathbf{xx}}^{-1/2}\mathbf{x}_i$, $\eta_{ik}=\mathbf{v}_k{}^{\top}S_{\mathbf{yy}}^{-1/2}\mathbf{y}_i$,

  所有样本的第k对典则变量$\boldsymbol{\xi}_k=\begin{pmatrix}\xi_{1k}\\ \vdots\\ \xi_{nk}\end{pmatrix}=XS_{\mathbf{xx}}^{-1/2}\mathbf{u}_{\mathrm{k}}$, $\boldsymbol{\eta}_k=\begin{pmatrix}\eta_{1k}\\ \vdots\\ \eta_{nk}\end{pmatrix}=YS_{\mathbf{yy}}^{-1/2}\mathbf{v}_{\mathrm{k}}$.

- 所有典则变量

$$X_{cca}=(\boldsymbol{\xi}_1,\ldots,\boldsymbol{\xi}_r)=XS_{\mathbf{xx}}^{-1/2}U,\ Y_{cca}=(\boldsymbol{\eta}_1,\ldots,\boldsymbol{\eta}_r)=YS_{\mathbf{yy}}^{-1/2}V.$$

样本CCA可理解为内积矩阵的对角化：

$$Y^\top X = (n-1)S_{\mathbf{yx}} \xrightarrow{cca} Y_{cca}^\top X_{cca} = (n-1)D$$

检验$\mathbf{x}, \mathbf{y}$不相关

$$H_0: \Sigma_{\mathbf{xy}} = 0 \Leftrightarrow \sqrt{\lambda_1} = 0$$

H_0下，近似地

$$n\log\left(\frac{|\hat{\Sigma}_0|}{|\hat{\Sigma}_1|}\right) = -n\log\left(\prod_{i=1}^{r}(1-\lambda_i)\right) \sim \chi_{pq}^2$$

例3. 某公司希望了解销售人员的能力测试成绩与销售业绩是否有关（50名销售人员）。销售业绩/响应包括（尺度相同）

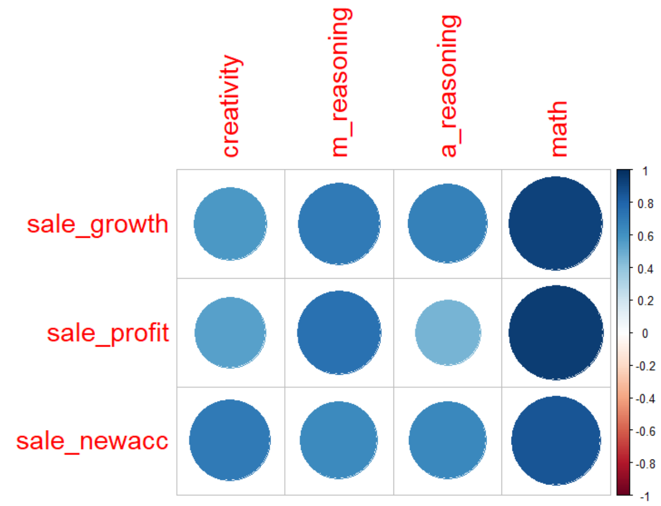
$$y_1 = \text{sale\_growth}, y_2 = \text{sale\_profit}, y_3 = \text{sale\_newacc},$$

能力测试成绩（自变量）包括4项：

x_1 = creativity （1-20）, x_2 = mechanical_reasoning（物理/动手能力，1-20）, x_3 = abstract_reasoning（抽象判断力，1-20）, x_4 = math (1-50)

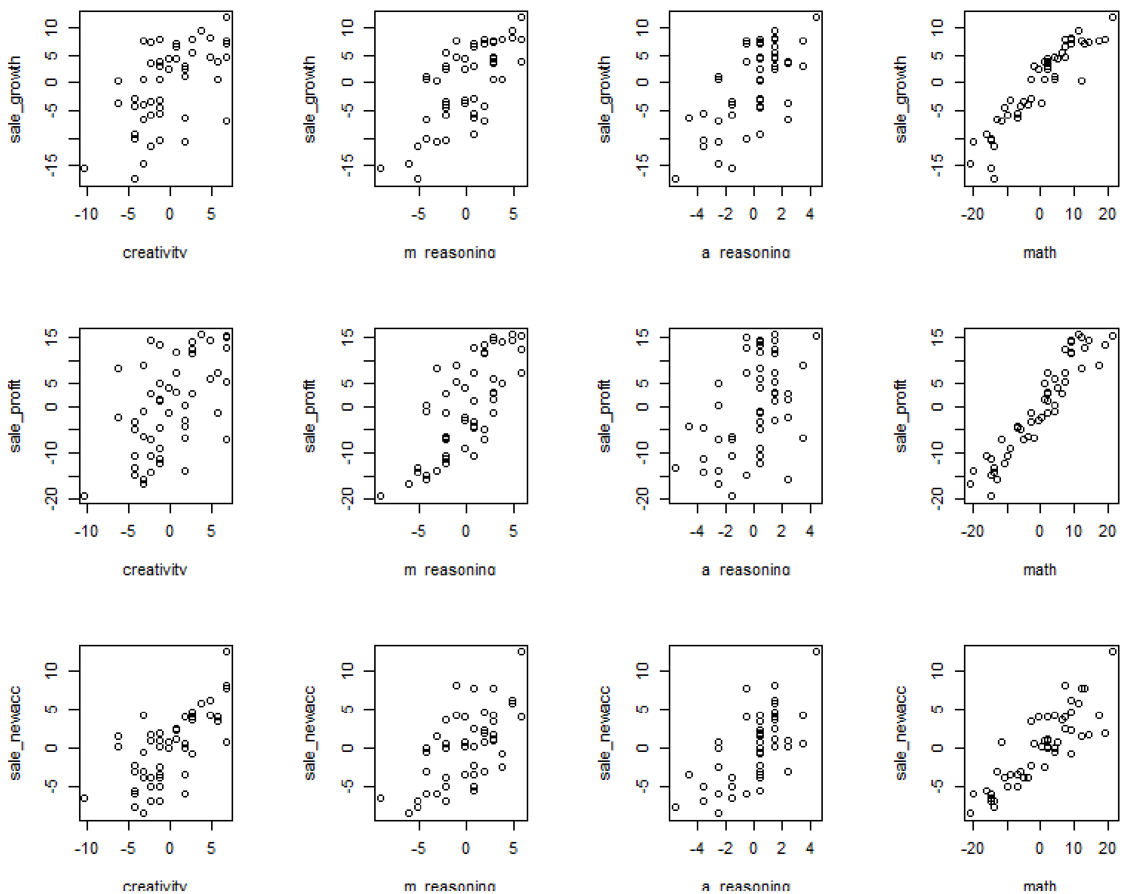
1. 相关系数

	creativity	m_reasoning	a_reasoning	math
sale_growth	0.572	0.708	0.674	0.927
sale_profit	0.542	0.746	0.465	0.944
sale_newacc	0.700	0.637	0.641	0.853




```
library(corrplot)
corrplot(r)
```

2. 散点图，线性回归



第1行是$y_1 \sim x_1, x_2, x_3, x_4$散点图，

第2行是$y_2 \sim x_1, x_2, x_3, x_4$散点图，

存在一定的线性关系。

R: lm(y~x),

y是50× 3响应矩阵，

x是50× 4自变量矩阵

应用R中的lm函数做多元回归，它实际上将三个响应分别对所有自变量做一元回归，并没有考虑响应之间的相关性。

3. 典则相关分析

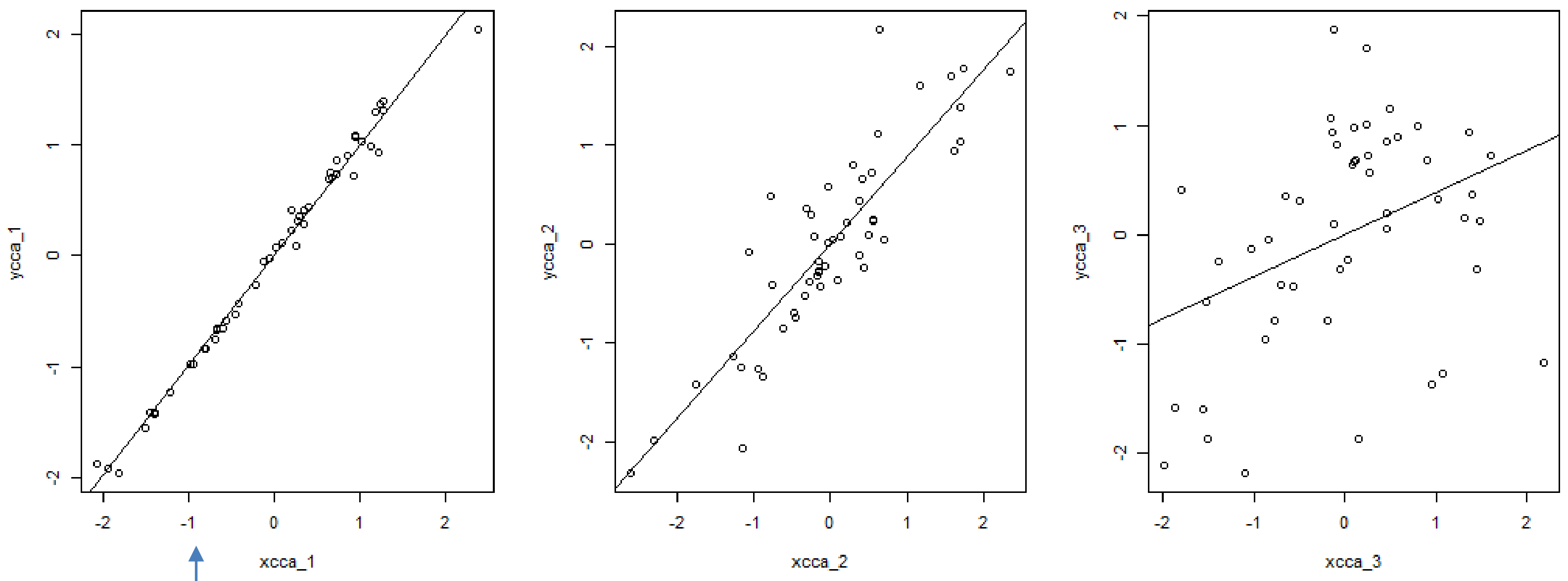
- 三个典则相关系数： 0.997, 0.937, 0.619。第一甚至第二典则相关系数接近于1，说明各种能力$x's$与销售成绩$y's$是强相关的。
- 第一对典则变量：
  $\xi_1 = 0.07x_1 + 0.03x_2 + 0.09x_3 + 0.06x_4$
  $\eta_1 = 0.06y_1 + 0.02y_2 + 0.08y_3$

  系数都是正数，典则变量 ξ_1，η_1分别是各项能力、各项销售成绩的加权平均。
  - ξ_1中，物理判断能力x_2的系数最小， 说明该变量与销售关系不大。考虑到math(x_4)总分比其它三项大5/2倍，math在ξ_1中分量是最重的。
  - η_1中销售利润y_2的系数最小，这说明销售利润与各项能力$x's$关系不大，
- 第二对典则变量：
  $\xi_2 = -0.19x_1 + 0.20x_2 - 0.50x_3 + 0.07x_4$
  $\eta_2 = -0.17y_1 + 0.24y_2 - 0.24y_3$

  系数有正有负，第二典则变量代表了两种不同或负相关的变量：
  - ξ_2中数学x_4的0.07系数≈ 0， ξ_2代表实际操作能力与抽象思维能力之差。
  - η_2代表销售利润与销售增量（新增户头）之差

  第二典则相关系数=0.937, 说明实际操作能力与销售利润强相关，抽象能力与销售增量强相关。

三对典则变量的散点图如下，线性依赖关系依次降低，
三个拟合直线的斜率等于典则相关系数，分别是0.997，0.937, 0.619，
它们的平方即决定系数，分别为0.994， 0.878，0.384



第一典则变量(ξ_1, η_1) 散点图，
其中的线性关系比上页任何
一个原始变量的图都强很多。

总结

CCA动机: $\mathrm{cov}(\mathbf{u}^\top \mathbf{x}, \mathbf{v}^\top \mathbf{y}) = \mathbf{u}^\top \Sigma_{\mathbf{xy}} \mathbf{v} = \max!$

CCA结果: $\mathrm{cov}(\mathbf{x}, \mathbf{y}) = \Sigma_{\mathbf{xy}} \xrightarrow{\text{cca}} \mathrm{cov}(\mathbf{x}_{cca}, \mathbf{y}_{cca})$ =diagonal

讨论

- ❑ 选取几对典则变量？如何决定？

  通常选取一对，把原始q维和p维随机向量压缩成两个一元随机变量。典则相关分析的目的本来就是把不好理解的随机向量之间的相关性用两个一元随机变量之间的相关性表示出来，两对以上的典则变量依旧不好理解。

- ❑ 是否可以将原始维数q和p减少为不同的维数？

  比如$q > 1$情形， 我们希望将 $\mathbf{y}$（响应）的个数q减少为1， 自变量$\mathbf{x}$的个数 p 减少为$r < p$。