第十六讲 典则相关分析CCA

2025.5.7

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: $\mathrm{cov}(\mathbf{x}, \mathbf{y}) = \Sigma_{\mathbf{xy}} \xrightarrow{\text{cca}} \mathrm{cov}(\boldsymbol{\xi}, \boldsymbol{\eta}) =$diagonal

Recap

SVD

奇异值分解(SVD，Singular value decomposition):

定理1：任何秩为 r 的 $n\times p$ 矩阵 X 有奇异值分解

$$X = UDV^{\top},$$

其中 $D_{r\times r} = \mathrm{diag}(\sqrt{\lambda_1},\ldots,\sqrt{\lambda_r})$，$\lambda_1 \ge \ldots \ge \lambda_r > 0$ 为 $X^{\top}X$ 或 $X^{\top}X$ 的非0特征根，$U^{\top}U = I_r, V^{\top}V = I_r$，其中 $U_{q\times r} = (\mathbf{u}_1,\ldots,\mathbf{u}_r), V_{p\times r} = (\mathbf{v}_1,\ldots,\mathbf{v}_r)$ 的列向量分别是 $X^{\top}X$ 或 $X^{\top}X$ 非0特征根对应的正交单位特征向量。

有时，将U,V扩张成正交矩阵应用起来更方便：

U,V正交阵情形

奇异值分解 $X_{n\times p} = U_{n\times r}D_{r\times r}V^{\top}$ 中，若将 U,V 补全为正交方阵 $\tilde{U},\tilde{V}$，D 补全为与 X 大小相同的矩阵 $\tilde{D}_{n\times p} = \begin{pmatrix} D & 0 \\ 0 & 0 \end{pmatrix}$，则奇异值分解可表示为

$$X_{n\times p} = \tilde{U}_{n\times n}\tilde{D}_{n\times p}\tilde{V}^{\top}{}_{p\times p}$$

其中$\tilde{U},\tilde{V}$ 分别是 n, p 阶正交矩阵。

最小二乘

奇异值分解的最优性：

定理2：平方误差意义下，$X_k = U_k D_k V_k^\top = \sum_{i=1}^{k} \sqrt{\lambda_i}\, \mathbf{u}_i \mathbf{v}_i^\top$ 是 X 的最优秩 k 逼近：

$$\min_{rank(W)=k} \|W - X\|_F^2 = \|X_k - X\|_F^2 = \sum_{i=k+1}^{r} \lambda_i$$

其中 $U_k = (\mathbf{u}_1, \dots, \mathbf{u}_k), V_k = (\mathbf{v}_1, \dots, \mathbf{v}_k)$ 分别是 $XX^\top$，$X^\top X$ 前 k 个共同最大特征根对应的特征向量。

注意当 $k \geq r$ 时，上述误差等于0，此时 $X = X_k = U_k D_k V_k^\top$，这样我们从逼近的角度证明了SVD（定理15.2的证明过程实质上并没有应用定理15.1）

SVD⇔正交对角化

$$X = UDV^\top = \sum_{i=1}^{r} \sqrt{\lambda_i}\, \mathbf{u}_i \mathbf{v}_i^\top$$

$$\Leftrightarrow U^\top X V = D$$

$$\Leftrightarrow \mathbf{u}_i^\top X \mathbf{v}_i = \sqrt{\lambda_i},\ \mathbf{u}_i^\top X \mathbf{v}_j = 0,\ 1 \leq i \neq j \leq r$$

双线性函数极大化

我们将在典则相关分析、配列、聚类中用到双线性函数 $\mathbf{u}^\top A\mathbf{v}$ 的极大化问题，该问题与奇异值分解的最优性有关。

定理3. 假设矩阵 $X_{n\times p}$ 的秩为 r，其奇异值分解为 $X = UDV^\top$，其中 $U = (\mathbf{u}_1, \dots, \mathbf{u}_r)$, $V = (\mathbf{v}_1, \dots, \mathbf{v}_r)$, $D = diag(\sqrt{\lambda_1}, \dots, \sqrt{\lambda_r})$, $\lambda_1 \geq \cdots \geq \lambda_r > 0$, 考虑极大化

$$\max_{\|\mathbf{u}\|=\|\mathbf{v}\|=1} \mathbf{u}^\top X\mathbf{v} \qquad (*)$$

(1) 当 $\mathbf{u} = \mathbf{u}_1$， $\mathbf{v} = \mathbf{v}_1$ 时(*)达到最大，最大值等于最大奇异值 $\sqrt{\lambda_1}$;

(2) 对任何 $k \leq r$, 限制 $\mathbf{u} \perp \mathbf{u}_1, \dots, \mathbf{u}_{k-1}$, $\mathbf{v} \perp \mathbf{v}_1, \dots, \mathbf{v}_{k-1}$, 则在 $\mathbf{u} = \mathbf{u}_k$, $\mathbf{v} = \mathbf{v}_k$ 时 (*)达到最大, 最大值等于 $\sqrt{\lambda_k}$.

证明1： (1) 由Cauchy-Schwartz不等式

$$\mathbf{u}^\top X\mathbf{v} \leq \sqrt{\mathbf{u}^\top\mathbf{u}}\sqrt{\mathbf{v}^\top X^\top X\mathbf{v}} = \sqrt{\mathbf{v}^\top X^\top X\mathbf{v}} \leq \sqrt{\lambda_1}$$

当 $\mathbf{v} = \mathbf{v}_1$ 时后一个等号成立，当 $\mathbf{u} \propto X\mathbf{v}$ 时前一个等号成立， 此时，$\mathbf{u} \propto X\mathbf{v}_1 \propto \mathbf{u}_1$. (2)证明类似。

证明2（拉格朗日乘子法）令

$$f(\mathbf{u},\mathbf{v}) = \mathbf{u}^\top X\mathbf{v} - \frac{\eta_1}{2}(\mathbf{v}^\top\mathbf{v} - 1) - \frac{\eta_2}{2}(\mathbf{u}^\top\mathbf{u} - 1),$$

令 $\frac{\partial f}{\partial \mathbf{u}} = X\mathbf{v} - \eta_2\mathbf{u} = 0$, $\frac{\partial f}{\partial \mathbf{v}} = X^\top\mathbf{u} - \eta_1\mathbf{v} = 0$, 即

$$X\mathbf{v} = \eta_2\mathbf{u}，X^\top\mathbf{u} = \eta_1\mathbf{v}$$

故 $\mathbf{u},\mathbf{v}$ 是奇异值分解中的一对特征向量，且 $\mathbf{u}^\top X\mathbf{v} = \eta_1 = \eta_2$ 最大为 $\sqrt{\lambda_1}$，故 $\mathbf{u} = \mathbf{u}_1$， $\mathbf{v} = \mathbf{v}_1$。

证明3: $A = UDV^\top$, 令 $\mathbf{a} = U^\top\mathbf{u}, \mathbf{b} = V^\top\mathbf{v}$, 因为 $UU^\top \le I_p$, 则

$\|\mathbf{a}\| = \mathbf{u}^\top UU^\top\mathbf{u} \le \mathbf{u}^\top\mathbf{u} = 1$, $\|\mathbf{b}\| \le 1$, 所以

$$\mathbf{u}^\top A\mathbf{v} = \mathbf{u}^\top UDV^\top\mathbf{v} = \mathbf{a}^\top D\mathbf{b} = \sum_{i=1}^k a_i b_i\sqrt{\lambda_i} \le \sqrt{\lambda_1}\sum_{i=1}^k |a_i b_i| \le \sqrt{\lambda_1},$$

若约束 $\mathbf{u} \perp \mathbf{u}_1$， $\mathbf{v} \perp \mathbf{v}_1$， 则 $\mathbf{u}^\top U = (0, \mathbf{u}^\top U_{-1}), V^\top\mathbf{v} = \begin{pmatrix} 0 \\ V_{-1}{}^\top\mathbf{v} \end{pmatrix}$,

其中 $U_{-1} = (\mathbf{u}_2,\ldots,\mathbf{u}_k), V_{-1} = (\mathbf{v}_2,\ldots,\mathbf{v}_k)$, 则 $\mathbf{u}^\top A\mathbf{v} = \sum_{i=2}^k a_i b_i\sqrt{\lambda_i} \le \sqrt{\lambda_2}$， 等等。

推论1. 假设矩阵 $A_{q\times p}$ 的秩为 r, 其奇异值分解 $A = UDV^\top$, $D = \mathrm{diag}(\sqrt{\lambda_1},\ldots,\sqrt{\lambda_r})$, $\lambda_1 \geq \ldots \geq \lambda_r > 0$. 则对任何 $k \leq r$，$P \in R^{p\times k}$, $Q \in R^{q\times k}$，$P^\top P = Q^\top Q = I_k$,

$$\max_{P,Q} trP^\top AQ = trU_k^\top AV_k,$$

其中 $U_k = (\mathbf{u}_1,\ldots,\mathbf{u}_k), V_k = (\mathbf{v}_1,\ldots,\mathbf{v}_k)$ 分别为 U,V 的前 k 列。

证明：记 $P = (\mathbf{p}_1,\ldots,\mathbf{p}_k), Q = (\mathbf{q}_1,\ldots,\mathbf{q}_k), trP^\top AQ = \sum_{i=1}^{k}\mathbf{p}_i^\top A\mathbf{q}_i$，应用定理3即得。

典则相关分析（CCA）

典则相关分析（CCA:Canonical correlation analysis）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：

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

- 阅读 $\mathbf{x} = (x_1, x_2)^{\mathsf{T}}$ 与数学 $\mathbf{y} = (y_1, y_2)^{\mathsf{T}}$ 的协方差 $\boldsymbol{\Sigma}_{\mathbf{xy}}$ 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?）

CCA问题的提出

一般地，我们考虑如下典则相关分析问题：

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

求解方向 $\mathbf{a}\in R^p$，$\mathbf{b}\in R^q$使得 $\mathbf{a}^\top\mathbf{x}$，$\mathbf{b}^\top\mathbf{y}$ 的协方差最大

$$\mathrm{cov}(\mathbf{a}^\top\mathbf{x},\ \mathbf{b}^\top\mathbf{y})=\mathbf{a}^\top\Sigma_{\mathbf{xy}}\mathbf{b}$$

这是有意义的问题，但传统上人们经常考虑相关系数最大化.

求解方向 $\mathbf{a}\in R^p$，$\mathbf{b}\in R^q$使得第一对典则变量 $\mathbf{a}^\top\mathbf{x}$，$\mathbf{b}^\top\mathbf{y}$ 的相关系数最大：

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

该问题没有唯一解。因为是研究相关系数，典则相关分析约束投影坐标方差为1：

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

在该约束下极大化

$$\mathrm{cov}(\mathbf{a}^\top\mathbf{x},\ \mathbf{b}^\top\mathbf{y})=\mathbf{a}^\top\Sigma_{\mathbf{xy}}\mathbf{b}。$$

CCA问题框架

$$\mathrm{var}\begin{pmatrix}\mathbf{x}_{p\times 1}\\ \mathbf{y}_{q\times 1}\end{pmatrix}=\begin{pmatrix}\Sigma_{\mathbf{xx}} & \Sigma_{\mathbf{xy}}\\ \Sigma_{\mathbf{yx}} & \Sigma_{\mathbf{yy}}\end{pmatrix},\quad \mathrm{cov}(\mathbf{a}^\top\mathbf{x},\ \mathbf{b}^\top\mathbf{y})=\mathbf{a}^\top\Sigma_{\mathbf{xy}}\mathbf{b},$$

优化目标： $\max\limits_{\mathbf{a}^\top\Sigma_{\mathbf{xx}}\mathbf{a}=1,\,\mathbf{b}^\top\Sigma_{\mathbf{yy}}\mathbf{b}=1} \mathbf{a}^\top\Sigma_{\mathbf{xy}}\mathbf{b}$

假设 $A=\Sigma_{\mathbf{xx}}^{-1/2}\Sigma_{\mathbf{xy}}\Sigma_{\mathbf{yy}}^{-1/2}$，其SVD: $A=UDV^\top=\sum_{i=1}^{r}\sqrt{\lambda_i}\,\mathbf{u}_i\mathbf{v}_i^\top$。

后续SVD出现的时候其中的记号含义不再一一说明：

$A=\Sigma_{\mathbf{xx}}^{-1/2}\Sigma_{\mathbf{xy}}\Sigma_{\mathbf{yy}}^{-1/2}\in R^{p\times q}$ 的奇异值分解：

假设 $rank(A)=r$, $A=UDV^\top=\sum_{i=1}^{r}\sqrt{\lambda_i}\,\mathbf{u}_i\mathbf{v}_i^\top$, 其中$D=\mathrm{diag}(\sqrt{\lambda_1},\dots,\sqrt{\lambda_r})$，$\lambda_1\ge\cdots\ge\lambda_r>0$为$AA^\top$，$A^\top A$的共同非0特征根,$U_k=(\mathbf{u}_1,\dots,\mathbf{u}_k)$, $V_k=(\mathbf{v}_1,\dots,\mathbf{v}_k)$ 分别是$AA^\top$，$A^\top A$前k个共同最大特征根对应的正交单位特征向量， $U^\top U=V^\top V=I_r$ 。

CCA求解过程

第1步:

令$\mathbf{u}=\Sigma_{\mathbf{xx}}^{1/2}\mathbf{a}$，$\mathbf{v}=\Sigma_{\mathbf{yy}}^{1/2}\mathbf{b}$, $\|\mathbf{u}\|=1$, $\|\mathbf{v}\|=1$, 由定理3,

$$\max \mathbf{a}^\top\Sigma_{\mathbf{xy}}\mathbf{b}=\max_{\|\mathbf{u}\|=1,\|\mathbf{v}\|=1}\mathbf{u}^\top\Sigma_{\mathbf{xx}}^{-1/2}\Sigma_{\mathbf{xy}}\Sigma_{\mathbf{yy}}^{-1/2}\mathbf{v}=\sqrt{\lambda_1}$$

最大值在$\mathbf{u}=\mathbf{u}_1$, $\mathbf{v}=\mathbf{v}_1$处达到，此时$\mathbf{a}_1=\Sigma_{\mathbf{xx}}^{-1/2}\mathbf{u}_1$, $\mathbf{b}_1=\Sigma_{\mathbf{yy}}^{-1/2}\mathbf{v}_1$.
我们称最优组合

$$\xi_1=\mathbf{a}_1^\top\mathbf{x}=\mathbf{u}_1^\top\Sigma_{\mathbf{xx}}^{-1/2}\mathbf{x},\ \ \eta_1=\mathbf{b}_1^\top\mathbf{y}=\mathbf{v}_1^\top\Sigma_{\mathbf{yy}}^{-1/2}\mathbf{y}$$

为第一对典则变量，最大值$\sqrt{\lambda_1}$为第一典则相关系数。

$$var\begin{pmatrix}\xi_1\\ \eta_1\end{pmatrix}=\begin{pmatrix}1 & \sqrt{\lambda_1}\\ \sqrt{\lambda_1} & 1\end{pmatrix}$$

给定前$k-1$个典则变量$\xi_1,\dots,\xi_{k-1}$； $\eta_1,\dots,\eta_{k-1}$

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, 由定理3

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

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

$$= \max_{\substack{\|\mathbf{u}\|=1,\|\mathbf{v}\|=1 \\ \mathbf{u}\perp\mathbf{u}_1,\dots,\mathbf{u}_{k-1};\mathbf{v}\perp\mathbf{v}_1,\dots,\mathbf{v}_{k-1}}} \mathbf{u}^\top\Sigma_{\mathbf{xx}}^{-1/2}\Sigma_{\mathbf{xy}}\Sigma_{\mathbf{yy}}^{-1/2}\mathbf{v} = \sqrt{\lambda_k}$$

最大值在 $\mathbf{u} = \mathbf{u}_k, \mathbf{v} = \mathbf{v}_{\boldsymbol{k}}$ 处达到，此时 $\mathbf{a}_k = \Sigma_{\mathbf{xx}}^{-1/2}\mathbf{u}_k, \mathbf{b}_k = \Sigma_{\mathbf{yy}}^{-1/2}\mathbf{v}_{\boldsymbol{k}}$。
我们称

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

为第 k 对典则变量，$\sqrt{\lambda_k}$ 为第 k 个典则相关系数。

且$var\begin{pmatrix}\xi_k \\ \eta_k\end{pmatrix} = \begin{pmatrix}1 & \sqrt{\lambda_k} \\ \sqrt{\lambda_k} & 1\end{pmatrix}$，$cov(\xi_k, \xi_j) = 0, cov(\eta_k, \eta_j) = 0$

CCA结论

综上，典则变量和典则相关系数定义如下：

假设$\mathrm{var}\begin{pmatrix}\mathbf{x}\\ \mathbf{y}\end{pmatrix} = \begin{pmatrix}\Sigma_{\mathbf{xx}} & \Sigma_{\mathbf{xy}}\\ \Sigma_{\mathbf{yx}} & \Sigma_{\mathbf{yy}}\end{pmatrix}$, $A = \Sigma_{\mathbf{xx}}^{-1/2}\Sigma_{\mathbf{xy}}\Sigma_{\mathbf{yy}}^{-1/2}$的奇异值分解为

$$A = UDV^\top = \sum_{i=1}^{r}\sqrt{\lambda_i}\,\mathbf{u}_i\mathbf{v}_i^\top, r = rank(\Sigma_{\mathbf{xy}}),$$

- 第k对典则变量：$\xi_k = \mathbf{u}_k^\top\Sigma_{\mathbf{xx}}^{-1/2}\mathbf{x}, \eta_k = \mathbf{v}_k^\top\Sigma_{\mathbf{yy}}^{-1/2}\mathbf{y}$,
- 第k典则相关系数：$\sqrt{\lambda_k}$，$k = 1, \dots, r$
- 所有典则变量: $\boldsymbol{\xi} = (\xi_1, \dots, \xi_r)^\top = U^\top\Sigma_{\mathbf{xx}}^{-1/2}\mathbf{x}$,
  $\boldsymbol{\eta} = (\eta_1, \dots, \eta_r)^\top = V^\top\Sigma_{\mathbf{yy}}^{-1/2}\mathbf{y}$
- 方差-协方差：$var\begin{pmatrix}\boldsymbol{\xi}\\ \boldsymbol{\eta}\end{pmatrix} = \begin{pmatrix}I_r & D\\ D & I_r\end{pmatrix}$

例如, $\mathrm{var}\begin{pmatrix}\xi_1\\ \xi_2\\ \eta_1\\ \eta_2\end{pmatrix} = \begin{pmatrix}1 & 0 & \sqrt{\lambda_1} & 0\\ 0 & 1 & 0 & \sqrt{\lambda_2}\\ \sqrt{\lambda_1} & 0 & 1 & 0\\ 0 & \sqrt{\lambda_2} & 0 & 1\end{pmatrix}.$

CCA简要步骤

$\mathbf{x}, \mathbf{y}$典则相关分析简要步骤：

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

相应地，协方差矩阵标准化：

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

❖ 奇异值分解：$A = UDV^\top$

❖ 典则变换（类似于主成分变换，A对角化）：

$$\boldsymbol{\xi} = U^\top\tilde{\mathbf{x}}, \quad \boldsymbol{\eta} = V^\top\tilde{\mathbf{y}}。$$

$$\operatorname{cov}(\tilde{\mathbf{x}}, \tilde{\mathbf{y}}) = \Sigma_{\mathbf{xx}}^{-1/2}\Sigma_{\mathbf{xy}}\Sigma_{\mathbf{yy}}^{-1/2} = UDV^\top$$

对角化：

两边同时左乘$U^\top$，右乘V

$$\operatorname{cov}(U^\top\tilde{\mathbf{x}}, V^\top\tilde{\mathbf{y}}) = D$$

典则变量

CCA与PCA

CCA本质上与PCA原理一致，都是求解子空间/方向上的投影，最大可能地保留协方差或方差结构。PCA考虑一个随机向量内部方差结构，利用方差矩阵的谱分解；CCA考虑两个随机向量之间的协方差结构，利用协方差矩阵的奇异值分解。

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

  $$\text{谱分解}: \Sigma_{\mathbf{xx}} = V\Lambda V^\top \Rightarrow \mathrm{cov}(\underbrace{V^\top \mathbf{x}, V^\top \mathbf{x}}_{\text{主成分}}) = \Lambda,$$

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

  $$\text{奇异值分解}: \Sigma_{\mathbf{xy}} = UDV^\top \Rightarrow \mathrm{cov}(\underbrace{U^\top \mathbf{x}, V^\top \mathbf{y}}_{\text{典则变量}}) = D$$

我们知道，相关分析与线性回归联系密切，典则相关分析也是。

CCA与决定系数

A, Φ, Ψ的含义？为什么其特征根小于1?

$A = \Sigma_{\mathbf{xx}}^{-1/2}\Sigma_{\mathbf{xy}}\Sigma_{\mathbf{yy}}^{-1/2}$: $\Sigma_{\mathbf{xy}}$的标准化

$\Phi = A^{\mathsf{T}}A = \Sigma_{\mathbf{yy}}^{-1/2}\Sigma_{\mathbf{yx}}\Sigma_{\mathbf{xx}}^{-1}\Sigma_{\mathbf{xy}}\Sigma_{\mathbf{yy}}^{-1/2} \le I_q$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：

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

两边同时取方差, $\mathrm{var}(B^{\mathsf{T}}\mathbf{x}) = \Sigma_{\mathbf{yx}}\Sigma_{\mathbf{xx}}^{-1}\Sigma_{\mathbf{xy}}$, $\mathrm{var}(\boldsymbol{\varepsilon}) = \Sigma_{\mathbf{yy}\bullet\mathbf{x}}$

$$\Sigma_{\mathbf{yy}} = \Sigma_{\mathbf{yx}}\Sigma_{\mathbf{xx}}^{-1}\Sigma_{\mathbf{xy}} + \Sigma_{\mathbf{yy}\bullet\mathbf{x}} \ge \Sigma_{\mathbf{yx}}\Sigma_{\mathbf{xx}}^{-1}\Sigma_{\mathbf{xy}}$$

$$I_q = \Sigma_{\mathbf{yy}}^{-1/2}\Sigma_{\mathbf{yx}}\Sigma_{\mathbf{xx}}^{-1}\Sigma_{\mathbf{xy}}\Sigma_{\mathbf{yy}}^{-1/2} + \Sigma_{\mathbf{yy}}^{-1/2}\Sigma_{\mathbf{yy}\bullet\mathbf{x}}\Sigma_{\mathbf{yy}}^{-1/2} = \Phi + \Sigma_{\mathbf{yy}}^{-1/2}\Sigma_{\mathbf{yy}\bullet\mathbf{x}}\Sigma_{\mathbf{yy}}^{-1/2}$$

$\Phi \le I_q$，代表$\mathbf{y}$中$\mathbf{x}$能被解释的"比例"， $\sqrt{\lambda_1(\Phi)} \le 1.$

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。 $p = 1$时类似。

CCA与消元法/对角化

CCA可看作是对多元线性模型（多个方程）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：

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

线性模型：$\mathbf{y}=B^{\top}\mathbf{x}+\boldsymbol{\varepsilon}=\Sigma_{\mathbf{yx}}\Sigma_{\mathbf{xx}}^{-1}+\boldsymbol{\varepsilon}$

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

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

SVD：$A=UDV^{\top}$

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

记$\boldsymbol{\eta}=V^{\top}\tilde{\mathbf{y}}, \boldsymbol{\xi}=U^{\top}\tilde{\mathbf{x}}$

$$\boldsymbol{\eta}=D\boldsymbol{\xi}+\boldsymbol{\delta}\Leftrightarrow\begin{cases}\eta_1=\sqrt{\lambda_1}\xi_1+\delta_1\\ \vdots\\ \eta_r=\sqrt{\lambda_r}\xi_r+\delta_r\end{cases}$$

重要性

$\eta_i=\sqrt{\lambda_i}\xi_i+\delta_i$ 中η_i，ξ_i都是标准化的，$\sqrt{\lambda_i}=\mathrm{corr}(\eta_i,\xi_i)$.

CCA 与分块对角化

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

分块对角化：

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

  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：$(X_{n\times p}, Y_{n\times q}) = \begin{pmatrix} \mathbf{x}_1^\top & \mathbf{y}_1^\top \\ \vdots & \vdots \\ \mathbf{x}_n^\top & \mathbf{y}_n^\top \end{pmatrix}$，假设 X, Y 都已中心化

  样本方差-协方差矩阵：$S = \begin{pmatrix} S_{\mathbf{xx}} & S_{\mathbf{xy}} \\ S_{\mathbf{yx}} & S_{\mathbf{yy}} \end{pmatrix} = \frac{1}{n-1}\begin{pmatrix} X^\top X & X^\top Y \\ Y^\top X & Y^\top Y \end{pmatrix}$,

  令$S_{\mathbf{xy}}$的标准化：$A = S_{\mathbf{xx}}^{-1/2} S_{\mathbf{xy}} S_{\mathbf{yy}}^{-1/2}$,

- 奇异值分解($rank(A) = r$):

$$A = UDV^\top$$

$SVD: A = S_{\mathbf{yy}}^{-1/2} S_{\mathbf{yx}} S_{\mathbf{xx}}^{-1/2} = UDV^\top$,
其中$U = (\mathbf{u}_1, ..., \mathbf{u}_r), V = (\mathbf{v}_1, ..., \mathbf{v}_r)$,
$U^\top U = V^\top V = I_r$,
$D = diag(\lambda_1, ..., \lambda), \lambda_1 \geq ... \geq \lambda_r > 0.$

- 典则变量：$\boldsymbol{\xi}_i = U^\top S_{\mathbf{xx}}^{-1/2} \mathbf{x}_i, \boldsymbol{\eta}_i = V^\top S_{\mathbf{yy}}^{-1/2} \mathbf{y}_i$， $i = 1, \dots, n$

以上过程与总体CCA完全相同.

- 所有数据的典则变量：

$$X_{\mathrm{cca}} = (\boldsymbol{\xi}_1, \dots, \boldsymbol{\xi}_n)^\top = X S_{\mathbf{xx}}^{-1/2} U,$$
$$Y_{\mathrm{cca}} = (\boldsymbol{\eta}_1, \dots, \boldsymbol{\eta}_n)^\top = Y S_{\mathbf{yy}}^{-1/2} V$$

$$X_{cca}^\top Y_{cca}/(n-1) = D$$

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

典则相关系数: $\sqrt{\lambda(P_X P_Y)}$

$$A = S_{\mathbf{xx}}^{-1/2} S_{\mathbf{xy}} S_{\mathbf{yy}}^{-1/2} = (X^\top X)^{-1/2} X^\top Y (Y^\top Y)^{-1/2} \Rightarrow$$

$$\Phi = A^\top A = \underline{(Y^\top Y)^{-1/2} Y^\top}\ \underline{X(X^\top X)^{-1} X^\top Y (Y^\top Y)^{-1/2}}$$

下划线部分交换次序不改变特征根

$$\lambda(\Phi) = \lambda(X(X^\top X)^{-1} X^\top Y (Y^\top Y)^{-1} Y^\top) = \lambda(P_X P_Y)$$

故典则相关系数是$P_X P_Y$特征根的平方根。

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

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

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：

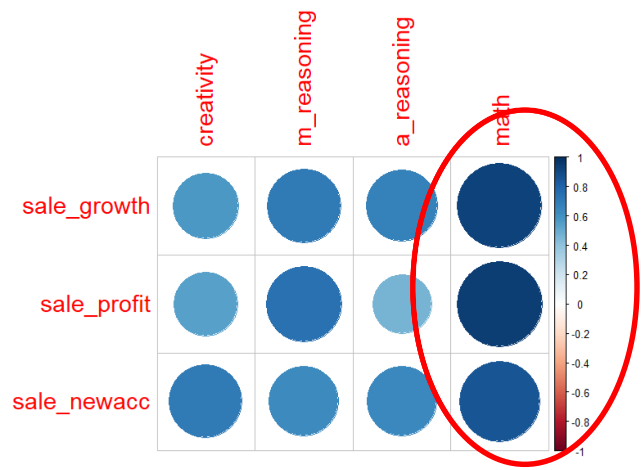
x_1 = creativity (1-20), x_2 = mechanical_reasoning (机械推理,1-20), x_3 = abstract_reasoning (抽象推理,1-20), x_4 = math (1-50)

1. 相关系数

x

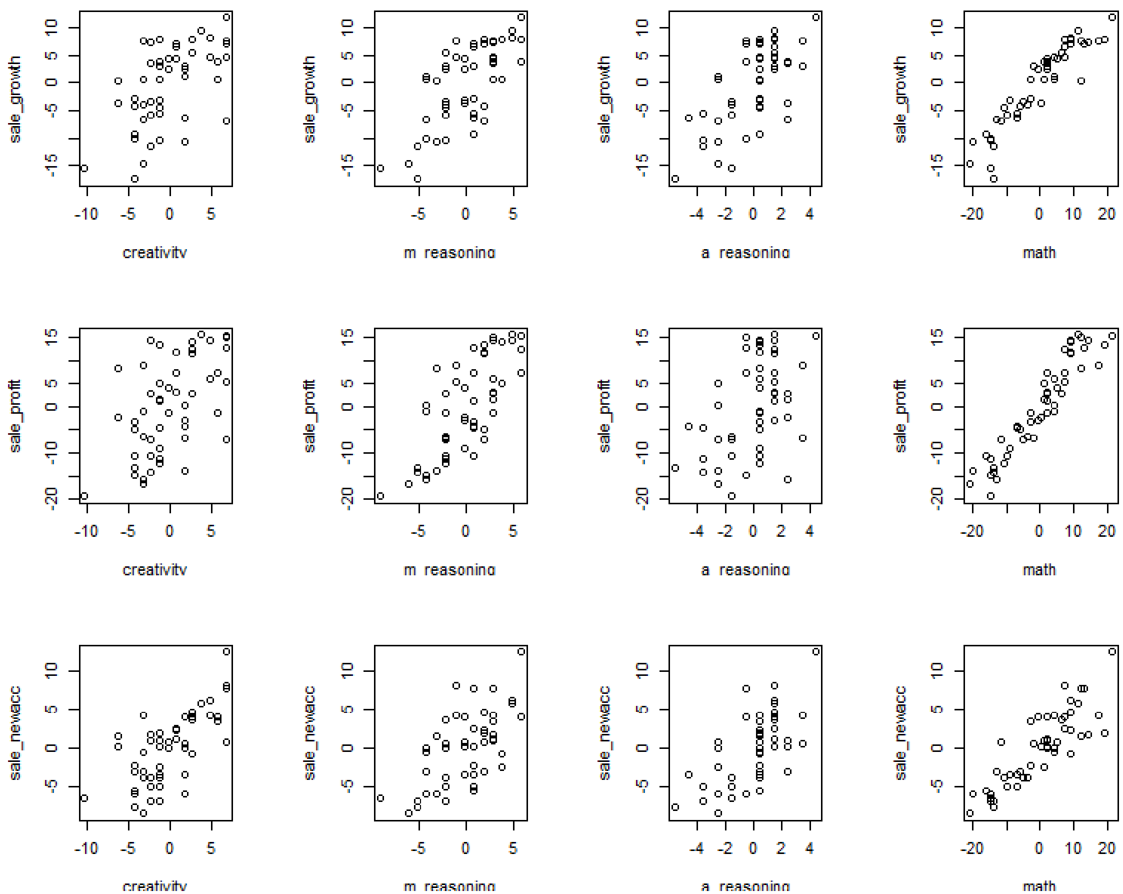
y

	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



x, y之间的相关性主要体现在数学math与销售总成绩的相关性？

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。

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

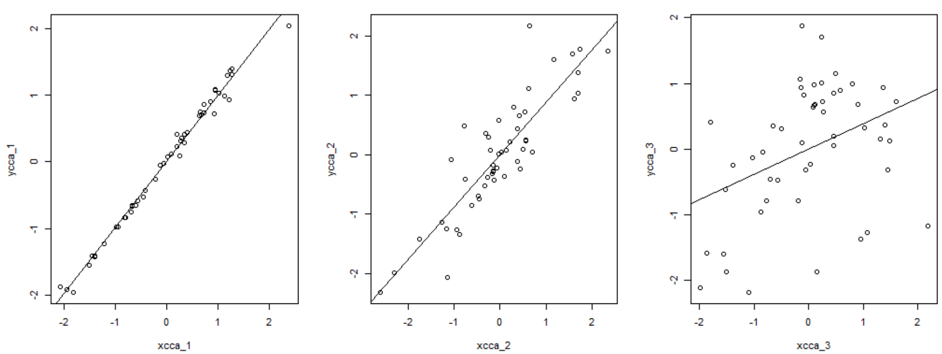
❖ 三个典则相关系数： 0.997, 0.937, 0.619。第一甚至第二典则相关系数接近于1，说明各种能力$x's$与销售成绩$y's$是强相关的。

❖ 第一对典则变量：
ξ_1=0.07x_1 + 0.03x_2 +0.09x_3 + 0.06x_4
η_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，机械推理能力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，

❖ 下面左边第一个图是第一典则变量(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（直线斜率等于第一典则相关系数0.997）。



- 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代表机械推理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. 说明机械推理与销售利润强相关，抽象能力与销售增量相关。

上页的第二个图是第二对典则变量的散点图如下，线性依赖关系稍弱于第一个图。拟合直线的斜率等于第二典则相关系数0.937。

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$r < p$。