Jordan Lecture Note-11: 典型相关分析(Canonical Correlation Analysis, CCA).

典型相关分析

（一）引入

典型相关分析（Canonical Correlation Analysis）是研究两组变量之间相关关系的一种多元统计方法。他能够揭示出两组变量之间的内在联系。

我们知道，在一元统计分析中，用相关系数来衡量两个随机变量的线性相关关系，用复相关系数研究一个随机变量与多个随机变量的线性相关关系。然而，这些方法均无法用于研究两组变量之间的相关关系，于是提出了CCA。其基本思想和主成分分析非常相似。首先，在每组变量中寻找出变量的线性组合，使得两组的线性组合之间具有最大的相关系数；然后选取和已经挑选出的这对线性组合不相关的另一对线性组合，并使其相关系数最大，如此下去，直到两组变量的相关性被提取完毕为止。被选出的线性组合配对称为典型变量，它们的相关系数称为典型相关系数。

（二）分析

设有两组随机变量$\mathbf{X}=(x_1,x_2,\cdots,x_p)'$和$\mathbf{Y}=(y_1,y_2,\cdots,y_q)'$，不妨设$p\leq q$。设第一组变量均值为$\mathbb{E}\mathbf{X}=\mu_1$，方差为$Var(\mathbf{X})=cov(\mathbf{X},\mathbf{X})=\Sigma_{11}$。第二组变量均值为$\mathbb{E}\mathbf{Y}=\mu_2$，方差为$Var(\mathbf{Y})=cov(\mathbf{Y},\mathbf{Y})=\Sigma_{22}$。第一组与第二组变量的协方差矩阵为$cov(\mathbf{X},\mathbf{Y})=\Sigma_{12}=\Sigma'_{21}$。

分别对两组变量做线性组合：

$$U=a_1x_1+a_2x_2+\cdots+a_px_p=a'\mathbf{X} \tag{1}$$
$$V=b_1x_1+b_2x_2+\cdots+b_qy_q=b'\mathbf{Y} \tag{2}$$

所以U,V的方差，协方差，相关系数为：

$$Var(U)=a'\ cov(\mathbf{X},\mathbf{X})a=a'\Sigma_{11}a \tag{3}$$
$$Var(V)=b'\ cov(\mathbf{Y},\mathbf{Y})b=b'\Sigma_{22}b \tag{4}$$
$$cov(U,V)=a'\ cov(\mathbf{X},\mathbf{Y})b=a'\Sigma_{12}b \tag{5}$$
$$\rho=corr(U,V)=\frac{a'\Sigma_{12}b}{\sqrt{a'\Sigma_{11}a}\sqrt{b'\Sigma_{12}b}} \tag{6}$$

其中U,V称为典型变量，它们之间的相关系数ρ称为典型相关系数。

CCA要解决的问题是，在所有线性组合U和V中选取典型相关系数最大的那对，即选取$a^{(1)}, b^{(1)}$使$U_1 = (a^{(1)})'\mathbf{X}$与$V_1 = (b^{(1)})'\mathbf{Y}$之间的相关系数最大，这里$(U_1, V_1)$称为第一对典型相关变量；然后在选取$a^{(2)}, b^{(2)}$使得$U_1 = (a^{(2)})'\mathbf{X}, V_2(b^{(2)})'\mathbf{Y}$，在与$U_1, V_1$不相关的情况下，使得$(U_2, V_2)$的相关系数最大，称为第二对典型相关变量；如此继续下去，直到所有分别与$(U_1, V_1), (U_2, V_2), \cdots, (U_{p-1}, V_{p-1})$都不相关的线性组合$(U_p, V_p)$为止，此时$p$为$\mathbf{X}$与$\mathbf{Y}$之间的协方差矩阵的秩。由上面的分析可得模型：

$$\max \quad \rho = \frac{a'\Sigma_{12}b}{\sqrt{a'\Sigma_{11}a}\sqrt{b'\Sigma_{22}b}} \tag{7}$$

由于收缩U和V的值并不会影响ρ，故我们可引入限制条件$a'\Sigma_{11}a = 1, b'\Sigma_{22}b = 1$将模型转化为：

$$\begin{aligned} \max \quad & a'\Sigma_{12}b \\ s.t. \quad & a'\Sigma_{11}a = 1 \\ & b'\Sigma_{22}b = 1 \end{aligned} \tag{8}$$

引入Lagrange乘子：

$$L(a, b, \lambda, \nu) = a'\Sigma_{12}b - \frac{\lambda}{2}(a'\Sigma_{11}a - 1) - \frac{\nu}{2}(b'\Sigma_{22}b - 1) \tag{9}$$

对Lagrange函数9求导得：

$$\frac{\partial L}{\partial a} = \Sigma_{12}b - \lambda\Sigma_{11}a = 0 \tag{10}$$

$$\frac{\partial L}{\partial b} = \Sigma_{21}a - \nu\Sigma_{22}b = 0 \tag{11}$$

将式子10左乘a'，式子11左乘b'得：

$$\begin{aligned} a'\Sigma_{12}b - \lambda a'\Sigma_{11}a = 0 \\ b'\Sigma_{21}a - \nu b'\Sigma_{22}b = 0 \end{aligned}$$

又因为$(a'\Sigma_{12}b)' = b'\Sigma_{21}a \Longrightarrow \lambda a'\Sigma_{11}a = \nu b'\Sigma_{22}b$。由限制条件知：$\lambda = \nu = \rho = a'\Sigma_{12}b$，即$\lambda$的值就是线性组合$U$和$V$的相关系数。我们重新将式子$10$和式子$11$写成：

$$-\lambda\Sigma_{11}a + \Sigma_{12}b = 0 \tag{12}$$

$$\Sigma_{21}a - \lambda\Sigma_{22}b = 0 \tag{13}$$

然后将式子13左乘$\Sigma_{12}\Sigma_{22}^{-1}$得：

$$\Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}a = \lambda\Sigma_{12}b \tag{14}$$

结合式子12得：

$$\Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}a=\lambda^2\Sigma_{11}a \Longrightarrow (\Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}-\lambda^2\Sigma_{11})a=0 \tag{15}$$

同理，将式子12左乘$\Sigma_{21}\Sigma_{11}^{-1}$,并将式子$13$代入式子$12$得：

$$(\Sigma_{21}\Sigma_{11}^{-1}\Sigma_{12}-\lambda^2\Sigma_{22})b=0 \tag{16}$$

将Σ_{11}^{-1}左乘式子15，Σ_{22}^{-1}左乘式子16得：

$$\begin{cases}(\Sigma_{11}^{-1}\Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}-\lambda^2)a=0\\(\Sigma_{22}^{-1}\Sigma_{21}\Sigma_{11}^{-1}\Sigma_{12}-\lambda^2)b=0\end{cases} \triangleq \begin{cases}\mathbf{A}a=\lambda^2 a\\ \mathbf{B}b=\lambda^2 b\end{cases} \tag{17}$$

说明：λ^2既是矩阵$\mathbf{A}$也是矩阵$\mathbf{B}$的特征值，a与b分别是对应的特征向量。所以我们的问题转化成求矩阵$\mathbf{A},\mathbf{B}$的最大特征值对应的特征向量，而特征值的平方根$\sqrt{\lambda}$为相关系数，从而求出第一对典型相关变量。

此时，我们可以得到如下的猜想：是否矩阵$\mathbf{A},\mathbf{B}$的所有非零特征值的平方跟都会是其对应的典型相关系数？接下去，我们来证明如下猜想：

设$\lambda_1^2\ge\lambda_2^2\ge\cdots\ge\lambda_p^2>0$为$\mathbf{A},\mathbf{B}$的特征值（$p$为矩阵$\Sigma_{12}$的秩），其对应的特征向量为$a^{(1)},a^{(2)},\cdots,a^{(p)}$和$b^{(1)},b^{(2)},\cdots,b^{(p)}$，于是得到$p$对线性组合：

$$\begin{aligned}U_1&=(a^{(1)})'\mathbf{X} & U_2&=(a^{(2)})'\mathbf{X} & \cdots \quad U_p&=(a^{(p)})'\mathbf{X}\\ V_1&=(b^{(1)})'\mathbf{Y} & V_2&=(b^{(2)})'\mathbf{Y} & \cdots \quad V_p&=(b^{(p)})'\mathbf{Y}\end{aligned} \tag{18}$$

可以证明$(U_1,V_1),(U_2,V_2),\cdots,(U_p,V_p)$就是其前$p$对典型变量，$\lambda_1\ge\lambda_2\ge\cdots\ge\lambda_p$为其典型相关系数。

首先，在求出第一对典型变量的基础上求第二对典型变量。由上述分析我们可以知道该模型为：

$$\begin{aligned}\max \quad & (a^{(2)})'\Sigma_{12}b^{(2)}\\ s.t. \quad & (a^{(2)})'\Sigma_{11}a^{(2)}=1\\ & (b^{(2)})'\Sigma_{22}b^{(2)}=1\\ & (a^{(1)})'\Sigma_{11}a^{(2)}=0 \qquad (19)\\ & (b^{(1)})'\Sigma_{22}b^{(2)}=0 \qquad (20)\end{aligned}$$

其中限制19和20是由于第二对典型变量必须与第一对典型变量无关。其Lagrange方程以及相应的偏导为：

$$\begin{aligned}L(a^{(1)},b^{(1)},a^{(2)},b^{(2)})=&(a^{(2)})'\Sigma_{12}b^{(2)}-\frac{\lambda}{2}((a^{(2)})'\Sigma_{11}a^{(2)}-1)-\frac{\nu}{2}((b^{(2)})'\Sigma_{22}b^{(2)}-1)\\&+\gamma(a^{(1)})'\Sigma_{11}a^{(2)}+\beta(b^{(1)})'\Sigma_{22}b^{(2)}\end{aligned}$$

$$\frac{\partial L}{\partial a^{(2)}} = \Sigma_{12}b^{(2)} - \lambda\Sigma_{11}a^{(2)} + \gamma\Sigma_{11}a^{(1)} = 0 \tag{21}$$
$$\frac{\partial L}{\partial b^{(2)}} = \Sigma_{21}a^{(2)} - \nu\Sigma_{22}b^{(2)} + \beta\Sigma_{22}b^{(1)} = 0 \tag{22}$$
$$\frac{\partial L}{\partial a1} = \gamma\Sigma_{11}a^{(2)} = 0 \tag{23}$$
$$\frac{\partial L}{\partial b^{(1)}} = \beta\Sigma_{22}b^{(2)} = 0 \tag{24}$$

将$(a^{(2)})'$左乘式子21，$(b^{(2)})'$左乘式子22得：

$$(a^{(2)})'\Sigma_{12}b^{(2)} - \lambda(a^{(2)})'\Sigma_{11}a^{(2)} + \gamma(a^{(2)})'\Sigma_{11}a^{(1)} = 0 \tag{25}$$
$$(b^{(2)})'\Sigma_{21}a^{(2)} - \nu(b^{(2)})'\Sigma_{22}b^{(2)} + \beta(b^{(2)})'\Sigma_{22}b^{(1)} = 0 \tag{26}$$

将$(a^{(1)})'$左乘式子23，$(b^{(2)})'$左乘式子24得：

$$(a^{(2)})'\Sigma_{11}a^{(1)} = 0 \tag{27}$$
$$(b^{(2)})'\Sigma_{22}b^{(2)} = 0 \tag{28}$$

将式子27和式子28代入式子25和26得：

$$(a^{(2)})'\Sigma_{12}b^{(2)} - \lambda(a^{(2)})'\Sigma_{11}a^{(2)} = 0 \tag{29}$$
$$(b^{(2)})'\Sigma_{21}a^{(2)} - \nu(b^{(2)})'\Sigma_{22}b^{(2)} = 0 \tag{30}$$

其中式子29和式子30与式子10和式子11有相同的形式，只是a, b换成$a^{(2)}, b^{(2)}$，故同样可以得到：

$$\begin{cases} \mathbf{A}a^{(2)} = \lambda^2 a^{(2)} \\ \mathbf{B}b^{(2)} = \lambda^2 b^{(2)} \end{cases} \tag{31}$$

此时$a^{(2)} \neq a, b^{(2)} \neq b$，否则不满足限制$19$和$20$，所以最大值为第二大特征值。以此类推，我们即可证明上述猜想。

注意：我们在求解上述普通特征值方程$\mathbf{A}a = \lambda^2 a$时，由于$\mathbf{A} = \Sigma_{11}^{-1}\Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}$而求逆过程的计算量大，精度低，故我们可以将其中的对称矩阵Σ_{11}进行Cholesky分解，即$\Sigma_{11} = \mathbf{R_1}\mathbf{R_1}'$，其中$\mathbf{R_1}$为下三角矩阵。于是方程可化为对称矩阵的求特征值问题：$\mathbf{R_1}^{-1}\Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}(\mathbf{R_1}^{-1})'u_x = \lambda^2 u_x$，其中$u_x = \mathbf{R_1}'a$。

分类: Jordan Lecture Note,机器学习

标签: KKT条件, 拉格朗日对偶, CCA

« 上一篇：PAT 1021
» 下一篇：王道论坛研究生机试练习赛（二）

最新IT新闻:

· 谷歌官方宣布Android平板暂不支持语音助手
· Ray Kurzweil预言计算机在12年内达到人类水平的智能
· 软银投资的第一家人工智能公司，到底有多少料?
· 飞利浦最新的Hue灯泡将带来不一样的烛光效果
· 特别策划 | 一图看懂摩拜和ofo们的今天和明天
» 更多新闻...

最新知识库文章:

· 为什么我要写自己的框架?
· 垃圾回收原来是这么回事
· 「代码家」的学习过程和学习经验分享
· 写给未来的程序媛
· 高质量的工程代码为什么难写
» 更多知识库文章...