hw6

6.1

假设数据矩阵为 $X_{n\times p}=(\mathbf{x}_1,\ldots,\mathbf{x}_n)^\top$, 且 X 已经中心化, 假设样本协方差阵 $S=X^\top X/(n-1)$ 的谱分解为

$$S=V\Lambda V^\top,$$

其中 $V^\top V=VV^\top=I_p,\Lambda=\operatorname{diag}(\lambda_1,\ldots,\lambda_p),\lambda_1\geq\ldots\geq\lambda_p\geq 0$ 。主成分矩阵定义为

$$Y_{n\times p}=XV,$$

Y 的第 j 列为所有 n 个样本点的第 j 主成分, $j=1,\ldots,p$ 。
(a) 证明 Y 是中心化的矩阵, 其样本协方差矩阵为 $Y^\top Y/(n-1)=\Lambda$ （即各个主成分之间不相关，第 j 主成分的样本方差为 λ_j)。
(b) 证明 Y 与 X 的样本协方差矩阵为 $Y^\top X/(n-1)=\Lambda V^\top$ （即第 j 主成分与第 k 变量的样本协方差为 $v_{kj}\lambda_j$)。
(c) 基于 $(a),(b)$, 证明第 j 主成分和第 k 个变量的样本相关系数

$$r_{jk}=v_{kj}\sqrt{\frac{\lambda_j}{s_{kk}}},$$

其中 s_{kk} 为第 k 个变量的样本方差, 即 S 的 (k,k) 元。
Solution:
(a) 因为 X 已经中心化，故 $X^\top\mathbf{1}_n=0$, 所以 $Y^\top\mathbf{1}_n=V^\top X^\top\mathbf{1}_n=0$，即 Y 是 (列) 中心化的。另外，

$$\begin{aligned}\frac{1}{n-1}Y^\top Y&=\frac{1}{n-1}V^\top X^\top XV\\&=V^\top SV\\&=V^\top\left(V\Lambda V^\top\right)V\\&=\Lambda\end{aligned}$$

(b)

$$\begin{aligned}\frac{1}{n-1}Y^\top X&=\frac{1}{n-1}V^\top X^\top X\\&=V^\top S\\&=V^\top V\Lambda V^\top\\&=\Lambda V^\top\end{aligned}$$

(c) 由 (b), 第 j 主成分与第 k 个变量之间的样本协方差为 $Y^\top X/(n-1)=\Lambda V^\top$ 的 (j,k) 元 $\lambda_j v_{jk}$, 所以样本相关系数

$$r_{jk}=\frac{v_{kj}\lambda_j}{\sqrt{s_{kk}}\sqrt{\lambda_j}}=v_{kj}\sqrt{\frac{\lambda_j}{s_{kk}}}.$$

6.2

假设 $\mathbf{x}_1,\ldots,\mathbf{x}_n\in R^p$ 的样本方差-协方差矩阵具有如下形式

$$S=\sigma^2 I_p+\mathbf{1}_p\mathbf{1}_p^\top+\mathbf{e}\mathbf{e}^\top,$$

其中 $\mathbf{1}_p=(1,1,\ldots,1)^\top,\mathbf{e}=(1,-1,0,\ldots,0)^\top$. 求 $\mathbf{x}_i,i=1,\ldots,n$ 的第一主成分和第二主成分, 以及它们的方差在总方差中所占的累积比例。
Solution:

$$S = \sigma^2 I_p + \mathbf{1}_p \mathbf{1}_p^\top + ee^\top$$
$$S\mathbf{1}_p = \sigma^2 \mathbf{1}_p + p\mathbf{1}_p = \left(\sigma^2 + p\right)\mathbf{1}_p$$
$$Se = \sigma^2 e + 2e = \left(\sigma^2 + 2\right)e$$

对任何满足 $\mathbf{v} \perp \mathbf{1}$, $\mathbf{v} \perp \mathbf{e}$ 的向量 $\mathbf{v} \in C(\mathbf{1}, \mathbf{e})^\perp$ (即 $\mathbf{1}, \mathbf{e}$ 张成空间的正交补空间)，都有

$$S\mathbf{v} = \sigma^2 \mathbf{v}$$

因为 $dim(C(\mathbf{1}, \mathbf{e})^\perp) = p - 2$, 所以 σ^2 是特征根，重数为 $p - 2$, 所以 S 的所有特征根为$\sigma^2 + p \geq \sigma^2 + 2 > \sigma^2 = ... = \sigma^2$
第一主成分

$$y_1 = \frac{1}{\sqrt{p}} \mathbf{1}_p^\top (x_1, \ldots, x_n) = \frac{1}{\sqrt{p}} \mathbf{1}_p^\top X^\top$$

第二主成分

$$y_2 = \frac{1}{\sqrt{2}} \mathbf{e}^\top (x_1, \ldots, x_n) = \frac{1}{\sqrt{2}} \mathbf{e}^\top X^\top$$

前 2 个主成分在总方差中所占累积比例为 $\frac{\sigma^2+p+\sigma^2+2}{\sigma^2+p+\sigma^2+2+(p-2)\sigma^2} = \frac{2\sigma^2+p+2}{p\sigma^2+p+2}$

6.3

1988 年汉城奥运会女子七项全能 (heptathlon) 的 25 个运动员的成绩统计如下:

	hurdles	highjump	shot	run200m	longjump	javelin	run800m	score
Joyner-Kersee (USA)	12.69	1.86	15.80	22.56	7.27	45.66	128.51	7291
John (GDR)	12.85	1.80	16.23	23.65	6.71	42.56	126.12	6897
Behmer (GDR)	13.20	1.83	14.20	23.10	6.68	44.54	124.20	6858
Sablovskaite (URS)	13.61	1.80	15.23	23.92	6.25	42.78	132.24	6540
Choubenkova (URS)	13.51	1.74	14.76	23.93	6.32	47.46	127.90	6540
Schulz (GDR)	13.75	1.83	13.50	24.65	6.33	42.82	125.79	6411
Fleming (AUS)	13.38	1.80	12.88	23.59	6.37	40.28	132.54	6351
Greiner (USA)	13.55	1.80	14.13	24.48	6.47	38.00	133.65	6297
Lajbnerova (CZE)	13.63	1.83	14.28	24.86	6.11	42.20	136.05	6252
Bouraga (URS)	13.25	1.77	12.62	23.59	6.28	39.06	134.74	6252
Wijnsma (HOL)	13.75	1.86	13.01	25.03	6.34	37.86	131.49	6205
Dimitrova (BUL)	13.24	1.80	12.88	23.59	6.37	40.28	132.54	6171
Scheider (SWI)	13.85	1.86	11.58	24.87	6.05	47.50	134.93	6137
Braun (FRG)	13.71	1.83	13.16	24.78	6.12	44.58	142.82	6109
Ruotsalainen (FIN)	13.79	1.80	12.32	24.61	6.08	45.44	137.06	6101
Yuping (CHN)	13.93	1.86	14.21	25.00	6.40	38.60	146.67	6087
Hagger (GB)	13.47	1.80	12.75	25.47	6.34	35.76	138.48	5975
Brown (USA)	14.07	1.83	12.69	24.83	6.13	44.34	146.43	5972
Mulliner (GB)	14.39	1.71	12.68	24.92	6.10	37.76	138.02	5746
Hautenauve (BEL)	14.04	1.77	11.81	25.61	5.99	35.68	133.90	5734
Kytola (FIN)	14.31	1.77	11.66	25.69	5.75	39.48	133.35	5686
Geremias (BRA)	14.23	1.71	12.95	25.50	5.50	39.64	144.02	5508
Hui-Ing (TAI)	14.85	1.68	10.00	25.23	5.47	39.14	137.30	5290
Jeong-Mi (KOR)	14.53	1.71	10.83	26.61	5.50	39.26	139.17	5289
Launa (PNG)	16.42	1.50	11.78	26.16	4.88	46.38	163.43	4566

最后一列 score 是运动员得分 (不参与 PCA 分析), 七个项目的相关系数矩阵如下:

	hurdles	highjump	shot	run200m	longjump	javelin	run800m
hurdles	1.000	-0.811	-0.651	0.774	-0.912	-0.008	0.779
highjump	-0.811	1.000	0.441	-0.488	0.782	0.002	-0.591
shot	-0.651	0.441	1.000	-0.683	0.743	0.269	-0.420
run200m	0.774	-0.488	-0.683	1.000	-0.817	-0.333	0.617
longjump	-0.912	0.782	0.743	-0.817	1.000	0.067	-0.700
javelin	-0.008	0.002	0.269	-0.333	0.067	1.000	0.020
run800m	0.779	-0.591	-0.420	0.617	-0.700	0.020	1.000

对于标准化的数据，主成分分析的主要结果汇总如下:

```
Call:
princomp(x = scale(heptathlon[, 3:9]))

Standard deviations:
Comp.1 Comp.2 Comp.3 Comp.4 Comp.5 Comp.6 Comp.7
2.069  1.071  0.707  0.662  0.485  0.265  0.217

7  variables and  25 observations.
----
Loadings: (空白处为 0)
             Comp.1 Comp.2 Comp.3 Comp.4 Comp.5 Comp.6 Comp.7
 hurdles      0.453  0.158                       0.783  0.380
 highjump    -0.377 -0.248 -0.368 -0.680               0.434
 shot        -0.363  0.289  0.676 -0.124  0.512        0.218
 run200m      0.408 -0.260        -0.361  0.650       -0.453
 longjump    -0.456         0.139 -0.111 -0.184  0.590 -0.612
 javelin             0.842 -0.472 -0.121  0.135       -0.173
 run800m      0.375  0.224  0.396 -0.603 -0.504 -0.156
```

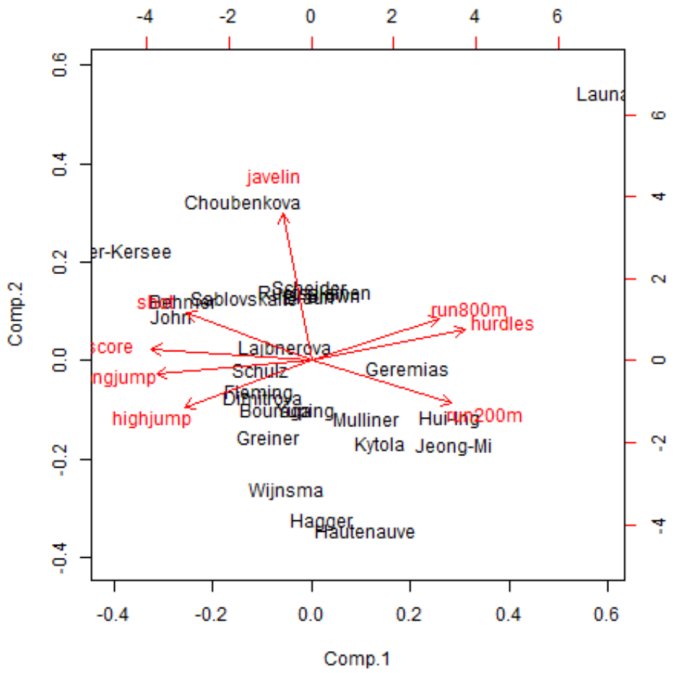
(a) 200 米短跑 (run200m) 成绩与跳远成绩的相关系数是多少？是正相关还是负相关？为什么？

(b) 七个项目成绩的相关系数矩阵的最大特征根是多少？

(c) 计算第一、二主成分的方差及其在总方差中的累积比例。

(d) 根据上面给出的载荷矩阵 (loadings, 空白处的值为 0), 结合下面的 PC1-PC2 散点图（特别关注其中比较特殊的运动员），试解释第一、二主成分的含义。

(e) 试分析奖牌获得者 (score 最大的 3 个运动员) 获胜的原因，你认为她们在哪些方面比较有优势？最后一名 Launa 有什么特点？



Solution:

(a)-0.817，负相关，跳远成绩越远越好，跑步成绩用时越短越好

(b) 最大特征根为 $2.069^2 = 4.281$.

(c) 第一主成分的方差为 4.281 , 第二主成分的方差为 1.147. 相应的在总方差中的累计比例为

$$\frac{2.069^2}{2.069^2+1.071^2+0.707^2+0.662^2+0.485^2+0.265^2+0.217^2} = 63.7\%$$

$$\frac{1.071^2+2.069^2}{2.069^2+1.071^2+0.707^2+0.662^2+0.485^2+0.265^2+0.217^2} = 80.8\%$$

注意：分母上的平方和即总方差可以直接使用相关系数矩阵的 trace，即 7。(但 R 计算出的 standard deviations 的平方和与 7 略有差异，这可能是因为将 loading 中部分载荷 round 成 0 导致的)。

(d) 第一主成分为除了标枪之外所有项目的总成绩（也可以认为是速度类项目的运动能力)；第二主成分代表（上肢）力量类运动能力。

(e) 奖牌获得者在 run200m 和 shot 上比较有优势，可能在这两项运动更容易拉开分差。Launa 的跳远、跳高、800 米和跨栏比较差，与其他几个项目包括标枪、铅球和 200 米等几乎正交即不相关，在这几个项目上表现一般，没有强项。

6.4

我们应用主成分分析方法研究欧洲 23 个国家的首都城市的气温数据集 temperature。23 个首都的信息如下表所示，其中地理位置 W、E、S、N 分别表示欧洲西、东、南、北部，每个城市记录了历史上 12 个月份 (Jan-Dec) 的平均气温。

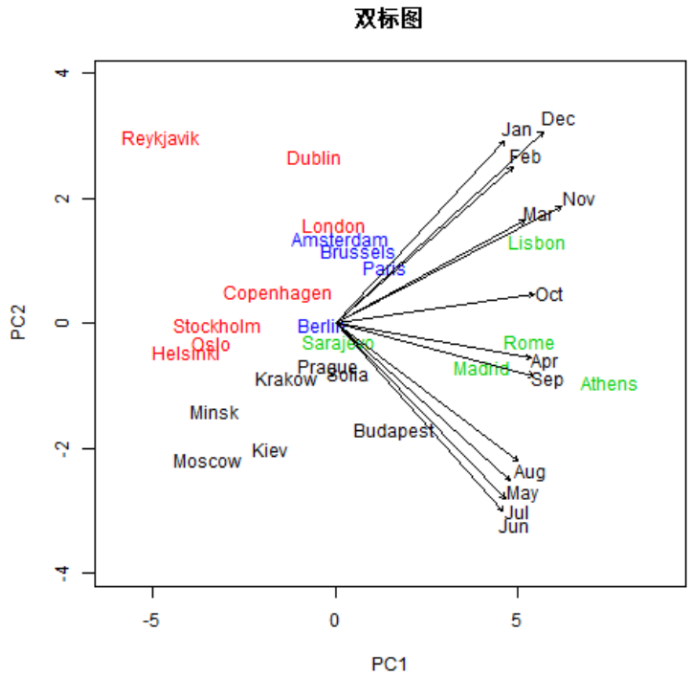
首都城市	国家	位置	首都城市	国家	位置	首都城市	国家	位置
Amsterdam (阿姆斯特丹)	荷兰	W	Athens (雅典)	希腊	S	Berlin (柏林)	德国	W
Brussels (布鲁塞尔)	比利时	W	Budapest (布达佩斯)	匈牙利	E	Copenhagen (哥本哈根)	丹麦	N
Dublin (都柏林)	爱尔兰	N	Helsinki (赫尔辛基)	芬兰	N	Kiev (基辅)	乌克兰	E
Krakow (克拉科夫)	波兰旧都	E	Lisbon (里斯本)	葡萄牙	S	London (伦敦)	英国	N
Madrid (马德里)	西班牙	S	Minsk (明斯克)	白俄罗斯	E	Moscow (莫斯科)	俄罗斯	E
Oslo (奥斯陆)	挪威	N	Paris (巴黎)	法国	W	Prague (布拉格)	捷克	E
Reykjavik (雷克雅未克)	冰岛	N	Rome (罗马)	意大利	S	Sarajevo (萨拉热窝)	波黑	S
Sofia (索菲亚)	保加利亚	E	Stockholm (斯德哥尔摩)	瑞典	N			

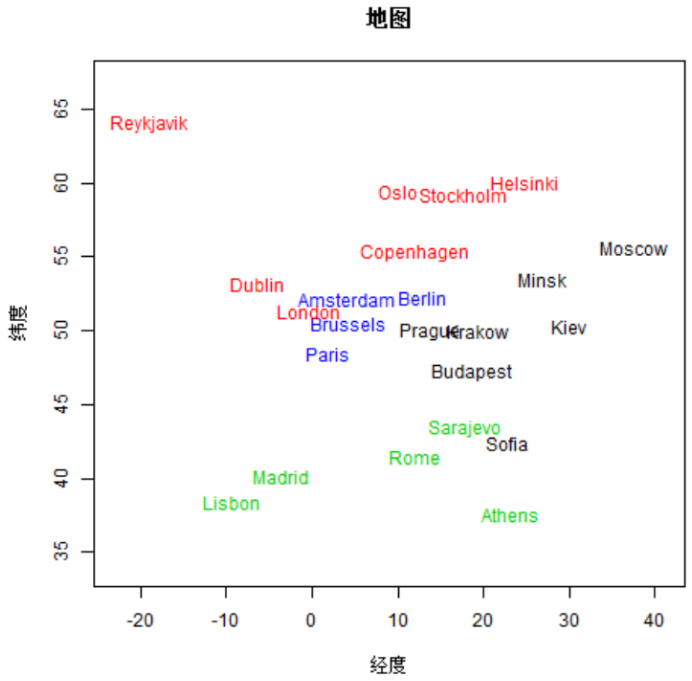
对标准化后的数据应用主成分分析方法 (R: princomp)，部分输出结果如下所示 (方框内)，包括主成分的标准差 (Standard deviations)，载荷或旋转矩阵 (Loadings) 的前两列，以及主成分分析的双标图 (biplot，左下图)。作为参考，右下图给出了这 23 个城市的经纬度坐标图。

```
Call: princomp(x =  temperature, cor=T)

Standard deviations:
Comp.1  Comp.2  Comp.3  Comp.4  Comp.5  Comp.6  Comp.7  Comp.8  Comp.9 Comp.10 Comp.11 Comp.12
3.16    1.36    0.36    0.20    0.13    0.11    0.08    0.05    0.03    0.03    0.02    0.01

Loadings:
     Comp.1 Comp.2
Jan   0.27   0.39
Feb   0.28   0.34
Mar   0.30   0.21
Apr   0.31  -0.07
May   0.28  -0.34
Jun   0.26  -0.40
Jul   0.27  -0.37
Aug   0.29  -0.30
Sep   0.31  -0.11
Oct   0.31   0.06
Nov   0.30   0.21
Dec   0.28   0.35
```





(a) 试计算第一、二主成分在总方差中的占比。根据载荷数据解释第一、二主成分的含义，它们与地理位置有什么关系?
(b) 四月份和哪个月份的气温相关系数最大? 哪个或哪几个月份的温度最能反映南北差异?
(c) 哪个城市或地区春秋两季 (四月和九月) 的温度相对较高? 萨拉热窝 (Sarajevo) 是南欧城市，其气温与其它南欧城市 (地图中的绿色标记城市) 相差较远，该城市与其它南欧城市在地理环境上有什么差异?

Solution:

(a) 第一、第二主成分在总方差中的占比分别为:

$$\frac{3.16^2}{12} = 0.829, \quad \frac{1.36^2}{12} = 0.154.$$

第一主成分代表平均温度，平均温度在南北方向上变化较大，所以我们也可以认为第一主成分代表南北或者纬度。
第二主成分代表温差也代表了东西方向。
(b)2 月和 9 月相关性最大 (最相似)，4、9、10 月接近平行于第一主成分方向 (代表南北), 这几个月份的温度男比差异较大。
(c)Athens 以及 Roma,Madrid 在 4 月和 9 月温度相对较高。Sarajevo 距离地中海较远

6.5

假设随机向量 $\mathbf{x} = (x_1, x_2, x_3)^\top$ 的协方差矩阵 (相关系数矩阵) 为

$$\Sigma = \begin{pmatrix} 1 & 0.63 & 0.45 \\ 0.63 & 1 & 0.35 \\ 0.45 & 0.35 & 1 \end{pmatrix}$$

假设 $\mathbf{x} = (x_1, x_2, x_3)^\top$ 由如下单因子模型模型产生

$$x_1 = 0.9F + \epsilon_1$$
$$x_2 = 0.7F + \epsilon_2$$
$$x_3 = 0.5F + \epsilon_3$$

其中 $\mathrm{var}(F)=1, \mathrm{cov}(F,\boldsymbol{\epsilon})=\mathbf{0}$, 写出载荷矩阵 L 并求 $\Psi=\mathrm{var}(\boldsymbol{\epsilon})$ 。

Solution:

$$L=(0.9,0.7,0.5)^T, \Psi=\Sigma-LL^T=\begin{pmatrix}0.19 & 0 & 0\\ 0 & 0.51 & 0\\ 0 & 0 & 0.75\end{pmatrix}$$

6.6

假设 $\mathbf{x}=(x_1,\ldots,x_6)^\top$ 的相关系数矩阵如下

$$R=\begin{pmatrix}1.000 & & & & & \\ 0.505 & 1.000 & & & & \\ 0.569 & 0.422 & 1.000 & & & \\ 0.602 & 0.467 & 0.926 & 1.000 & & \\ 0.621 & 0.482 & 0.877 & 0.874 & 1.000 & \\ 0.603 & 0.450 & 0.878 & 0.894 & 0.937 & 1.000\end{pmatrix}.$$

对两因子模型, 极大似然法得到如下载荷

变量	F_1	F_2
x_1	0.478	0.417
x_2	0.371	0.323
x_3	0.593	0.727
x_4	0.514	0.855
x_5	0.859	0.506
x_6	0.735	0.604

试计算

(a) 特殊方差.

(b) 公共方差.

(c) 每个因子解释的方差比例.

(d) 每个变量被 F_1, F_2 解释的方差的比例.

(e) 残差矩阵 $R-LL^\top-\Psi$.

Solution:

(a) 特殊方差 $\Psi_i=\sigma_{ii}^2-\left(l_{i1}^2+\cdots+l_{im}^2\right), i=1,\cdots,6$ 代入数据为 $(0.598,0.759,0.120,0.005,0.006,0.095)$

(b) 公共方差 $h_i=l_{i1}^2+\cdots+l_{im}^2, i=1,\cdots,6$ 代入数据为 $(0.402,0.241,0.880,0.995,0.994,0.905)$

(c) 记 L 的各列: $L=\left(\mathbf{l}_{(1)},\mathbf{l}_{(2)}\right)$ $SS_j=\left\|\mathbf{l}_{(j)}\right\|^2=l_{1j}^2+\cdots+l_{pj}^2$ 为 L 第 j 列的平方和第 j 个因子的方差贡献率率为 $SS_j/\mathrm{tr}(R), j=1,2$ 代入数据得: $(37.63\%, 35.98\%)$

(d) 第 i 个变量被 F_j 解释的方差的比例记为 $\beta_{ij}:=\frac{h_i^2}{\sigma_{ii}^2}$ 代入数据得

$$\beta_{i1}=(0.228,0.137,0.351,0.264,0.737,0.540)$$
$$\beta_{i2}=(0.174,0.104,0.528,0.731,0.256,0.365)$$

(e)

$$R - LL^T - \Psi = \begin{pmatrix} 0 & 0.192971 & -0.01761 & -0.00023 & -0.0006 & -0.0002 \\ 0.192971 & 0 & -0.03282 & 0.000141 & -0.00013 & -0.01778 \\ -0.01761 & -0.03282 & 0 & -0.00039 & -0.00025 & 0.003037 \\ -0.00023 & 0.000141 & -0.00039 & 0 & -0.00016 & -0.00021 \\ -0.0006 & -0.00013 & -0.00025 & -0.00016 & 0 & 0.000011 \\ -0.0002 & -0.01778 & 0.003037 & -0.00021 & 0.000011 & 0 \end{pmatrix}$$

6.7

假设 $n \times 2$ 数据矩阵 $X = (\mathbf{x}_1, \mathbf{x}_2), \|\mathbf{x}_1\| = a, \|\mathbf{x}_2\| = b, a > b.$ 假设 $\mathbf{x}_1^\top \mathbf{x}_2 = 0$, 求 X 的奇异值分解

Solution:

$$X^T X = \begin{pmatrix} \mathbf{x}_1^T \mathbf{x}_1 & \mathbf{x}_1^T \mathbf{x}_2 \\ \mathbf{x}_2^T \mathbf{x}_1 & \mathbf{x}_2^T \mathbf{x}_2 \end{pmatrix} = \begin{pmatrix} a^2 & 0 \\ 0 & b^2 \end{pmatrix}, \text{ 特征根}a^2, b^2 \text{ 特征向量为}(1,0)^\top, (0,1)^\top,$$

$$XX^T = \mathbf{x}_1 \mathbf{x}_1^T + \mathbf{x}_2 \mathbf{x}_2^T, \text{ 特征根}a^2, b^2 \text{ 特征向量为}\mathbf{x_1}, \mathbf{x}_2,$$

由上可知 $X = \begin{pmatrix} \mathbf{x}_1/a & \mathbf{x}_2/b \end{pmatrix} \begin{pmatrix} a & 0 \\ 0 & b \end{pmatrix} I_2.$

6.8

假设 A 是 $n \times m$ 矩阵, $\mathbf{u} \in R^n, \mathbf{v} \in R^m$, 假设 $A\mathbf{v} = \alpha \mathbf{u}, A^\top \mathbf{u} = \beta \mathbf{v}$, 其中实数 $\alpha, \beta \neq 0$ 。证明 α 和 β 同号且 $\sqrt{\alpha\beta}$ 是 A 的一个奇异值。

Solution:

$$AA^T u = \alpha\beta u$$

$$A^T A v = \alpha\beta v.$$

$A^T A$ 半正定，所以 $\alpha\beta \geq 0$，α 和 β 同号。由奇异值分解可知, $\sqrt{\alpha\beta}$ 是 A 的一个奇异值。

6.9

假设 y 是随机变量, $\mathbf{x}$ 是 $p \times 1$ 随机向量, 假设它们的协方差矩阵为

$$\Sigma = \operatorname{cov}\begin{pmatrix} \mathbf{x} \\ y \end{pmatrix} = \begin{pmatrix} \Sigma_{\mathbf{xx}} & \Sigma_{\mathbf{x}y} \\ \Sigma_{y\mathbf{x}} & \Sigma_{yy} \end{pmatrix} > 0$$

其中 $\Sigma_{yy} = \sigma_y^2$ 是标量（正实数）, $\Sigma_{\mathbf{x}y}$ 是 $p \times 1$ 向量。

(a) 试求 $\mathbf{x}, y$ 的第一典则相关系数 $\sqrt{\lambda_1}$ (也是唯一的非 0 典则相关系数) 和第一对 (也是唯一的一对) 典则变量 (ξ_1, η_1) 。

(b) 假设 $y, \mathbf{x}$ 均值都为 0 , 满足线性模型

$$y = \beta^\top \mathbf{x} + \epsilon, \epsilon \sim \left(0, \sigma^2\right), \epsilon \perp\!\!\!\perp \mathbf{x},$$

证明该模型等价于

$$\eta_1 = \sqrt{\lambda_1}\xi_1 + \delta_1, \delta_1 \sim \left(0, 1 - \lambda_1\right), \delta_1 \perp\!\!\!\perp \xi_1$$

(验证: $\beta^\top \mathbf{x} = \sqrt{\lambda_1}\sigma_y \times \xi_1, y = \sigma_y \times \eta_1$).

Solution:

(a) $A=\Sigma_{\mathbf{xx}}^{-\frac{1}{2}}\Sigma_{\mathbf{x}y}\Sigma_{yy}^{-\frac{1}{2}}$ 是 $p\times 1$ 向量

$$\Rightarrow A^\top A=\Sigma_{yy}^{-1/2}\Sigma_{y\mathbf{x}}\Sigma_{\mathbf{xx}}^{-1}\Sigma_{\mathbf{x}y}\Sigma_{yy}^{-1/2}=\frac{\Sigma_{y\mathbf{x}}\Sigma_{\mathbf{xx}}^{-1}\Sigma_{\mathbf{x}y}}{\Sigma_{yy}},$$

这是一个标量（实数），其唯一非 0 特征根 λ_1 是它本身，所以第一典则相关系数

$$\sqrt{\lambda_1}=\sqrt{A^\top A}=\sqrt{\frac{\Sigma_{y\mathbf{x}}\Sigma_{\mathbf{xx}}^{-1}\Sigma_{\mathbf{x}y}}{\Sigma_{yy}}}.$$

标量 $A^\top A$ 的单位长特征向量为 $v_1=1$ （或-1)。另外，对于 $p\times 1$ 向量 A, $AA^\top$ 的唯一非 0 特征根为 λ_1, 对应的特征向量为 A, 其模长为 $\sqrt{A^\top A}=\sqrt{\lambda_1}$, 所以模长为 1 的特征向量为

$$\mathbf{u}_1=A/\sqrt{\lambda_1}=\Sigma_{\mathbf{xx}}^{-\frac{1}{2}}\Sigma_{\mathbf{x}y}/\sqrt{\Sigma_{y\mathbf{x}}\Sigma_{\mathbf{xx}}^{-1}\Sigma_{\mathbf{x}y}},$$

综上，A 的 SVD 为 $A=u_1\sqrt{\lambda_1}v_1$. 故第一典则变量为

$$\xi_1=\mathbf{u}_1^\top\Sigma_{\mathbf{xx}}^{-1/2}\mathbf{x}=\Sigma_{y\mathbf{x}}\Sigma_{\mathbf{xx}}^{-1}\mathbf{x}/\sqrt{\Sigma_{y\mathbf{x}}\Sigma_{\mathbf{xx}}^{-1}\Sigma_{\mathbf{x}y}},\ \eta_1=v_1^\top\Sigma_{yy}^{-1/2}y=y/\sqrt{\Sigma_{yy}}=y/\sigma_y.$$

(b)
由于 $y,\mathbf{x}$ 均值都为 0，$\beta^T=\Sigma_{\mathbf{yx}}\Sigma_{\mathbf{xx}}^{-1}$ 两侧同时除以 σ_y，得到

$$\eta_1=\sqrt{\lambda_1}\xi_1+\frac{\epsilon}{\sigma_y}=\sqrt{\lambda_1}\xi_1+\delta_1$$

由于 $1-\lambda_1=\frac{\Sigma_{yy\bullet x}}{\sigma_y^2}$ 和 $\sigma^2=\Sigma_{yy\bullet x}$ 可知

$$\delta_1\sim(0,1-\lambda_1)$$

独立性易得

6.10

(非典型 CCA) 假设 $\mathbf{x},\mathbf{y}$ 分别是 $p\times 1, q\times 1$ 随机向量, 它们的协方差矩阵

$$\Sigma=\operatorname{cov}\begin{pmatrix}\mathbf{x}\\ \mathbf{y}\end{pmatrix}=\begin{pmatrix}\Sigma_{\mathbf{xx}} & \Sigma_{\mathbf{xy}}\\ \Sigma_{\mathbf{yx}} & \Sigma_{\mathbf{yy}}\end{pmatrix}>0,$$

假设奇异值分解 $\Sigma_{\mathbf{xy}}=UDV^\top$ 。求解方向（单位模长的常数向量）$\mathbf{u}\in R^p,\mathbf{v}\in R^q,\|\mathbf{u}\|=\|\mathbf{v}\|=1$, 使得 $\operatorname{cov}\left(\mathbf{u}^\top\mathbf{x},\mathbf{v}^\top\mathbf{y}\right)$ 最大 (最优解以 Σ 及 U,D,V 表示)。
注解：典则相关分析 (CCA) 极大化两个随机向量 $\mathbf{x},\mathbf{y}$ 的线性组合（投影坐标）之间的相关系数, 第一对典则相关变量最大可能地保留了两个原始随机向量之间的相关性, 这里“相关性”以相关系数度量。非典型 CCA 是 CCA 的简化版本: 求解两个最优的方向, 使得随机向量 $\mathbf{x},\mathbf{y}$ 在这两个方向上分别投影得到的投影坐标之间的协方差最大, 我们同样要求最优解能最大可能地保留两个原始随机向量 $\mathbf{x},\mathbf{y}$ 之间的相关性, 只不过这里我们以协方差度量“相关性”。
Solution:

$$\operatorname{cov}(u^T\mathbf{x},v^T\mathbf{x})=u^T\Sigma_{\mathbf{xy}}v=u^TUDVv^T$$

$$max\operatorname{cov}(u^T\mathbf{x},v^T\mathbf{x})=\sqrt{\lambda_1(UDV^TVDU^T)}=\sqrt{\lambda_1(UD^2U^T)}=\sqrt{\lambda_1(D^2)}$$

记 $U=(u_1,\cdots,u_r),V=(v_1,\cdots,v_r)$，在 $u=u_1,v=v_1$ 时达到 $\operatorname{cov}(u^T\mathbf{x},v^T\mathbf{x})$ 最大值