hw7

7.1

假设随机向量 $\binom{\mathbf{y}_{q\times 1}}{\mathbf{x}_{p\times 1}}$ 的协方差矩阵为相关系数矩阵 $R=(\rho_{ij})$, 其中 $\rho_{ii}=1, i=1,\ldots,p+q$ 。证明 $\mathbf{x}$ 和 $\mathbf{y}$ 的第一典则相关系数 $\sqrt{\lambda_1}\geq\rho_{ij}, i=1,\ldots,q; j=q+1,\ldots,q+p$ 。

Solution:

$$
\begin{aligned}
\sqrt{\lambda_1} &= \max_{a,b}\rho\left(a^\top x, b^\top y\right)\\
&\geqslant \rho\left(e_i^\top x, e_j^\top y\right)\\
&= \rho_{ij}
\end{aligned}
$$

7.2

假设随机向量 $\binom{\mathbf{y}_{q\times 1}}{\mathbf{x}_{p\times 1}}$ 的协方差矩阵为 $\Sigma=(1-\rho)I_m+\rho\mathbf{1}_m\mathbf{1}_m^\top, m=p+q, \rho>-1/(m-1)$, 试求 $\mathbf{x},\mathbf{y}$ 的第一典则相关系数和典则变量 (其中 I_m 是 m 阶单位阵, $\mathbf{1}_m$ 是长度为 m 的元素全为 1 的向量)。

Solution:

$$
\frac{a^\top\Sigma_{xy}b}{\sqrt{a^\top\Sigma_{xx}a}\sqrt{b^\top\Sigma_{yy}b}}=\frac{\rho(\sum_{i=1}^p a_i)(\sum_{i=1}^q b_j)}{\sqrt{(1-\rho)\sum_{i=1}^p a_i^2+\rho(\sum_{i=1}^p a_i)^2}\sqrt{(1-\rho)\sum_{i=1}^q b_j^2+\rho(\sum_{i=1}^q b_j)^2}}
$$

极大化上式可得第一典则相关系数不妨设 $\|a\|=\|b\|=1$

$$
\begin{aligned}
&\frac{\sum_{i=1}^p a_i}{\sqrt{(1-\rho)+\rho\left(\sum_{i=1}^p a_i\right)^2}}\\
=&\frac{sign(\sum_{i=1}^p a_i)}{\sqrt{(1-\rho)\frac{1}{\left(\sum_{i=1}^p a_i\right)^2}+\rho}}
\end{aligned}
$$

极小化 $\sum_{i=1}^p a_i$ 在 $\sum_{i=1}^p a_i^2=1$ 条件下

$$
\begin{aligned}
L(a,\lambda) &= \sum_{i=1}^p a_i+\lambda\left(\sum_{i=1}^p a_i^2-1\right)\\
\frac{\partial L}{\partial a_i} &= 1+2\lambda a_i=0 \quad \Rightarrow a_i=-\frac{1}{2\lambda}\\
\frac{\partial L}{\partial\lambda} &= \sum_{i=1}^p a_i^2-1=0
\end{aligned}
$$

$$
\begin{aligned}
&\Rightarrow \frac{1}{4\lambda^2}=\frac{1}{p}\\
&\Rightarrow a_i^2=\frac{1}{p}\\
\text{同理} &\Rightarrow b_j^2=\frac{1}{q}
\end{aligned}
$$

所以第一典则相关系数为 $\frac{\rho\sqrt{p}\sqrt{q}}{\sqrt{(1-\rho)+\rho p}\sqrt{(1-\rho)+\rho q}}$，第一典则变量 $(\xi_1,\eta_1)=(\frac{1}{\sqrt{p}}\mathbf{1}_p^\top\mathbf{x},\frac{1}{\sqrt{q}}\mathbf{1}_q^\top\mathbf{y})$或$(-\frac{1}{\sqrt{p}}\mathbf{1}_p^\top\mathbf{x},-\frac{1}{\sqrt{q}}\mathbf{1}_q^\top\mathbf{y})$

7.3

证明在 p 维欧氏空间中至多存在 $p+1$ 个点, 它们两两之间的欧氏距离全相等。

Solution:

假设 p 维空间存在 $p+2$ 个点，它们之间的欧氏距离为 1

$$D=(d_ij)=\begin{pmatrix}0&1&1&\cdots&1\\1&0&1&\cdots&1\\1&1&0&\cdots&1\\\vdots&\vdots&\vdots&\ddots&\vdots\\1&1&1&\cdots&0\end{pmatrix}_{(p+2)\times(p+2)}$$

$$\tilde{D}=\left(-\frac{1}{2}d_{ij}^2\right)=-\frac{1}{2}D$$

$$J_{p+2}=I_{p+2}-\mathbf{1}_{p+2}\mathbf{1}_{p+2}^\top/(p+2)$$

$$\begin{aligned}J_{p+2}\tilde{D}J_{p+2}&=-\frac{1}{2}\left(D-\mathbf{1}_{p+2}\left(\frac{p+1}{p+2}\mathbf{1}_{p+2}\right)^\top-\left(\frac{p+1}{p+2}\mathbf{1}_{p+2}\right)\mathbf{1}_{p+2}^\top+\frac{(p+2)(p+1)}{(p+2)^2}\mathbf{1}_{p+2}\mathbf{1}_{p+2}^\top\right)\\&=\tilde{D}-\frac{p+1}{p+2}\mathbf{1}_{p+2}\mathbf{1}_{p+2}^\top\\&=\begin{pmatrix}-\frac{p+1}{p+2}&\frac{1}{p+2}&\cdots&\frac{1}{p+2}\\\frac{1}{p+2}&-\frac{p+1}{p+2}&\cdots&\frac{1}{p+2}\\\vdots&\vdots&\ddots&\vdots\\\frac{1}{p+2}&\frac{1}{p+2}&\cdots&-\frac{p+1}{p+2}\end{pmatrix}\end{aligned}$$

显然 $J_{p+2}\tilde{D}J_{p+2}$ 不是半正定矩阵（对角元为负数），根据 lecture19 命题 2，D 不是欧氏距离矩阵

7.4

假设 $S=(s_{ij})$ 是 $n\times n$ 对称矩阵, $s_{ij}\geq 0$, 记 $d_i=\sum_{j=1}^n s_{ij}, D=\operatorname{diag}(d_1,\ldots,d_n)$, 假设 $d_i>0, i=1,\ldots,n$ 。令拉普拉斯矩阵（Laplacian）$L\triangleq D-S$, 令 $B=D^{-1}S$ 。

(a) 证明 L 是双向中心化的（即行和、列和都为 0), 且 0 一定是 L 的最小特征根。

(b) 对任何 $\mathbf{x}=(x_1,\ldots,x_n)^\top\in R^n$,

$$\mathbf{x}^\top L\mathbf{x}=\frac{1}{2}\sum_{1\leq i,j\leq n}s_{ij}\left(x_i-x_j\right)^2$$

(c) 说明 B 的所有特征根 λ 都是实数, 证明 $|\lambda|\leq 1$, 且 1 是最大特征根。

(d)（选）若 $s_{ij}>0$, 我们称节点 i,j 是相邻的, 记作 $i\sim j$ 。证明 $\lambda=-1$ 是 B 的特征根当且仅当图所有 n 个节点可由两种颜色染色, 使得相邻的节点不同色。

(e)（选）假设 $n\times n$ 对称矩阵 $S=(s_{ij})$ 中 $s_{ii}=1, 1\leq i\leq n, s_{i,i+1}=s_{i+1,i}=\rho, 1\leq i\leq n-1, 0<\rho\leq 1$ 。其它元素为 0 。试证明或否定 L 的最小非 0 特征根对应的特征向量的所有分量是单调的。

Solution:

(a)

$$L=(l_{ij})=\begin{pmatrix}d_1&&\\&\ddots&\\&&d_n\end{pmatrix}-\begin{pmatrix}s_{11}&\cdots&s_{1n}\\\vdots&&\vdots\\s_{n1}&\cdots&s_{nn}\end{pmatrix}$$

$$\sum_{i=1}^n l_{ij}=d_j-\sum_{i=1}s_{ij}=0$$

$$\sum_{j=1}^n l_{ij}=d_i-\sum_{j=1}s_{ij}=0$$

所以 L 是双向中心化的。
由于 L 双向中心化，可知

$$L\mathbf{1}=D\mathbf{1}-S\mathbf{1}=0$$

0 为 L 的特征根，再由 (b) 知 L 半正定，所以 0 为 L 的最小特征根。
(b)

$$\forall \mathbf{x}=\left(\begin{array}{c} x_1 \\ \vdots \\ x_n \end{array}\right) \in \mathbb{R}^{n\times 1}$$

$$\begin{aligned}
x^{\top}Lx &= x^{\top}(D-S)x \\
&= \sum_{i=1}^{n} d_i x_i^2 - \left(\sum_{i=1}^{n} x_i s_{i1}, \ldots, \sum_{i=1}^{n} x_i s_{in}\right)\left(\begin{array}{c} x_1 \\ \vdots \\ x_n \end{array}\right) \\
&= \sum_{i=1}^{n}\sum_{j=1}^{n} s_{ij}x_i^2 - \sum_{j=1}^{n}\sum_{i=1}^{n} x_j x_i s_{ij} \\
&= \frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n} s_{ij}\left(x_i-x_j\right)^2
\end{aligned}$$

(c) 因为对任何方阵 A, B, AB 和 BA 有相同的非 0 特征根, 故 $B=D^{-1}S$ 与 $D^{-1/2}SD^{-1/2}$ 有相同的非 0 特征根, 后者是对称矩阵, 特征根为实数, 所以 B 的特征根全是实数。假设 λ 是 B 的一个特征根, 对应的特征向量为 $\mathbf{x}=(x_1,...,x_n)^{\top}\neq\mathbf{0}$, 则有特征方程

$$B\mathbf{x}=D^{-1}S\mathbf{x}=\lambda\mathbf{x},$$

记 $|x_k|=\max_{1\leq i\leq n}|x_i|$, 必定 $x_j\neq 0$。考察特征方程两边的第 k 个分量

$$|\lambda x_k| = |\sum_{j=1}^{n} s_{kj}x_j/d_k| \leq \sum_{j=1}^{n} s_{kj}|x_j|/d_k \leq \sum_{j=1}^{n} s_{kj}|x_k|/d_k = |x_k|$$

由此得 $|\lambda|\leq 1$ (因为 $x_k\neq 0$)。另外，由 (a), $D\mathbf{1}=S\mathbf{1}\Rightarrow D^{-1}S\mathbf{1}=\mathbf{1}$, 这表明 1 是 B 的特征根。

(d) 若 -1 是 B 的特征根，假设其特征向量为 $\mathbf{x}$, 则有

$$D^{-1}S\mathbf{x}=-\mathbf{x}\Rightarrow(D+S)\mathbf{x}=0\Rightarrow\mathbf{x}^{\top}(D+S)\mathbf{x}=0,$$

与 (b) 类似，我们有

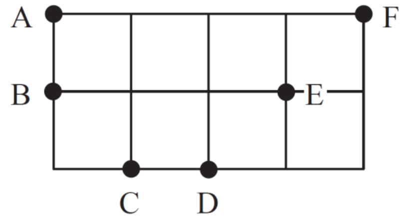
$$0=\mathbf{x}^{\top}(D+S)\mathbf{x}=\frac{1}{2}\sum_{i,j}s_{ij}(x_i+x_j)^2$$

这说明，若 $s_{ij}>0$ 即 $i\sim j$, 则 $x_i=-x_j$，进一步，若 $j\sim k$, 则 $x_k=-x_j=x_i$。不妨假设所有节点是全连通的即对任何 i,j，存在一条路径连接它们: $i\sim i_1\sim ...\sim i_k\sim j$），路径上的 x 值的绝对值全部相等，相邻的符号相反，这表明节点可染成两种颜色，相邻的节点不同色。

(e) （still open)

7.5

考虑下图平面方格上的 5 个点 A-F, 计算它们之间的 Manhattan 距离矩阵, 应用单连结层次聚集聚类法进行聚类。画出树图。



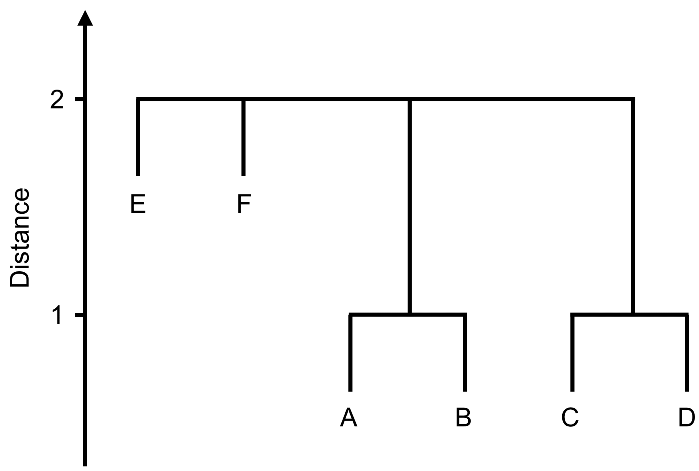
Solution:

(以下只展示结果，考试时需要写清楚过程)

Manhattan 距离矩阵为:

$$\begin{pmatrix} 0 \\ 1 & 0 \\ 3 & 2 & 0 \\ 4 & 3 & 1 & 0 \\ 4 & 3 & 3 & 2 & 0 \\ 4 & 5 & 5 & 4 & 2 & 0 \end{pmatrix}$$

树图为:



7.6

假设物件 $a-e$ 的距离矩阵如下: 我们的目标是应用 K-中心方法将 $a-e$ 聚集为 $K=2$ 类, 假设初始指定两类的中心 (medoids) 各为 a 和 b, 试应用 K-中心 (K-medoid) 方法进行聚类。

$$D = \begin{array}{c} \\ a \\ b \\ c \\ d \\ e \end{array} \begin{array}{c} \begin{array}{ccccc} a & b & c & d & e \end{array} \\ \begin{pmatrix} 0 & 1 & 3 & 4 & 3 \\ 1 & 0 & 2 & 3 & 2 \\ 3 & 2 & 0 & 1 & 2 \\ 4 & 3 & 1 & 0 & 1 \\ 3 & 2 & 2 & 1 & 0 \end{pmatrix} \end{array}$$

Solution:

（以下只展示结果，考试时需要写清楚过程）

b，d 为 medoids，ab 被分为一类，cde 被分为另一类。