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

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

2. 假设随机向量 $\begin{pmatrix} \mathbf{y}_{q\times 1} \\ \mathbf{x}_{p\times 1} \end{pmatrix}$ 的协方差矩阵为 $\Sigma = (1-\rho)I_m + \rho \mathbb{1}_m \mathbb{1}_m^\top$, $m = p+q$, $\rho > -1/(m-1)$, 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，$\mathbb{1}_m$ 是长度为 m 的元素全为 1 的向量)。

配列 (seriation), **多维标度法**（MDS）：

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。

4. 假设 $S = (s_{ij})$ 是 $n \times n$ 对称矩阵, $s_{ij} \geq 0$，记 $d_i = \sum_{j=1}^n s_{ij}$, $D = diag(d_1, ..., d_n)$, 假设 $d_i > 0, i = 1, ..., n$。令拉普拉斯矩阵（Laplacian）$L \stackrel{\Delta}{=} D - S$，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 $\mathbf{x} = (x_1, ..., x_n)^\top \in R^n$,
   $$\mathbf{x}^\top L \mathbf{x} = \frac{1}{2} \sum_{1 \leq i,j \leq n} w_{ij}(x_i - x_j)^2.$$

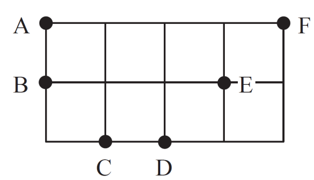
   (c) 说明 B 的所有特征根 λ 都是实数，证明 $|\lambda| \leq 1$, 且 1 是最大特征根。

   (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。

聚类分析 (Cluster analysis)：

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。



6. 假设物件 a-e 的距离矩阵如下：

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

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。