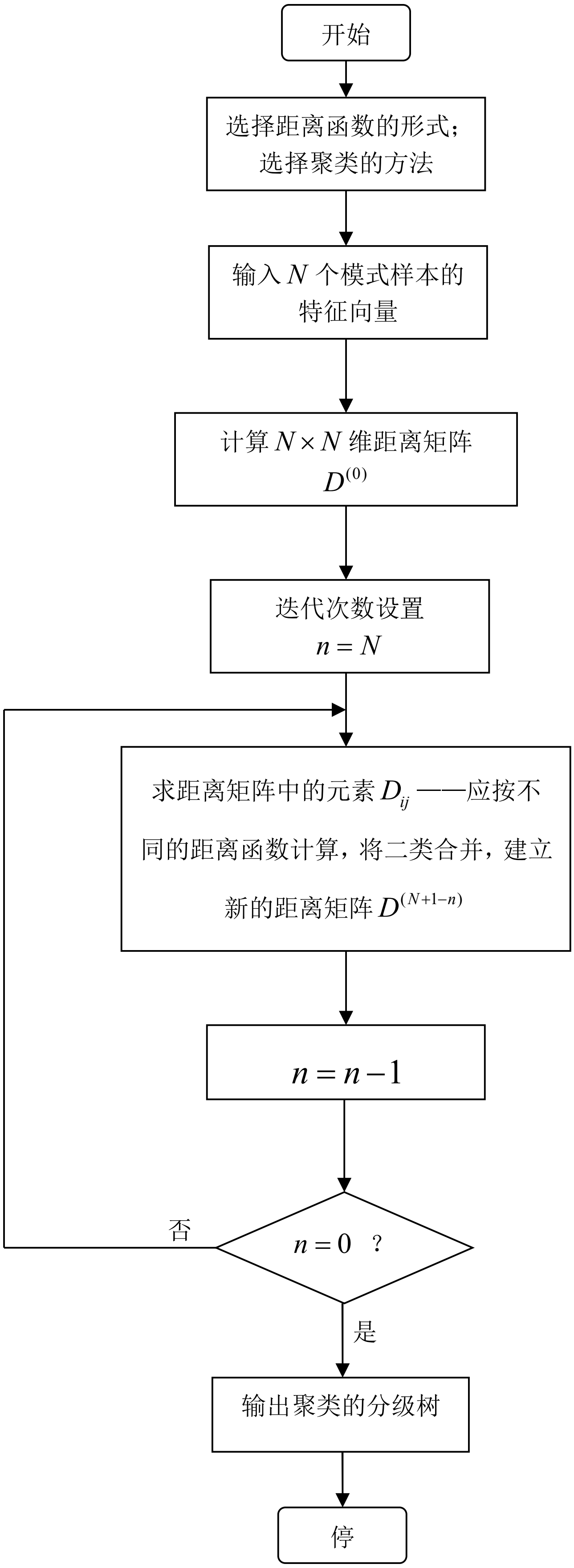
题 1：画出给定迭代次数为 n 的系统聚类法的算法流程框图



题 2：对如下 5 个 6 维模式样本，用最小聚类准则进行系统聚类分析

x1: 0, 1, 3, 1, 3, 4
x2: 3, 3, 3, 1, 2, 1
x3: 1, 0, 0, 0, 1, 1
x4: 2, 1, 0, 2, 2, 1
x5: 0, 0, 1, 0, 1, 0

第 1 步：将每一样本看成单独一类，得

$G_1^{(0)}=\{x_1\},G_2^{(0)}=\{x_2\},G_3^{(0)}=\{x_3\}$

$G_4^{(0)}=\{x_4\},G_5^{(0)}=\{x_5\}$

计算各类之间的欧式距离，可得距离矩阵 $D^{(0)}$

	$G_1^{(0)}$	$G_2^{(0)}$	$G_3^{(0)}$	$G_4^{(0)}$	$G_5^{(0)}$
$G_1^{(0)}$	0				
$G_2^{(0)}$	$\sqrt{23}$	0			
$G_3^{(0)}$	$\sqrt{25}$	$\sqrt{24}$	0		
$G_4^{(0)}$	$\sqrt{24}$	$\sqrt{15}$	$\sqrt{7}$	0	
$G_5^{(0)}$	$\sqrt{26}$	$\sqrt{25}$	$\sqrt{3}$	$\sqrt{12}$	0

第 2 步：矩阵 $D^{(0)}$ 中最小元素为 $\sqrt{3}$，它是 $G_3^{(0)}$ 和 $G_5^{(0)}$ 之间的距离，将他们合并为一类，得新的分类为

$G_1^{(1)}=\{G_1^{(0)}\},G_2^{(1)}=\{G_2^{(0)}\},G_3^{(1)}=\{G_3^{(0)},G_5^{(0)}\},G_4^{(1)}=\{G_4^{(0)}\}$

计算聚类后的距离矩阵 $D^{(1)}$

	$G_1^{(1)}$	$G_2^{(1)}$	$G_3^{(1)}$	$G_4^{(1)}$
$G_1^{(1)}$	0			
$G_2^{(1)}$	$\sqrt{23}$	0		
$G_3^{(1)}$	$\sqrt{25}$	$\sqrt{24}$	0	
$G_4^{(1)}$	$\sqrt{24}$	$\sqrt{15}$	$\sqrt{7}$	0

第 3 步：由于$D^{(1)}$中距离最小者为$\sqrt{7}$，它是$G_3^{(1)}$与$G_4^{(1)}$之间的距离，于是合并$G_3^{(1)}$和$G_4^{(1)}$，得新的分类为

$$G_1^{(2)}=\{G_1^{(1)}\},G_2^{(2)}=\{G_2^{(2)}\},G_3^{(2)}=\{G_3^{(1)},G_4^{(1)}\}$$

同样，按最小距离准则计算距离矩阵$D^{(2)}$，得

	$G_1^{(2)}$	$G_2^{(2)}$	$G_3^{(2)}$
$G_1^{(2)}$	0		
$G_2^{(2)}$	$\sqrt{23}$	0	
$G_3^{(2)}$	$\sqrt{24}$	$\sqrt{15}$	0

第 4 步：同理得

$$G_1^{(3)}=\{G_1^{(2)}\},G_2^{(3)}=\{G_2^{(2)},G_3^{(2)}\}$$

满足聚类要求，如聚为 2 类，聚类完毕。

题 3：选 $k=2$， $z_1(1)=x_1, z_2(1)=x_{10}$，用 K-均值算法进行聚类分析

第一步：选取 $z_1(1)=x_1=\begin{pmatrix}0\\0\end{pmatrix}, z_2(1)=x_{10}=\begin{pmatrix}7\\6\end{pmatrix}$

第二步：根据聚类中心进行聚类，得到

$$S_1(1)=\{x_1,x_2,x_3,x_4,x_5,x_6,x_7,x_8\}$$
$$S_2(1)=\{x_9,x_{10},x_{11},x_{12},\cdots x_{20}\}$$

第三步：计算新的聚类中心

$$z_1(2)=\frac{1}{N_1}\sum_{x\in S_1(1)}x=\frac{1}{8}(x_1+x_2+\cdots+x_8)=\begin{pmatrix}1.2500\\1.1250\end{pmatrix}$$
$$z_2(2)=\frac{1}{N_2}\sum_{x\in S_2(1)}x=\frac{1}{12}(x_9+x_{10}+\cdots+x_{20})=\begin{pmatrix}7.6667\\7.3333\end{pmatrix}$$

第四步：因 $z_j(2)\neq z_j(1), j=1,2$，故回到第二步

第二步：根据新的聚类中心重新进行聚类，得到

$$S_1(2)=\{x_1,x_2,x_3,x_4,x_5,x_6,x_7,x_8\}$$
$$S_2(2)=\{x_9,x_{10},x_{11},x_{12},\cdots x_{20}\}$$

第三步：计算新的聚类中心

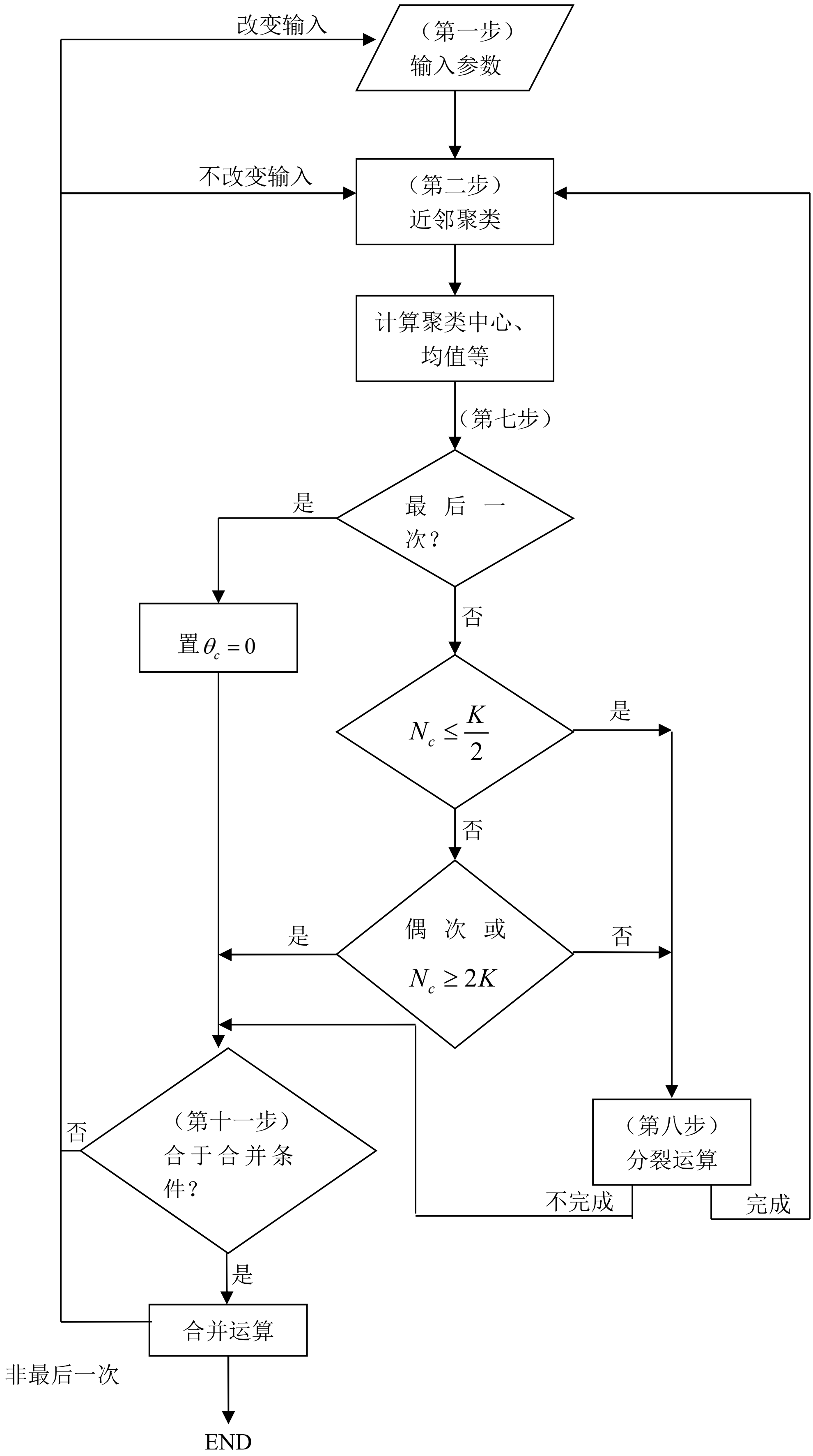
$$z_1(3)=\frac{1}{N_1}\sum_{x\in S_1(2)}x=\frac{1}{8}(x_1+x_2+\cdots+x_8)=\begin{pmatrix}1.2500\\1.1250\end{pmatrix}$$
$$z_2(3)=\frac{1}{N_2}\sum_{x\in S_2(2)}x=\frac{1}{12}(x_9+x_{10}+\cdots+x_{20})=\begin{pmatrix}7.6667\\7.3333\end{pmatrix}$$

第四步： $z_j(3)=z_j(2), j=1,2$，所以算法收敛，得聚类中心为

$$z_1=\begin{pmatrix}1.2500\\1.1250\end{pmatrix}, z_2=\begin{pmatrix}7.6667\\7.3333\end{pmatrix}$$

迭代结束。

题 4：画出 ISODATA 算法的流程框图



题 5：试用 ISODATA 算法对如下模式分布进行聚类分析：
{x1(0，0)，x2(3,8)，x3(2,2)，x4(1,1)，x5(5,3)，x6(4,8)，x7(6,3)，x8(5,4)，x9(6,4)，x10(7,5)}

从题目中我们可知，$N=10$。假如取初始值 $N_c=1$，$z_1=\begin{pmatrix}0\\0\end{pmatrix}$，则具体运算步骤为：

第一步：取参数 $K=2,\theta_N=1,\theta_s=1,\theta_c=4,L=0,I=4$。

第二步：因只有一个聚类中心，故 $S_1=\{x_1,x_2,\cdots,x_{10}\}$ 和 $N_1=10$。

第三步：因 $N_1>\theta_N$，无子集可抛弃。

第四步：修改聚类中心

$$z_1=\frac{1}{N_1}\sum_{x\in S_1}x=\begin{pmatrix}3.90\\3.80\end{pmatrix}$$

第五步：计算 $\bar{D}_j$

$$\bar{D}_1=\frac{1}{N_1}\sum_{x\in S_1}\|x-z_1\|=3.0749$$

第六步：计算 $\bar{D}$

$$\bar{D}=\bar{D}_1=3.0749$$

第七步：因还不是最后一次迭代，且 $N_c=K/2$，故进入第八步。

第八步：求 S_1 中的标准差向量

$$\sigma_1=\begin{pmatrix}2.2113\\2.5219\end{pmatrix}$$

第九步：σ_1 中最大分量是 2.5219，因此 $\sigma_{1\max}=2.5219$

第十步：因 $\sigma_{1\max}>\theta_S$ 且 $N_c=K/2$，可将 z_1 分裂成两个新的聚类。设 $r_j=0.5\sigma_{j\max}=1.26$，则

$$z_1=\begin{pmatrix}3.90\\5.06\end{pmatrix}，\quad z_2=\begin{pmatrix}3.90\\2.54\end{pmatrix}，$$ N_c 增加 1，跳回到第二步。

第二步：新的样本集为

$$S_1=\{x_2,x_6,x_8,x_9,x_{10}\}，\quad S_2=\{x_1,x_3,x_4,x_5,x_7\}$$

则 $N_1 = N_2 = 5$。

第三步：因 N_1 和 N_2 都大于 θ_N，无子集可抛弃。

第四步：修改聚类中心

$$z_1 = \begin{pmatrix} 5.00 \\ 5.80 \end{pmatrix},\quad z_2 = \begin{pmatrix} 2.80 \\ 1.80 \end{pmatrix}$$

第五步：计算 $\bar{D}_j, j = 1,2$

$$\bar{D}_1 = 2.28\text{，}\quad \bar{D}_2 = 2.41$$

第六步：计算 $\bar{D}$

$$\bar{D} = 2.3450$$

第七步：因这是偶迭代次数，因此进入第十一步

第十一步：因 $L = 0$，故聚类中心不发生合并，转至第十四步

第十四步：因还不是最后一次迭代，且经判断不需要修改给定的参数，回到第二步

第二步：新的样本集为

$$S_1 = \{x_2, x_6, x_7, x_8, x_9, x_{10}\}\text{，}\quad S_2 = \{x_1, x_3, x_4, x_5\}$$

则 $N_1 = 6, N_2 = 4$。

第三步：因 N_1 和 N_2 都大于 θ_N，无子集可抛弃。

第四步：修改聚类中心

$$z_1 = \begin{pmatrix} 5.17 \\ 5.33 \end{pmatrix},\quad z_2 = \begin{pmatrix} 2.00 \\ 1.50 \end{pmatrix}$$

第五步：计算 $\bar{D}_j, j = 1,2$

$$\bar{D}_1 = 2.27\text{，}\quad \bar{D}_2 = 1.87$$

第六步：计算 $\bar{D}$

$$\bar{D} = 2.11$$

第七步：该步中没有一种情况可被满足，继续执行第八步。

第八步：计算 $S_1 = \{x_2, x_6, x_7, x_8, x_9, x_{10}\}$ 和 $S_2 = \{x_1, x_3, x_4, x_5\}$ 中的标准差

$$\sigma_1 = \begin{pmatrix} 1.34 \\ 1.97 \end{pmatrix}, \sigma_2 = \begin{pmatrix} 1.87 \\ 1.12 \end{pmatrix}$$

第九步：$\sigma_{1\max} = 1.97 > \theta_S$，且 $\bar{D}_1 > \bar{D}$ 和 $N_1 > 2(\theta_N + 1)$

则将 z_1 分裂成两个新的聚类。设 $r_j = 0.5\sigma_{j\max} = 0.99$，则

$$z_1 = \begin{pmatrix} 5.17 \\ 6.32 \end{pmatrix}, z_2 = \begin{pmatrix} 5.17 \\ 4.35 \end{pmatrix}, z_3 = \begin{pmatrix} 2.00 \\ 1.50 \end{pmatrix}$$，N_c 增加 1，跳回到第二步。

第二步：新的样本集为：

$S_1 = \{x_2, x_6\}$，$S_2 = \{x_5, x_7, x_8, x_9, x_{10}\}$，$S_3 = \{x_1, x_3, x_4\}$

$N_1 = 2, N_2 = 5, N_3 = 3$

第三步：因 N_1，N_2 和 N_3 都大于 θ_N，无子集可抛弃。

第四步：修改聚类中心

$$z_1 = \begin{pmatrix} 3.50 \\ 8.00 \end{pmatrix}, z_2 = \begin{pmatrix} 5.80 \\ 3.80 \end{pmatrix}, z_3 = \begin{pmatrix} 1.00 \\ 1.00 \end{pmatrix}$$

第五步：计算 $\bar{D}_j, j = 1,2,3$

$\bar{D}_1 = 0.50$，$\bar{D}_2 = 0.95$，$\bar{D}_3 = 0.94$

第六步：计算 $\bar{D}$

$\bar{D} = 0.86$

第七步：因为最后一次迭代，跳到第十四步

第十四步：最后一次迭代，故算法结束。

最终将原样本集聚成三类

$S_1 = \{x_2, x_6\}$，$S_2 = \{x_5, x_7, x_8, x_9, x_{10}\}$，$S_3 = \{x_1, x_3, x_4\}$。