半监督学习

1. 主动学习：利用已标记样本训练模型，询问专家获取新样本加入训练

目的：更少的查询，更优的性能。

2. 半监督学习：让学习器不依赖外界交互，自动地利用未标记样本来提升学习性能。

聚类假设：同一簇的样本属于同一类别

流形假设：数据分布在一个流形结构上，邻近样本有相似的输出 → 相似样本，相似输出

纯半监督：未标记样本不是预测数据

直推学习：未标记样本是预测数据

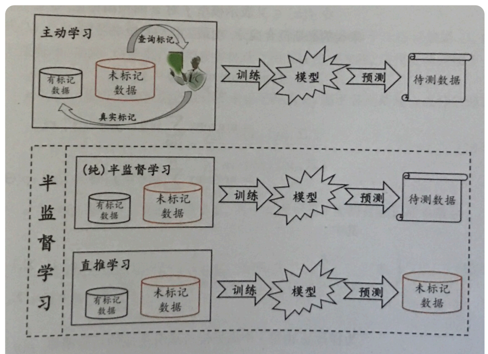


图 13.2 主动学习、(纯)半监督学习、直推学习

3. 生成式方法：基于EM算法的极大似然估计

假设：数据由同一潜在模型生成，未标记样本即模型的缺失参数。

不同模型假设会产生不同的方法。

假设样本由高斯混合模型生成：$p(x)=\sum_{i=1}^{N}\underbrace{\alpha_i}_{\text{混合系数}}\cdot p(x\mid \underbrace{\mu_i,\Sigma_i}_{\text{高斯模型参数}})$

预测样本类别：$f(x)=\underset{j\in y}{\arg\max}\ p(y=j\mid x)=\underset{j\in y}{\arg\max}\sum_{i=1}^{N}p(y=j,\Theta=i\mid x)$

$$=\underset{j\in y}{\arg\max}\sum_{i=1}^{N}p(y=j\mid\Theta=i,x)\,p(\Theta=i\mid x)$$

$p(\Theta=i\mid x)$ 表示样本 x 由 Θ_i 生成的概率。

$$p(\Theta=i\mid x)=\frac{\alpha_i\cdot p(x\mid\mu_i,\Sigma_i)}{\sum_{i=1}^{N}\alpha_i\cdot p(x\mid\mu_i,\Sigma_i)}$$

$D_l\cup D_u$ 标记与未标记的对数似然为

$$\sum_{(x_j,y_j)\in D_l}\ln\left(\sum_{i=1}^{N}\alpha_i\,p(x_j\mid\mu_i,\Sigma_i)\cdot p(y_j\mid\Theta=i,x_j)\right)+\sum_{x_j\in D_u}\ln\left(\sum_{i=1}^{N}\alpha_i\cdot p(x_j\mid\mu_i,\Sigma_i)\right)$$

① E步：计算未标记样本 x_j 属于各高斯成分的概率

$$\gamma_{ji}=\frac{\alpha_i\cdot p(x_j\mid\mu_i,\Sigma_i)}{\sum_{i=1}^{N}\alpha_i\cdot p(x_j\mid\mu_i,\Sigma_i)}$$

② M步：基于 γ_{ji} 更新模型参数。（l_i：第 i 类有标记样本数。

$$\mu_i=\frac{1}{\sum_{x_j\in D_u}\gamma_{ji}+l_i}\left(\sum_{x_j\in D_u}\gamma_{ji}x_j+\sum_{(x_j,y_j)\in D_l\wedge y_j=i}x_j\right)$$

$$\Sigma_i=\frac{1}{\sum_{x_j\in D_u}\gamma_{ji}+l_i}\left(\sum_{x_j\in D_u}\gamma_{ji}(x_j-\mu_i)(x_j-\mu_i)^T+\sum_{(x_j,y_j)\in D_l\wedge y_j=i}(x_j-\mu_i)(x_j-\mu_i)^T\right)$$

$$\alpha_i=\frac{1}{m}\left(\sum_{x_j\in D_u}\gamma_{ji}+l_i\right)$$

①，② 循环，迭代至收敛。最后根据 $f(x)$ 确定样本类别。

4. 半监督 SVM

S3VM：考虑未标记样本，将两类样本分开，并穿过低密度区域

TSVM：对未标记样本进行标记指派，寻找所有样本间隔最大化的超平面。

输入: 有标记样本集 $D_l=\{(\boldsymbol{x}_1,y_1),(\boldsymbol{x}_2,y_2),\ldots,(\boldsymbol{x}_l,y_l)\}$;
未标记样本集 $D_u=\{\boldsymbol{x}_{l+1},\boldsymbol{x}_{l+2},\ldots,\boldsymbol{x}_{l+u}\}$;
折中参数 C_l, C_u.

过程:
1: 用 D_l 训练一个 SVM_l;
2: 用 SVM_l 对 D_u 中样本进行预测, 得到 $\hat{\boldsymbol{y}}=(\hat{y}_{l+1},\hat{y}_{l+2},\ldots,\hat{y}_{l+u})$;
3: 初始化 $C_u \ll C_l$;
4: **while** $C_u < C_l$ **do**
5: 基于 $D_l, D_u, \hat{\boldsymbol{y}}, C_l, C_u$ 求解式(13.9), 得到 $(\boldsymbol{w}, b)$, $\boldsymbol{\xi}$;
6: **while** $\exists\{i,j \mid (\hat{y}_i\hat{y}_j<0)\wedge(\xi_i>0)\wedge(\xi_j>0)\wedge(\xi_i+\xi_j>2)\}$ **do**
7: $\hat{y}_i=-\hat{y}_i$;
8: $\hat{y}_j=-\hat{y}_j$;
9: 基于 $D_l, D_u, \hat{\boldsymbol{y}}, C_l, C_u$ 重新求解式(13.9), 得到 $(\boldsymbol{w}, b)$, $\boldsymbol{\xi}$
10: **end while**
11: $C_u=\min\{2C_u, C_l\}$
12: **end while**

输出: 未标记样本的预测结果: $\hat{\boldsymbol{y}}=(\hat{y}_{l+1},\hat{y}_{l+2},\ldots,\hat{y}_{l+u})$

图 13.4 TSVM 算法

形式化地说, 给定 $D_l=\{(\boldsymbol{x}_1,y_1),(\boldsymbol{x}_2,y_2),\ldots,(\boldsymbol{x}_l,y_l)\}$ 和 $D_u=\{\boldsymbol{x}_{l+1},\boldsymbol{x}_{l+2},\ldots,\boldsymbol{x}_{l+u}\}$, 其中 $y_i\in\{-1,+1\}$, $l\ll u$, $l+u=m$. TSVM 的学习目标是为 D_u 中的样本给出预测标记 $\hat{\boldsymbol{y}}=(\hat{y}_{l+1},\hat{y}_{l+2},\ldots,\hat{y}_{l+u})$, $\hat{y}_i\in\{-1,+1\}$, 使得

$$\min_{\boldsymbol{w},b,\hat{\boldsymbol{y}},\boldsymbol{\xi}} \frac{1}{2}\|\boldsymbol{w}\|_2^2 + C_l\sum_{i=1}^{l}\xi_i + C_u\sum_{i=l+1}^{m}\xi_i \tag{13.9}$$

$$\text{s.t.}\quad y_i(\boldsymbol{w}^{\mathrm{T}}\boldsymbol{x}_i+b)\geqslant 1-\xi_i,\quad i=1,2,\ldots,l,$$
$$\hat{y}_i(\boldsymbol{w}^{\mathrm{T}}\boldsymbol{x}_i+b)\geqslant 1-\xi_i,\quad i=l+1,l+2,\ldots,m,$$
$$\xi_i\geqslant 0,\quad i=1,2,\ldots,m,$$

其中, $(\boldsymbol{w},b)$ 确定了一个划分超平面; $\boldsymbol{\xi}$ 为松弛向量, $\xi_i\ (i=1,2,\ldots,l)$ 对应于有标记样本, $\xi_i\ (i=l+1,l+2,\ldots,m)$ 对应于未标记样本; C_l 与 C_u 是由用户指定的用于平衡模型复杂度、有标记样本与未标记样本重要程度的折中参数.

思路：局部搜索，先利用已标记样本学习一个 SVM，对未标记样本进行标记指派得到伪标记，代入右图，优化 SVM。

此时令 $C_u < C_l$ 强化已标记样本的重要性，找出错误的伪标记，交换标记。在这个过程中逐步增加 C_u，直至 $C_u = C_l$ 为止

① 对于类别不平衡：$C_u \to C_u^+, C_u^-$，$C_u^+ = \frac{u^-}{u^+} C_u^-$

② 对于未标样本 x_i, x_j，伪标记 $\hat{y}_i, \hat{y}_j$ 若 $\xi_i + \xi_j > 2$，交换标记。

5. 图半监督：样本为图中结点，相似度高的样本结点之间有边，标记样本结点到未标记结点的扩散。

图 $G=(V,E)$ 其中
$$\begin{cases} V=\{x_1,\cdots x_l,\ x_{l+1},\cdots,x_{l+u}\} \\ E \text{ 表示为 } (W)_{ij}=\begin{cases}\exp\left(\dfrac{-\|x_i-x_j\|_2^2}{2\sigma^2}\right), & \text{若 } i\neq j \\ 0, & \text{otherwise}\end{cases}\end{cases}$$

目标 $f: V\to \mathbb{R}$，
$$\begin{cases} y_i=\mathrm{sign}(f(x_i)) \\ y_i\in\{-1,+1\}\end{cases}$$

$$\longrightarrow E(f)=\frac{1}{2}\sum_{i=1}^{m}\sum_{j=1}^{m}(W)_{ij}(f(x_i)-f(x_j))^2$$
$$=\sum_{i=1}^{m}d_i f^2(x_i)-\sum_{i=1}^{m}\sum_{j=1}^{m}(W)_{ij}f(x_i)f(x_j)$$
$$=f^{\mathrm{T}}(D-W)f$$

$(f_l^{\mathrm{T}};f_u^{\mathrm{T}})$　对角阵 $d_i=\sum_{j=1}^{l+u}(W)_{ij}$

$$=(f_l^{\mathrm{T}}\ f_u^{\mathrm{T}})\left(\begin{bmatrix}D_{ll} & 0_{lu}\\ 0_{ul} & D_{uu}\end{bmatrix}-\begin{bmatrix}W_{ll} & W_{lu}\\ W_{ul} & W_{uu}\end{bmatrix}\right)\begin{bmatrix}f_l\\ f_u\end{bmatrix}$$
$$=f_l^{\mathrm{T}}(D_{ll}-W_{ll})\cdot f_l-2f_u^{\mathrm{T}}W_{ul}f_l+f_u^{\mathrm{T}}(D_{uu}-W_{uu})f_u$$

$\boxed{\begin{array}{l} \min[E(f)] \text{ 有 } f(x_i)=y_i\ (i=1,\cdots,l) \\ \Delta f=0, \\ \Delta=D-W,\ \text{其中 } W \text{ 可写成 } \begin{bmatrix}W_{ll} & W_{lu}\\ W_{ul} & W_{uu}\end{bmatrix} \\ D \text{ 可写成 } \begin{bmatrix}D_{ll} & 0_{lu}\\ 0_{ul} & D_{uu}\end{bmatrix}\end{array}}$

由 $\frac{\partial E(f)}{\partial f_u}=0$ 得 $f_u=(D_{uu}-W_{uu})^{-1}W_{ul}f_l$

令 $P=D^{-1}W=\begin{bmatrix}D_{ll}^{-1} & 0_{lu}\\ 0_{ul} & D_{uu}^{-1}\end{bmatrix}\begin{bmatrix}W_{ll} & W_{lu}\\ W_{ul} & W_{uu}\end{bmatrix}=\begin{bmatrix}D_{ll}^{-1}W_{ll} & D_{ll}^{-1}W_{lu}\\ \underbrace{D_{uu}^{-1}W_{ul}}_{P_{ul}} & \underbrace{D_{uu}^{-1}W_{uu}}_{P_{uu}}\end{bmatrix}$

$$f_u=(D_{uu}(I-D_{uu}^{-1}W_{uu}))^{-1}W_{ul}f_l=(I-D_{uu}^{-1}W_{uu})^{-1}D_{uu}^{-1}W_{ul}f_l$$
$$=(I-P_{uu})^{-1}P_{ul}f_l$$

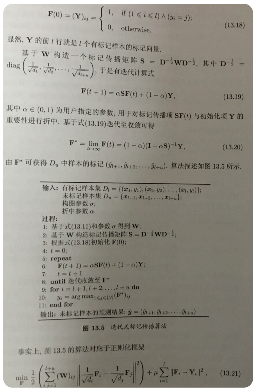
二分类

多分类：定义一个 $(l+u)\times|y|$ 的非负矩阵 $F=(F_1^T, F_2^T, \ldots, F_{l+u}^T)^T$

$F_i=((F)_{i1}, (F)_{i2}, \ldots, (F)_{i|y|})$ 为示例 x_i 的标记向量：初始化条件

$$F(0)=(Y)_{ij}=\begin{cases}1, & \text{if } (1\le i\le l)\wedge(y_i=j).\\ 0, & \text{otherwise}\end{cases}$$

另有 $y_i=\arg\max_{1\le j\in|y|}(F)_{ij}$.



$$\mathbf{F}(0)=(\mathbf{Y})_{ij}=\begin{cases}1, & \text{if } (1\leqslant i\leqslant l)\wedge(y_i=j);\\ 0, & \text{otherwise.}\end{cases} \quad (13.18)$$

显然, $\mathbf{Y}$ 的前 l 行就是 l 个有标记样本的标记向量.

基于 $\mathbf{W}$ 构造一个标记传播矩阵 $\mathbf{S}=\mathbf{D}^{-\frac{1}{2}}\mathbf{W}\mathbf{D}^{-\frac{1}{2}}$, 其中 $\mathbf{D}^{-\frac{1}{2}}=\mathrm{diag}\left(\frac{1}{\sqrt{d_1}},\frac{1}{\sqrt{d_2}},\ldots,\frac{1}{\sqrt{d_{l+u}}}\right)$, 于是有迭代计算式

$$\mathbf{F}(t+1)=\alpha\mathbf{S}\mathbf{F}(t)+(1-\alpha)\mathbf{Y}, \quad (13.19)$$

其中 $\alpha\in(0,1)$ 为用户指定的参数, 用于对标记传播项 $\mathbf{S}\mathbf{F}(t)$ 与初始化项 $\mathbf{Y}$ 的重要性进行折中. 基于式(13.19)迭代至收敛可得

$$\mathbf{F}^*=\lim_{t\mapsto\infty}\mathbf{F}(t)=(1-\alpha)(\mathbf{I}-\alpha\mathbf{S})^{-1}\mathbf{Y}, \quad (13.20)$$

由 $\mathbf{F}^*$ 可获得 D_u 中样本的标记 $(\hat{y}_{l+1},\hat{y}_{l+2},\ldots,\hat{y}_{l+u})$. 算法描述如图 13.5 所示.

输入: 有标记样本集 $D_l=\{(\boldsymbol{x}_1,y_1),(\boldsymbol{x}_2,y_2),\ldots,(\boldsymbol{x}_l,y_l)\}$;
未标记样本集 $D_u=\{\boldsymbol{x}_{l+1},\boldsymbol{x}_{l+2},\ldots,\boldsymbol{x}_{l+u}\}$;
构图参数 σ;
折中参数 α.

过程:
基于式(13.11)和参数 σ 得到 $\mathbf{W}$;
基于 $\mathbf{W}$ 构造标记传播矩阵 $\mathbf{S}=\mathbf{D}^{-\frac{1}{2}}\mathbf{W}\mathbf{D}^{-\frac{1}{2}}$;
根据式(13.18)初始化 $\mathbf{F}(0)$;
$t=0$;
repeat
 $\mathbf{F}(t+1)=\alpha\mathbf{S}\mathbf{F}(t)+(1-\alpha)\mathbf{Y}$;
 $t=t+1$
until 迭代收敛至 $\mathbf{F}^*$
for $i=l+1,l+2,\ldots,l+u$ **do**
 $y_i=\arg\max_{1\leqslant j\leqslant|\mathcal{Y}|}(\mathbf{F}^*)_{ij}$
end for

输出: 未标记样本的预测结果: $\hat{\boldsymbol{y}}=(\hat{y}_{l+1},\hat{y}_{l+2},\ldots,\hat{y}_{l+u})$

图 13.5 迭代式标记传播算法

事实上, 图 13.5 的算法对应于正则化框架

$$\min_{\mathbf{F}}\frac{1}{2}\left(\sum_{i,j=1}^{l+u}(\mathbf{W})_{ij}\left\|\frac{1}{\sqrt{d_i}}\mathbf{F}_i-\frac{1}{\sqrt{d_j}}\mathbf{F}_j\right\|^2\right)+\mu\sum_{i=1}^{l}\|\mathbf{F}_i-\mathbf{Y}_i\|^2, \quad (13.21)$$

当 $\mu=\frac{1-\alpha}{\alpha}$ 时 13.21 最优解为 F^*

其中第一项：相近样本有相似属性

第二项：有标记样本的准确性.

特点：存储开销大，接收新样本要额外调整.

6. 基于分歧的方法：使用多个学习器的分歧、协同训练

利用多视图的互补性.

对于单视图：不同算法，不同数据分布，不同参数.

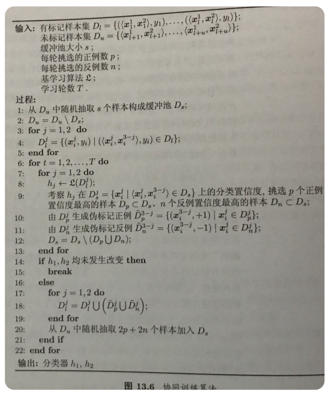
充分：每个视图重产生最优学习器

条件独立：① 每个视图基于有标记样本训练分类器

② 挑选未标记、产生伪标记，提供给另一分类器作新增标记.

置信度：朴素贝叶斯：后验概率

SVM：间隔大小



输入: 有标记样本集 $D_l=\{(\langle\boldsymbol{x}_1^1,\boldsymbol{x}_1^2\rangle,y_1),\ldots,(\langle\boldsymbol{x}_l^1,\boldsymbol{x}_l^2\rangle,y_l)\}$;
未标记样本集 $D_u=\{\langle\boldsymbol{x}_{l+1}^1,\boldsymbol{x}_{l+1}^2\rangle,\ldots,\langle\boldsymbol{x}_{l+u}^1,\boldsymbol{x}_{l+u}^2\rangle\}$;
缓冲池大小 s;
每轮挑选的正例数 p;
每轮挑选的反例数 n;
基学习算法 $\mathfrak{L}$;
学习轮数 T.

过程:
从 D_u 中随机抽取 s 个样本构成缓冲池 D_s;
$D_u=D_u\setminus D_s$;
for $j=1,2$ **do**
 $D_l^j=\{(\boldsymbol{x}_i^j,y_i)\mid(\langle\boldsymbol{x}_i^j,\boldsymbol{x}_i^{3-j}\rangle,y_i)\in D_l\}$;
end for
for $t=1,2,\ldots,T$ **do**
 for $j=1,2$ **do**
  $h_j\leftarrow\mathfrak{L}(D_l^j)$;
  考察 h_j 在 $D_s^j=\{\boldsymbol{x}_i^j\mid\langle\boldsymbol{x}_i^j,\boldsymbol{x}_i^{3-j}\rangle\in D_s\}$ 上的分类置信度, 挑选 p 个正例置信度最高的样本 $D_p\subset D_s$、n 个反例置信度最高的样本 $D_n\subset D_s$;
  由 D_p^j 生成伪标记正例 $\tilde{D}_p^{3-j}=\{(\boldsymbol{x}_i^{3-j},+1)\mid\boldsymbol{x}_i^j\in D_p^j\}$;
  由 D_n^j 生成伪标记反例 $\tilde{D}_n^{3-j}=\{(\boldsymbol{x}_i^{3-j},-1)\mid\boldsymbol{x}_i^j\in D_n^j\}$;
  $D_s=D_s\setminus(D_p\bigcup D_n)$;
 end for
 if h_1,h_2 均未发生改变 **then**
  break
 else
  for $j=1,2$ **do**
   $D_l^j=D_l^j\bigcup\left(\tilde{D}_p^j\bigcup\tilde{D}_n^j\right)$;
  end for
  从 D_u 中随机抽取 $2p+2n$ 个样本加入 D_s
 end if
end for

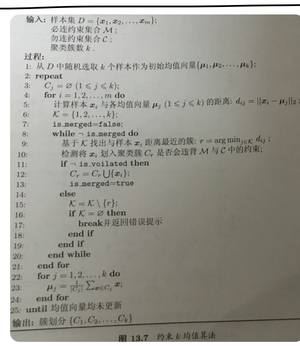
输出: 分类器 h_1,h_2

图 13.6 协同训练算法

7. 半监督聚类

第一类监督信息：聚类过程中保持 M 与 C 的约束

第二类监督信息：少量有标记样本，作种子，初始化聚类中心



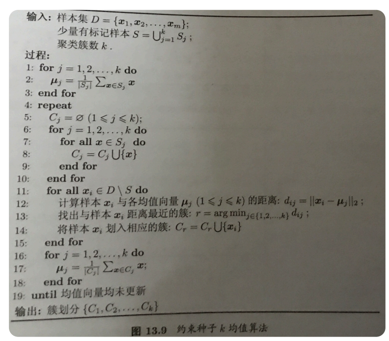
输入: 样本集 $D=\{\boldsymbol{x}_1,\boldsymbol{x}_2,\ldots,\boldsymbol{x}_m\}$;
必连约束集合 $\mathcal{M}$;
勿连约束集合 $\mathcal{C}$;
聚类簇数 k.

过程:
从 D 中随机选取 k 个样本作为初始均值向量$\{\boldsymbol{\mu}_1,\boldsymbol{\mu}_2,\ldots,\boldsymbol{\mu}_k\}$;
repeat
 $C_j=\varnothing\ (1\leqslant j\leqslant k)$;
 for $i=1,2,\ldots,m$ **do**
  计算样本 $\boldsymbol{x}_i$ 与各均值向量 $\boldsymbol{\mu}_j\ (1\leqslant j\leqslant k)$ 的距离: $d_{ij}=\|\boldsymbol{x}_i-\boldsymbol{\mu}_j\|_2$;
  $\mathcal{K}=\{1,2,\ldots,k\}$;
  is_merged=false;
  while $\neg$ is_merged **do**
   基于 $\mathcal{K}$ 找出与样本 $\boldsymbol{x}_i$ 距离最近的簇: $r=\arg\min_{j\in\mathcal{K}}d_{ij}$;
   检测将 $\boldsymbol{x}_i$ 划入聚类簇 C_r 是否会违背 $\mathcal{M}$ 与 $\mathcal{C}$ 中的约束;
   if $\neg$ is_voilated **then**
    $C_r=C_r\bigcup\{\boldsymbol{x}_i\}$;
    is_merged=true
   else
    $\mathcal{K}=\mathcal{K}\setminus\{r\}$;
    if $\mathcal{K}=\varnothing$ **then**
     break并返回错误提示
    end if
   end if
  end while
 end for
 for $j=1,2,\ldots,k$ **do**
  $\boldsymbol{\mu}_j=\frac{1}{|C_j|}\sum_{\boldsymbol{x}\in C_j}\boldsymbol{x}$;
 end for
until 均值向量均未更新

输出: 簇划分 $\{C_1,C_2,\ldots,C_k\}$

图 13.7 约束 k 均值算法

①



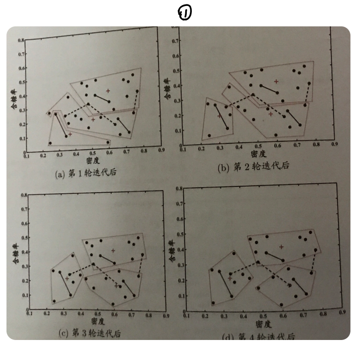
输入: 样本集 $D=\{\boldsymbol{x}_1,\boldsymbol{x}_2,\ldots,\boldsymbol{x}_m\}$;
少量有标记样本 $S=\bigcup_{j=1}^k S_j$;
聚类簇数 k.

过程:
for $j=1,2,\ldots,k$ **do**
 $\boldsymbol{\mu}_j=\frac{1}{|S_j|}\sum_{\boldsymbol{x}\in S_j}\boldsymbol{x}$
end for
repeat
 $C_j=\varnothing\ (1\leqslant j\leqslant k)$;
 for $j=1,2,\ldots,k$ **do**
  for all $\boldsymbol{x}\in S_j$ **do**
   $C_j=C_j\bigcup\{\boldsymbol{x}\}$
  end for
 end for
 for all $\boldsymbol{x}_i\in D\setminus S$ **do**
  计算样本 $\boldsymbol{x}_i$ 与各均值向量 $\boldsymbol{\mu}_j\ (1\leqslant j\leqslant k)$ 的距离: $d_{ij}=\|\boldsymbol{x}_i-\boldsymbol{\mu}_j\|_2$;
  找出与样本 $\boldsymbol{x}_i$ 距离最近的簇: $r=\arg\min_{j\in\{1,2,\ldots,k\}}d_{ij}$;
  将样本 $\boldsymbol{x}_i$ 划入相应的簇: $C_r=C_r\bigcup\{\boldsymbol{x}_i\}$
 end for
 for $j=1,2,\ldots,k$ **do**
  $\boldsymbol{\mu}_j=\frac{1}{|C_j|}\sum_{\boldsymbol{x}\in C_j}\boldsymbol{x}$;
 end for
until 均值向量均未更新

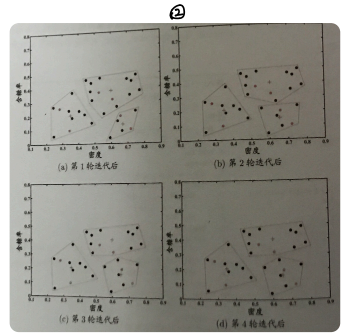
输出: 簇划分 $\{C_1,C_2,\ldots,C_k\}$

图 13.9 约束种子 k 均值算法

②



(a) 第 1 轮迭代后 (b) 第 2 轮迭代后 (c) 第 3 轮迭代后 (d) 第 4 轮迭代后



(a) 第 1 轮迭代后 (b) 第 2 轮迭代后 (c) 第 3 轮迭代后 (d) 第 4 轮迭代后