中国科学技术大学
University of Science and Technology of China

第三次和第八次作业答案

第三次作业

6.5 分别用带有前向检验、MRV和最少约束值启发式的回溯算法手工求解**图6.2**中的密码算数问题

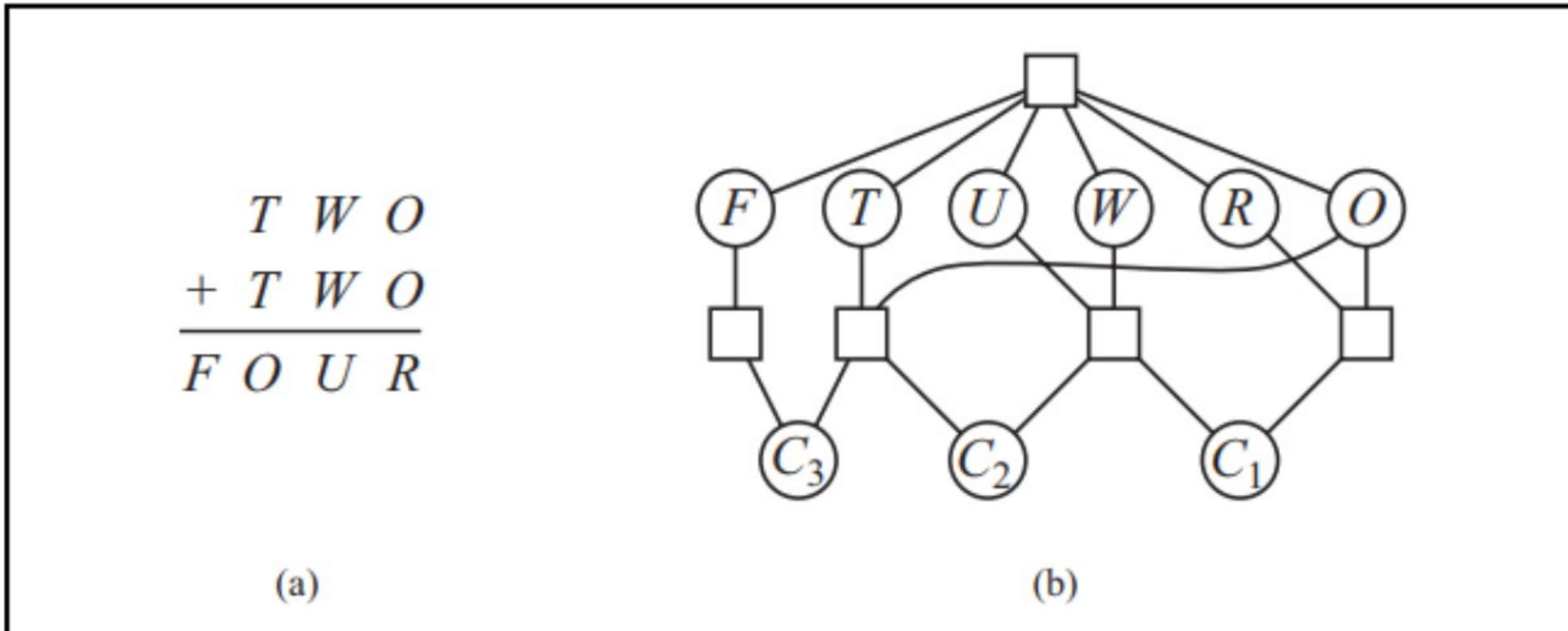


Figure 6.2 (a) A cryptarithmetic problem. Each letter stands for a distinct digit; the aim is to find a substitution of digits for letters such that the resulting sum is arithmetically correct, with the added restriction that no leading zeroes are allowed. (b) The constraint hypergraph for the cryptarithmetic problem, showing the *Alldiff* constraint (square box at the top) as well as the column addition constraints (four square boxes in the middle). The variables C_1, C_2, and C_3 represent the carry digits for the three columns.

第三次作业

约束条件：

$O+O=R+10C_1$

$C_1+W+W=U+10C_2$

$C_2+T+T=O+10C_3$

$C_3=F$

$T\neq 0$

$F\neq 0$

F,T,U,W,R,O 数字各不相同

变量：

$F,T,U,W,R,O\in\{0,1,2..,9\}$

$C_1,C_2,C_3\in\{0,1\}$

求解：

	C_3	C_2	C_1	F	T	W	O	U	R
$Init$	{1}	{0,1}	{0,1}	{1}	{1,..,9}	{0,..,9}	{0,..,9}	{0,..,9}	{0,..,9}
$C_3=1$	1	{0,1}	{0,1}	{1}	{5,...,9}	{0,..,9}	{0,..,9}	{0,..,9}	{0,..,9}
$F=1$	1	{0,1}	{0,1}	1	{5,...,9}	{0,..,9}	{0,..,9}	{0,..,9}	{0,..,9}
$C_2=0$	1	0	{0,1}	1	{5,...,9}	{0,..,4}	{0,2,4,6,8}	{0,..,9}	{0,..,9}
$C_1=0$	1	0	0	1	{5,6,7}	{0,2,3,4}	{0,2,4}	{0,4,6,8}	{0,4,8}
$O=4$	1	0	0	1	{7}	{0,2,3,4}	4	{0,4,6,8}	{8}
$R=8$	1	0	0	1	{7}	{0,3}	4	{0,6}	8
$T=7$	1	0	0	1	7	{0,3}	4	{0,6}	8
$W=0$	1	0	0	1	7	3	4	{6}	8
U=6	1	0	0	1	7	3	4	6	8

$C_3=\{0,1\}\Longrightarrow C_3=1$(如果$C_3=0$，那么$F=0$，导致出错)$\Rightarrow F=1$

$C_2=\{0,1\}$，选择$C_2=0$，$C_1=\{0,1\}$，选择$C_1=0$

O和R是偶数($T+T=O+10$，$O+O=R$)，因此$O=\{2,4\}$，选择$O=4$，那么$R=8,T=7$，U是偶数，因此$U=\{2,6\}$，而$U=2$时，$W=1$，不符，因此$U=6$，$W=3$，得到一组解：

TWO=734,FOUR=1468

第三次作业

6.11 用**AC-3**算法说明弧相容对**图6.1**中问题能够检测出部分赋值，$WA = green, V = red$ 的不相容。

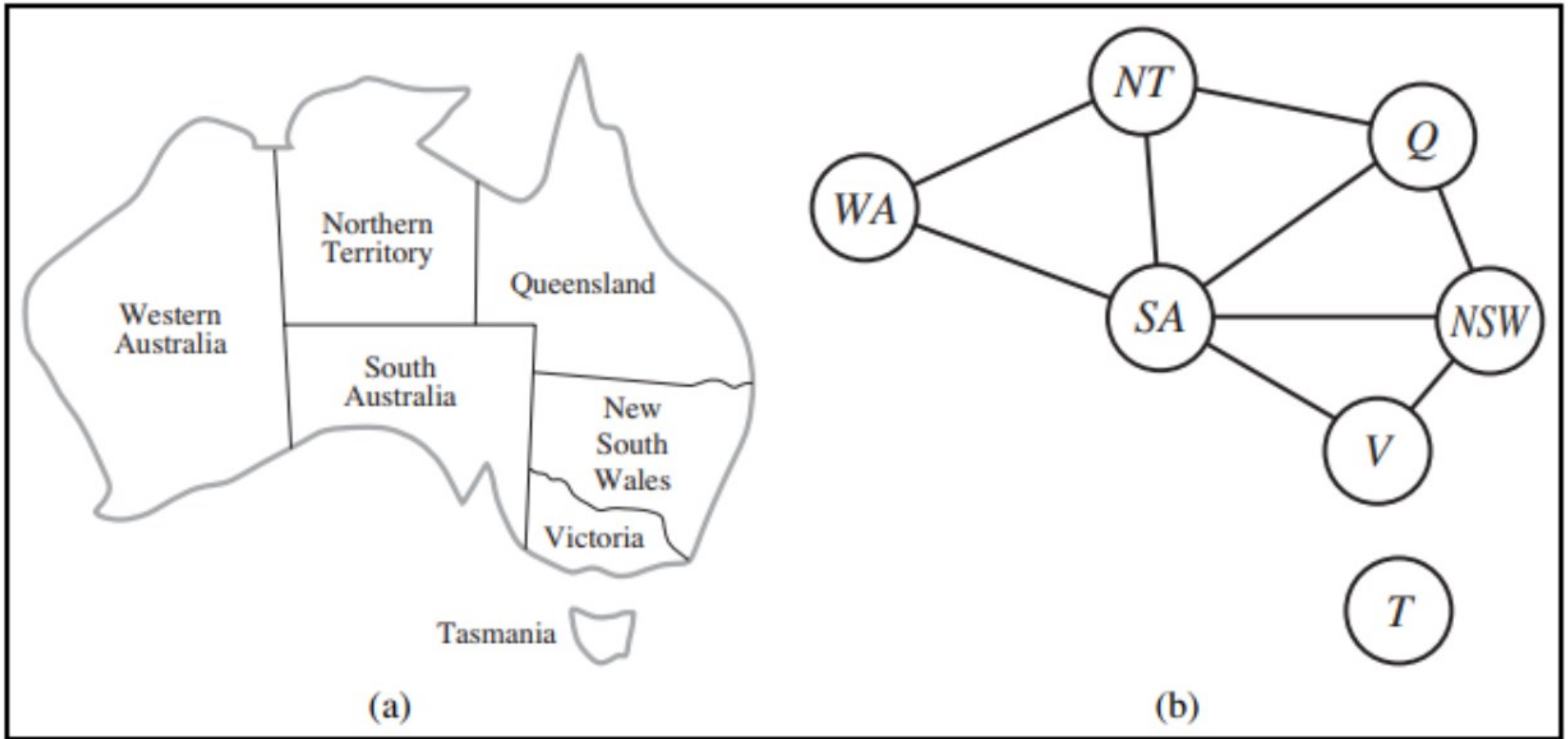
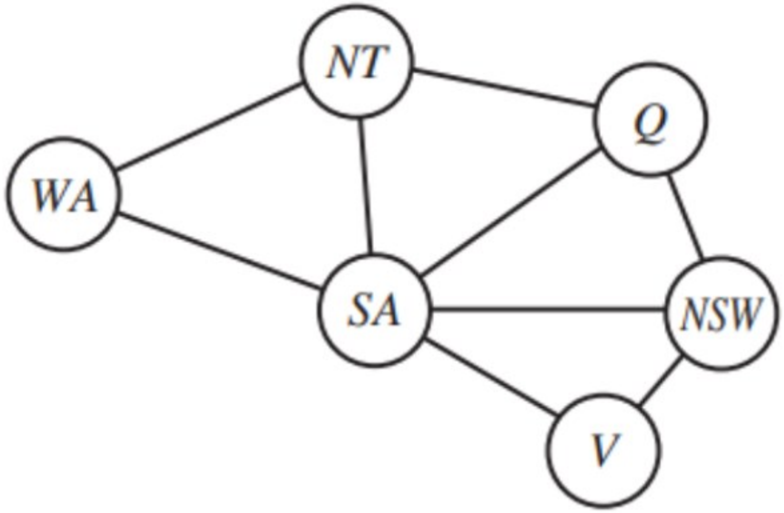


Figure 6.1 (a) The principal states and territories of Australia. Coloring this map can be viewed as a constraint satisfaction problem (CSP). The goal is to assign colors to each region so that no neighboring regions have the same color. (b) The map-coloring problem represented as a constraint graph.

第三次作业

	SA	WA	NT	Q	NSW	V
$Init$	{RGB}	{G}	{RGB}	{RGB}	{RGB}	{R}
{SA,WA}	{RB}	{G}	{RGB}	{RGB}	{RGB}	{R}
{SA,V}	{B}	{G}	{RGB}	{RGB}	{RGB}	{R}
{NT,SA},{Q,SA},{NSW,SA}	{B}	{G}	{RG}	{RG}	{RG}	{R}
{NT,WA}	{B}	{G}	{R}	{RG}	{RG}	{R}
{NSW,V}	{B}	{G}	{R}	{RG}	{G}	{R}
{Q,NT}	{B}	{G}	{R}	{G}	{G}	{R}
{Q,NSW}	{B}	{G}	{R}	{}	{G}	{R}



第三次作业

6.12 用**AC-3**算法求解树结构**CSP**在最坏情况下的复杂度是多少?

- 采用逆拓扑序检验，保证每条弧只需要检验一次
- 假设每个顶点的值域最多有 d 个取值，则每条弧检验复杂度为 $O(d^2)$
- 假设有 n 个顶点，则用**AC-3**算法在最坏情况下的复杂度是 $O(nd^2)$

第八次作业

- 试证明对于不含冲突数据集（即特征向量完全相同但标记不同）的训练集，必存在与训练集一致（即训练误差为 0）的决策树。

反证法。假设不存在与训练集一致的决策树，那么训练集训练得到的决策树 上至少有一个节点存在无法划分的多个数据（即，若节点上没有冲突数据的 话，则必然能够将数据划分开）。这与前提“不含冲突数据”相矛盾，因此必有与训练集一致的决策树

第八次作业

- 最小二乘学习方法在求解 $\min_{\mathbf{w}}(\mathbf{X}\mathbf{w}-\mathbf{y})^2$ 问题后得到闭式解 $\mathbf{w}^* = (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{y}$（为简化问题，我们忽略偏差项 b）。如果我们知道数据中部分特征有较大的误差，在不修改损失函数的情况下，引入规范化项 $\lambda\mathbf{w}^T\mathbf{D}\mathbf{w}$，其中 $\mathbf{D}$ 为对角矩阵，由我们取值。相应的最小二乘分类学习问题转换为以下形式的优化问题：

$$\min_{\mathbf{w}}(\mathbf{X}\mathbf{w}-\mathbf{y})^2 + \lambda\mathbf{w}^T\mathbf{D}\mathbf{w}$$

(1) 请说明选择规范化项 $\mathbf{w}^T\mathbf{D}\mathbf{w}$ 而非 L2 规范化项 $\mathbf{w}^T\mathbf{w}$ 的理由是什么。$\mathbf{D}$ 的对角线元素 D_{ii} 有何意义，它的取值越大意味着什么？

(2) 请对以上问题进行求解。

第八次作业

(1) 使用ww^T进行规范化时所有特征权重相同。由于部分特征的误差较大，需要降低其权重。使用wDw^T作为规范化项可以通过改变D，对w的不同分量的给予不同的规范化限制,从而影响其权重。

D的对角线元素D_{ii}是w第i分量的权重惩罚系数。

D_{ii}取值越大表示规范化项中对w_i分量的限制越大,对应特征的权重受到的惩罚越大。

(2) 对$(Xw-y)^2+\lambda w^T Dw$，求导后取极值点:

解得$w^*=(X^TX+\lambda D)^{-1}X^Ty$

第八次作业

- 假设有 n 个数据点 $\mathbf{x}_1, \ldots, \mathbf{x}_n$ 以及一个映射 φ: $\mathbf{x} \to \varphi(\mathbf{x})$，以此定义核函数 $K(\mathbf{x}, \mathbf{x}') = \varphi(\mathbf{x}) \cdot \varphi(\mathbf{x}')$。试证明由该核函数所决定的核矩阵 K： $K_{i,j} = K(\mathbf{x}_i, \mathbf{x}_j)$ 有以下性质:
  (1) K 是一个对称矩阵；
  (2) K 是一个半正定矩阵，即 $\forall \mathbf{z} \in \mathbf{R}^n$，$\mathbf{z}^T K \mathbf{z} \geq 0$。

(1) $K_{i,j} = K(x_i, x_j) = \varphi(x_i) \cdot \varphi(x_j) = \varphi(x_j) \cdot \varphi(x_i) = K_{j,i}$

(2) $K = \varphi^T(x)\varphi(x)$

$$\forall z \in R^n,\ z^T K z = z^T \varphi^T \varphi z = (z\varphi)^T (z\varphi) \geq 0$$

第八次作业

- 已知正例点 $x_1 = (1, 2)^T$, $x_2 = (2, 3)^T$, $x_3 = (3, 3)^T$，负例点 $x_4 = (2, 1)^T$, $x_5 = (3, 2)^T$，试求 Hard Margin SVM 的最大间隔分离超平面和分类决策函数，并在图上画出分离超平面、间隔边界及支持向量。

第八次作业

假设超平面 $w^T x + b = 0$

$$\begin{cases} \min \frac{1}{2}\|w\|^2 \\ s.t.\ \ y_i(w^T x_i + b) \geq 1 \end{cases}$$

$x_1 = \binom{1}{2}$ $x_2 = \binom{2}{3}$ $x_3 = \binom{3}{3}$

$x_4 = \binom{2}{1}$ $x_5 = \binom{3}{2}$ $y = (1,1,1,-1,-1)$

拉格朗日对偶

$$\max_{\alpha} \sum_{i=1}^{n} \alpha_i - \frac{1}{2} \sum_{i,j=1}^{n} \alpha_i \alpha_j y_i y_j ({x_i}^T x_j)$$

subject to $\alpha_i \geq 0$, $\sum_{i=1}^{n} \alpha_i y_i = 0$

$\alpha_1 + \alpha_2 + \alpha_3 - \alpha_4 - \alpha_5 = 0$ 代入目标函数，有

$$\max \ 2(\alpha_1 + \alpha_2 + \alpha_3) - \frac{1}{2}(4{\alpha_1}^2 + 2\,{\alpha_2}^2 + {\alpha_3}^2 + 2{\alpha_4}^2 + 4\alpha_1\alpha_2 + 6\alpha_1\alpha_4 + 2\alpha_2\alpha_3 + 2\alpha_3\alpha_4)$$

第八次作业

首先，求解上式各偏导/梯度 = 0

$$\alpha_1 = -0.4 < 0,\ \ \alpha_2 = 1.2,\ \ \alpha_3 = \alpha_4 = 0$$

不满足约束条件，故最大值一定在可行域的边界处取到；又从上述目标函数可以直观地看出最大值处 α_4 必为0 （因为含有α_4的项为 $-\alpha_4{}^2 - 3\alpha_1\alpha_4 - \alpha_3\alpha_4$ ）
因此上式简化为

$$\max\ \ 2(\alpha_1+\alpha_2+\alpha_3) - \frac{1}{2}(4\alpha_1{}^2 + 2\,\alpha_2{}^2 + \alpha_3{}^2 + 4\alpha_1\alpha_2 + 2\alpha_2\alpha_3)$$

求解上式各偏导/梯度=0，无解；故α_1 , α_2 α_3至少有一个为0，即边界值

α_1 =0时， 求解偏导为0，可得 α_2 =0， α_3 =2，f=2

α_2=0时， 求解偏导为0，可得 α_1 =0.5， α_3 =2，f=2.5

α_3 =0时， 求解偏导为0，可得 α_1 =0， $\alpha_2 = 1$，f=1

第八次作业

$$\alpha_1 = \frac{1}{2} \quad \alpha_2 = 0 \quad \alpha_3 = 2 \quad \alpha_4 = 0 \quad \alpha_5 = \frac{5}{2}$$

$$w = \sum \alpha_i y_i x_i = \begin{pmatrix} -1 \\ 2 \end{pmatrix} \quad b = -2$$

决策函数：f(x)=sign(-$x_1 + 2x_2 - 2$)

第八次作业

- 计算 $\frac{\partial}{\partial w_j} L_{CE}(\mathbf{w}, b)$，其中

$$L_{CE}(\mathbf{w}, b) = -[y \log \sigma(\mathbf{w} \cdot \mathbf{x} + b) + (1 - y) \log(1 - \sigma(\mathbf{w} \cdot \mathbf{x} + b))]$$

为 Logistic Regression 的 Loss Function。

  - 已知

$$\begin{aligned} \frac{\partial}{\partial z}\sigma(z) = \frac{\partial}{\partial z}\frac{1}{1+e^{-z}} &= -\left(\frac{1}{1+e^{-z}}\right)^2 \times (-e^{-z}) \\ &= \sigma^2(z)\left(\frac{1-\sigma(z)}{\sigma(z)}\right) = \sigma(z)(1-\sigma(z)) \end{aligned}$$

第八次作业

$$L_{CE}(\mathbf{w}, b) = -[y \log \sigma(\mathbf{w} \cdot \mathbf{x} + b) + (1 - y) \log(1 - \sigma(\mathbf{w} \cdot \mathbf{x} + b))]$$

$$\begin{aligned}
\frac{\partial}{\partial \boldsymbol{w}_j} L_{CE}(\boldsymbol{w}, b) &= -\left[y \frac{\partial}{\partial \boldsymbol{w}_j} \log \sigma(\boldsymbol{w} \cdot \boldsymbol{x} + b) + (1 - y) \frac{\partial}{\partial \boldsymbol{w}_j} \log\big(1 - \sigma(\boldsymbol{w} \cdot \boldsymbol{x} + b)\big)\right] \\
&= -\left[y \frac{1}{\sigma(\boldsymbol{w} \cdot \boldsymbol{x} + b)} \frac{\partial}{\partial \boldsymbol{w}_j} \sigma(\boldsymbol{w} \cdot \boldsymbol{x} + b) + (1 - y) \frac{-1}{1 - \sigma(\boldsymbol{w} \cdot \boldsymbol{x} + b)} \frac{\partial}{\partial \boldsymbol{w}_j} \sigma(\boldsymbol{w} \cdot \boldsymbol{x} + b)\right] \\
&= \left[\frac{-y}{\sigma(\boldsymbol{w} \cdot \boldsymbol{x} + b)} + \frac{1 - y}{1 - \sigma(\boldsymbol{w} \cdot \boldsymbol{x} + b)}\right] \frac{\partial}{\partial \boldsymbol{w}_j} \sigma(\boldsymbol{w} \cdot \boldsymbol{x} + b) \\
&= \left[\frac{-y + y\sigma(\boldsymbol{w} \cdot \boldsymbol{x} + b) + \sigma(\boldsymbol{w} \cdot \boldsymbol{x} + b) - y\sigma(\boldsymbol{w} \cdot \boldsymbol{x} + b)}{\sigma(\boldsymbol{w} \cdot \boldsymbol{x} + b)\big(1 - \sigma(\boldsymbol{w} \cdot \boldsymbol{x} + b)\big)}\right] \sigma(\boldsymbol{w} \cdot \boldsymbol{x} + b)\big(1 - \sigma(\boldsymbol{w} \cdot \boldsymbol{x} + b)\big) \boldsymbol{x}_j \\
&= (\sigma(\boldsymbol{w} \cdot \boldsymbol{x} + b) - y) \boldsymbol{x}_j
\end{aligned}$$

$$\frac{\partial}{\partial \boldsymbol{w}_j} L_{CE}(\boldsymbol{w}, b) = (\sigma(\boldsymbol{w} \cdot \boldsymbol{x} + b) - y) \boldsymbol{x}_j$$

- K-means 算法是否一定会收敛？如果是，给出证明过程；如果不是，给出说明。

K-means 算法一定会收敛
考虑到每次迭代都使得目标函数值变小，而目标函数至少有下界0，因此当迭代次数趋于无穷时必定会收敛

6. K- means 算法会收敛，

目标函数 $J = \sum_{i=1}^{N}\sum_{k=1}^{K} r_{ik}\cdot\|x_i - c_k\|^2$，

其中 $r_{ik} = \begin{cases}1, & i点在聚类k中 \\ 0, & i点不在聚类k中\end{cases}$，

c_k 为聚类 k 的中心点，

对于K-means两种过程：

① 重新选择中心点，

由 $\nabla_{c_k} J = -\sum_{i=1}^{N}\sum_{k=1}^{K} r_{ik}\cdot 2(x_i - c_k) = 0$ 得 $c_k^* = \frac{\sum_{i=1}^{N}\sum_{k=1}^{K} r_{ik}x_i}{\sum_{i=1}^{N} r_{ik}}$，

即新的中心点正好是 J 的驻点，这一步会使 J 非增，

② 重新选择所在聚类，

$r_{ik}^* = \begin{cases}1, & k = \arg\min_k \|x_i - c_k\|^2 \\ 0, & otherwise\end{cases}$

$\sum_{k=1}^{K} r_{ik}\|x_i - \mu_k\|^2 \geq \sum_{k=1}^{K} r_{ik}^*\|x_i - \mu_k\|$.

所以这一步同样非增，

又由于 $J \geq 0$，所以在进行有限步后会收敛.