题目

假设数据由混合专家(mixture of experts)模型生成，即数据是基于k个成分混合的概率密度生成：$p(x|\theta)=\sum\limits_{i=1}^{k}a_i\cdot p(x|\theta_i)$，其中$\theta=\{\theta_1,\theta_2,\cdots,\theta_k\}$是模型参数，$p(x|\theta_i)$是第$i$个混合成分的概率密度，混合系数$\alpha_i\geq 0$，$\sum\limits_{i=1}^{k}\alpha_i=1$。假设每个混合成分对应一种类别，但每个类别可能包含多个混合成分。试推导出生成式半监督学习算法。

解答

答：首先需要假定：

> 1. 数据集 X 包括 M 个样本：$X=\{x_j\},j=1,\cdots,M$其中 l 个标记样本，u 个未标记样本：$M=l+u$
> 2. 样本里共包括 $|C|$ 个类别：$y_j\in C$
> 3. 混合模型含有N个混合成分，样本X_j可能的混合成分由m_j表示：$\{m_j=i\},i=1,\cdots,N$若θ_i表示对应混合成分的模型参数，则对应模型可表示为：$f(x_j|\theta_i)=p(x_j|m_j=i,\theta_i)=p(x_j|\theta_i)$

最大似然估计

针对给定标记样本集$D_l=\{(x_1,y_1),(x_2,y_2),\cdots,(x_l,y_l)\}$和未标定样本集$D_u=\{x_{l+1},x_{l+2},\cdots,x_u\}$。用极大似然法来估计高斯混合模型的参数$\{(\alpha_i,\mu_i,\Sigma_i)|1\leq i\leq N\}$，$D_l\cup D_u$的对数似然是：

$$\begin{aligned}LL(D_l\cup D_u)&=\sum_{(xi,yj)\in D_l}\ln p(x_j,y_j\mid\theta)+\sum_{x_i\in D_u}\ln p(x_j\mid\theta)\\&=\sum_{(\mathbf{x}_i,c_j)\in D_l}\ln\sum_{i=1}^{N}\alpha_i p(c_j\mid\mathbf{x}_j,m_j=i,\theta_i)p(\mathbf{x}_j\mid m_j=i,\theta_i)+\sum_{\mathbf{x}_i\in D_u}\ln\sum_{i=1}^{N}\alpha_i p(\mathbf{x}_j\mid m_j=i,\theta_i)\\&=\sum_{(\mathbf{x}_i,c_j)\in D_l}\ln\sum_{i=1}^{N}\alpha_i p(c_j\mid\mathbf{x}_j,m_j=i,\theta_i)f(\mathbf{x}_j\mid\theta_i)+\sum_{\mathbf{x}_i\in D_u}\ln\sum_{i=1}^{N}\alpha_i f(\mathbf{x}_j\mid\theta_i)\end{aligned}\quad(1)$$

接下来介绍一下题目中所说的 **每个类别可包含多个混合成分**的混合模型的具体表示：

> 首先，我们知道：
>
> $$p(m_j=i\mid\mathbf{x}_j)=\frac{\alpha_i\cdot p(\mathbf{x}_j\mid\theta_i)}{\sum_{i=1}^{N}\alpha_i\cdot p(\mathbf{x}_j\mid\theta_i)}\quad(2)$$
>
> 根据(D. J. Miller and H. s. Uyar, 1996)的观点，主要有两种混合方法：
>
> **划分混合模型(The “Partitioned" Mixture Model, PM):**
>
> 混合组分与各个类别具有硬划分的关系，即 $M_i\in C_k$，其中 M_i 代表混合组分 i，也就是说各个类别是由特定的混合组分组合而成，C_k 代表类别 k 具有的混合组分形成的集合，混合模型后验概率为：
>
> $$p(c_j=k\mid\mathbf{x}_j)=\frac{\sum_{i=1\wedge M_i\in C_k}^{N}\alpha_i\cdot p(\mathbf{x}_j\mid\theta_i)}{\sum_{i=1}^{N}\alpha_i\cdot p(\mathbf{x}_j\mid\theta_i)}\quad(3)$$
>
> **广义混合模型(The Generalized Mixture Model, GM):**
>
> 每个混合组分 $i\in\{1,\ldots,K\}$ 都有可能是形成某个类别 k 的一个混合成分，定义：
>
> $$p(c_j\mid m_j,\mathbf{x}_j)=p(c_j\mid m_j)=\beta_{c_j|m_j}\quad(4)$$
>
> 其中第二项成立是因为 $\beta_{c_j|m_j}$ 与具体的 $\mathbf{x}_j$ 取值无关。在此基础上可知，混合模型后验概率为:
>
> $$p(c_j\mid\mathbf{x}_j)=\frac{\sum_{i=1}^{N}(\alpha_i\cdot p(\mathbf{x}_j\mid\theta_i))\beta_{c_j|i}}{\sum_{i=1}^{N}\alpha_i\cdot p(\mathbf{x}_j\mid\theta_i)}\quad(5)$$
>
> 显然，令 GM中真正属于 c_j 的混合成分 i 均为 $\beta_{c_j|i}=1$，其他 $\beta_{c_{j|i}}=0$，则此时广义混合模型退化为 PM $_e$

在这里，我们采用 GM，采用高斯分布作为混合成分，来推导 `EM` 算法的更新参数。

显然，此时：

$$f(\mathbf{x}_j\mid\theta_i)=p(\mathbf{x}_j\mid\theta_i)=p(\mathbf{x}_j\mid\mu_i,\Sigma_i)\quad(*)$$

则 (1) 变为:

$$LL\left(D_l \cup D_u\right)=\sum_{\left(\mathbf{x}_i, c_j\right) \in D_l} \ln \sum_{i=1}^N \alpha_i p\left(c_j \mid \mathbf{x}_j, m_j=i, \mu_i, \Sigma_i\right) p\left(\mathbf{x}_j \mid \mu_i, \Sigma_i\right)+\sum_{\mathbf{x}_i \in D_u} \ln \sum_{i=1}^N \alpha_i p\left(\mathbf{x}_j \mid \mu_i, \Sigma_i\right) \tag{6}$$

(4) 带入 (6) 可得:

$$LL\left(D_l \cup D_u\right)=\sum_{\left(\mathbf{x}_i, c_j\right) \in D_l} \ln \sum_{i=1}^N \alpha_i \beta_{c_j \mid i} p\left(\mathbf{x}_j \mid \mu_i, \Sigma_i\right)+\sum_{\mathbf{x}_i \in D_u} \ln \sum_{i=1}^N \alpha_i p\left(\mathbf{x}_j \mid \mu_i, \Sigma_i\right) \tag{7}$$

我们的目的是要求得最优的 $\alpha_i, \beta_{cj \mid i}, \mu_i, \Sigma_i$ 使上式 (7) 取得最大值。

> 在这里，依据对数据完整性的不同看法，可有两种 EM 算法:
>
> **EM-1(假定不含类标记):**
>
> 对于 $\left(\mathbf{x}_j, c_j\right) \in D_l, \mathbf{x}_j \in D_u$, 均缺乏混合成分 m_j 信息，相应的完整数据为 $\left\{\left(\mathbf{x}_j, c_j, m_j\right)\right\}$ 和 $\left\{\left(\mathbf{x}_j, m_j\right)\right\}$，也就是说不用推断 $\mathbf{x}_j \in D_u$ 的类标记。
>
> **EM-II(假定含类标记):**
>
> 对于 D_l 定义同上，但对于 $\mathbf{x}_j \in D_u$，认定其缺少 m_j, c_j，因此对应于 $\mathbf{x}_j \in D_u$ 的完整数据为 $\left\{\left(\mathbf{x}_j, c_j, m_j\right)\right\}$，也就是说既要推断 $\mathbf{x}_j \in D_u$ 的类标记，还要推断 $\mathbf{x}_j \in D_u$ 的混合成分。

EM-I

对于混合系数 α_i, 除了要最大化 $LL\left(D_l \cup D_u\right)$，还应满足隐含条件: $\alpha_i \geq 0, \sum_{i=1}^N \alpha_i=1$， 因此考虑对 $LL\left(D_l \cup D_u\right)$ 使用拉格朗日乘子法，变为优化:

$$LL\left(D_l \cup D_u\right)+\lambda\left(\sum_{i=1}^N \alpha_i-1\right) \tag{8}$$

将 (7) 带入 (8)，并令 (8) 对 α_i 的导数为 0，得到:

$$\frac{\partial LL\left(D_l \cup D_u\right)}{\partial \alpha_i}=\sum_{\mathbf{x} j \in D_l} \frac{\beta_{cj \mid i} \cdot p\left(\mathbf{x}_j \mid \mu_i, \Sigma_i\right)}{\sum_{i=1}^N \alpha_i \cdot \beta_{cj \mid i} \cdot p\left(\mathbf{x}_j \mid \mu_i, \Sigma_i\right)}+\sum_{\mathbf{x}_j \in D_u} \frac{p\left(\mathbf{x}_j \mid \mu_i, \Sigma_i\right)}{\sum_{i=1}^N \alpha_i \cdot p\left(\mathbf{x}_j \mid \mu_i, \Sigma_i\right)}+\lambda=0 \tag{9}$$

令:

$$p\left(m_j=i \mid c_j, \mathbf{x}_j, \mu_i, \Sigma_i\right)=\frac{\alpha_i \cdot \beta_{cji} \cdot p\left(\mathbf{x}_j \mid \mu_i, \Sigma_i\right)}{\sum_{i=1}^N \alpha_i \cdot \beta_{cj \mid i} \cdot p\left(\mathbf{x}_j \mid \mu_i, \Sigma_i\right)} \tag{10}$$

同时，将高斯模型 $(*)$ 带入 (2) 可得:

$$p\left(m_j=i \mid \mathbf{x}_j, \mu_i, \Sigma_i\right)=\frac{\alpha_i \cdot p\left(\mathbf{x}_j \mid \mu_i, \Sigma_i\right)}{\sum_{i=1}^N \alpha_i \cdot p\left(\mathbf{x}_j \mid \mu_i, \Sigma_i\right)} \tag{11}$$

对 (9) 两边同时乘以 α_i，将 (10), (11) 代入可得:

$$0=\sum_{\mathrm{x} j \in D_l} p\left(m_j=i \mid c_j, \mathbf{x}_j, \mu_i, \Sigma_i\right)+\sum_{\mathrm{x} j \in D_u} p\left(m_j=i \mid \mathbf{x}_j, \mu_i, \Sigma_i\right)+\alpha_i \cdot \lambda \tag{12}$$

令 (12) 对所有高斯混合成分求和:

$$\begin{aligned}
0 & =\sum_{\mathrm{x} j \in D_l} \sum_{i=1}^N p\left(m_j=i \mid c_j, \mathbf{x}_j, \mu_i, \Sigma_i\right)+\sum_{\mathrm{x} j \in D_u} \sum_{i=1}^N p\left(m_j=i \mid \mathbf{x}_j, \mu_i, \Sigma_i\right)+\alpha_i \cdot \lambda \\
& =\sum_{\mathbf{x}_j \in D_l} 1+\sum_{\mathbf{x}_j \in D_u} 1+\lambda \\
& =M+\lambda
\end{aligned} \tag{13}$$

由 (13) 可得， $\lambda=-M$，将其带入 (12) 可得:

$$\alpha_i=\frac{1}{M} \cdot\left(\sum_{\mathrm{x} j \in D_l} p\left(m_j=i \mid c_j, \mathbf{x}_j, \mu_i, \Sigma_i\right)+\sum_{\mathbf{x}_j \in D_u} p\left(m_j=i \mid \mathbf{x}_j, \mu_i, \Sigma_i\right)\right) \tag{14}$$

对于高斯分布，其偏导具有如下性质:

$$\frac{\partial p\left(\mathbf{x} \mid \mu_i, \Sigma_i\right)}{\partial \mu_i}=p\left(\mathbf{x} \mid \mu_i, \Sigma_i\right) \cdot \Sigma_i^{-1} \cdot\left(\mu_i-\mathbf{x}\right) \tag{15}$$

$$\frac{\partial p\left(\mathbf{x} \mid \mu_i, \Sigma_i\right)}{\partial \Sigma_i}=p\left(\mathbf{x} \mid \mu_i, \Sigma_i\right) \cdot \Sigma_i^{-2} \cdot\left(\left(\mathbf{x}-\mu_i\right)\left(\mathbf{x}-\mu_i\right)^{\top}-\Sigma_i\right) \tag{16}$$

求 (7) 对 μ_i 的偏导，结合 (15), (10), (11) 可得:

$$\begin{aligned}
\frac{\partial LL\left(D_l \cup D_u\right)}{\partial \mu_i} &=\sum_{\mathbf{x}_j \in D_l} \frac{\alpha_i \cdot \beta_{cji} \cdot p\left(\mathbf{x}_j \mid \mu_i, \Sigma_i\right)}{\sum_{i=1}^N \alpha_i \cdot \beta_{cj|i} \cdot p\left(\mathbf{x}_j \mid \mu_i, \Sigma_i\right)} \cdot \Sigma_i^{-1} \cdot\left(\mu_i-\mathbf{x}_j\right)+\sum_{\mathbf{x}_j \in D_u} \frac{\alpha_i \cdot p\left(\mathbf{x}_j \mid \mu_i, \Sigma_i\right)}{\sum_{i=1}^N \alpha_i \cdot p\left(\mathbf{x}_j \mid \mu_i, \Sigma_i\right)} \cdot \Sigma_i^{-1} \cdot\left(\mu_i-\mathbf{x}_j\right) \\
&=\sum_{\mathbf{x}_j \in D_l} p\left(m_j=i \mid c_j, \mathbf{x}_j, \mu_i, \Sigma_i\right) \cdot \Sigma_i^{-1} \cdot\left(\mu_i-\mathbf{x}_j\right)+\sum_{\mathbf{x}_j \in D_u} p\left(m_j=i \mid \mathbf{x}_j, \mu_i, \Sigma_i\right) \cdot \Sigma_i^{-1} \cdot\left(\mu_i-\mathbf{x}_j\right) \\
&=\Sigma_i^{-1} \cdot\left(\sum_{\mathbf{x}_j \in D_l} p\left(m_j=i \mid c_j, \mathbf{x}_j, \mu_i, \Sigma_i\right) \cdot\left(\mu_i-\mathbf{x}_j\right)+\sum_{\mathbf{x}_j \in D_u} p\left(m_j=i \mid \mathbf{x}_j, \mu_i, \Sigma_i\right) \cdot\left(\mu_i-\mathbf{x}_j\right)\right)
\end{aligned} \tag{17}$$

令 $(17)=0$，将 (14) 带入可得:

$$\mu_i=\frac{1}{M \alpha_i} \cdot\left(\sum_{x_j \in D_l} \mathbf{x}_j \cdot p\left(m_j=i \mid c_j, \mathbf{x}_j, \mu_i, \Sigma_i\right)+\sum_{\mathbf{x}_j \in D_u} \mathbf{x}_j \cdot p\left(m_j=i \mid \mathbf{x}_j, \mu_i, \Sigma_i\right)\right) \tag{18}$$

同样地，求 (7) 对 Σ_i 的偏导，结合 (16), (10), (11) 可得:

$$\begin{aligned}
\frac{\partial LL\left(D_l \cup D_u\right)}{\partial \Sigma_i} &=\sum_{\mathbf{x}_j \in D_l} \frac{\alpha_i \cdot \beta_{cj|i} \cdot p\left(\mathbf{x}_j \mid \mu_i, \Sigma_i\right)}{\sum_{i=1}^N \alpha_i \cdot \beta_{cj|i} \cdot p\left(\mathbf{x}_j \mid \mu_i, \Sigma_i\right)} \cdot \Sigma_i^{-2} \cdot\left(\left(\mathbf{x}_j-\mu_i\right)\left(\mathbf{x}_j-\mu_i\right)^{\top}-\Sigma_i\right) \\
&\quad+\sum_{\mathbf{x}_j \in D_u} \frac{\alpha_i \cdot p\left(\mathbf{x}_j \mid \mu_i, \Sigma_i\right)}{\sum_{i=1}^N \alpha_i \cdot p\left(\mathbf{x}_j \mid \mu_i, \Sigma_i\right)} \cdot \Sigma_i^{-2} \cdot\left(\left(\mathbf{x}_j-\mu_i\right)\left(\mathbf{x}_j-\mu_i\right)^{\top}-\Sigma_i\right) \\
&=\sum_{\mathbf{x}_j \in D_l} p\left(m_j=i \mid c_j, \mathbf{x}_j, \mu_i, \Sigma_i\right) \cdot \Sigma_i^{-2} \cdot\left(\left(\mathbf{x}_j-\mu_i\right)\left(\mathbf{x}_j-\mu_i\right)^{\top}-\Sigma_i\right) \\
&\quad+\sum_{\mathbf{x}_j \in D_u} p\left(m_j=i \mid \mathbf{x}_j, \mu_i, \Sigma_i\right) \cdot \Sigma_i^{-2} \cdot\left(\left(\mathbf{x}_j-\mu_i\right)\left(\mathbf{x}_j-\mu_i\right)^{\top}-\Sigma_i\right)
\end{aligned} \tag{19}$$

令 $(19)=0$，将 (14) 带入可得:

$$\begin{aligned}
\Sigma_i=\frac{1}{M \alpha_i} \cdot & \left(\sum_{\mathbf{x}_j \in D_l} p\left(m_j=i \mid c_j, \mathbf{x}_j, \mu_i, \Sigma_i\right) \cdot\left(\left(\mathbf{x}_j-\mu_i\right)\left(\mathbf{x}_j-\mu_i\right)^{\top}\right)\right. \\
& \left.+\sum_{\mathbf{x}_j \in D_u} p\left(m_j=i \mid \mathbf{x}_j, \mu_i, \Sigma_i\right) \cdot\left(\left(\mathbf{x}_j-\mu_i\right)\left(\mathbf{x}_j-\mu_i\right)^{\top}\right)\right)
\end{aligned} \tag{20}$$

对于系数 $\beta_{k|i}$，除了要最大化 $LL\left(D_l \cup D_u\right)$，还应满足隐含条件: $\beta_{k|i} \geq 0, \sum_{k=1}^{|\mathcal{C}|} \beta_{k|i}=1$，因此考虑对 $LL\left(D_l \cup D_u\right)$ 使用拉格朗日乘子法，变为优化:

$$LL\left(D_l \cup D_u\right)+\lambda\left(\sum_{k=1}^{|\mathcal{C}|} \beta_{k|i}-1\right) \tag{21}$$

将 (7) 带入 (21)，并令 (21) 对 $\beta_{k|i}$ 的导数为 0，得到:

$$\frac{\partial LL\left(D_l \cup D_u\right)}{\partial \beta_{k|i}}=\sum_{\mathbf{x}_j \in D_l \wedge cj=k} \frac{\alpha_i \cdot p\left(\mathbf{x}_j \mid \mu_i, \Sigma_i\right)}{\sum_{i=1}^N \alpha_i \cdot \beta_{cj|i} \cdot p\left(\mathbf{x}_j \mid \mu_i, \Sigma_i\right)}+\lambda=0 \tag{22}$$

两边同时乘以 $\beta_{k|i}$, 结合 (10) 得:

$$\begin{aligned}
0 &=\sum_{\mathbf{x}_j \in D_l \wedge cj=k} \frac{\alpha_i \cdot \beta_{k|i} \cdot p\left(\mathbf{x}_j \mid \mu_i, \Sigma_i\right)}{\sum_{i=1}^N \alpha_i \cdot \beta_{cj|i} \cdot p\left(\mathbf{x}_j \mid \mu_i, \Sigma_i\right)}+\beta_{k|i} \cdot \lambda \\
&=\sum_{\mathbf{x}_j \in D_l \wedge cj=k} p\left(m_j=i \mid c_j, \mathbf{x}_j, \mu_i, \Sigma_i\right)+\beta_{k|i} \cdot \lambda
\end{aligned} \tag{23}$$

令 (23) 对所有卷标记求和:

$$0=\sum_{k=1}^{|\mathcal{C}|} \sum_{\mathbf{x} j \in D l \wedge c j=k} p\left(m_j=i \mid c_j, \mathbf{x}_j, \mu_i, \Sigma_i\right)+\sum_{k=1}^{|\mathcal{C}|} \beta_{k|i} \cdot \lambda \tag{24}$$

$$= \sum_{\mathbf{x}_j \in D_l} p\left(m_j = i \mid c_j, \mathbf{x}_j, \mu_i, \Sigma_i\right) + \lambda \tag{24}$$

即:

$$\lambda = -\sum_{\mathbf{x}_j \in D_l} p\left(m_j = i \mid c_j, \mathbf{x}_j, \mu_i, \Sigma_i\right) \tag{25}$$

将 (25) 带入 (23) 可得:

$$\beta_{k|i} = \frac{\sum_{\mathbf{x}j \in D_l \wedge cj=k} p\left(m_j = i \mid c_j, \mathbf{x}_j, \mu_i, \Sigma_i\right)}{\sum_{\mathbf{x}j \in D_l} p\left(m_j = i \mid c_j, \mathbf{x}_j, \mu_i, \Sigma_i\right)} \tag{26}$$

EM-II

对于EM-II，由于需要预测 $\mathbf{x}_j \in D_u$ 下的 c_j，根据贝叶斯定理，(7)变为:

$$\begin{aligned} LL\left(D_l \cup D_u\right) &= \sum_{(\mathbf{x}_i, cj) \in D_l} \ln \sum_{i=1}^{N} \alpha_i \beta_{cj|i} p\left(\mathbf{x}_j \mid \mu_i, \Sigma_i\right) + \sum_{\mathbf{x}_i \in D_u} \ln \sum_{i=1}^{N} \alpha_i p\left(\mathbf{x}_j \mid \mu_i, \Sigma_i\right) \\ &= \sum_{(\mathbf{x}_i, c_j) \in D_l} \ln \sum_{i=1}^{N} \alpha_i \beta_{cj|i} p\left(\mathbf{x}_j \mid \mu_i, \Sigma_i\right) + \sum_{\mathbf{x}_i \in D_u} \ln \sum_{i=1}^{N} \sum_{k=1}^{|\mathcal{C}|} \alpha_i p\left(c_j = k \mid \mathbf{x}_j, m_j = i, \mu_i, \Sigma_i\right) p\left(\mathbf{x}_j \mid \mu_i, \Sigma_i\right) \\ &= \sum_{(\mathbf{x}i, cj) \in D_l} \ln \sum_{i=1}^{N} \alpha_i \beta_{cj|i} p\left(\mathbf{x}_j \mid \mu_i, \Sigma_i\right) + \sum_{\mathbf{x}_i \in D_u} \ln \sum_{i=1}^{N} \sum_{k=1}^{|\mathcal{C}|} \alpha_i \beta_{k|i} p\left(\mathbf{x}_j \mid \mu_i, \Sigma_i\right) \end{aligned} \tag{27}$$

显然，此时的模型参数 $\alpha_i, \mu_i, \Sigma_i$ 与 EM-I一致，对于 $\beta_{k|i}$，同样满足隐含条件: $\beta_{k|i} \geq 0, \sum_{k=1}^{|\mathcal{C}|} \beta_{k|i} = 1$，因此同样将 (27) 带入 (21) 求偏导，并令 (21) 对 $\beta_{k|i}$ 的导数为 0，得到

$$\frac{\partial LL\left(D_l \cup D_u\right)}{\partial \beta_{k|i}} = \sum_{\mathbf{x}_j \in D_l \wedge c_j = k} \frac{\alpha_i \cdot p\left(\mathbf{x}_j \mid \mu_i, \Sigma_i\right)}{\sum_{i=1}^{N} \alpha_i \cdot \beta_{cj|i} \cdot p\left(\mathbf{x}_j \mid \mu_i, \Sigma_i\right)} + \sum_{\mathbf{x}_j \in D_u} \frac{\alpha_i \cdot p\left(\mathbf{x}_j \mid \mu_i, \Sigma_i\right)}{\sum_{i=1}^{N} \alpha_i \cdot p\left(\mathbf{x}_j \mid \mu_i, \Sigma_i\right)} + \lambda = 0 \tag{28}$$

$$p\left(m_j = i, c_j = k \mid \mathbf{x}_j, \mu_i, \Sigma_i\right) = \frac{\alpha_i \cdot \beta_{k|i} \cdot p\left(\mathbf{x}_j \mid \mu_i, \Sigma_i\right)}{\sum_{i=1}^{N} \alpha_i \cdot p\left(\mathbf{x}_j \mid \mu_i, \Sigma_i\right)} \tag{29}$$

对 (28) 两边同乘 $\beta_{k|i}$， 结合 $(10), (29)$ 可得:

$$0 = \sum_{\mathbf{x}j \in Dl \wedge cj = k} p\left(m_j = i \mid c_j, \mathbf{x}_j, \mu_i, \Sigma_i\right) + \sum_{\mathbf{x}j \in D_u} p\left(m_j = i, c_j = k \mid \mathbf{x}_j, \mu_i, \Sigma_i\right) + \beta_{k|i} \lambda \tag{30}$$

对所有类标记求和可得:

$$\lambda = -M\alpha_i \tag{31}$$

最后，将(31)带入(30)即可解得:

$$\beta_{k|i} = \frac{1}{M\alpha_i}\left(\sum_{\mathbf{x}_j \in D_l \wedge c_j = k} p\left(m_j = i \mid c_j, \mathbf{x}_j, \mu_i, \Sigma_i\right) + \sum_{\mathbf{x}_j \in D_u} p\left(m_j = i, c_j = k \mid \mathbf{x}_j, \mu_i, \Sigma_i\right)\right) \tag{32}$$

由此，我们得到了EM-I和EM-II算法下的模型参数 $\alpha_i, \mu_i, \Sigma_i, \beta_{k|i}$ 的更新公式。