$$\sum_{j=1}^{m} rac{lpha_i \cdot p\left(oldsymbol{x_j} | oldsymbol{\mu}_i, oldsymbol{\Sigma}_i
ight)}{\sum_{l=1}^{k} lpha_l \cdot p(oldsymbol{x_j} | oldsymbol{\mu}_l, oldsymbol{\Sigma}_l)} (oldsymbol{x_j} - oldsymbol{\mu_i}) = 0$$

[推导]:根据公式(9.28)可知:

$$p(oldsymbol{x_j}|oldsymbol{\mu_i}, oldsymbol{\Sigma_i}) = rac{1}{(2\pi)^{rac{n}{2}}|oldsymbol{\Sigma_i}|^{rac{1}{2}}}e^{-rac{1}{2}(oldsymbol{x_j}-oldsymbol{\mu_i})^Toldsymbol{\Sigma}_i^{-1}(oldsymbol{x_j}-oldsymbol{\mu_i})}$$

又根据公式(9.32),由

$$\frac{\partial LL(D)}{\partial u_i} = 0$$

$$\frac{\partial LL(D)}{\partial \mu_{i}} = \frac{\partial}{\partial \mu_{i}} \sum_{j=1}^{m} ln \left( \sum_{i=1}^{k} \alpha_{i} \cdot p(\boldsymbol{x}_{j} | \mu_{i}, \boldsymbol{\Sigma}_{i}) \right)$$

$$= \sum_{j=1}^{m} \frac{\partial}{\partial \mu_{i}} ln \left( \sum_{i=1}^{k} \alpha_{i} \cdot p(\boldsymbol{x}_{j} | \mu_{i}, \boldsymbol{\Sigma}_{i}) \right)$$

$$= \sum_{j=1}^{m} \frac{\partial}{\partial \mu_{i}} (p(\boldsymbol{x}_{j} | \mu_{i}, \boldsymbol{\Sigma}_{i}))$$

$$= \sum_{j=1}^{m} \frac{\alpha_{i} \cdot \frac{\partial}{\partial \mu_{i}} (p(\boldsymbol{x}_{j} | \mu_{i}, \boldsymbol{\Sigma}_{i}))}{\sum_{l=1}^{k} \alpha_{l} \cdot p(\boldsymbol{x}_{j} | \mu_{l}, \boldsymbol{\Sigma}_{l})}$$

$$= \sum_{j=1}^{m} \frac{\alpha_{i} \cdot \frac{1}{2} \frac{1}{2} e^{-\frac{1}{2}(\boldsymbol{x}_{j} - \mu_{i})^{T} \boldsymbol{\Sigma}_{i}^{-1}(\boldsymbol{x}_{j} - \mu_{i})}}{\sum_{l=1}^{k} \alpha_{i} \cdot p(\boldsymbol{x}_{j} | \mu_{l}, \boldsymbol{\Sigma}_{l})} \left( -\frac{1}{2} \right) \left( \left( \boldsymbol{\Sigma}_{i}^{-1} + \left( \boldsymbol{\Sigma}_{i}^{-1} \right)^{T} \right) \cdot (\boldsymbol{x}_{j} - \mu_{i}) \cdot (-1) \right)$$

$$= \sum_{j=1}^{m} \frac{\alpha_{i} \cdot p(\boldsymbol{x}_{j} | \mu_{i}, \boldsymbol{\Sigma}_{l})}{\sum_{l=1}^{k} \alpha_{i} \cdot p(\boldsymbol{x}_{j} | \mu_{l}, \boldsymbol{\Sigma}_{l})} \left( -\frac{1}{2} \right) \left( -\boldsymbol{\Sigma}_{i}^{-1} \left( \boldsymbol{x}_{j} - \mu_{i} \right) - \left( \boldsymbol{\Sigma}_{i}^{-1} \right)^{T} \left( \boldsymbol{x}_{j} - \mu_{i} \right) \right)$$

$$= \sum_{j=1}^{m} \frac{\alpha_{i} \cdot p(\boldsymbol{x}_{j} | \mu_{i}, \boldsymbol{\Sigma}_{l})}{\sum_{l=1}^{k} \alpha_{l} \cdot p(\boldsymbol{x}_{j} | \mu_{l}, \boldsymbol{\Sigma}_{l})} \left( -\frac{1}{2} \right) \left( -\boldsymbol{\Sigma}_{i}^{-1} \left( \boldsymbol{x}_{j} - \mu_{i} \right) - \left( \boldsymbol{\Sigma}_{i}^{-1} \right)^{T} \left( \boldsymbol{x}_{j} - \mu_{i} \right) \right)$$

$$= \sum_{j=1}^{m} \frac{\alpha_{i} \cdot p(\boldsymbol{x}_{j} | \mu_{i}, \boldsymbol{\Sigma}_{l})}{\sum_{l=1}^{k} \alpha_{l} \cdot p(\boldsymbol{x}_{j} | \mu_{l}, \boldsymbol{\Sigma}_{l})} \left( -\frac{1}{2} \right) \left( -\boldsymbol{\Sigma}_{i}^{-1} \left( \boldsymbol{x}_{j} - \mu_{i} \right) - \boldsymbol{\Sigma}_{i}^{-1} \left( \boldsymbol{x}_{j} - \mu_{i} \right) \right)$$

$$= \sum_{j=1}^{m} \frac{\alpha_{i} \cdot p(\boldsymbol{x}_{j} | \mu_{i}, \boldsymbol{\Sigma}_{l})}{\sum_{l=1}^{k} \alpha_{l} \cdot p(\boldsymbol{x}_{j} | \mu_{l}, \boldsymbol{\Sigma}_{l})} \left( -\frac{1}{2} \right) \left( -\boldsymbol{\Sigma}_{i}^{-1} \left( \boldsymbol{x}_{j} - \mu_{i} \right) - \boldsymbol{\Sigma}_{i}^{-1} \left( \boldsymbol{x}_{j} - \mu_{i} \right) \right)$$

$$= \sum_{j=1}^{m} \frac{\alpha_{i} \cdot p(\boldsymbol{x}_{j} | \mu_{i}, \boldsymbol{\Sigma}_{l})}{\sum_{l=1}^{k} \alpha_{l} \cdot p(\boldsymbol{x}_{j} | \mu_{l}, \boldsymbol{\Sigma}_{l})} \sum_{l=1}^{-1} \left( -\frac{1}{2} \right) \left( -2 \cdot \boldsymbol{\Sigma}_{i}^{-1} \left( \boldsymbol{x}_{j} - \mu_{i} \right) \right)$$

$$= \sum_{j=1}^{m} \frac{\alpha_{i} \cdot p(\boldsymbol{x}_{j} | \mu_{i}, \boldsymbol{\Sigma}_{l})}{\sum_{l=1}^{k} \alpha_{i} \cdot p(\boldsymbol{x}_{j} | \mu_{l}, \boldsymbol{\Sigma}_{l})} \sum_{l=1}^{-1} \alpha_{l} \cdot p(\boldsymbol{x}_{j} | \mu_{l}, \boldsymbol{\Sigma}_{l}) \right)$$

$$= \sum_{j=1}^{m} \frac{\alpha_{i} \cdot p(\boldsymbol{x}_{j} | \mu_{i}, \boldsymbol{\Sigma}_{l})}{\sum_{l=1}^{k} \alpha_{i} \cdot p(\boldsymbol{x}_{j} | \mu_{l}, \boldsymbol{\Sigma}_{l})} \sum_{l=1}^{-1} \alpha_{l} \cdot p(\boldsymbol{x}_{j} | \mu_{l}, \boldsymbol{\Sigma}_{l})$$

$$= \sum_{j=1}^{m} \frac{\alpha_{i} \cdot p(\boldsymbol{x$$

9.35

$$oldsymbol{\Sigma}_i = rac{\sum_{j=1}^m \gamma_{ji} (oldsymbol{x_j} - oldsymbol{\mu_i}) (oldsymbol{x_j} - oldsymbol{\mu_i})^T}{\sum_{j=1}^m} \gamma_{ji}$$

[推导]:根据公式(9.28)可知:

$$p(oldsymbol{x_j}|oldsymbol{\mu_i}, oldsymbol{\Sigma_i}) = rac{1}{(2\pi)^{rac{n}{2}}|oldsymbol{\Sigma}_i|^{rac{1}{2}}}e^{-rac{1}{2}(oldsymbol{x_j}-oldsymbol{\mu_i})^Toldsymbol{\Sigma}_i^{-1}(oldsymbol{x_j}-oldsymbol{\mu_i})}$$

又根据公式(9.32),由

$$\frac{\partial LL(D)}{\partial \Sigma_i} = 0$$

$$egin{aligned} rac{\partial LL(D)}{\partial oldsymbol{\Sigma}_i} &= rac{\partial}{\partial oldsymbol{\Sigma}_i} \sum_{j=1}^m ln igg( \sum_{i=1}^k lpha_i \cdot p(oldsymbol{x_j} | oldsymbol{\mu}_i, oldsymbol{\Sigma}_i) igg) \ &= \sum_{j=1}^m rac{\partial}{\partial oldsymbol{\Sigma}_i} ln igg( \sum_{i=1}^k lpha_i \cdot p(oldsymbol{x_j} | oldsymbol{\mu}_i, oldsymbol{\Sigma}_i) igg) \ &= \sum_{j=1}^m rac{lpha_i \cdot rac{\partial}{\partial oldsymbol{\Sigma}_i} p(oldsymbol{x_j} | oldsymbol{\mu}_i, oldsymbol{\Sigma}_i)}{\sum_{l=1}^k lpha_l \cdot p(oldsymbol{x_j} | oldsymbol{\mu}_l, oldsymbol{\Sigma}_l)} \ &= \sum_{j=1}^m rac{lpha_i \cdot rac{\partial}{\partial oldsymbol{\Sigma}_i} rac{1}{(2\pi)^{rac{n}{2}} |oldsymbol{\Sigma}_i|^{rac{1}{2}}} e^{-rac{1}{2}(oldsymbol{x_j} - oldsymbol{\mu}_l)^T oldsymbol{\Sigma}_i^{-1}(oldsymbol{x_j} - oldsymbol{\mu}_l)} \ &= \sum_{j=1}^m rac{lpha_i \cdot rac{\partial}{\partial oldsymbol{\Sigma}_i} rac{1}{(2\pi)^{rac{n}{2}} |oldsymbol{\Sigma}_i|^{rac{1}{2}}} e^{-rac{1}{2}(oldsymbol{x_j} - oldsymbol{\mu}_l)^T oldsymbol{\Sigma}_i^{-1}(oldsymbol{x_j} - oldsymbol{\mu}_l)} \ &= \sum_{j=1}^m rac{lpha_i \cdot rac{\partial}{\partial oldsymbol{\Sigma}_i} rac{1}{(2\pi)^{rac{n}{2}} |oldsymbol{\Sigma}_l|^{rac{1}{2}}} e^{-rac{1}{2}(oldsymbol{x_j} - oldsymbol{\mu}_l)^T oldsymbol{\Sigma}_i^{-1}(oldsymbol{x_j} - oldsymbol{\mu}_l)} \ &= \sum_{j=1}^m rac{lpha_i \cdot rac{\partial}{\partial oldsymbol{\Sigma}_i} rac{1}{(2\pi)^{rac{n}{2}} |oldsymbol{\Sigma}_l|^{rac{1}{2}}} e^{-rac{1}{2}(oldsymbol{x_j} - oldsymbol{\mu}_l)^T oldsymbol{\Sigma}_i^{-1}(oldsymbol{x_j} - oldsymbol{\mu}_l)} \ &= \sum_{j=1}^m rac{lpha_i \cdot rac{\partial}{\partial oldsymbol{\Sigma}_l} rac{1}{(2\pi)^{rac{n}{2}} |oldsymbol{\Sigma}_l|^{rac{n}{2}}} e^{-rac{1}{2}(oldsymbol{X_j} - oldsymbol{\mu}_l)} \ &= \sum_{j=1}^m rac{lpha_i \cdot rac{\partial}{\partial oldsymbol{\Sigma}_l} rac{1}{(2\pi)^{rac{n}{2}} |oldsymbol{\Sigma}_l|^{rac{n}{2}}} e^{-rac{1}{2}(oldsymbol{X_j} - oldsymbol{\Sigma}_l)} \ &= \sum_{j=1}^m rac{a_j \cdot rac{\partial}{\partial oldsymbol{\Sigma}_l} rac{1}{(2\pi)^{rac{n}{2}} |oldsymbol{\Sigma}_l|^{rac{n}{2}}} e^{-rac{1}{2}(oldsymbol{\Sigma}_l)} \ &= \sum_{j=1}^m rac{a_j \cdot rac{\partial}{\partial oldsymbol{\Sigma}_l} rac{a_j \cdot oldsymbol{\Sigma}_l}{(2\pi)^{n}} \ &= \sum_{j=1}^m rac{a_j \cdot oldsymbol{\Sigma}_l} rac{a_j \cdot oldsymbol{\Sigma}_l}{(2\pi)^{n}} \ &=$$

可得

$$\begin{split} &\sum_{l=1}^{k} \alpha_{l} \cdot p(\boldsymbol{x}_{j}, |\boldsymbol{\mu}_{l}, \boldsymbol{\Sigma}_{l}) \\ &= \sum_{j=1}^{m} \frac{\alpha_{i} \cdot \frac{\partial}{\partial \boldsymbol{\Sigma}_{i}} e^{-\frac{1}{2}(\boldsymbol{x}_{j} - \boldsymbol{\mu}_{i})^{T} \boldsymbol{\Sigma}_{i}^{-1}(\boldsymbol{x}_{j} - \boldsymbol{\mu}_{i})}}{\sum_{l=1}^{k} \alpha_{l} \cdot p(\boldsymbol{x}_{j}, |\boldsymbol{\mu}_{l}, \boldsymbol{\Sigma}_{l})} \\ &= \sum_{j=1}^{m} \frac{\alpha_{i} \cdot \frac{\partial}{\partial \boldsymbol{\Sigma}_{i}} e^{-\frac{n}{2} ln(2\pi) - \frac{1}{2} ln(|\boldsymbol{\Sigma}_{i}|) - \frac{1}{2}(\boldsymbol{x}_{j} - \boldsymbol{\mu}_{i})^{T} \boldsymbol{\Sigma}_{i}^{-1}(\boldsymbol{x}_{j} - \boldsymbol{\mu}_{i})}}{\sum_{l=1}^{k} \alpha_{l} \cdot p(\boldsymbol{x}_{j}, |\boldsymbol{\mu}_{l}, \boldsymbol{\Sigma}_{l})} \\ &= \sum_{j=1}^{m} \frac{\alpha_{i} \cdot p(\boldsymbol{x}_{j} | \boldsymbol{\mu}_{i}, \boldsymbol{\Sigma}_{i})}{\sum_{l=1}^{k} \alpha_{l} \cdot p(\boldsymbol{x}_{j}, |\boldsymbol{\mu}_{l}, \boldsymbol{\Sigma}_{l})} \frac{\partial}{\partial \boldsymbol{\Sigma}_{i}} \left( -\frac{n}{2} ln(2\pi) - \frac{1}{2} ln(|\boldsymbol{\Sigma}_{i}|) - \frac{1}{2}(\boldsymbol{x}_{j} - \boldsymbol{\mu}_{i})^{T} \boldsymbol{\Sigma}_{i}^{-1}(\boldsymbol{x}_{j} - \boldsymbol{\mu}_{i}) \right) \\ &= \sum_{j=1}^{m} \frac{\alpha_{i} \cdot p(\boldsymbol{x}_{j} | \boldsymbol{\mu}_{l}, \boldsymbol{\Sigma}_{l})}{\sum_{l=1}^{k} \alpha_{l} \cdot p(\boldsymbol{x}_{j}, |\boldsymbol{\mu}_{l}, \boldsymbol{\Sigma}_{l})} \left( -\frac{1}{2} (\boldsymbol{\Sigma}_{i}^{-1})^{T} - \frac{1}{2} \frac{\partial}{\partial \boldsymbol{\Sigma}_{i}} (\boldsymbol{x}_{j} - \boldsymbol{\mu}_{i})^{T} \boldsymbol{\Sigma}_{i}^{-1}(\boldsymbol{x}_{j} - \boldsymbol{\mu}_{i}) \right) \end{split}$$

为求得

$$rac{\partial}{\partial oldsymbol{\Sigma}_i} (oldsymbol{x_j} - oldsymbol{\mu_i})^T oldsymbol{\Sigma}_i^{-1} \left(oldsymbol{x_j} - oldsymbol{\mu_i}
ight)$$

首先分析对 $\Sigma_i$ 中单一元素的求导,用r代表矩阵 $\Sigma_i$ 的行索引,c代表矩阵 $\Sigma_i$ 的列索引,则

$$egin{aligned} rac{\partial}{\partial \Sigma_{i_{rc}}} (oldsymbol{x_j} - oldsymbol{\mu_i})^T oldsymbol{\Sigma}_i^{-1} \left(oldsymbol{x_j} - oldsymbol{\mu_i}
ight)^T oldsymbol{\Sigma}_{i_{rc}}^{-1} \left(oldsymbol{x_j} - oldsymbol{\mu_i}
ight) \ &= -(oldsymbol{x_j} - oldsymbol{\mu_i})^T oldsymbol{\Sigma}_{i_{rc}}^{-1} rac{\partial oldsymbol{\Sigma}_i}{\partial \Sigma_{i_{rc}}} oldsymbol{\Sigma}_i^{-1} \left(oldsymbol{x_j} - oldsymbol{\mu_i}
ight) \end{aligned}$$

$$= -(oldsymbol{x_j} - oldsymbol{\mu_i})^T oldsymbol{\Sigma}_{i_{rc}}^{-1} rac{\partial oldsymbol{\Sigma}_i}{\partial \Sigma_{i_{rc}}} oldsymbol{\Sigma}_{i_{rc}}^{-1} \left(oldsymbol{x_j} - oldsymbol{\mu_i}
ight) \end{aligned}$$

$$= -(oldsymbol{x_j} - oldsymbol{\mu_i})^T oldsymbol{\Sigma}_{i_{rc}}^{-1} \left(oldsymbol{x_j} - oldsymbol{\mu_i}
ight)$$

$$B^T = \left(\mathbf{\Sigma}_i^{-1} \left(\mathbf{x}_j - \mathbf{\mu}_i\right)\right)^T$$
 $= \left(\mathbf{x}_j - \mathbf{\mu}_i\right)^T \left(\mathbf{\Sigma}_i^{-1}\right)^T$  所以  $\frac{\partial}{\partial \Sigma_{i_{rc}}} \left(\mathbf{x}_j - \mathbf{\mu}_i\right)^T \mathbf{\Sigma}_i^{-1} \left(\mathbf{x}_j - \mathbf{\mu}_i\right) = -B^T \frac{\partial \mathbf{\Sigma}_i}{\partial \Sigma_{i_{rc}}} B$  其中 $B$ 为 $n \times 1$ 阶矩阵, $\frac{\partial \mathbf{\Sigma}_i}{\partial \Sigma_{i_{rc}}}$ 为  $= \left(\mathbf{x}_j - \mathbf{\mu}_i\right)^T \mathbf{\Sigma}_i^{-1}$   $n$ 阶方阵,且 $\frac{\partial \mathbf{\Sigma}_i}{\partial \Sigma_{i_{rc}}}$ 仅在 $(r,c)$ 位置处的元素值为1,其它位置处的元素值均为0,所以

$$\frac{\partial}{\partial \Sigma_{i_{rc}}} (\boldsymbol{x_j} - \boldsymbol{\mu_i})^T \boldsymbol{\Sigma}_i^{-1} (\boldsymbol{x_j} - \boldsymbol{\mu_i}) = -B^T \frac{\partial \boldsymbol{\Sigma}_i}{\partial \Sigma_{i_{rc}}} B = -B_r \cdot B_c = -\left(B \cdot B^T\right)_{rc} = \left(-B \cdot B^T\right)_{rc}$$
即对 $\boldsymbol{\Sigma}_i$ 中特定位置

的元素的求导结果对应于 $(-B \cdot B^T)$ 中相同位置的元素值,所以

$$\begin{split} \frac{\partial}{\partial \boldsymbol{\Sigma}_i} (\boldsymbol{x}_j - \boldsymbol{\mu}_i)^T \boldsymbol{\Sigma}_i^{-1} \left( \boldsymbol{x}_j - \boldsymbol{\mu}_i \right) &= -B \cdot B^T \\ &= -\boldsymbol{\Sigma}_i^{-1} \left( \boldsymbol{x}_j - \boldsymbol{\mu}_i \right) \left( \boldsymbol{\Sigma}_i^{-1} \left( \boldsymbol{x}_j - \boldsymbol{\mu}_i \right) \right)^T \\ &= -\boldsymbol{\Sigma}_i^{-1} \left( \boldsymbol{x}_j - \boldsymbol{\mu}_i \right) \left( \boldsymbol{x}_j - \boldsymbol{\mu}_i \right)^T \boldsymbol{\Sigma}_i^{-1} \end{split}$$

因此最终结果为

$$\frac{\partial LL(D)}{\partial \boldsymbol{\Sigma}_{i}} = \sum_{j=1}^{m} \frac{\alpha_{i} \cdot p(\boldsymbol{x}_{j} | \boldsymbol{\mu}_{i}, \boldsymbol{\Sigma}_{i})}{\sum_{l=1}^{k} \alpha_{l} \cdot p(\boldsymbol{x}_{j}, | \boldsymbol{\mu}_{l}, \boldsymbol{\Sigma}_{l})} \bigg( -\frac{1}{2} \bigg( \boldsymbol{\Sigma}_{i}^{-1} - \boldsymbol{\Sigma}_{i}^{-1} \left( \boldsymbol{x}_{j} - \boldsymbol{\mu}_{i} \right) \left( \boldsymbol{x}_{j} - \boldsymbol{\mu}_{i} \right)^{T} \boldsymbol{\Sigma}_{i}^{-1} \bigg) \bigg) = 0$$

整理可得

$$oldsymbol{\Sigma}_i = rac{\sum_{j=1}^m \gamma_{ji} (oldsymbol{x_j} - oldsymbol{\mu_i}) (oldsymbol{x_j} - oldsymbol{\mu_i})^T}{\sum_{j=1}^m \gamma_{ji}} \gamma_{ji}$$

9.38

$$lpha_i = rac{1}{m} \sum_{i=1}^m \gamma_{ji}$$

[推导]:基于公式(9.37)进行恒等变形:

$$egin{aligned} &\sum_{j=1}^{m} rac{p(oldsymbol{x}_{j}|oldsymbol{\mu_{i}},oldsymbol{\Sigma}_{i})}{\sum_{l=1}^{k} lpha_{l} \cdot p(oldsymbol{x}_{j}|oldsymbol{\mu_{l}},oldsymbol{\Sigma}_{l})} + \lambda = 0 \end{aligned} \ \Rightarrow \sum_{j=1}^{m} rac{lpha_{i} \cdot p(oldsymbol{x}_{j}|oldsymbol{\mu_{i}},oldsymbol{\Sigma}_{l})}{\sum_{l=1}^{k} lpha_{l} \cdot p(oldsymbol{x}_{j}|oldsymbol{\mu_{l}},oldsymbol{\Sigma}_{l})} + lpha_{i}\lambda = 0 \end{aligned}$$

对所有混合成分进行求和:

$$\Rightarrow \sum_{i=1}^{k} \left( \sum_{j=1}^{m} \frac{\alpha_{i} \cdot p(\boldsymbol{x}_{j} | \boldsymbol{\mu}_{i}, \boldsymbol{\Sigma}_{i})}{\sum_{l=1}^{k} \alpha_{l} \cdot p(\boldsymbol{x}_{j} | \boldsymbol{\mu}_{l}, \boldsymbol{\Sigma}_{l})} + \alpha_{i} \lambda \right) = 0$$

$$\Rightarrow \sum_{i=1}^{k} \sum_{j=1}^{m} \frac{\alpha_{i} \cdot p(\boldsymbol{x}_{j} | \boldsymbol{\mu}_{i}, \boldsymbol{\Sigma}_{i})}{\sum_{l=1}^{k} \alpha_{l} \cdot p(\boldsymbol{x}_{j} | \boldsymbol{\mu}_{l}, \boldsymbol{\Sigma}_{l})} + \sum_{i=1}^{k} \alpha_{i} \lambda = 0$$

$$\Rightarrow \lambda = -\sum_{i=1}^{k} \sum_{j=1}^{m} \frac{\alpha_{i} \cdot p(\boldsymbol{x}_{j} | \boldsymbol{\mu}_{i}, \boldsymbol{\Sigma}_{l})}{\sum_{l=1}^{k} \alpha_{l} \cdot p(\boldsymbol{x}_{j} | \boldsymbol{\mu}_{l}, \boldsymbol{\Sigma}_{l})} = -m$$

又

$$egin{aligned} \sum_{j=1}^{m} rac{lpha_i \cdot p(oldsymbol{x}_j | oldsymbol{\mu}_i, oldsymbol{\Sigma}_i)}{\sum_{l=1}^{k} lpha_l \cdot p(oldsymbol{x}_j | oldsymbol{\mu}_l, oldsymbol{\Sigma}_l)} + lpha_i \lambda = 0 \end{aligned} \ \Rightarrow \sum_{j=1}^{m} \gamma_{ji} + lpha_i \lambda = 0 \ \Rightarrow lpha_i = -rac{\sum_{j=1}^{m} \gamma_{ji}}{\lambda} = rac{1}{m} \sum_{j=1}^{m} \gamma_{ji} \end{aligned}$$

## 附录

参考公式

$$egin{aligned} rac{\partial oldsymbol{x}^T B oldsymbol{x}}{\partial oldsymbol{x}} &= \left(B + B^T
ight) oldsymbol{x} \ rac{\partial}{\partial A} ln |A| &= \left(A^{-1}
ight)^T \ rac{\partial}{\partial x} \left(A^{-1}
ight) &= -A^{-1} rac{\partial A}{\partial x} A^{-1} \end{aligned}$$

## 参考资料

Petersen, K. B. & Pedersen, M. S. *The Matrix Cookbook*.

Bishop, C. M. (2006). Pattern Recognition and Machine Learning. Springer.