# 9.33

$$\sum_{j=1}^{m} \frac{\alpha_i \cdot p\left(\boldsymbol{x_j} | \boldsymbol{\mu}_i, \boldsymbol{\Sigma}_i\right)}{\sum_{l=1}^{k} \alpha_l \cdot p(\boldsymbol{x_j} | \boldsymbol{\mu}_l, \boldsymbol{\Sigma}_l)}(\boldsymbol{x_j}-\boldsymbol{\mu_i})=0$$

[推导]：根据公式(9.28)可知：

$$p(\boldsymbol{x_j} | \boldsymbol{\mu_i}, \boldsymbol{\Sigma_i})=\frac{1}{(2\pi)^{\frac{n}{2}}\left|\boldsymbol{\Sigma}_i\right|^{\frac{1}{2}}} e^{-\frac{1}{2}(\boldsymbol{x}_j-\boldsymbol{\mu}_i)^T \boldsymbol{\Sigma}_i^{-1}(\boldsymbol{x}_j-\boldsymbol{\mu}_i)}$$

又根据公式(9.32)，由

$$\frac{\partial LL(D)}{\partial \boldsymbol{\mu}_i}=0$$

可得

$$\begin{aligned}
\frac{\partial LL(D)}{\partial \boldsymbol{\mu}_i} &= \frac{\partial}{\partial \boldsymbol{\mu}_i} \sum_{j=1}^{m} ln\left(\sum_{i=1}^{k} \alpha_i \cdot p(\boldsymbol{x_j} | \boldsymbol{\mu}_i, \boldsymbol{\Sigma}_i)\right) \\
&= \sum_{j=1}^{m} \frac{\partial}{\partial \boldsymbol{\mu}_i} ln\left(\sum_{i=1}^{k} \alpha_i \cdot p(\boldsymbol{x_j} | \boldsymbol{\mu}_i, \boldsymbol{\Sigma}_i)\right) \\
&= \sum_{j=1}^{m} \frac{\alpha_i \cdot \frac{\partial}{\partial \boldsymbol{\mu_i}}(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_{j=1}^{m} \frac{\alpha_i \cdot \frac{1}{(2\pi)^{\frac{n}{2}}\left|\boldsymbol{\Sigma}_i\right|^{\frac{1}{2}}} 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})} \frac{\partial}{\partial \boldsymbol{\mu}_i}\left(-\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_i}, \boldsymbol{\Sigma_i})}{\sum_{l=1}^{k} \alpha_l \cdot p(\boldsymbol{x}_j | \boldsymbol{\mu_l}, \boldsymbol{\Sigma_l})}(-\frac{1}{2})\left(\left(\boldsymbol{\Sigma}_i^{-1}+\left(\boldsymbol{\Sigma}_i^{-1}\right)^T\right) \cdot (\boldsymbol{x_j}-\boldsymbol{\mu_i}) \cdot (-1)\right) \\
&= \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{1}{2})\left(-\boldsymbol{\Sigma}_i^{-1}(\boldsymbol{x_j}-\boldsymbol{\mu_i})-\left(\boldsymbol{\Sigma}_i^{-1}\right)^T(\boldsymbol{x_j}-\boldsymbol{\mu_i})\right) \\
&= \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{1}{2})\left(-\boldsymbol{\Sigma}_i^{-1}(\boldsymbol{x_j}-\boldsymbol{\mu_i})-\left(\boldsymbol{\Sigma}_i^T\right)^{-1}(\boldsymbol{x_j}-\boldsymbol{\mu_i})\right) \\
&= \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{1}{2})\left(-\boldsymbol{\Sigma}_i^{-1}(\boldsymbol{x_j}-\boldsymbol{\mu_i})-\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_i}, \boldsymbol{\Sigma_i})}{\sum_{l=1}^{k} \alpha_l \cdot p(\boldsymbol{x}_j | \boldsymbol{\mu_l}, \boldsymbol{\Sigma_l})}(-\frac{1}{2})\left(-2 \cdot \boldsymbol{\Sigma}_i^{-1}(\boldsymbol{x_j}-\boldsymbol{\mu_i})\right) \\
&= \sum_{j=1}^{m} \frac{\alpha_i \cdot p\left(\boldsymbol{x_j} | \boldsymbol{\mu}_i, \boldsymbol{\Sigma}_i\right)}{\sum_{l=1}^{k} \alpha_l \cdot p(\boldsymbol{x_j} | \boldsymbol{\mu}_l, \boldsymbol{\Sigma}_l)} \boldsymbol{\Sigma}_i^{-1}(\boldsymbol{x_j}-\boldsymbol{\mu_i}) \\
&= \sum_{j=1}^{m} \frac{\alpha_i \cdot p\left(\boldsymbol{x_j} | \boldsymbol{\mu}_i, \boldsymbol{\Sigma}_i\right)}{\sum_{l=1}^{k} \alpha_l \cdot p(\boldsymbol{x_j} | \boldsymbol{\mu}_l, \boldsymbol{\Sigma}_l)}(\boldsymbol{x_j}-\boldsymbol{\mu_i})=0
\end{aligned}$$

# 9.35

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

[推导]：根据公式(9.28)可知：

$$p(\boldsymbol{x_j}|\boldsymbol{\mu_i},\boldsymbol{\Sigma_i})=\frac{1}{(2\pi)^{\frac{n}{2}}\left|\boldsymbol{\Sigma}_i\right|^{\frac{1}{2}}}e^{-\frac{1}{2}(\boldsymbol{x}_j-\boldsymbol{\mu}_i)^T\boldsymbol{\Sigma}_i^{-1}(\boldsymbol{x_j}-\boldsymbol{\mu_i})}$$

又根据公式(9.32)，由

$$\frac{\partial LL(D)}{\partial \boldsymbol{\Sigma}_i}=0$$

$$\begin{aligned}
\frac{\partial LL(D)}{\partial \boldsymbol{\Sigma}_i}&=\frac{\partial}{\partial \boldsymbol{\Sigma}_i}\sum_{j=1}^{m}ln\left(\sum_{i=1}^{k}\alpha_i\cdot p(\boldsymbol{x_j}|\boldsymbol{\mu}_i,\boldsymbol{\Sigma}_i)\right)\\
&=\sum_{j=1}^{m}\frac{\partial}{\partial \boldsymbol{\Sigma}_i}ln\left(\sum_{i=1}^{k}\alpha_i\cdot p(\boldsymbol{x_j}|\boldsymbol{\mu}_i,\boldsymbol{\Sigma}_i)\right)\\
&=\sum_{j=1}^{m}\frac{\alpha_i\cdot\frac{\partial}{\partial \boldsymbol{\Sigma}_i}p(\boldsymbol{x}_j|\boldsymbol{\mu}_i,\boldsymbol{\Sigma}_i)}{\sum_{l=1}^{k}\alpha_l\cdot p(\boldsymbol{x}_j,|\boldsymbol{\mu}_l,\boldsymbol{\Sigma}_l)}\\
&=\sum_{j=1}^{m}\frac{\alpha_i\cdot\frac{\partial}{\partial \boldsymbol{\Sigma}_i}\frac{1}{(2\pi)^{\frac{n}{2}}\left|\boldsymbol{\Sigma}_i\right|^{\frac{1}{2}}}e^{-\frac{1}{2}(\boldsymbol{x_j}-\boldsymbol{\mu}_i)^T\boldsymbol{\Sigma}_i^{-1}(\boldsymbol{x_j}-\boldsymbol{\mu_i})}}{\sum_{l=1}^{k}\alpha_l\cdot p(\boldsymbol{x}_j,|\boldsymbol{\mu}_l,\boldsymbol{\Sigma}_l)}
\end{aligned}$$

可得

$$\begin{aligned}
&=\sum_{j=1}^{m}\frac{\alpha_i\cdot\frac{\partial}{\partial \boldsymbol{\Sigma}_i}e^{ln\left(\frac{1}{(2\pi)^{\frac{n}{2}}\left|\boldsymbol{\Sigma}_i\right|^{\frac{1}{2}}}e^{-\frac{1}{2}(\boldsymbol{x_j}-\boldsymbol{\mu}_i)^T\boldsymbol{\Sigma}_i^{-1}(\boldsymbol{x_j}-\boldsymbol{\mu_i})}\right)}}{\sum_{l=1}^{k}\alpha_l\cdot p(\boldsymbol{x}_j,|\boldsymbol{\mu}_l,\boldsymbol{\Sigma}_l)}\\
&=\sum_{j=1}^{m}\frac{\alpha_i\cdot\frac{\partial}{\partial \boldsymbol{\Sigma}_i}e^{-\frac{n}{2}ln(2\pi)-\frac{1}{2}ln(|\boldsymbol{\Sigma}_i|)-\frac{1}{2}(\boldsymbol{x_j}-\boldsymbol{\mu_i})^T\boldsymbol{\Sigma}_i^{-1}(\boldsymbol{x_j}-\boldsymbol{\mu_i})}}{\sum_{l=1}^{k}\alpha_l\cdot p(\boldsymbol{x}_j,|\boldsymbol{\mu}_l,\boldsymbol{\Sigma}_l)}\\
&=\sum_{j=1}^{m}\frac{\alpha_i\cdot p(\boldsymbol{x}_j|\boldsymbol{\mu}_i,\boldsymbol{\Sigma}_i)}{\sum_{l=1}^{k}\alpha_l\cdot p(\boldsymbol{x}_j,|\boldsymbol{\mu}_l,\boldsymbol{\Sigma}_l)}\frac{\partial}{\partial \boldsymbol{\Sigma}_i}\left(-\frac{n}{2}ln\left(2\pi\right)-\frac{1}{2}ln\left(|\boldsymbol{\Sigma}_i|\right)-\frac{1}{2}(\boldsymbol{x_j}-\boldsymbol{\mu_i})^T\boldsymbol{\Sigma}_i^{-1}\left(\boldsymbol{x_j}-\boldsymbol{\mu_i}\right)\right)\\
&=\sum_{j=1}^{m}\frac{\alpha_i\cdot p(\boldsymbol{x}_j|\boldsymbol{\mu}_i,\boldsymbol{\Sigma}_i)}{\sum_{l=1}^{k}\alpha_l\cdot p(\boldsymbol{x}_j,|\boldsymbol{\mu}_l,\boldsymbol{\Sigma}_l)}\left(-\frac{1}{2}\left(\boldsymbol{\Sigma}_i^{-1}\right)^T-\frac{1}{2}\frac{\partial}{\partial \boldsymbol{\Sigma}_i}(\boldsymbol{x_j}-\boldsymbol{\mu_i})^T\boldsymbol{\Sigma}_i^{-1}\left(\boldsymbol{x_j}-\boldsymbol{\mu_i}\right)\right)
\end{aligned}$$

为求得

$$\frac{\partial}{\partial \boldsymbol{\Sigma}_i}(\boldsymbol{x_j}-\boldsymbol{\mu_i})^T\boldsymbol{\Sigma}_i^{-1}\left(\boldsymbol{x_j}-\boldsymbol{\mu_i}\right)$$

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

$$\begin{aligned}
\frac{\partial}{\partial \Sigma_{i_{rc}}}(\boldsymbol{x_j}-\boldsymbol{\mu_i})^T\boldsymbol{\Sigma}_i^{-1}\left(\boldsymbol{x_j}-\boldsymbol{\mu_i}\right)&=(\boldsymbol{x_j}-\boldsymbol{\mu_i})^T\frac{\partial \boldsymbol{\Sigma}_i^{-1}}{\partial \Sigma_{i_{rc}}}(\boldsymbol{x_j}-\boldsymbol{\mu_i})\\
&=-(\boldsymbol{x_j}-\boldsymbol{\mu_i})^T\boldsymbol{\Sigma}_i^{-1}\frac{\partial \boldsymbol{\Sigma}_i}{\partial \Sigma_{i_{rc}}}\boldsymbol{\Sigma}_i^{-1}\left(\boldsymbol{x_j}-\boldsymbol{\mu_i}\right)
\end{aligned}$$

设$B=\boldsymbol{\Sigma}_i^{-1}\left(\boldsymbol{x_j}-\boldsymbol{\mu_i}\right)$，则

$$\begin{aligned}
B^T&=\left(\boldsymbol{\Sigma}_i^{-1}\left(\boldsymbol{x_j}-\boldsymbol{\mu_i}\right)\right)^T\\
&=\left(\boldsymbol{x_j}-\boldsymbol{\mu_i}\right)^T\left(\boldsymbol{\Sigma}_i^{-1}\right)^T\\
&=\left(\boldsymbol{x_j}-\boldsymbol{\mu_i}\right)^T\boldsymbol{\Sigma}_i^{-1}
\end{aligned}$$

所以 $\frac{\partial}{\partial \Sigma_{i_{rc}}}(\boldsymbol{x_j}-\boldsymbol{\mu_i})^T\boldsymbol{\Sigma}_i^{-1}\left(\boldsymbol{x_j}-\boldsymbol{\mu_i}\right)=-B^T\frac{\partial \boldsymbol{\Sigma}_i}{\partial \Sigma_{i_{rc}}}B$ 其中$B$为$n\times 1$阶矩阵，$\frac{\partial \boldsymbol{\Sigma}_i}{\partial \Sigma_{i_{rc}}}$为$n$阶方阵，且$\frac{\partial \boldsymbol{\Sigma}_i}{\partial \Sigma_{i_{rc}}}$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

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

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

$$\begin{aligned}\frac{\partial}{\partial \boldsymbol{\Sigma}_i}\left(\boldsymbol{x_j}-\boldsymbol{\mu_i}\right)^T \boldsymbol{\Sigma}_i^{-1}\left(\boldsymbol{x_j}-\boldsymbol{\mu_i}\right) &= -B\cdot B^T \\ &= -\boldsymbol{\Sigma}_i^{-1}\left(\boldsymbol{x_j}-\boldsymbol{\mu_i}\right)\left(\boldsymbol{\Sigma}_i^{-1}\left(\boldsymbol{x_j}-\boldsymbol{\mu_i}\right)\right)^T \\ &= -\boldsymbol{\Sigma}_i^{-1}\left(\boldsymbol{x_j}-\boldsymbol{\mu_i}\right)\left(\boldsymbol{x_j}-\boldsymbol{\mu_i}\right)^T \boldsymbol{\Sigma}_i^{-1}\end{aligned}$$

因此最终结果为

$$\frac{\partial LL(D)}{\partial \boldsymbol{\Sigma}_i}=\sum_{j=1}^m \frac{\alpha_i\cdot p(\boldsymbol{x}_j|\boldsymbol{\mu}_i,\boldsymbol{\Sigma}_i)}{\sum_{l=1}^k \alpha_l\cdot p(\boldsymbol{x}_j,|\boldsymbol{\mu}_l,\boldsymbol{\Sigma}_l)}\left(-\frac{1}{2}\left(\boldsymbol{\Sigma}_i^{-1}-\boldsymbol{\Sigma}_i^{-1}\left(\boldsymbol{x_j}-\boldsymbol{\mu_i}\right)\left(\boldsymbol{x_j}-\boldsymbol{\mu_i}\right)^T\boldsymbol{\Sigma}_i^{-1}\right)\right)=0$$

整理可得

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

# 9.38

$$\alpha_i=\frac{1}{m}\sum_{j=1}^m \gamma_{ji}$$

[推导]：基于公式(9.37)进行恒等变形：

$$\sum_{j=1}^m \frac{p(\boldsymbol{x}_j|\boldsymbol{\mu_i},\boldsymbol{\Sigma_i})}{\sum_{l=1}^k \alpha_l\cdot p(\boldsymbol{x}_j|\boldsymbol{\mu_l},\boldsymbol{\Sigma_l})}+\lambda=0$$

$$\Rightarrow \sum_{j=1}^m \frac{\alpha_i\cdot p(\boldsymbol{x}_j|\boldsymbol{\mu_i},\boldsymbol{\Sigma_i})}{\sum_{l=1}^k \alpha_l\cdot p(\boldsymbol{x}_j|\boldsymbol{\mu_l},\boldsymbol{\Sigma_l})}+\alpha_i\lambda=0$$

对所有混合成分进行求和：

$$\Rightarrow \sum_{i=1}^k\left(\sum_{j=1}^m \frac{\alpha_i\cdot p(\boldsymbol{x}_j|\boldsymbol{\mu_i},\boldsymbol{\Sigma_i})}{\sum_{l=1}^k \alpha_l\cdot p(\boldsymbol{x}_j|\boldsymbol{\mu_l},\boldsymbol{\Sigma_l})}+\alpha_i\lambda\right)=0$$

$$\Rightarrow \sum_{i=1}^k\sum_{j=1}^m \frac{\alpha_i\cdot p(\boldsymbol{x}_j|\boldsymbol{\mu_i},\boldsymbol{\Sigma_i})}{\sum_{l=1}^k \alpha_l\cdot p(\boldsymbol{x}_j|\boldsymbol{\mu_l},\boldsymbol{\Sigma_l})}+\sum_{i=1}^k\alpha_i\lambda=0$$

$$\Rightarrow \lambda=-\sum_{i=1}^k\sum_{j=1}^m \frac{\alpha_i\cdot p(\boldsymbol{x}_j|\boldsymbol{\mu_i},\boldsymbol{\Sigma_i})}{\sum_{l=1}^k \alpha_l\cdot p(\boldsymbol{x}_j|\boldsymbol{\mu_l},\boldsymbol{\Sigma_l})}=-m$$

又

$$\sum_{j=1}^m \frac{\alpha_i\cdot p(\boldsymbol{x}_j|\boldsymbol{\mu_i},\boldsymbol{\Sigma_i})}{\sum_{l=1}^k \alpha_l\cdot p(\boldsymbol{x}_j|\boldsymbol{\mu_l},\boldsymbol{\Sigma_l})}+\alpha_i\lambda=0$$

$$\Rightarrow \sum_{j=1}^m \gamma_{ji}+\alpha_i\lambda=0$$

$$\Rightarrow \alpha_i=-\frac{\sum_{j=1}^m \gamma_{ji}}{\lambda}=\frac{1}{m}\sum_{j=1}^m \gamma_{ji}$$

# 附录

参考公式

$$\frac{\partial \boldsymbol{x}^T B \boldsymbol{x}}{\partial \boldsymbol{x}} = \left(B + B^T\right) \boldsymbol{x}$$

$$\frac{\partial}{\partial A} ln|A| = \left(A^{-1}\right)^T$$

$$\frac{\partial}{\partial x}\left(A^{-1}\right) = -A^{-1} \frac{\partial A}{\partial x} A^{-1}$$

参考资料

Petersen, K. B. & Pedersen, M. S. *The Matrix Cookbook*.
Bishop, C. M. (2006). *Pattern Recognition and Machine Learning*. Springer.