机器学习半监督学习作业

161910126 赵安

13.1

13.5

由高斯混合模型的定义式可知：

$$\mathrm{p}(\boldsymbol{x})=\sum_{\mathrm{i}=1}^{\mathrm{N}} \alpha_{\mathrm{i}} \cdot \mathrm{p}\left(\boldsymbol{x} \mid \boldsymbol{\mu} * \mathrm{i}, \boldsymbol{\Sigma} * \mathrm{i}\right)$$

因此：

$$\begin{aligned}
\mathrm{p}(\Theta=\mathrm{i} \mid \boldsymbol{x}) &= \frac{\mathrm{p}(\Theta=\mathrm{i}, \boldsymbol{x})}{\mathrm{P}(\mathrm{x})} \\
&= \frac{\alpha_{\mathrm{i}} \cdot \mathrm{p}\left(\boldsymbol{x} \mid \boldsymbol{\mu} * \mathrm{i}, \boldsymbol{\Sigma} * \mathrm{i}\right)}{\sum_{\mathrm{i}=1}^{\mathrm{N}} \alpha_{\mathrm{i}} \cdot \mathrm{p}\left(\boldsymbol{x} \mid \boldsymbol{\mu} * \mathrm{i}, \boldsymbol{\Sigma} * \mathrm{i}\right)}
\end{aligned}$$

又因为：

$$\lambda_{ij}=P(\Theta=i|x)$$

所以：

$$\lambda_{ij}=\frac{\alpha_{\mathrm{i}} \cdot \mathrm{p}\left(\boldsymbol{x} \mid \boldsymbol{\mu} * \mathrm{i}, \boldsymbol{\Sigma} * \mathrm{i}\right)}{\sum_{\mathrm{i}=1}^{\mathrm{N}} \alpha_{\mathrm{i}} \cdot \mathrm{p}\left(\boldsymbol{x} \mid \boldsymbol{\mu} * \mathrm{i}, \boldsymbol{\Sigma} * \mathrm{i}\right)}$$

13.6

令

$$\frac{\partial LL(D_l \cup D_u)}{\partial \mu_i}=0$$

其中

$$\mathrm{LL}\left(\mathrm{D} \cdot 1 \cup \mathrm{D} \cdot \mathrm{u}\right)=\sum_{(\mathrm{xj}, \mathrm{y} \cdot \mathrm{j}) \in \mathrm{D} \cdot 1} \ln \left(\sum_{\mathrm{i}=1}^{\mathrm{N}} \alpha_{\mathrm{i}} \cdot \mathrm{p}\left(\boldsymbol{x} \cdot \mathrm{j} \mid \boldsymbol{\mu} \cdot \mathrm{i}, \boldsymbol{\Sigma} \cdot \mathrm{i}\right) \cdot \mathrm{p}\left(\mathrm{y} \cdot \mathrm{j} \mid \Theta=\mathrm{i}, \boldsymbol{x}\mathrm{j}\right)\right)+\sum \cdot \mathrm{x} \cdot \mathrm{j} \in \mathrm{D} \cdot \mathbf{u} \ln \left(\sum_{\mathrm{i}=1}^{\mathrm{N}} \alpha_{\mathrm{i}} \cdot \mathrm{p}\left(\boldsymbol{x} \cdot \mathrm{j} \mid \boldsymbol{\mu} \cdot \mathrm{i}, \boldsymbol{\Sigma}_{\mathrm{i}}\right)\right)$$

将式$\mathrm{LL}\left(\mathrm{D} \cdot 1 \cup \mathrm{D} \cdot \mathrm{u}\right)$的两项分别记为：

$$\begin{aligned}
\mathrm{LL}\left(\mathrm{D} \cdot 1\right) &= \sum \cdot\left(x_j, y_{\mathrm{j}} \in \mathrm{D} \cdot 1\right) \ln \left(\sum \cdot \mathrm{s}=1^{\mathrm{N}} \alpha_{\mathrm{s}} \cdot \mathrm{p}\left(\boldsymbol{x} \cdot j \mid \boldsymbol{\mu} \cdot \mathrm{s}, \boldsymbol{\Sigma} \cdot \mathrm{s}\right) \cdot \mathrm{p}\left(\mathrm{y} \cdot \mathrm{i} \mid \boldsymbol{\Theta}=\mathrm{s}, \boldsymbol{x} \cdot j\right)\right) \\
&= \sum \cdot\left(x_j, \mathrm{yj} \in \mathrm{D} \cdot 1\right) \ln \left(\sum \cdot s=1^{\mathrm{N}} \alpha_{\mathrm{yj}} \cdot \mathrm{p}\left(\boldsymbol{x} \cdot j \mid \boldsymbol{\mu} \cdot \mathrm{yj}, \boldsymbol{\Sigma} \cdot \mathrm{y} \cdot \mathrm{j}\right)\right) \\
\mathrm{LL}\left(\mathrm{D} \cdot \mathrm{u}\right) &= \sum \cdot \boldsymbol{x} \cdot j \in \mathrm{D} \cdot \mathrm{u} \ln \left(\alpha_{\mathrm{s}} \cdot \mathrm{p}\left(\boldsymbol{x} \cdot j \mid \boldsymbol{\mu} \cdot \mathrm{s}, \boldsymbol{\Sigma}_{\mathrm{s}}\right)\right)
\end{aligned}$$

对第一项求导

$$\begin{aligned}
\frac{\partial \mathrm{LL}\left(\mathrm{D} \cdot \mathrm{l}\right)}{\partial \boldsymbol{\mu} \cdot \mathrm{i}} &= \sum_{(\boldsymbol{x} \cdot \mathrm{j}, \mathrm{y} \cdot \mathrm{j}) \in \mathrm{D} \cdot 1 \wedge \mathrm{y} \cdot \mathrm{j}=\mathrm{i}} \frac{\partial \ln \left(\alpha_{\mathrm{i}} \cdot \mathrm{p}\left(\boldsymbol{x} \cdot \mathrm{j} \mid \boldsymbol{\mu} \cdot \mathrm{i}, \boldsymbol{\Sigma} \cdot \mathrm{i}\right)\right)}{\partial \boldsymbol{\mu} \cdot \mathrm{i}} \\
&= \sum_{(\boldsymbol{x} \cdot \mathrm{j}, \mathrm{y} \cdot \mathrm{j}) \in \mathrm{D} \cdot 1 \wedge \mathrm{y} \cdot \mathrm{j}=\mathrm{i}} \frac{1}{\mathrm{p}\left(\boldsymbol{x} \cdot \mathrm{j} \mid \boldsymbol{\mu} \cdot \mathrm{i}, \boldsymbol{\Sigma} \cdot \mathrm{i}\right)} \cdot \frac{\partial \mathrm{p}\left(\boldsymbol{x} \cdot \mathrm{j} \mid \boldsymbol{\mu} \cdot \mathrm{i}, \boldsymbol{\Sigma} \cdot \mathrm{i}\right)}{\partial \boldsymbol{\mu}_{\mathrm{i}}}
\end{aligned}$$

其中

$$p\left(\boldsymbol{x} * j \mid \boldsymbol{\mu} * i, \boldsymbol{\Sigma} * i\right)=\frac{1}{(2 \pi)^{\frac{n}{2}}\left|\boldsymbol{\Sigma} * i\right|^{\frac{1}{2}}} e^{-\frac{1}{2}\left(\boldsymbol{x} * j-\boldsymbol{\mu} * i\right)^{\top} \boldsymbol{\Sigma} * i^{-1}\left(\boldsymbol{x} * j-\boldsymbol{\mu}_i\right)}$$

$$
\begin{aligned}
\frac{\partial p\left(\boldsymbol{x}\cdot j \mid \boldsymbol{\mu}\cdot i, \boldsymbol{\Sigma}\cdot i\right)}{\partial \boldsymbol{\mu}\cdot i} &= \frac{\partial \frac{1}{(2\pi)^{\frac{n}{2}}\left|\Sigma_i\right|^{\frac{1}{2}}} e^{-\frac{1}{2}(\boldsymbol{x}\cdot j-\boldsymbol{\mu}\cdot i)^{\top} \boldsymbol{\Sigma}\cdot i^{-1}(\boldsymbol{x}\cdot j-\boldsymbol{\mu}\cdot i)}}{\partial \boldsymbol{\mu}\cdot i} \\
&= \frac{1}{(2\pi)^{\frac{n}{2}}\left|\Sigma_i\right|^{\frac{1}{2}}} \cdot \frac{\partial e^{-\frac{1}{2}(\boldsymbol{x}\cdot j-\boldsymbol{\mu}\cdot i)^{\top} \boldsymbol{\Sigma}\cdot i^{-1}(\boldsymbol{x}*j-\boldsymbol{\mu}\cdot i)}}{\partial \boldsymbol{\mu}\cdot i} \\
&= \frac{1}{(2\pi)^{\frac{n}{2}}\left|\Sigma_i\right|^{\frac{1}{2}}} \cdot e^{-\frac{1}{2}(\boldsymbol{x}\cdot j-\boldsymbol{\mu}\cdot i)^{\top} \boldsymbol{\Sigma}\cdot i^{-1}(\boldsymbol{x}\cdot j-\boldsymbol{\mu}\cdot i)} \cdot -\frac{1}{2} \frac{\partial(\boldsymbol{x}\cdot j-\boldsymbol{\mu}\cdot i)^{\top} \boldsymbol{\Sigma}\cdot i^{-1}\left(\boldsymbol{x}\cdot j-\boldsymbol{\mu}\cdot i\right)}{\partial \boldsymbol{\mu}_i}
\end{aligned}
$$

其中协方差矩阵的逆矩阵$\boldsymbol{\Sigma}_i^{-1}$是对称阵，根据矩阵求导公式，当W为对称矩阵时，有

$$
\frac{\partial}{\partial \mathbf{s}}(\mathbf{x}-\mathbf{s})^{T} \mathbf{W}(\mathbf{x}-\mathbf{s})=-2 \mathbf{W}(\mathbf{x}-\mathbf{s})
$$

所以

$$
\begin{aligned}
\frac{\partial p\left(\boldsymbol{x}\cdot j \mid \boldsymbol{\mu}\cdot i, \boldsymbol{\Sigma}\cdot i\right)}{\partial \boldsymbol{\mu}\cdot i} &= \frac{1}{(2\pi)^{\frac{n}{2}}\left|\Sigma_i\right|^{\frac{1}{2}}} \cdot e^{-\frac{1}{2}(\boldsymbol{x}\cdot j-\boldsymbol{\mu}\cdot i)^{\top} \boldsymbol{\Sigma}\cdot i^{-1}(\boldsymbol{x}\cdot j-\boldsymbol{\mu}\cdot i)} \cdot -\frac{1}{2} \cdot 2 \boldsymbol{\Sigma}\cdot i^{-1}\left(\boldsymbol{\mu}\cdot i-\boldsymbol{x}\cdot j\right) \\
&= \frac{1}{(2\pi)^{\frac{n}{2}}\left|\Sigma_i\right|^{\frac{1}{2}}} \cdot e^{-\frac{1}{2}(\boldsymbol{x}\cdot j-\boldsymbol{\mu}\cdot i)^{\top} \boldsymbol{\Sigma}\cdot i^{-1}(\boldsymbol{x}\cdot j-\boldsymbol{\mu}\cdot i)} \cdot \boldsymbol{\Sigma}\cdot i^{-1}\left(\boldsymbol{x}\cdot i-\boldsymbol{\mu}\cdot j\right) \\
&= \mathrm{p}\left(\boldsymbol{x}\cdot \mathrm{j} \mid \boldsymbol{\mu}\cdot \mathrm{i}, \boldsymbol{\Sigma}\cdot \mathrm{i}\right) \cdot \boldsymbol{\Sigma}\cdot \mathrm{i}^{-1}\left(\boldsymbol{x}\cdot \mathrm{j}-\boldsymbol{\mu}\cdot \mathrm{i}\right)
\end{aligned}
$$

所以

$$
\begin{aligned}
\frac{\partial \mathrm{LL}\left(\mathrm{D}\cdot \mathrm{l}\right)}{\partial \boldsymbol{\mu}\cdot \mathrm{i}} &= \sum_{(\boldsymbol{x}\cdot \mathrm{j}, \mathrm{yj}) \in \mathrm{D}\cdot 1 \wedge \mathrm{yj}=\mathrm{i}} \frac{1}{\mathrm{p}\left(\boldsymbol{x}\cdot \mathrm{j} \mid \boldsymbol{\mu}\cdot \mathrm{i}, \boldsymbol{\Sigma}\cdot \mathrm{i}\right)} \cdot \mathrm{p}\left(\boldsymbol{x}\cdot \mathrm{j} \mid \boldsymbol{\mu}\cdot \mathrm{i}, \boldsymbol{\Sigma}\cdot \mathrm{i}\right) \cdot \boldsymbol{\Sigma}\cdot \mathrm{i}^{-1}\left(\boldsymbol{x}\cdot \mathrm{j}-\boldsymbol{\mu}\cdot \mathrm{i}\right) \\
&= \sum \cdot(\boldsymbol{x}\cdot \mathrm{j}, \mathrm{y}\cdot \mathrm{j}) \in \mathrm{D}\cdot 1 \wedge \mathrm{y}\cdot \mathrm{j}=\mathrm{i} \boldsymbol{\Sigma}\cdot \mathrm{i}^{-1}\left(\boldsymbol{x}\cdot \mathrm{j}-\boldsymbol{\mu}_{\mathrm{i}}\right)
\end{aligned}
$$

对第二项求导

$$
\begin{aligned}
\frac{\partial LL(D_u)}{\partial \boldsymbol{\mu}\cdot i} &= \frac{\partial LL(D\cdot u)}{\partial p\left(\boldsymbol{x}\cdot j \mid \boldsymbol{\mu}\cdot i, \boldsymbol{\Sigma}\cdot i\right)} \cdot \frac{\partial p\left(\boldsymbol{x}\cdot j \mid \boldsymbol{\mu}\cdot i, \boldsymbol{\Sigma}\cdot i\right)}{\partial \boldsymbol{\mu}\cdot i} \\
&= \frac{\partial \sum \cdot j=1^{m} \ln \left(\sum_{l=1}^{k} \alpha_l \cdot p\left(\boldsymbol{x}\cdot j \mid \boldsymbol{\mu}\cdot l, \boldsymbol{\Sigma}\cdot l\right)\right)}{\partial p\left(\boldsymbol{x}\cdot j \mid \boldsymbol{\mu}\cdot i, \boldsymbol{\Sigma}\cdot i\right)} \cdot \frac{\partial p\left(\boldsymbol{x}\cdot j \mid \boldsymbol{\mu}\cdot i, \boldsymbol{\Sigma}\cdot i\right)}{\partial \boldsymbol{\mu}\cdot i} \\
&= \sum_{j=1}^{m} \frac{\partial \ln \left(\sum_{l=1}^{k} \alpha_l \cdot p\left(\boldsymbol{x}\cdot j \mid \boldsymbol{\mu}\cdot l, \boldsymbol{\Sigma}\cdot l\right)\right)}{\partial p\left(\boldsymbol{x}\cdot j \mid \boldsymbol{\mu}\cdot i, \boldsymbol{\Sigma}\cdot i\right)} \cdot \frac{\partial p\left(\boldsymbol{x}\cdot j \mid \boldsymbol{\mu}\cdot i, \boldsymbol{\Sigma}\cdot i\right)}{\partial \boldsymbol{\mu}\cdot i} \\
&= \sum_{j=1}^{m} \frac{\alpha_i}{\sum_{l=1}^{k} \alpha_l \cdot p\left(\boldsymbol{x}\cdot j \mid \boldsymbol{\mu}\cdot l, \boldsymbol{\Sigma}\cdot l\right)} \cdot \frac{\partial p\left(\boldsymbol{x}\cdot \mid \boldsymbol{\mu}\cdot i, \boldsymbol{\Sigma}\cdot i\right)}{\partial \boldsymbol{\mu}_i}
\end{aligned}
$$

因为

$$
\frac{\partial p\left(\boldsymbol{x}\cdot j \mid \boldsymbol{\mu}\cdot i, \boldsymbol{\Sigma}\cdot i\right)}{\partial \boldsymbol{\mu}\cdot i}=\mathrm{p}\left(\boldsymbol{x}\cdot \mathrm{j} \mid \boldsymbol{\mu}\cdot \mathrm{i}, \boldsymbol{\Sigma}\cdot \mathrm{i}\right) \cdot \boldsymbol{\Sigma}\cdot \mathrm{i}^{-1}\left(\boldsymbol{x}\cdot \mathrm{j}-\boldsymbol{\mu}\cdot \mathrm{i}\right)
$$

所以

$$
\begin{aligned}
\frac{\partial LL(D_u)}{\partial \boldsymbol{\mu}\cdot i} &= \sum \cdot j=1^{m} \frac{\alpha_i}{\sum_{l=1}^{N} \alpha_l \cdot p\left(\boldsymbol{x}\cdot j \mid \boldsymbol{\mu}\cdot l, \boldsymbol{\Sigma}\cdot l\right)} \cdot \frac{\partial p\left(\boldsymbol{x}\cdot j \mid \boldsymbol{\mu}\cdot i, \boldsymbol{\Sigma}\cdot i\right)}{\partial \boldsymbol{\mu}\cdot i} \\
&= \sum \cdot j=1^{m} \frac{\alpha_i}{\sum_{l=1}^{N} \alpha_l \cdot p\left(\boldsymbol{x}\cdot j \mid \boldsymbol{\mu}\cdot l, \boldsymbol{\Sigma}\cdot l\right)} \cdot \mathrm{p}\left(\boldsymbol{x}\cdot \mathrm{j} \mid \boldsymbol{\mu}\cdot \mathrm{i}, \boldsymbol{\Sigma}\cdot \mathrm{i}\right) \cdot \boldsymbol{\Sigma}\cdot \mathrm{i}^{-1}\left(\boldsymbol{x}\cdot \mathrm{j}-\boldsymbol{\mu}\cdot \mathrm{i}\right) \\
&= \sum \cdot x_{\mathrm{j}} \in \mathrm{D}\cdot \mathrm{u} \gamma * \mathrm{ji} \cdot \boldsymbol{\Sigma}\cdot \mathrm{i}^{-1}\left(\boldsymbol{x}\cdot \mathrm{j}-\boldsymbol{\mu}_{\mathrm{i}}\right)
\end{aligned}
$$

因此

$$
\begin{aligned}
\frac{\partial \mathrm{LL}(\mathrm{D}\cdot 1\cup \mathrm{D}\cdot \mathrm{u})}{\partial \boldsymbol{\mu}\cdot \mathrm{i}} &= \sum \cdot(\boldsymbol{x}\cdot \mathrm{j}, \mathrm{y}\cdot \mathrm{j}) \in \mathrm{D}\cdot 1 \wedge \mathrm{y}\cdot \mathrm{j} = \mathrm{i}\boldsymbol{\Sigma}\cdot \mathrm{i}^{-1}(\boldsymbol{x}\cdot \mathrm{j}-\boldsymbol{\mu}\cdot \mathrm{i}) + \sum \cdot \boldsymbol{x}\cdot \mathrm{j} \in \mathrm{D}\cdot \mathrm{u}\gamma_{\mathrm{ji}}\cdot \boldsymbol{\Sigma}\cdot \mathrm{i}^{-1}(\boldsymbol{x}\cdot \mathrm{j}-\boldsymbol{\mu}\cdot \mathrm{i}) \\
&= \boldsymbol{\Sigma}\cdot \mathrm{i}^{-1}\left(\sum_{(x_{\mathrm{j}},\mathrm{y}\cdot \mathrm{j})\in \mathrm{D}\cdot 1\wedge \mathrm{y}\cdot \mathrm{j}=\mathrm{i}} (\boldsymbol{x}\cdot \mathrm{j}-\boldsymbol{\mu}\cdot \mathrm{i}) + \sum \cdot \boldsymbol{x}\cdot \mathrm{j} \in \mathrm{D}\cdot \mathrm{u}\gamma_{\mathrm{ji}}\cdot (\boldsymbol{x}\cdot \mathrm{j}-\boldsymbol{\mu}\cdot \mathrm{i})\right) \\
&= \boldsymbol{\Sigma}\cdot \mathrm{i}^{-1}\left(\sum \cdot(x_{\mathrm{j}}, \mathrm{y}\cdot \mathrm{j}) \in \mathrm{D}\cdot 1 \wedge \mathrm{y}\cdot \mathrm{j} = \mathrm{i}\boldsymbol{x}\cdot \mathrm{j} + \sum_{\boldsymbol{x}\cdot \mathrm{j}\in \mathrm{D}\cdot \mathrm{u}} \gamma_{\mathrm{ji}}\cdot \boldsymbol{x}\cdot \mathrm{j} - \sum \cdot(\boldsymbol{x}\cdot \mathrm{j}, \mathrm{y}\cdot \mathrm{j}) \in \mathrm{D}\cdot 1 \wedge \mathrm{y}\cdot \mathrm{j} = \mathrm{i}\boldsymbol{\mu}\cdot \mathrm{i} - \sum \cdot \boldsymbol{x}\cdot \mathrm{j} \in \mathrm{D}\cdot \mathrm{u}\gamma_{\mathrm{ji}}\cdot \boldsymbol{\mu}_{\mathrm{i}}\right) \\
&= 0
\end{aligned}
$$

所以

$$
\sum_{x_{\mathrm{j}}\in \mathrm{D}\cdot \mathrm{u}} \gamma\cdot \mathrm{ji}\cdot \boldsymbol{\mu}\cdot \mathrm{i} + \sum \cdot(x_{\mathrm{j}}, \mathrm{y}\cdot \mathrm{j}) \in \mathrm{D}\cdot 1 \wedge \mathrm{yj} = \mathrm{i}\boldsymbol{\mu}\cdot \mathrm{i} = \sum \cdot x_{\mathrm{j}} \in \mathrm{D}\cdot \mathrm{u}\gamma\cdot \mathrm{ji}\cdot \boldsymbol{x}\cdot \mathrm{j} + \sum \cdot(x_{\mathrm{j}}, \mathrm{yj}) \in \mathrm{D}\cdot 1 \wedge \mathrm{yj}\cdot \mathrm{j} = \mathrm{i}\boldsymbol{x}_{\mathrm{j}}
$$

$$
\left(\sum_{x_j\in D_u} \gamma_{ji} + \sum_{(x_j,y_j)\in D_1\wedge y_j=i} 1\right)\mu_{\mathrm{i}} = \sum_{x_j\in D_u} \gamma_{ji}\cdot x_j + \sum_{(x_j,y_j)\in D_1\wedge y_j=i} x_j
$$

因为

$$
\sum_{(\mathrm{x}\cdot \mathrm{j},\mathrm{yj}\cdot \mathrm{j})\in \mathrm{D}\cdot 1\wedge \mathrm{yj}=\mathrm{i}} 1 = \mathrm{l}\cdot \mathrm{i}
$$

所以

$$
\boldsymbol{\mu}\cdot \mathrm{i} = \frac{1}{\sum \cdot x_{\mathrm{j}} \in \mathrm{D}\cdot \mathrm{u}\gamma\cdot \mathrm{ji} + l_{\mathrm{i}}}\left(\sum_{x_{\mathrm{j}}\in \mathrm{D}\cdot \mathrm{u}} \gamma\cdot \mathrm{ji}\boldsymbol{x}\cdot \mathrm{j} + \sum \cdot(\boldsymbol{x}\cdot \mathrm{j}, \mathrm{yj}) \in \mathrm{D}\cdot 1 \wedge \mathrm{yj} = \mathrm{i}\boldsymbol{x}_{\mathrm{j}}\right)
$$

13.7

$LL(D_1)$对$\sum_i$求偏导:

$$
\begin{aligned}
\frac{\partial \mathrm{LL}(\mathrm{D}\cdot 1)}{\partial \boldsymbol{\Sigma}\cdot \mathrm{i}} &= \sum_{(x_{\mathrm{j}},y_{\mathrm{j}})\in \mathrm{D}\cdot 1\wedge \mathrm{y}\cdot \mathrm{j}=\mathrm{i}} \frac{\partial \ln(\alpha_{\mathrm{i}}\cdot \mathrm{p}(\boldsymbol{x}\cdot \mathrm{j}\mid \boldsymbol{\mu}\cdot \mathrm{i}, \boldsymbol{\Sigma}\cdot \mathrm{i}))}{\partial \boldsymbol{\Sigma}\cdot \mathrm{i}} \\
&= \sum_{(x_{\mathrm{j}},\mathrm{yj})\in \mathrm{D}\cdot 1\wedge y\cdot \mathrm{j}=\mathrm{i}} \frac{1}{\mathrm{p}(\boldsymbol{x}\cdot \mathrm{j}\mid \boldsymbol{\mu}\cdot \mathrm{i}, \boldsymbol{\Sigma}\cdot \mathrm{i})}\cdot \frac{\partial \mathrm{p}(\boldsymbol{x}\cdot \mathrm{j}\mid \boldsymbol{\mu}\cdot \mathrm{i}, \boldsymbol{\Sigma}\cdot \mathrm{i})}{\partial \boldsymbol{\Sigma}_{\mathrm{i}}}
\end{aligned}
$$

其中

$$
\begin{aligned}
\frac{\partial \mathrm{p}(\boldsymbol{x}\cdot \mathrm{j}\mid \boldsymbol{\mu}\cdot \mathrm{i}, \boldsymbol{\Sigma}\cdot \mathrm{i})}{\partial \boldsymbol{\Sigma}\cdot \mathrm{i}} &= \frac{\partial}{\partial \boldsymbol{\Sigma}\cdot \mathrm{i}}\left[\frac{1}{(2\pi)^{\frac{n}{2}}|\boldsymbol{\Sigma}\cdot \mathrm{i}|^{\frac{1}{2}}\ \exp}\left(-\frac{1}{2}(\boldsymbol{x}\cdot \mathrm{j}-\boldsymbol{\mu}\cdot \mathrm{i})^{\mathrm{T}}\boldsymbol{\Sigma}\cdot \mathrm{i}^{-1}(\boldsymbol{x}\cdot \mathrm{j}-\boldsymbol{\mu}\cdot \mathrm{i})\right)\right] \\
&= \frac{\partial}{\partial \boldsymbol{\Sigma}\cdot \mathrm{i}}\left\{\exp\left[\ln\left(\frac{1}{(2\pi)^{\frac{n}{2}}|\boldsymbol{\Sigma}\cdot \mathrm{i}|^{\frac{1}{2}}}\exp\left(-\frac{1}{2}(\boldsymbol{x}\cdot \mathrm{j}-\boldsymbol{\mu}\cdot \mathrm{i})^{\mathrm{T}}\boldsymbol{\Sigma}\cdot \mathrm{i}^{-1}(\boldsymbol{x}\cdot \mathrm{j}-\boldsymbol{\mu}\cdot \mathrm{i})\right)\right)\right]\right\} \\
&= \mathrm{p}(\boldsymbol{x}\cdot \mathrm{j}\mid \boldsymbol{\mu}\cdot \mathrm{i}, \boldsymbol{\Sigma}\cdot \mathrm{i})\cdot \frac{\partial}{\partial \boldsymbol{\Sigma}\cdot \mathrm{i}}\left[\ln\left(\frac{1}{(2\pi)^{\frac{n}{2}}|\boldsymbol{\Sigma}\cdot \mathrm{i}|^{\frac{1}{2}}}\exp\left(-\frac{1}{2}(\boldsymbol{x}\cdot \mathrm{j}-\boldsymbol{\mu}\cdot \mathrm{i})^{\mathrm{T}}\boldsymbol{\Sigma}\cdot \mathrm{i}^{-1}(\boldsymbol{x}\cdot \mathrm{j}-\boldsymbol{\mu}\cdot \mathrm{i})\right)\right)\right] \\
&= \mathrm{p}(\boldsymbol{x}\cdot \mathrm{j}\mid \boldsymbol{\mu}\cdot \mathrm{i}, \boldsymbol{\Sigma}\cdot \mathrm{i})\cdot \frac{\partial}{\partial \boldsymbol{\Sigma}\cdot \mathrm{i}}\left[\ln\frac{1}{(2\pi)^{\frac{n}{2}}} - \frac{1}{2}\ln|\boldsymbol{\Sigma}\cdot \mathrm{i}| - \frac{1}{2}(\boldsymbol{x}\cdot \mathrm{j}-\boldsymbol{\mu}\cdot \mathrm{i})^{\mathrm{T}}\boldsymbol{\Sigma}\cdot \mathrm{i}^{-1}(\boldsymbol{x}\cdot \mathrm{j}-\boldsymbol{\mu}\cdot \mathrm{i})\right] \\
&= \mathrm{p}(\boldsymbol{x}\cdot \mathrm{j}\mid \boldsymbol{\mu}\cdot \mathrm{i}, \boldsymbol{\Sigma}\cdot \mathrm{i})\cdot \left[-\frac{1}{2}\frac{\partial(\ln|\boldsymbol{\Sigma}\cdot \mathrm{i}|)}{\partial \boldsymbol{\Sigma}\cdot \mathrm{i}} - \frac{1}{2}\frac{\partial\left[(\boldsymbol{x}\cdot \mathrm{j}-\boldsymbol{\mu}\cdot \mathrm{i})^{\mathrm{T}}\boldsymbol{\Sigma}\cdot \mathrm{i}^{-1}(\boldsymbol{x}\cdot \mathrm{j}-\boldsymbol{\mu}\cdot \mathrm{i})\right]}{\partial \boldsymbol{\Sigma}_{\mathrm{i}}}\right]
\end{aligned}
$$

又因为

$$
\frac{\partial|\mathbf{X}|}{\partial \mathbf{X}} = |\mathbf{X}|\cdot\left(\mathbf{X}^{-1}\right)^{\mathrm{T}}, \frac{\partial \boldsymbol{a}^{\mathrm{T}}\mathbf{X}^{-1}\boldsymbol{b}}{\partial \mathbf{X}} = -\mathbf{X}^{-\mathrm{T}}\boldsymbol{a}\boldsymbol{b}^{\mathrm{T}}\mathbf{X}^{-\mathrm{T}}
$$

$$
\frac{\partial}{\partial \boldsymbol{\Sigma}\cdot \mathrm{i}}(\mathrm{p}(\boldsymbol{x}\cdot \mathrm{j}\mid \boldsymbol{\mu}\cdot \mathrm{i}, \boldsymbol{\Sigma}\cdot \mathrm{i})) = \mathrm{p}(\boldsymbol{x}\cdot \mathrm{j}\mid \boldsymbol{\mu}\cdot \mathrm{i}, \boldsymbol{\Sigma}\cdot \mathrm{i})\cdot\left[-\frac{1}{2}\boldsymbol{\Sigma}\cdot \mathrm{i}^{-1} + \frac{1}{2}\boldsymbol{\Sigma}\cdot \mathrm{i}^{-1}(\boldsymbol{x}\cdot \mathrm{j}-\boldsymbol{\mu}\cdot \mathrm{i})(\boldsymbol{x}\cdot \mathrm{j}-\boldsymbol{\mu}\cdot \mathrm{i})^{\mathrm{T}}\boldsymbol{\Sigma}\cdot \mathrm{i}^{-1}\right]
$$

所以

$$\frac{\partial \mathrm{LL}(\mathrm{D}\cdot 1)}{\partial \boldsymbol{\Sigma}\cdot \mathrm{i}}=\sum_{(x_{\mathrm{j}},\mathrm{y}\cdot \mathrm{j})\in \mathrm{D}\cdot 1\wedge \mathrm{y}\cdot \mathrm{j}=\mathrm{i}}\frac{1}{\mathrm{p}\left(\boldsymbol{x}\cdot \mathrm{j}\mid \boldsymbol{\mu}\cdot \mathrm{i},\boldsymbol{\Sigma}\cdot \mathrm{i}\right)}\cdot \mathrm{p}\left(\boldsymbol{x}\cdot \mathrm{j}\mid \boldsymbol{\mu}\cdot \mathrm{i},\boldsymbol{\Sigma}\cdot \mathrm{i}\right)\cdot\left(\boldsymbol{\Sigma}\cdot \mathrm{i}^{-1}\left(\boldsymbol{x}\cdot \mathrm{j}-\boldsymbol{\mu}\cdot \mathrm{i}\right)\left(\boldsymbol{x}\cdot \mathrm{j}-\boldsymbol{\mu}\cdot \mathrm{i}\right)^{\top}-\boldsymbol{I}\right)\cdot\frac{1}{2}\boldsymbol{\Sigma}\cdot \mathrm{i}^{-1}$$
$$=\sum\cdot(x_{\mathrm{j}},\mathrm{y}\cdot \mathrm{j})\in \mathrm{D}\cdot 1\wedge y_j=\mathrm{i}\left(\boldsymbol{\Sigma}\cdot \mathrm{i}^{-1}\left(\boldsymbol{x}\cdot \mathrm{j}-\boldsymbol{\mu}\cdot \mathrm{i}\right)\left(\boldsymbol{x}\cdot \mathrm{j}-\boldsymbol{\mu}\cdot \mathrm{i}\right)^{\top}-\boldsymbol{I}\right)\cdot\frac{1}{2}\boldsymbol{\Sigma}\cdot \mathrm{i}^{-1}$$

$LL(D_u)$对$\sum_i$求偏导:

$$\frac{\partial LL(D_u)}{\partial \boldsymbol{\Sigma}\cdot \mathrm{i}}=\sum\cdot j=1^m\frac{\alpha_i}{\sum_{l=1}^{N}\alpha_l\cdot p\left(\boldsymbol{x}\cdot j\mid \boldsymbol{\mu}\cdot l,\boldsymbol{\Sigma}\cdot l\right)}\cdot\frac{\partial p\left(\boldsymbol{x}\cdot j\mid \boldsymbol{\mu}\cdot i,\boldsymbol{\Sigma}\cdot i\right)}{\partial \boldsymbol{\Sigma}_{\mathrm{i}}}$$

因为

$$\frac{\partial}{\partial \boldsymbol{\Sigma}\cdot \mathrm{i}}\left(\mathrm{p}\left(\boldsymbol{x}\cdot \mathrm{j}\mid \boldsymbol{\mu}\cdot \mathrm{i},\boldsymbol{\Sigma}\cdot \mathrm{i}\right)\right)=\mathrm{p}\left(\boldsymbol{x}\cdot \mathrm{j}\mid \boldsymbol{\mu}\cdot \mathrm{i},\boldsymbol{\Sigma}\cdot \mathrm{i}\right)\cdot\left[-\frac{1}{2}\boldsymbol{\Sigma}\cdot \mathrm{i}^{-1}+\frac{1}{2}\boldsymbol{\Sigma}\cdot \mathrm{i}^{-1}\left(\boldsymbol{x}\cdot \mathrm{j}-\boldsymbol{\mu}\cdot \mathrm{i}\right)\left(\boldsymbol{x}\cdot \mathrm{j}-\boldsymbol{\mu}\cdot \mathrm{i}\right)^{\mathrm{T}}\boldsymbol{\Sigma}\cdot \mathrm{i}^{-1}\right]$$

所以

$$\frac{\partial \mathrm{LL}(\mathrm{D_u})}{\partial \boldsymbol{\Sigma}\cdot \mathrm{i}}=\sum\cdot \mathrm{j}=1^{\mathrm{m}}\frac{\alpha_{\mathrm{i}}\cdot \mathrm{p}\left(\boldsymbol{x}\cdot \mathrm{j}\mid \boldsymbol{\mu}\cdot \mathrm{i},\boldsymbol{\Sigma}\cdot \mathrm{i}\right)}{\sum\cdot \mathrm{l}=1^{\mathrm{k}}\alpha_1\cdot \mathrm{p}\left(\boldsymbol{x}\cdot \mathrm{j}\mid \boldsymbol{\mu}\cdot \mathrm{l},\boldsymbol{\Sigma}\cdot 1\right)}\cdot\left[-\frac{1}{2}\boldsymbol{\Sigma}\cdot \mathrm{i}^{-1}+\frac{1}{2}\boldsymbol{\Sigma}\cdot \mathrm{i}^{-1}\left(\boldsymbol{x}\cdot \mathrm{j}-\boldsymbol{\mu}\cdot \mathrm{i}\right)\left(\boldsymbol{x}\cdot \mathrm{j}-\boldsymbol{\mu}\cdot \mathrm{i}\right)^{\mathrm{T}}\boldsymbol{\Sigma}\cdot \mathrm{i}^{-1}\right]$$

所以

$$\frac{\partial \mathrm{LL}(\mathrm{D}\cdot \mathrm{u})}{\partial \boldsymbol{\Sigma}\cdot \mathrm{i}}=\sum_{x_{\mathrm{j}}\in \mathrm{D}\cdot \mathrm{u}}\gamma\cdot \mathrm{ji}\cdot\left(\boldsymbol{\Sigma}\cdot \mathrm{i}^{-1}\left(\boldsymbol{x}\cdot \mathrm{j}-\boldsymbol{\mu}\cdot \mathrm{i}\right)\left(\boldsymbol{x}\cdot \mathrm{j}-\boldsymbol{\mu}\cdot \mathrm{i}\right)^{\top}-\boldsymbol{I}\right)\cdot\frac{1}{2}\boldsymbol{\Sigma}\cdot \mathrm{i}^{-1}$$

$$\frac{\partial \mathrm{LL}(\mathrm{D}\cdot 1\cup \mathrm{D}\cdot \mathrm{u})}{\partial \boldsymbol{\Sigma}\cdot \mathrm{i}}=\sum\cdot x_{\mathrm{i}}\in \mathrm{D}\cdot \mathrm{u}\gamma\cdot \mathrm{ji}\cdot\left(\boldsymbol{\Sigma}\cdot \mathrm{i}^{-1}\left(\boldsymbol{x}\cdot \mathrm{j}-\boldsymbol{\mu}\cdot \mathrm{i}\right)\left(\boldsymbol{x}\cdot \mathrm{j}-\boldsymbol{\mu}\cdot \mathrm{i}\right)^{\top}-\boldsymbol{I}\right)\cdot\frac{1}{2}\boldsymbol{\Sigma}\cdot \mathrm{i}^{-1}$$
$$+\sum_{(\boldsymbol{x}\cdot \mathrm{i},y\cdot \mathrm{j})\in \mathrm{D}\cdot \mathrm{i}\wedge \mathrm{yj}=\mathrm{i}}\left(\boldsymbol{\Sigma}\cdot \mathrm{i}^{-1}\left(\boldsymbol{x}\cdot \mathrm{j}-\boldsymbol{\mu}\cdot \mathrm{i}\right)\left(\boldsymbol{x}\cdot \mathrm{j}-\boldsymbol{\mu}\cdot \mathrm{i}\right)^{\top}-\boldsymbol{I}\right)\cdot\frac{1}{2}\boldsymbol{\Sigma}\cdot \mathrm{i}^{-1}$$
$$=\left(\sum\cdot \boldsymbol{x}\boldsymbol{j}\in \mathrm{D}\cdot \mathrm{u}\gamma\cdot \mathrm{ji}\cdot\left(\boldsymbol{\Sigma}\cdot \mathrm{i}^{-1}\left(\boldsymbol{x}\cdot \mathrm{j}-\boldsymbol{\mu}\cdot \mathrm{i}\right)\left(\boldsymbol{x}\cdot \mathrm{j}-\boldsymbol{\mu}\cdot \mathrm{i}\right)^{\top}-\boldsymbol{I}\right)\right.$$
$$\left.+\sum\cdot(\boldsymbol{x}\mathrm{j},\mathrm{y}\cdot \mathrm{j})\in \mathrm{D}\backslash\wedge \mathrm{y}\cdot \mathrm{i}=\mathrm{i}\left(\boldsymbol{\Sigma}\cdot \mathrm{i}^{-1}\left(\boldsymbol{x}\cdot \mathrm{j}-\boldsymbol{\mu}\cdot \mathrm{i}\right)\left(\boldsymbol{x}\cdot \mathrm{j}-\boldsymbol{\mu}\cdot \mathrm{i}\right)^{\top}-\boldsymbol{I}\right)\right)\cdot\frac{1}{2}\boldsymbol{\Sigma}\cdot \mathrm{i}^{-1}=0$$

所以

$$\sum_{x_j\in D_u}\gamma_{ji}\cdot\Sigma_i^{-1}\left(x_j-\mu_i\right)\left(x_j-\mu_i\right)^{\top}+\sum_{(x_j,yj\in Dl\wedge yj=i}\Sigma_i^{-1}\left(x_j-\mu_i\right)\left(x_j-\mu_i\right)^{\top}$$
$$=\sum_{x_j\in D_u}\gamma_{ji}\cdot I+\sum_{(x_j,y_j)\in D_1\wedge y_j=i}I$$
$$=\left(\sum_{x_j\in D_u}\gamma_{ji}+l_i\right)I$$

$$\sum_{x_j\in D_u}\gamma_{ji}\cdot\left(x_j-\mu_i\right)\left(x_j-\mu_i\right)^{\top}+\sum_{(x_j,y_j)\in D_1\wedge y_j=i}\left(x_j-\mu_i\right)\left(x_j-\mu_i\right)^{\top}=\left(\sum_{x_j\in D_u}\gamma_{ji}+l_i\right)\boldsymbol{\Sigma}_{\mathrm{i}}$$

所以

$$\boldsymbol{\Sigma}\cdot \mathrm{i}=\frac{1}{\sum\cdot x\mathrm{j}\in \mathrm{Du}\gamma_{\mathrm{ji}}+\mathrm{l}\cdot \mathrm{i}}\left(\sum\cdot x_{\mathrm{j}}\in \mathrm{D}\cdot \mathrm{u}\gamma\cdot \mathrm{ji}\cdot\left(\boldsymbol{x}\cdot \mathrm{j}-\boldsymbol{\mu}\cdot \mathrm{i}\right)\left(\boldsymbol{x}\cdot \mathrm{j}-\boldsymbol{\mu}\cdot \mathrm{i}\right)^{\top}+\sum_{(\boldsymbol{x}\cdot \mathrm{j},\mathrm{y}\cdot \mathrm{j})\in \mathrm{D}\cdot 1\wedge \mathrm{y}\cdot \mathrm{j}=\mathrm{i}}\left(\boldsymbol{x}\cdot \mathrm{j}-\boldsymbol{\mu}\cdot \mathrm{i}\right)\left(\boldsymbol{x}\cdot \mathrm{j}-\boldsymbol{\mu}\cdot \mathrm{i}\right)^{\top}\right)$$

13.8

$LL(D_l\cup D_u)$的拉格朗日形式

$$\mathcal{L}\left(\mathrm{D}\cdot 1\cup \mathrm{D}\cdot \mathrm{u},\lambda\right)=\mathrm{LL}\left(\mathrm{D}\cdot 1\cup \mathrm{D}\cdot \mathrm{u}\right)+\lambda\left(\sum_{\mathrm{s}=1}^{\mathrm{N}}\alpha_{\mathrm{s}}-1\right)=\mathrm{LL}\left(\mathrm{D}\cdot 1\right)+\mathrm{LL}\left(\mathrm{D}\cdot \mathrm{u}\right)+\lambda\left(\sum_{\mathrm{s}=1}^{\mathrm{N}}\alpha_{\mathrm{s}}-1\right)$$

$LL(D_u)$对α_i求偏导

$$\frac{\partial \mathrm{LL}(\mathrm{D}\cdot\mathrm{u})}{\partial\alpha\cdot\mathrm{i}} = \sum_{x_{\mathrm{j}}\in\mathrm{D}\cdot\mathrm{u}} \frac{1}{\sum\cdot\mathrm{s}=1^{\mathrm{N}}\alpha_{\mathrm{s}}\cdot\mathrm{p}\left(\boldsymbol{x}\cdot\mathrm{j}\mid\boldsymbol{\mu}\cdot\mathrm{s},\boldsymbol{\Sigma}\cdot\mathrm{s}\right)}\cdot\mathrm{p}\left(\boldsymbol{x}\cdot\mathrm{j}\mid\boldsymbol{\mu}\cdot\mathrm{i},\boldsymbol{\Sigma}\cdot\mathrm{i}\right)$$

$LL(D_l)$对α_i求偏导

$$\begin{aligned}
\frac{\partial \mathrm{LL}(\mathrm{D}\cdot\mathrm{l})}{\partial\alpha\cdot\mathrm{i}} &= \sum_{(\mathrm{x}\cdot\mathrm{j},\mathrm{y}\cdot\mathrm{j})\in\mathrm{D}\cdot 1\wedge\mathrm{y}\cdot\mathrm{j}=\mathrm{i}} \frac{\partial\ln(\alpha_{\mathrm{i}}\cdot\mathrm{p}\left(\boldsymbol{x}\cdot\mathrm{j}\mid\boldsymbol{\mu}\cdot\mathrm{i},\boldsymbol{\Sigma}\cdot\mathrm{i}\right))}{\partial\alpha\cdot\mathrm{i}} \\
&= \sum_{(\boldsymbol{x}\cdot\mathrm{j},\mathrm{y}\cdot\mathrm{j})\in\mathrm{D}\cdot 1\wedge\mathrm{yj}=\mathrm{i}} \frac{1}{\alpha\cdot\mathrm{i}\cdot\mathrm{p}\left(\boldsymbol{x}\cdot\mathrm{j}\mid\boldsymbol{\mu}\cdot\mathrm{i},\boldsymbol{\Sigma}\cdot\mathrm{i}\right)}\cdot\frac{\partial\left(\alpha\cdot\mathrm{i}\cdot\mathrm{p}\left(\boldsymbol{x}\cdot\mathrm{j}\mid\boldsymbol{\mu}\cdot\mathrm{i},\boldsymbol{\Sigma}\cdot\mathrm{i}\right)\right)}{\partial\alpha\cdot\mathrm{i}} \\
&= \sum_{(\boldsymbol{x}\cdot\mathrm{j},\mathrm{y}\cdot\mathrm{j})\in\mathrm{D}\cdot 1\wedge\mathrm{y}\cdot\mathrm{j}=\mathrm{i}} \frac{1}{\alpha_{\mathrm{i}}\cdot\mathrm{p}\left(\boldsymbol{x}\cdot\mathrm{j}\mid\boldsymbol{\mu}\cdot\mathrm{i},\boldsymbol{\Sigma}\cdot\mathrm{i}\right)}\cdot\mathrm{p}\left(\boldsymbol{x}\cdot\mathrm{j}\mid\boldsymbol{\mu}\cdot\mathrm{i},\boldsymbol{\Sigma}\cdot\mathrm{i}\right) \\
&= \sum_{(\boldsymbol{x}\cdot\mathrm{j},\mathrm{y}\cdot\mathrm{j})\in\mathrm{D}\cdot 1\wedge\mathrm{y}\cdot\mathrm{j}=\mathrm{i}} \frac{1}{\alpha_{\mathrm{i}}} = \frac{1}{\alpha_{\mathrm{i}}}\cdot\sum_{(\boldsymbol{x}\cdot\mathrm{j},\mathrm{y}\cdot\mathrm{j})\in\mathrm{D}\cdot 1\wedge\mathrm{y}\cdot\mathrm{j}=\mathrm{i}} 1 = \frac{\mathrm{l}\cdot\mathrm{i}}{\alpha\cdot\mathrm{i}}
\end{aligned}$$

$\alpha_i, p(x_j|\mu_i,\Sigma_i)$是常量，所以

$$\frac{\partial\mathcal{L}\left(\mathrm{D}\cdot 1\cup\mathrm{D}\cdot\mathrm{u},\lambda\right)}{\partial\alpha_{\mathrm{i}}} = \frac{\mathrm{l}\cdot\mathrm{i}}{\alpha\cdot\mathrm{i}} + \sum_{x_{\mathrm{j}}\in\mathrm{D}\cdot\mathrm{u}} \frac{\mathrm{p}\left(\boldsymbol{x}\cdot\mathrm{j}\mid\boldsymbol{\mu}\cdot\mathrm{i},\boldsymbol{\Sigma}\cdot\mathrm{i}\right)}{\sum_{\mathrm{s}=1}^{\mathrm{N}}\alpha_{\mathrm{s}}\cdot\mathrm{p}\left(\boldsymbol{x}\cdot\mathrm{j}\mid\boldsymbol{\mu}\cdot\mathrm{s},\boldsymbol{\Sigma}_{\mathrm{s}}\right)} + \lambda = 0$$

两边同乘以α_i

$$\alpha_{\mathrm{i}}\cdot\frac{\mathrm{l}\cdot\mathrm{i}}{\alpha\cdot\mathrm{i}} + \sum_{x_{\mathrm{j}}\in\mathrm{D}\cdot\mathrm{u}} \frac{\alpha\cdot\mathrm{i}\cdot\mathrm{p}\left(\boldsymbol{x}\cdot\mathrm{j}\mid\boldsymbol{\mu}\cdot\mathrm{i},\boldsymbol{\Sigma}\cdot\mathrm{i}\right)}{\sum\cdot\mathrm{s}=1^{\mathrm{N}}\alpha_{\mathrm{s}}\cdot\mathrm{p}\left(\boldsymbol{x}\cdot\mathrm{j}\mid\boldsymbol{\mu}\cdot\mathrm{s},\boldsymbol{\Sigma}\cdot\mathrm{s}\right)} + \lambda\cdot\alpha\cdot\mathrm{i} = 0$$
$$\mathrm{l}\cdot\mathrm{i} + \sum\cdot\mathrm{x}\cdot\mathrm{i}\in\mathrm{D}\cdot\mathrm{u}\gamma_{\mathrm{ji}} + \lambda\alpha_{\mathrm{i}} = 0$$

对所有混合成分求和

$$\sum_{i=1}^{N} l_i + \sum_{i=1}^{N}\sum_{x_i\in D_u}\gamma_{ji} + \sum_{i=1}^{N}\lambda\alpha_i = 0$$

因为

$$\begin{aligned}
&\sum_{i=1}^{N}\gamma_{ji} = \sum_{i=1}^{N}\frac{\alpha_i\cdot p\left(x_j\mid\mu_i,\Sigma_i\right)}{\Sigma_{s=1}^{N}\alpha_s\cdot p\left(x_j\mid\mu_s,\Sigma_s\right)} = \frac{\sum_{i=1}^{N}\alpha_i\cdot p\left(x_j\mid\mu_i,\Sigma_i\right)}{\sum_{s=1}^{N}\alpha_s\cdot p\left(x_j\mid\mu_s,\Sigma_s\right)} = 1 \\
&\Rightarrow \sum_{i=1}^{N}\sum_{x_i\in D_u}\gamma_{ji} = \sum_{x_i\in D_u}\sum_{i=1}^{N}\gamma_{ji} = \sum_{x_i\in D_u}1 = u
\end{aligned}$$

因为 $\sum_{x_j\in D_u}$ 形式与 $\sum_{j=1}^{u}$ 等价，$\sum_{i=1}^{N} l_i = l$其中l为有标记样本集的样本个数；代入

$$\sum_{i=1}^{N} l_i + \sum_{i=1}^{N}\sum_{x_i\in D_u}\gamma_{ji} + \sum_{i=1}^{N}\lambda\alpha_i = 0$$

解得

$$\begin{gathered}
\begin{cases} l+u+\lambda=0 \\ l+u=m \end{cases} \\
\Rightarrow \mathrm{l}\cdot\mathrm{i} + \sum\cdot\mathrm{x}\cdot\mathrm{j}\in\mathrm{D}\cdot\mathrm{u}\gamma_{\mathrm{ji}} - \lambda\alpha_{\mathrm{i}} = 0
\end{gathered}$$

所以

$$\alpha_{\mathrm{i}} = \frac{1}{\mathrm{m}}\left(\sum_{\boldsymbol{x}\cdot\mathrm{j}\in\mathrm{D}\cdot\mathrm{u}}\gamma_{\mathrm{ji}} + \mathrm{l}_{\mathrm{i}}\right)$$

13.5

具体代码见“***半监督学习.py***”，在此仅列出部分代码

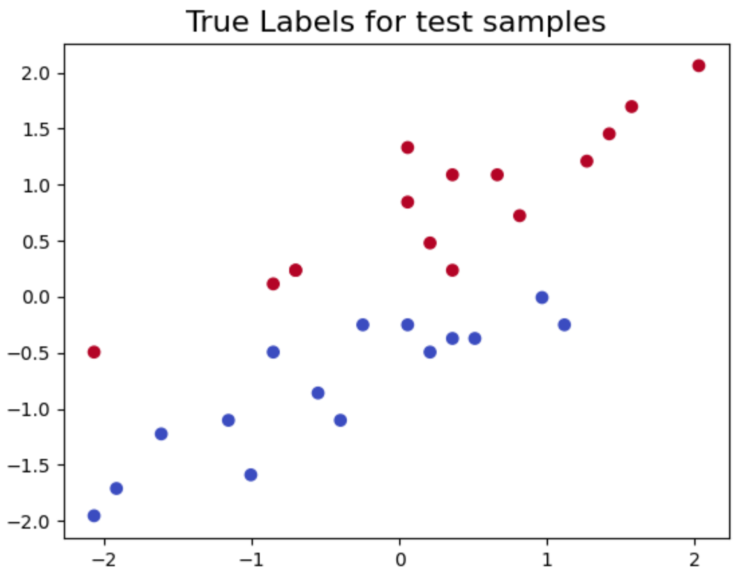
```
# 定义TSVM函数
def TSVM(test_feature, test_label, X1, Y1, X2, X3):
    # 调用sklearn中的SVM
    clf_svm = svm.SVC(C=1, kernel='linear')
    clf_svm.fit(X1, Y1)
    Y3_svm = clf_svm.predict(X3)

    clf_TSVM = svm.SVC(C=1, kernel='linear')
    clf_TSVM.fit(X1, Y1)

    # 用训练好的 SVM对Du进行预测
    Y2 = clf_TSVM.predict(X2)

    # 初始化cu,cl
    cu = 0.001
    cl = 1

    # 样本权重， 直接让有标签数据的权重为Cl,无标签数据的权重为Cu
    sample_weight = np.ones(len(X1) + len(X2))
    sample_weight[len(X1):] = cu

    # 初始化 id 数组
    id_set = np.arange(len(X2))

    while cu < cl:
        Y3 = np.concatenate((Y1, Y2))  # 合并有标签样本和无标签样本
        clf_TSVM.fit(X3, Y3, sample_weight=sample_weight)  # 对TSVM模型进行训练
        while True:
            Y2 = clf_TSVM.predict(X2)
            X2_dist = clf_TSVM.decision_function(X2)  # 计算无标签样本的距离
            norm_weight = np.linalg.norm(clf_TSVM.coef_)  # 进行标准化
            epsilon = 1 - X2_dist * Y2 * norm_weight

            plus_set, plus_id = epsilon[Y2 > 0], id_set[Y2 > 0]  # 正标记（1）样本
            minus_set, minus_id = epsilon[Y2 < 0], id_set[Y2 < 0]  # 负标记（-1）样
本
            # 找到最大、最小值的索引
            plus_max_id, minus_max_id = plus_id[np.argmax(plus_set)], 
minus_id[np.argmax(minus_set)]
            a, b = epsilon[plus_max_id], epsilon[minus_max_id]

            if a > 0 and b > 0 and a + b > 2:
                # 将无标签样本的预测值进行翻转
                Y2[plus_max_id], Y2[minus_max_id] = -Y2[plus_max_id], -
Y2[minus_max_id]
                Y3 = np.concatenate((Y1, Y2))  # 合并有标签样本和无标签样本的预测值
                clf_TSVM.fit(X3, Y3, sample_weight=sample_weight)  # 对TSVM模型进
行训练
            else:
                break
        cu = min(cu * 2, cl)
        sample_weight[len(Y1):] = cu
```

（1）选用UCI数据集中的 iris 数据集

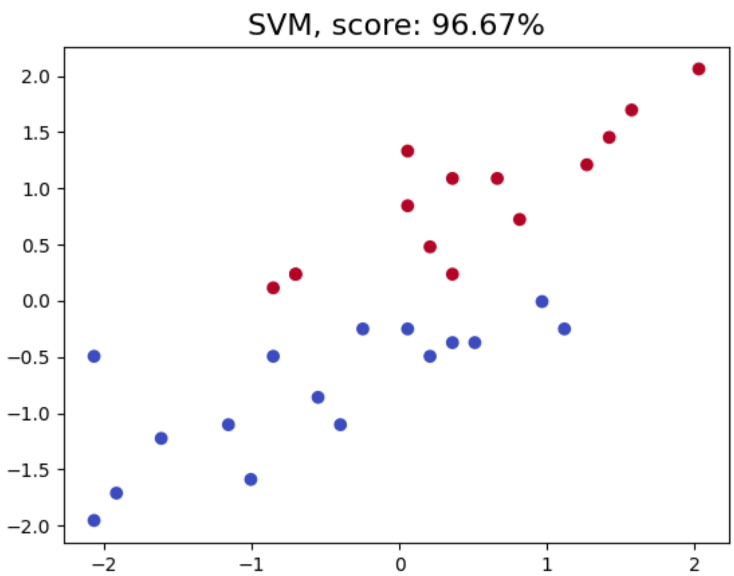
```
from sklearn.datasets import load_iris
iris = load_iris()
# 取iris数据集后100数据，即分类为2，3的数据集，构成一个二分类问题
feature, label = iris.data[50:, :], iris.target[50:] * 2 - 3
# 按题目要求分成测试集，有标签样本，无标签样本
# 特征均已归一化
```

经训练后，在测试集上得到如下结果：

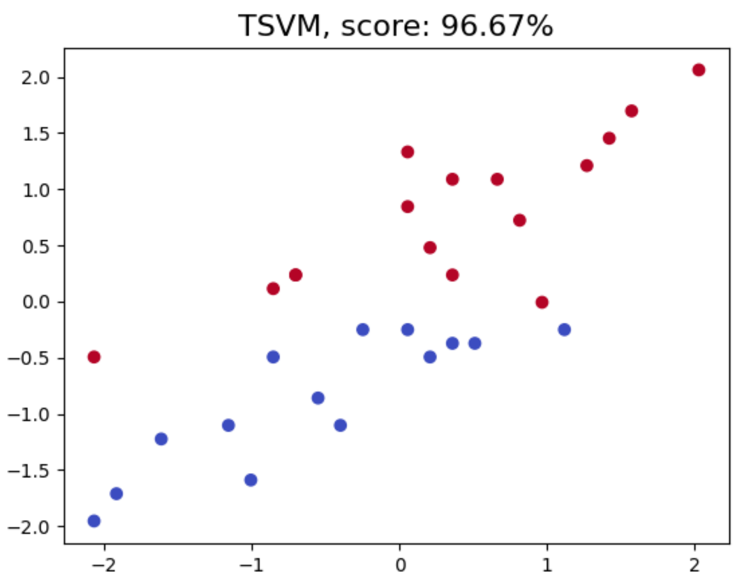
测试集样本：



在SVM上的结果：



在TSVM上的结果：



分析

二者的模型精度一致，可能是因为样测试样本较少的缘故；

在坐标（-2，-0.5）和（1.2，-0.2） 两处SVM与TSVM的预测结果不一致，并且都与真实情况相差一例；

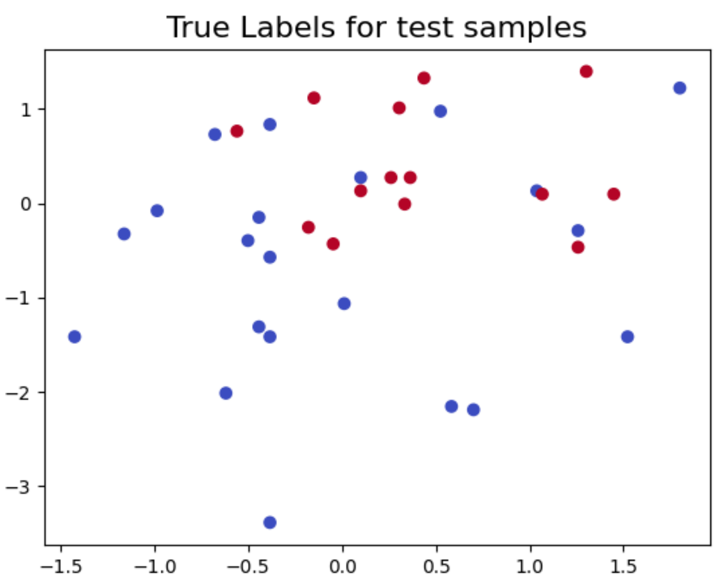
对基本训练模型svm采用高斯核效果一致。

（2）选用UCI数据集中的 wine 数据集

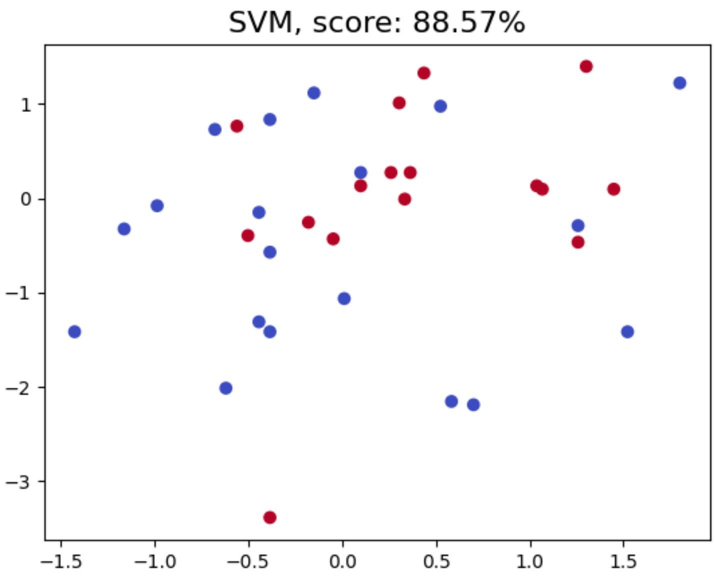
```
from sklearn.datasets import load_wine
iris = load_wine()
# 取wine数据集后119个数据，即分类为2，3的数据集，构成一个二分类问题
feature, label = wine.data[59:, :], wine.target[59:] * 2 - 3
# 按题目要求分成测试集，有标签样本，无标签样本
# 特征均已归一化
```

经训练后，在测试集上得到如下结果：

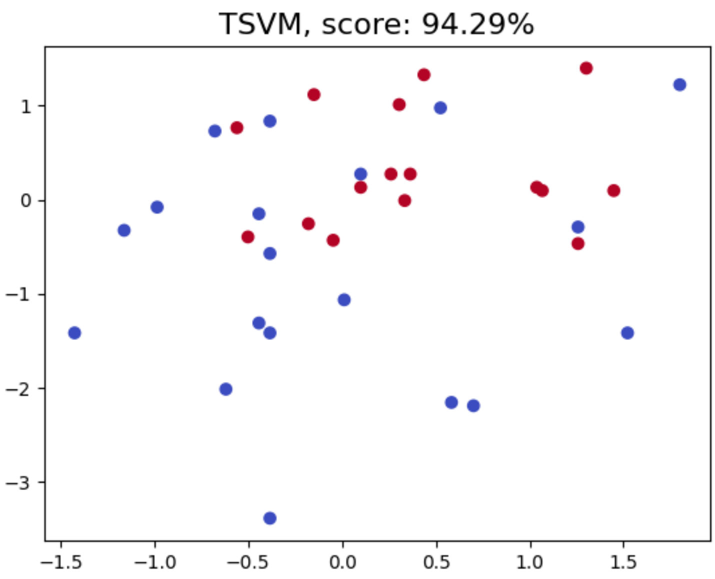
测试集样本：



在SVM上的结果：



在TSVM上的结果：



分析：

与普通SVM相比，TSVM的模型精度较高；

SVM与TSVM容易将正类模型预测为负，SVM预测为负的数量较多；

对基本训练模型svm采用高斯核效果略有提升。