宿題 1

ガウス混合分布の最尤推定量が満たす関係式を求める。

ガウス混合モデルは次のように表される。ここで，m は混合数である。

$$q(\boldsymbol{x};\ \boldsymbol{\theta}) = \sum_{j=1}^{m} w_j \phi(\boldsymbol{x};\ \boldsymbol{\mu}_j, \sigma_j) \tag{1}$$

$$w_j \geq 0, \quad \sum_{j=1}^{m} w_j = 1 \tag{2}$$

$$\boldsymbol{\theta} = \begin{bmatrix} w_1 & \cdots & w_m & \boldsymbol{\mu}_1^{\mathrm{T}} & \cdots & \boldsymbol{\mu}_m^{\mathrm{T}} & \sigma_1 & \cdots & \sigma_m \end{bmatrix}^{\mathrm{T}} \tag{3}$$

$$w_j \in \mathbb{R},\ \boldsymbol{\mu}_j \in \mathbb{R}^d,\ \sigma_j \in \mathbb{R}_{>0} \tag{4}$$

$$\phi(\boldsymbol{x};\ \boldsymbol{\mu}, \sigma) = \frac{1}{(2\pi\sigma^2)^{d/2}} \exp\left(-\frac{1}{2\sigma^2} (\boldsymbol{x} - \boldsymbol{\mu})^{\mathrm{T}} (\boldsymbol{x} - \boldsymbol{\mu}) \right) \tag{5}$$

いま，w_j を $\gamma_j \in \mathbb{R}$ を用いて次のように表すことで，式 (2) の w_j の拘束条件を自動的に満たすことができる。

$$w_j = \frac{\exp(\gamma_j)}{\sum_{j'=1}^{m} \exp(\gamma_{j'})} \tag{6}$$

いま，$q(\boldsymbol{x}_i;\ \boldsymbol{\theta}) = q_i$, $\phi(\boldsymbol{x}_i;\ \boldsymbol{\mu}_j, \sigma_j) = \phi_{ij}$ と表すと，対数尤度は次のように表される。

$$\log(L(\boldsymbol{\theta})) = \sum_{i=1}^{n} \log q_i \tag{7}$$

$$= \sum_{i=1}^{n} \log \left\{ \sum_{j=1}^{m} w_j \phi_{ij} \right\} \tag{8}$$

最尤推定量を与える $\boldsymbol{\theta}$ について，$\partial \log L \big/ \partial \boldsymbol{\theta} = \mathbf{0}$ が成り立つから，$j = 1, \cdots, m$ について次式が成り立つ。

$$\frac{\partial \log L}{\partial \gamma_j} = 0, \quad \frac{\partial \log L}{\partial \boldsymbol{\mu}_j} = \mathbf{0}, \quad \frac{\partial \log L}{\partial \sigma_j} = 0 \tag{9}$$

また，対数尤度の，$\boldsymbol{\theta}$ のうち一部のパラメータ $\boldsymbol{\psi}$ による偏微分は次のようになる。

$$\frac{\partial}{\partial \boldsymbol{\psi}} (\log(L(\boldsymbol{\theta}))) = \frac{\partial}{\partial \boldsymbol{\psi}} \left(\sum_{i=1}^{n} \log q_i \right) \tag{10}$$

$$= \sum_{i=1}^{n} \frac{1}{q_i} \frac{\partial q_i}{\partial \boldsymbol{\psi}} \tag{11}$$

まず，対数尤度の γ_j による偏微分を考える。いま，式 (6) から，

$$
\begin{align}
\frac{\partial w_j}{\partial \gamma_j} &= \frac{\partial}{\partial \gamma_j}\left(\frac{\exp(\gamma_j)}{\sum_{j'=1}^{m}\exp\left(\gamma_{j'}\right)}\right) \tag{12}\\
&= \exp(\gamma_j)\cdot\frac{1}{\sum_{j'=1}^{m}\exp\left(\gamma_{j'}\right)} + \exp(\gamma_j)\cdot\frac{-\exp(\gamma_j)}{\left(\sum_{j'=1}^{m}\exp\left(\gamma_{j'}\right)\right)^2} \tag{13}\\
&= \frac{\exp(\gamma_j)}{\sum_{j'=1}^{m}\exp\left(\gamma_{j'}\right)} - \left(\frac{\exp(\gamma_j)}{\sum_{j'=1}^{m}\exp\left(\gamma_{j'}\right)}\right)^2 \tag{14}\\
&= w_j - {w_j}^2 \tag{15}\\
&= w_j(1-w_j) \tag{16}
\end{align}
$$

$$
\begin{align}
\frac{\partial w_k}{\partial \gamma_j} &= \frac{\partial}{\partial \gamma_j}\left(\frac{\exp(\gamma_k)}{\sum_{j'=1}^{m}\exp\left(\gamma_{j'}\right)}\right) \tag{17}\\
&= \exp(\gamma_k)\left(\frac{-\exp(\gamma_j)}{\left(\sum_{j'=1}^{m}\exp\left(\gamma_{j'}\right)\right)^2}\right) \tag{18}\\
&= -\frac{\exp(\gamma_j)}{\sum_{j'=1}^{m}\exp\left(\gamma_{j'}\right)}\cdot\frac{\exp(\gamma_k)}{\sum_{j'=1}^{m}\exp\left(\gamma_{j'}\right)} \tag{19}\\
&= -w_j w_k \tag{20}
\end{align}
$$

よって，q_i の γ_j による偏微分は，

$$
\begin{align}
\frac{\partial q_i}{\partial \gamma_j} &= \frac{\partial}{\partial \gamma_j}\left(\sum_{j=1}^{m} w_j \phi_{ij}\right) \tag{21}\\
&= \sum_{k=1,k\neq j}^{m}(-w_j w_k)\phi_{ik} + w_j(1-w_j)\phi_{ij} \tag{22}\\
&= -w_j\sum_{k=1}^{m}(w_k\phi_{ik}) + w_j\phi_{ij} \tag{23}\\
&= w_j\phi_{ij} - w_j q_i \tag{24}
\end{align}
$$

従って，式 (11) を用いることで，対数尤度の γ_j による偏微分は次のようになる。

$$
\begin{align}
\frac{\partial}{\partial \gamma_j}(\log(L(\boldsymbol{\theta}))) &= \sum_{i=1}^{n}\frac{1}{q_i}\frac{\partial q_i}{\partial \gamma_j} \tag{25}\\
&= \sum_{i=1}^{n}\frac{1}{q_i}(w_j\phi_{ij} - w_j q_i) \tag{26}\\
&= \sum_{i=1}^{n}\left(\frac{w_j\phi_{ij}}{q_i} - w_j\right) \tag{27}
\end{align}
$$

ここで，

$$\eta_{ij} = \frac{w_j \phi_{ij}}{q_i} = \frac{w_j \phi(\boldsymbol{x}_i;\ \boldsymbol{\mu}_j, \sigma_j)}{q(\boldsymbol{x}_i;\ \boldsymbol{\theta})} \tag{28}$$

$$= \frac{w_j \phi(\boldsymbol{x}_i;\ \boldsymbol{\mu}_j, \sigma_j)}{\sum_{j'=1}^{b} w_{j'} \phi(\boldsymbol{x}_i;\ \boldsymbol{\mu}_{j'}, \sigma_{j'})} \tag{29}$$

なる η_{ij} を用いると，結局，

$$\frac{\partial}{\partial \gamma_j}(\log(L(\boldsymbol{\theta}))) = \sum_{i=1}^{n} \eta_{ij} - n w_j \tag{30}$$

これが 0 となるときに最尤推定量となるので，

$$\hat{w}_j = \frac{1}{n} \sum_{i=1}^{n} \hat{\eta}_{ij} \tag{31}$$

を得る。

次に，対数尤度の $\boldsymbol{\mu}_j$ による偏微分を考える。ϕ_{ij} の $\boldsymbol{\mu}_j$ による偏微分は次のようになる。

$$\frac{\partial \phi_{ij}}{\partial \boldsymbol{\mu}_j} = \frac{\partial}{\partial \boldsymbol{\mu}_j} \left(\frac{1}{(2\pi\sigma_j^2)^{d/2}} \exp\left(-\frac{1}{2\sigma_j^2} (\boldsymbol{x}_i - \boldsymbol{\mu}_j)^{\mathrm{T}} (\boldsymbol{x}_i - \boldsymbol{\mu}_j) \right) \right) \tag{32}$$

$$= \frac{1}{(2\pi\sigma_j^2)^{d/2}} \cdot \left(-\frac{1}{2\sigma_j^2} \right) \cdot 2(\boldsymbol{\mu}_j - \boldsymbol{x}_i) \cdot \exp\left(-\frac{1}{2\sigma_j^2} (\boldsymbol{x}_i - \boldsymbol{\mu}_j)^{\mathrm{T}} (\boldsymbol{x}_i - \boldsymbol{\mu}_j) \right) \tag{33}$$

$$= \frac{1}{\sigma_j^2} (\boldsymbol{x}_i - \boldsymbol{\mu}_j) \phi_{ij} \tag{34}$$

よって，q_i の $\boldsymbol{\mu}_j$ による偏微分は，

$$\frac{\partial q_i}{\partial \boldsymbol{\mu}_j} = \frac{\partial}{\partial \boldsymbol{\mu}_j} \left(\sum_{j=1}^{m} w_j \phi_{ij} \right) \tag{35}$$

$$= w_j \frac{\partial \phi_{ij}}{\partial \boldsymbol{\mu}_j} \tag{36}$$

$$= w_j \frac{1}{\sigma_j^2} (\boldsymbol{x}_i - \boldsymbol{\mu}_j) \phi_{ij} \tag{37}$$

$$= w_j \phi_{ij} \cdot \frac{1}{\sigma_j^2} (\boldsymbol{x}_i - \boldsymbol{\mu}_j) \tag{38}$$

式 (11) から，対数尤度の $\boldsymbol{\mu}_j$ による偏微分は次のようになる。

$$\frac{\partial}{\partial \boldsymbol{\mu}_j}(\log(L(\boldsymbol{\theta}))) = \sum_{i=1}^{n} \frac{1}{q_i} \frac{\partial q_i}{\partial \boldsymbol{\mu}_j} \tag{39}$$

$$= \sum_{i=1}^{n} \frac{1}{q_i} \left(w_j \phi_{ij} \cdot \frac{1}{\sigma_j^2} (\boldsymbol{x}_i - \boldsymbol{\mu}_j) \right) \tag{40}$$

$$= \frac{1}{\sigma_j^2} \left(\sum_{i=1}^{n} \frac{w_j \phi_{ij}}{q_i} \boldsymbol{x}_i - \boldsymbol{\mu}_j \sum_{i=1}^{n} \frac{w_j \phi_{ij}}{q_i} \right) \tag{41}$$

$$= \frac{1}{\sigma_j^2} \left(\sum_{i=1}^{n} \eta_{ij} \boldsymbol{x}_i - \boldsymbol{\mu}_j \sum_{i=1}^{n} \eta_{ij} \right) \tag{42}$$

これが $\mathbf{0}$ となるときに最尤推定量となるので，

$$\hat{\boldsymbol{\mu}}_j = \frac{\sum_{i=1}^{n} \hat{\eta}_{ij} \boldsymbol{x}_i}{\sum_{i=1}^{n} \hat{\eta}_{ij}} \tag{43}$$

を得る。

最後に，対数尤度の σ_j による偏微分を考える。ϕ_{ij} の σ_j による偏微分は次のようになる。

$$\frac{\partial \phi_{ij}}{\partial \sigma_j} = \frac{\partial}{\partial \sigma_j}\left(\frac{1}{(2\pi\sigma_j^2)^{d/2}} \exp\left(-\frac{1}{2\sigma_j^2}(\boldsymbol{x}_i - \boldsymbol{\mu}_j)^{\mathrm{T}}(\boldsymbol{x}_i - \boldsymbol{\mu}_j)\right)\right) \tag{44}$$

$$= \frac{1}{(2\pi)^{d/2}} \frac{-d}{\sigma_j{}^{d+1}} \exp\left(-\frac{1}{2\sigma_j^2}(\boldsymbol{x}_i - \boldsymbol{\mu}_j)^{\mathrm{T}}(\boldsymbol{x}_i - \boldsymbol{\mu}_j)\right) \tag{45}$$

$$+ \frac{1}{(2\pi)^{d/2}} \frac{1}{\sigma_j{}^d}\left\{-\frac{-2}{2\sigma_j{}^3}(\boldsymbol{x}_i - \boldsymbol{\mu}_j)^{\mathrm{T}}(\boldsymbol{x}_i - \boldsymbol{\mu}_j)\right\} \exp\left(-\frac{1}{2\sigma_j^2}(\boldsymbol{x}_i - \boldsymbol{\mu}_j)^{\mathrm{T}}(\boldsymbol{x}_i - \boldsymbol{\mu}_j)\right) \tag{46}$$

$$= \frac{1}{(2\pi\sigma_j^2)^{d/2}} \exp\left(-\frac{1}{2\sigma_j^2}(\boldsymbol{x}_i - \boldsymbol{\mu}_j)^{\mathrm{T}}(\boldsymbol{x}_i - \boldsymbol{\mu}_j)\right)\left\{-\frac{d}{\sigma_j} + \frac{1}{\sigma_j{}^3}(\boldsymbol{x}_i - \boldsymbol{\mu}_j)^{\mathrm{T}}(\boldsymbol{x}_i - \boldsymbol{\mu}_j)\right\} \tag{47}$$

$$= \phi_{ij}\left\{-\frac{d}{\sigma_j} + \frac{1}{\sigma_j{}^3}(\boldsymbol{x}_i - \boldsymbol{\mu}_j)^{\mathrm{T}}(\boldsymbol{x}_i - \boldsymbol{\mu}_j)\right\} \tag{48}$$

よって，q_i の σ_j による偏微分は，

$$\frac{\partial q_i}{\partial \sigma_j} = \frac{\partial}{\partial \sigma_j}\left(\sum_{j=1}^{m} w_j \phi_{ij}\right) \tag{49}$$

$$= w_j \frac{\partial \phi_{ij}}{\partial \sigma_j} \tag{50}$$

$$= w_j \phi_{ij}\left\{-\frac{d}{\sigma_j} + \frac{1}{\sigma_j{}^3}(\boldsymbol{x}_i - \boldsymbol{\mu}_j)^{\mathrm{T}}(\boldsymbol{x}_i - \boldsymbol{\mu}_j)\right\} \tag{51}$$

式 (11) から，対数尤度の σ_j による偏微分は次のようになる。

$$\frac{\partial}{\partial \sigma_j}(\log(L(\boldsymbol{\theta}))) = \sum_{i=1}^{n} \frac{1}{q_i} \frac{\partial q_i}{\partial \sigma_j} \tag{52}$$

$$= \sum_{i=1}^{n} \frac{1}{q_i} w_j \phi_{ij}\left\{-\frac{d}{\sigma_j} + \frac{1}{\sigma_j{}^3}(\boldsymbol{x}_i - \boldsymbol{\mu}_j)^{\mathrm{T}}(\boldsymbol{x}_i - \boldsymbol{\mu}_j)\right\} \tag{53}$$

$$= \frac{1}{\sigma_j{}^3} \sum_{i=1}^{n} \frac{w_j \phi_{ij}}{q_i}(\boldsymbol{x}_i - \boldsymbol{\mu}_j)^{\mathrm{T}}(\boldsymbol{x}_i - \boldsymbol{\mu}_j) - \frac{d}{\sigma_j} \sum_{i=1}^{n} \frac{w_j \phi_{ij}}{q_i} \tag{54}$$

$$= \frac{1}{\sigma_j{}^3} \sum_{i=1}^{n} \eta_{ij}(\boldsymbol{x}_i - \boldsymbol{\mu}_j)^{\mathrm{T}}(\boldsymbol{x}_i - \boldsymbol{\mu}_j) - \frac{d}{\sigma_j} \sum_{i=1}^{n} \eta_{ij} \tag{55}$$

これが $\mathbf{0}$ となるときに最尤推定量となるので，

$$\hat{\sigma}_j^2 = \frac{1}{d} \frac{\sum_{i=1}^{n} \hat{\eta}_{ij}(\boldsymbol{x}_i - \hat{\boldsymbol{\mu}}_j)^{\mathrm{T}}(\boldsymbol{x}_i - \hat{\boldsymbol{\mu}}_j)}{\sum_{i=1}^{n} \hat{\eta}_{ij}} \tag{56}$$

$$\hat{\sigma}_j = \sqrt{\frac{1}{d} \frac{\sum_{i=1}^{n} \hat{\eta}_{ij}(\boldsymbol{x}_i - \hat{\boldsymbol{\mu}}_j)^{\mathrm{T}}(\boldsymbol{x}_i - \hat{\boldsymbol{\mu}}_j)}{\sum_{i=1}^{n} \hat{\eta}_{ij}}} \tag{57}$$

を得る。

以上，式 (29)，(31)，(43)，(57) が，ガウス混合分布の最尤推定量の関係式である。