Intelligent Speech Distinguish Homework 01

Qiu Yihang

May.7-May.15, 2022

Parameters of **GMM-HMM**:

- Probabilities of state transition $\boldsymbol{A}$: $a_{ij} = \mathbf{Pr}\left(q_t = j | q_{t-1} = i\right), 1 \le i, j \le N$
- The distribution of state output $\boldsymbol{B}$: $b_j(\mathbf{o}) = \sum_{m=1}^{M_j} c_{jm} \cdot \mathcal{N}(\mathbf{o}|\boldsymbol{\mu}_{jm}, \mathbf{\Sigma}_{jm})$
  For GMM, parameters are $c_{jm}, \boldsymbol{\mu}_{jm}, \mathbf{\Sigma}_{jm}$ $(1 \le j \le N, 1 \le m \le M_j)$.

In conclusion, parameter set of GMM-HMM is $\{a_{ij}\}_{1 \le i \le N, 1 \le j \le N}$ and $\left\{c_{jm}, \boldsymbol{\mu}_{jm}, \mathbf{\Sigma}_{jm}\right\}_{1 \le j \le N, 1 \le m \le M_j}$.
The likelihood is given as follows. We maximize the likelihood to get best parameters.

$$\mathcal{L}(\theta) = \sum_{r=1}^{R} \log p\left(\mathbf{O}^{(r)}|\theta\right) = \sum_{r=1}^{R} \log\left(\sum_{\mathbf{q}} p(\mathbf{O}^{(r)}, \mathbf{q}|\theta)\right)$$

where θ is the set of parameters. $\mathbf{q}$ is the sequence of hidden states.
$\mathbf{O}^{(r)}$ are the data given $(1 \le r \le R)$.

How we use Expectation Maximization to update parameters in GMM-HMM is as follows.
The proof is also given below.

Use $\hat{\theta}$ to denote the initial value of θ. Use θ^* to denote what θ should be after an iteration of EM.

Task 0. Some preparations.

In class, we already proved that

$$\mathcal{L}(\theta) \ge \mathtt{H}\left(\mathbf{Pr}\left(\mathbf{q}|\mathbf{O}^{(r)}, \hat{\theta}\right)\right) + \sum_{r=1}^{R}\sum_{\mathbf{q}} \mathbf{Pr}\left(\mathbf{q}|\mathbf{O}^{(r)}, \hat{\theta}\right) \log p\left(\mathbf{O}^{(r)}, \mathbf{q}|\theta\right)$$

(where $\mathtt{H}(\cdot)$ is the information entropy.)

In class, we define occupancy as follows.

$$\begin{cases} \gamma_{(i,j)}^{(r)}(t) = \mathbf{Pr}\left(q_{t-1} = i, q_t = j | \mathbf{O}^{(r)}, \hat{\theta}\right) \\ \gamma_j^{(r)}(t) = \mathbf{Pr}\left(q_t = j | \mathbf{O}^{(r)}, \hat{\theta}\right) \end{cases}$$

Also, we define

$$\mathcal{Q}_A = \sum_{r=1}^{R}\sum_{t=1}^{T^{(r)}}\sum_{i=1}^{N}\sum_{j=1}^{N} \gamma_{(i,j)}^{(r)}(t) \log \mathbf{Pr}\left(q_t | q_{t-1}, \theta\right), \qquad \mathcal{Q}_B = \sum_{r=1}^{R}\sum_{t=1}^{T^{(r)}}\sum_{j=1}^{N} \gamma_j^{(r)}(t) \log p\left(\mathbf{o}_t^{(r)} | q_t = j, \theta\right)$$

which satisfies that $\mathcal{L}(\theta) = const + \mathcal{Q}_A + \mathcal{Q}_B$.

In class, we have proved that (here $\mathbf{O} = \mathbf{O}^{(r)}, \mathbf{o} = \mathbf{o}^{(r)}$)

$$\begin{aligned}
\gamma_{(i,j)}^{(r)}(t) = \mathbf{Pr}\left(q_{t-1} = i, q_t = j | \mathbf{O}_1^T, \hat{\theta}\right) &= \frac{p(q_{t-1} = i, q_t = j, \mathbf{O}_1^T | \hat{\theta})}{p(\mathbf{O}_1^T | \hat{\theta})} \\
&= \frac{p(q_{t+1} = i, \mathbf{O}_1^{t-1}) \mathbf{Pr}\left(q_t = j | q_{t-1} = i\right) p(\mathbf{o}_t | q_t = j) p(\mathbf{O}_{t+1}^T | q_t = j)}{p(\mathbf{O}_1^T | \hat{\theta})} \\
&= \frac{\alpha_i(t-1) \hat{a}_{ij} \hat{b}_j(\mathbf{o}_t) \beta_j(t)}{\alpha_N(T+1)} \\
&= \frac{\alpha_i(t-1) \hat{a}_{ij} \sum_{m=1}^{M_j} \hat{c}_{jm} \, \mathcal{N}(\mathbf{o}_t | \hat{\boldsymbol{\mu}}_{jm}, \hat{\boldsymbol{\Sigma}}_{jm}) \beta_j(t)}{\alpha_N(T+1)}
\end{aligned}$$

where

$$\alpha_j(t) = b_j(\mathbf{o}_t) \sum_{i=1}^{N-1} \hat{a}_{ij} \alpha_i(t-1) = \sum_{m=1}^{M_j} \hat{c}_{jm} \, \mathcal{N}(\mathbf{o}_t | \hat{\boldsymbol{\mu}}_{jm}, \hat{\boldsymbol{\Sigma}}_{jm}) \sum_{i=1}^{N-1} \hat{a}_{ij} \alpha_i(t-1),$$

$$\text{(for } 1 \le t \le T, 1 \le j \le N)$$

$$\text{with } \alpha_j(0) = \begin{cases} 1, & j = 1 \\ 0, & otherwise \end{cases}$$

$$\beta_j(t) = \sum_{i=1}^{N-1} b_i(\mathbf{o}_{t+1}) \hat{a}_{ji} \beta_i(t+1) = \sum_{i=1}^{N-1} \left(\sum_{m=1}^{M_i} \hat{c}_{im} \, \mathcal{N}(\mathbf{o}_{t+1} | \hat{\boldsymbol{\mu}}_{im}, \hat{\boldsymbol{\Sigma}}_{im}) \right) \hat{a}_{ji} \beta_i(t+1),$$

$$\text{(for } 1 \le t \le T, 1 \le j \le N)$$

$$\text{with } \beta_j(T) = a_{jN}, \ \beta_N(T+1) = 1.$$

Task 1. Now we consider how a_{ij} should be updated in the Maximization Step. □

We want to maximize $\mathcal{Q}_A$. Thus,

$$a_{ij}^* = \underset{a_{ij}}{\text{argmax}} \sum_{r=1}^{R} \sum_{t=1}^{T^{(r)}} \sum_{i=1}^{N} \sum_{j=1}^{N} \gamma_{(i,j)}^{(r)}(t) \log a_{ij} \quad \text{s.t.} \quad \begin{cases} \sum_{j=1}^{N} a_{ij} = 1, \\ 0 \le a_{ij} \le 1, \ (1 \le i \le N, 1 \le j \le N) \end{cases}$$

It is a constrained optimization problem. The Lagrangian

$$\mathcal{L}_{\mathcal{Q}_A}(\boldsymbol{A}, \lambda) = \sum_{r=1}^{R} \sum_{t=1}^{T^{(r)}} \sum_{i=1}^{N} \sum_{j=1}^{N} \gamma_{(i,j)}^{(r)}(t) \log a_{ij} + \sum_{i=1}^{N} \lambda_i \left(\sum_{j=1}^{N} a_{ij} - 1 \right)$$

Then we have

$$\begin{aligned}
\frac{\partial}{\partial a_{ij}} \mathcal{L}_{\mathcal{Q}_A} &= \sum_{r=1}^{R} \sum_{t=1}^{T^{(r)}} \frac{\gamma_{(i,j)}^{(r)}(t)}{a_{ij}} + \lambda_i \\
\frac{\partial}{\partial \lambda_i} \mathcal{L}_{\mathcal{Q}_A} &= \sum_{r=1}^{R} \sum_{t=1}^{T^{(r)}} \sum_{j=1}^{N} a_{ij} - 1
\end{aligned}$$

Set the gradient to 0. We have

$$\lambda_i = -\sum_{j=1}^{N}\sum_{r=1}^{R}\sum_{t=1}^{T^{(r)}} \gamma_{(i,j)}^{(r)}(t)$$

$$a_{ij}^* = -\frac{1}{\lambda_i}\sum_{r=1}^{R}\sum_{t=1}^{T^{(r)}} \gamma_{(i,j)}^{(r)}(t)$$

Thus, the update of parameter a_{ij} should be

$$a_{ij}^* = \frac{\sum_{r=1}^{R}\sum_{t=1}^{t^{(r)}} \gamma_{(i,j)}^{(r)}(t)}{\sum_{r=1}^{R}\sum_{t=1}^{t^{(r)}}\sum_{j=1}^{N} \gamma_{(i,j)}^{(r)}(t)} \qquad \blacksquare$$

Task 2. Now we consider how $c_{jm}, \boldsymbol{\mu}_{jm}, \boldsymbol{\Sigma}_{jm}$ should be updated in the Maximization Step. $\square$

$$\mathcal{Q}_B(\theta, \hat{\theta}) = \sum_{r=1}^{R}\sum_{t=1}^{T^{(r)}}\sum_{j=1}^{N} \gamma_j^{(r)}(t) \log p\left(\mathbf{o}_t^{(r)} | q_t = j, \theta\right)$$

We know

$$\begin{aligned}
\mathcal{Q}_B(\theta, \hat{\theta}) &= \sum_{r=1}^{R}\sum_{t=1}^{T^{(r)}}\sum_{j=1}^{N} \gamma_j^{(r)}(t) \log p\left(\mathbf{o}_t^{(r)} | q_t = j, \theta\right) = \sum_{r=1}^{R}\sum_{t=1}^{T^{(r)}}\sum_{j=1}^{N} \gamma_j^{(r)}(t) \log b_j\left(\mathbf{o}_t^{(r)}\right) \\
&= \sum_{r=1}^{R}\sum_{t=1}^{T^{(r)}}\sum_{j=1}^{N} \gamma_j^{(r)}(t) \log\left(\sum_{m=1}^{M_j} \mathbf{Pr}\left(g_t = m | q_t = j, \mathbf{O}_1^T, \hat{\theta}\right) c_{jm}\mathcal{N}\left(\mathbf{o}_t^{(r)} | \boldsymbol{\mu}_{jm}, \boldsymbol{\Sigma}_{jm}\right)\right) \\
&\geq \sum_{r=1}^{R}\sum_{t=1}^{T^{(r)}}\sum_{j=1}^{N}\sum_{m=1}^{M_j} \gamma_{jm}^{(r)}(t) \log\left(c_{jm}\mathcal{N}\left(\mathbf{o}_t^{(r)} | \boldsymbol{\mu}_{jm}, \boldsymbol{\Sigma}_{jm}\right)\right)
\end{aligned}$$

(By the convexity of log function, i.e. Jensen's Inequality)

where $\gamma_j^{(r)}(t)$ and $\gamma_{jm}^{(r)}(t)$ is given as follows. (Here $\mathbf{O} = \mathbf{O}^{(r)}$)

$$\gamma_j^{(r)}(t) = \mathbf{Pr}\left(q_t = j | \mathbf{O}_1^T, \hat{\theta}\right) = \frac{\mathbf{Pr}\left(q_t = j, \mathbf{O}_1^T | \hat{\theta}\right)}{p(\mathbf{O}_1^T | \hat{\theta})} = \frac{p(\mathbf{O}_1^t, q_t = j)p(\mathbf{O}_{t+1}^T | q_t = j)}{\alpha_N(T+1)} = \frac{\alpha_j(t)\beta_j(t)}{\alpha_N(T+1)}$$

$$\begin{aligned}
\gamma_{jm}^{(r)}(t) &= \mathbf{Pr}\left(q_t = j, g_t = m | \mathbf{O}_1^T, \hat{\theta}\right) = \mathbf{Pr}\left(q_t = j | \mathbf{O}_1^T, \hat{\theta}\right) \mathbf{Pr}\left(g_t = m | q_t = j, \mathbf{O}_1^T, \hat{\theta}\right) \\
&= \mathbf{Pr}\left(q_t = j | \mathbf{O}_1^T, \hat{\theta}\right) \mathbf{Pr}\left(g_t = m | \mathbf{O}_1^T, \hat{\theta}\right) \quad (q_t \text{ and } g_m \text{ are independent}) \\
&= \gamma_j^{(r)}(t) \cdot \gamma_m^{(r)}(t)
\end{aligned}$$

Define $\mathcal{Q}_B' = \sum_{r=1}^{R}\sum_{t=1}^{T^{(r)}}\sum_{j=1}^{N}\sum_{m=1}^{M_j} \gamma_{jm}^{(r)}(t) \log\left(c_{jm}\mathcal{N}\left(\mathbf{o}_t^{(r)} | \boldsymbol{\mu}_{jm}, \boldsymbol{\Sigma}_{jm}\right)\right)$.

We want to maximize $\mathcal{Q}_B'$.

1) For the update of $\boldsymbol{\mu}_{jm}$, we know

$$\frac{\partial}{\partial \boldsymbol{\mu}_{jm}} \mathcal{Q}_B' = \sum_{r=1}^{R}\sum_{t=1}^{T^{(r)}} \gamma_{jm}^{(r)}(t) \frac{c_{jm}}{c_{jm}\mathcal{N}\left(\mathbf{o}_t^{(r)} | \boldsymbol{\mu}_{jm}, \boldsymbol{\Sigma}_{jm}\right)} \frac{\partial}{\partial \boldsymbol{\mu}_{jm}} \mathcal{N}\left(\mathbf{o}_t^{(r)} | \boldsymbol{\mu}_{jm}, \boldsymbol{\Sigma}_{jm}\right)$$

$$
\begin{aligned}
&= \sum_{r=1}^{R} \sum_{t=1}^{T^{(r)}} \gamma_{jm}^{(r)}(t) \frac{c_{jm}\mathcal{N}\left(\mathbf{o}_t^{(r)} | \boldsymbol{\mu}_{jm}, \boldsymbol{\Sigma}_{jm}\right)}{c_{jm}\mathcal{N}\left(\mathbf{o}_t^{(r)} | \boldsymbol{\mu}_{jm}, \boldsymbol{\Sigma}_{jm}\right)} \boldsymbol{\Sigma}_{jm}^{-1} \left(\mathbf{o}_t^{(r)} - \boldsymbol{\mu}_{jm}\right) \\
&= \sum_{r=1}^{R} \sum_{t=1}^{T^{(r)}} \gamma_{jm}^{(r)}(t) \boldsymbol{\Sigma}_{jm}^{-1} \left(\mathbf{o}_t^{(r)} - \boldsymbol{\mu}_{jm}\right)
\end{aligned}
$$

Set the gradient to 0. We get

$$
\begin{aligned}
\sum_{r=1}^{R} \sum_{t=1}^{T^{(r)}} \gamma_{jm}^{(r)}(t) \boldsymbol{\Sigma}_{jm}^{-1} \left(\mathbf{o}_t^{(r)} - \boldsymbol{\mu}_{jm}^*\right) = 0 &\Longrightarrow \sum_{r=1}^{R} \sum_{t=1}^{T^{(r)}} \gamma_{jm}^{(r)}(t) \left(\mathbf{o}_t^{(r)} - \boldsymbol{\mu}_{jm}^*\right) = 0 \\
&\Longrightarrow \boldsymbol{\mu}_{jm}^* = \frac{\sum_{r=1}^{R} \sum_{t=1}^{T^{(r)}} \gamma_{jm}^{(r)}(t) \cdot \mathbf{o}_t^{(r)}}{\sum_{r=1}^{R} \sum_{t=1}^{T^{(r)}} \gamma_{jm}^{(r)}}
\end{aligned}
\qquad \blacksquare
$$

2) For the update of $\boldsymbol{\Sigma}_{jm}$, we know

$$
\frac{\partial \mathcal{Q}'_B}{\partial \boldsymbol{\Sigma}_{jm}} = \sum_{r=1}^{R} \sum_{t=1}^{T^{(r)}} \gamma_{jm}^{(r)}(t) \frac{c_{jm}}{c_{jm}\mathcal{N}\left(\mathbf{o}_t^{(r)} | \boldsymbol{\mu}_{jm}, \boldsymbol{\Sigma}_{jm}\right)} \frac{\partial}{\partial \boldsymbol{\Sigma}_{jm}} \mathcal{N}\left(\mathbf{o}_t^{(r)} | \boldsymbol{\mu}_{jm}, \boldsymbol{\Sigma}_{jm}\right)
$$

Considering that

$$
\begin{aligned}
\frac{\partial}{\partial \boldsymbol{\Sigma}} \mathcal{N}(\mathbf{x}|\boldsymbol{\mu}, \boldsymbol{\Sigma}) &= \frac{\partial}{\partial \boldsymbol{\Sigma}} \left(\frac{1}{\sqrt{(2\pi)^d |\boldsymbol{\Sigma}|}} \exp\left(-\frac{1}{2}(\mathbf{x} - \boldsymbol{\mu})^T \boldsymbol{\Sigma}^{-1} (\mathbf{x} - \boldsymbol{\mu})\right)\right) \\
&= \mathcal{N}(\mathbf{x}|\boldsymbol{\mu}, \boldsymbol{\Sigma}) \left(\sqrt{|\boldsymbol{\Sigma}|} \frac{\partial |\boldsymbol{\Sigma}|^{-\frac{1}{2}}}{\partial \boldsymbol{\Sigma}} - \frac{1}{2} \frac{\partial (\mathbf{x} - \boldsymbol{\mu})^T \boldsymbol{\Sigma}^{-1} (\mathbf{x} - \boldsymbol{\mu})}{\partial \boldsymbol{\Sigma}} \right) \\
&= \mathcal{N}(\mathbf{x}|\boldsymbol{\mu}, \boldsymbol{\Sigma}) \left(-\frac{1}{2} |\boldsymbol{\Sigma}|^{\frac{1}{2}} |\boldsymbol{\Sigma}|^{-\frac{3}{2}} \frac{\partial |\boldsymbol{\Sigma}|}{\partial \boldsymbol{\Sigma}} + \frac{1}{2} \boldsymbol{\Sigma}^{-1} (\mathbf{x} - \boldsymbol{\mu}) \left(\boldsymbol{\Sigma}^{-1} (\mathbf{x} - \boldsymbol{\mu})\right)^T \right) \\
&= \mathcal{N}(\mathbf{x}|\boldsymbol{\mu}, \boldsymbol{\Sigma}) \left(-\frac{1}{2} |\boldsymbol{\Sigma}|^{-1} |\boldsymbol{\Sigma}| \boldsymbol{\Sigma}^{-1} + \frac{1}{2} \boldsymbol{\Sigma}^{-1} (\mathbf{x} - \boldsymbol{\mu}) (\mathbf{x} - \boldsymbol{\mu})^T \left(\boldsymbol{\Sigma}^{-1}\right)^T \right) \\
&= -\frac{1}{2} \mathcal{N}(\mathbf{x}|\boldsymbol{\mu}, \boldsymbol{\Sigma}) \left(\boldsymbol{\Sigma}^{-1} - \boldsymbol{\Sigma}^{-1} (\mathbf{x} - \boldsymbol{\mu}) (\mathbf{x} - \boldsymbol{\mu})^T \boldsymbol{\Sigma}^{-1} \right) \\
&\qquad \text{(By the symmetry of } \boldsymbol{\Sigma} \text{ and } \boldsymbol{\Sigma}^{-1}\text{)}
\end{aligned}
$$

we know

$$
\frac{\partial \mathcal{Q}'_B}{\partial \boldsymbol{\Sigma}_{jm}} = -\frac{1}{2} \sum_{r=1}^{R} \sum_{t=1}^{T^{(r)}} \gamma_{jm}^{(r)}(t) \left(\boldsymbol{\Sigma}_{jm}^{-1} - \boldsymbol{\Sigma}_{jm}^{-1} \left(\mathbf{o}_t^{(r)} - \boldsymbol{\mu}_{jm}\right) \left(\mathbf{o}_t^{(r)} - \boldsymbol{\mu}_{jm}\right)^T \boldsymbol{\Sigma}_{jm}^{-1} \right)
$$

Set the gradient to 0. We have

$$
\begin{aligned}
&-\frac{1}{2} \sum_{r=1}^{R} \sum_{t=1}^{T^{(r)}} \gamma_{jm}^{(r)}(t) \left(\boldsymbol{\Sigma}_{jm}^{-1} - \boldsymbol{\Sigma}_{jm}^{-1} \left(\mathbf{o}_t^{(r)} - \boldsymbol{\mu}_{jm}\right) \left(\mathbf{o}_t^{(r)} - \boldsymbol{\mu}_{jm}\right)^T \boldsymbol{\Sigma}_{jm}^{-1} \right) = 0 \\
\Longrightarrow \; &\boldsymbol{\Sigma}_{jm} \left[\sum_{r=1}^{R} \sum_{t=1}^{T^{(r)}} \gamma_{jm}^{(r)}(t) \left(\boldsymbol{\Sigma}_{jm}^{-1} - \boldsymbol{\Sigma}_{jm}^{-1} \left(\mathbf{o}_t^{(r)} - \boldsymbol{\mu}_{jm}\right) \left(\mathbf{o}_t^{(r)} - \boldsymbol{\mu}_{jm}\right)^T \boldsymbol{\Sigma}_{jm}^{-1} \right) \right] \boldsymbol{\Sigma}_{jm} = 0 \\
\Longrightarrow \; &\sum_{r=1}^{R} \sum_{t=1}^{T^{(r)}} \gamma_{jm}^{(r)}(t) \boldsymbol{\Sigma}_{jm} \left[\boldsymbol{\Sigma}_{jm}^{-1} - \boldsymbol{\Sigma}_{jm}^{-1} \left(\mathbf{o}_t^{(r)} - \boldsymbol{\mu}_{jm}\right) \left(\mathbf{o}_t^{(r)} - \boldsymbol{\mu}_{jm}\right)^T \boldsymbol{\Sigma}_{jm}^{-1} \right] \boldsymbol{\Sigma}_{jm} = 0
\end{aligned}
$$

$$\Longrightarrow \qquad \sum_{r=1}^{R}\sum_{t=1}^{T^{(r)}} \gamma_{jm}^{(r)}(t)\left(\boldsymbol{\Sigma}_{jm} - \left(\mathbf{o}_t^{(r)} - \boldsymbol{\mu}_{jm}\right)\left(\mathbf{o}_t^{(r)} - \boldsymbol{\mu}_{jm}\right)^T\right) = 0$$

$$\Longrightarrow \qquad \boldsymbol{\Sigma}_{jm}^{*} = \frac{\sum_{r=1}^{R}\sum_{t=1}^{T^{(r)}} \gamma_{jm}^{(r)}(t)\left(\mathbf{o}_t^{(r)} - \boldsymbol{\mu}_{jm}^{*}\right)\left(\mathbf{o}_t^{(r)} - \boldsymbol{\mu}_{jm}^{*}\right)^T}{\sum_{r=1}^{R}\sum_{t=1}^{T^{(r)}} \gamma_{jm}^{(r)}(t)} \qquad \blacksquare$$

3) For the update of c_{jm}, the optimization problem is actually a constrained one.

$$c_{jm}^{*} = \underset{c_{jm}}{\operatorname{argmax}}\ \mathcal{Q}'_B \qquad \text{s.t.} \ \sum_{m=1}^{M_j} c_{jm} = 1.$$

The Lagrangian

$$\mathcal{L}_{\mathcal{Q}'_B} = \mathcal{Q}'_B + \xi\left(\sum_{m=1}^{M_j} c_{jm} - 1\right)$$

$$\frac{\partial \mathcal{L}_{\mathcal{Q}'_B}}{\partial c_{jm}} = \sum_{r=1}^{R}\sum_{t=1}^{T^{(r)}} \gamma_{jm}^{(r)}(t)\frac{\mathcal{N}\left(\mathbf{o}_t^{(r)}|\boldsymbol{\mu}_{jm}, \boldsymbol{\Sigma}_{jm}\right)}{c_{jm}\mathcal{N}\left(\mathbf{o}_t^{(r)}|\boldsymbol{\mu}_{jm}, \boldsymbol{\Sigma}_{jm}\right)} + \xi = \frac{1}{c_{jm}}\sum_{r=1}^{R}\sum_{t=1}^{T^{(r)}} \gamma_{jm}^{(r)}(t) + \xi.$$

$$\frac{\partial \mathcal{L}_{\mathcal{Q}'_B}}{\partial \xi} = \sum_{m=1}^{M_j} c_{jm} - 1.$$

Set the gradient to 0, we have

$$c_{jm} = -\frac{1}{\xi}\sum_{r=1}^{R}\sum_{t=1}^{T^{(r)}} \gamma_{jm}^{(r)}(t)$$

$$\xi = -\sum_{r=1}^{R}\sum_{t=1}^{T^{(r)}}\sum_{m=1}^{M_j} \gamma_{jm}^{(r)}(t)$$

Thus, the update of parameter c_{jm} should be

$$c_{jm}^{*} = \frac{\sum_{r=1}^{R}\sum_{t=1}^{T^{(r)}} \gamma_{jm}^{(r)}(t)}{\sum_{r=1}^{R}\sum_{t=1}^{T^{(r)}}\sum_{m=1}^{M_j} \gamma_{jm}^{(r)}(t)} \qquad \blacksquare$$

IN CONCLUSION,

We define

$$\gamma_{(i,j)} \triangleq \sum_{r=1}^{R}\sum_{t=1}^{T^{(r)}} \gamma_{(i,j)}^{(r)}(t), \qquad \gamma_{jm} \triangleq \sum_{r=1}^{R}\sum_{t=1}^{T^{(r)}} \gamma_{jm}^{(r)}(t),$$

$$\boldsymbol{\mu}_{jm}^{\mathsf{acc}} = \sum_{r=1}^{R}\sum_{t=1}^{T^{(r)}} \gamma_{jm}^{(r)}(t)\mathbf{o}_t^{(r)}, \qquad \boldsymbol{\Sigma}_{jm}^{\mathsf{acc}} \triangleq \sum_{r=1}^{R}\sum_{t=1}^{T^{(r)}} \gamma_{jm}^{(r)}(t)\mathbf{o}_t^{(r)}\left(\mathbf{o}_t^{(r)}\right)^T.$$

Then we can rewrite the update of all parameters as follows.

$$
\begin{aligned}
a_{ij}^* &= \frac{\sum_{r=1}^{R}\sum_{t=1}^{t^{(r)}}\gamma_{(i,j)}^{(r)}(t)}{\sum_{r=1}^{R}\sum_{t=1}^{t^{(r)}}\sum_{j=1}^{N}\gamma_{(i,j)}^{(r)}(t)} = \frac{\gamma_{(i,j)}}{\sum_{j=1}^{N}\gamma_{(i,j)}} \\
c_{jm}^* &= \frac{\sum_{r=1}^{R}\sum_{t=1}^{T^{(r)}}\gamma_{jm}^{(r)}(t)}{\sum_{r=1}^{R}\sum_{t=1}^{T^{(r)}}\sum_{m=1}^{M_j}\gamma_{jm}^{(r)}(t)} = \frac{\gamma_{jm}}{\sum_{m=1}^{M_j}\gamma_{jm}} \\
\boldsymbol{\mu}_{jm}^* &= \frac{\sum_{r=1}^{R}\sum_{t=1}^{T^{(r)}}\gamma_{jm}^{(r)}(t)\cdot\mathbf{o}_t^{(r)}}{\sum_{r=1}^{R}\sum_{t=1}^{T^{(r)}}\gamma_{jm}^{(r)}} = \frac{\boldsymbol{\mu}_{jm}^{\mathsf{acc}}}{\gamma_{jm}} \\
\boldsymbol{\Sigma}_{jm}^* &= \frac{\sum_{r=1}^{R}\sum_{t=1}^{T^{(r)}}\gamma_{jm}^{(r)}(t)\left(\mathbf{o}_t^{(r)}-\boldsymbol{\mu}_{jm}^*\right)\left(\mathbf{o}_t^{(r)}-\boldsymbol{\mu}_{jm}^*\right)^T}{\sum_{r=1}^{R}\sum_{t=1}^{T^{(r)}}\gamma_{jm}^{(r)}(t)} \\
&= \frac{\boldsymbol{\Sigma}_{jm}^{\mathsf{acc}}}{\gamma_{jm}} - \boldsymbol{\mu}_{jm}^*\frac{\sum_{r=1}^{R}\sum_{t=1}^{T^{(r)}}\gamma_{jm}^{(r)}(t)\mathbf{o}_t^{(r)T}}{\gamma_{jm}} - \frac{\sum_{r=1}^{R}\sum_{t=1}^{T^{(r)}}\gamma_{jm}^{(r)}(t)\mathbf{o}_t^{(r)}}{\gamma_{jm}}{\boldsymbol{\mu}_{jm}^*}^T + \frac{\boldsymbol{\mu}_{jm}^{\mathsf{acc}}\boldsymbol{\mu}_{jm}^{\mathsf{acc}T}}{\gamma_{jm}^2} \\
&= \frac{\boldsymbol{\Sigma}_{jm}^{\mathsf{acc}}}{\gamma_{jm}} - \frac{\boldsymbol{\mu}_{jm}^{\mathsf{acc}}}{\gamma_{jm}}\frac{\boldsymbol{\mu}_{jm}^{\mathsf{acc}T}}{\gamma_{jm}} - \frac{\boldsymbol{\mu}_{jm}^{\mathsf{acc}}}{\gamma_{jm}}\frac{\boldsymbol{\mu}_{jm}^{\mathsf{acc}T}}{\gamma_{jm}} + \frac{\boldsymbol{\mu}_{jm}^{\mathsf{acc}}\boldsymbol{\mu}_{jm}^{\mathsf{acc}T}}{\gamma_{jm}^2} \\
&= \frac{\boldsymbol{\Sigma}_{jm}^{\mathsf{acc}}}{\gamma_{jm}} - \frac{\boldsymbol{\mu}_{jm}^{\mathsf{acc}}\boldsymbol{\mu}_{jm}^{\mathsf{acc}T}}{\gamma_{jm}^2}
\end{aligned}
$$

Thus, the update of all parameter set $\theta^* = \left\{a_{ij}, c_{jm}, \boldsymbol{\mu}_{jm}, \boldsymbol{\Sigma}_{jm}\right\}$ is as follows.

$$
a_{ij}^* = \frac{\gamma_{(i,j)}}{\sum_{j=1}^{N}\gamma_{(i,j)}}, \quad c_{jm}^* = \frac{\gamma_{jm}}{\sum_{m=1}^{M_j}\gamma_{jm}}, \quad \boldsymbol{\mu}_{jm}^* = \frac{\boldsymbol{\mu}_{jm}^{\mathsf{acc}}}{\gamma_{jm}}, \quad \boldsymbol{\Sigma}_{jm}^* = \frac{\boldsymbol{\Sigma}_{jm}^{\mathsf{acc}}}{\gamma_{jm}} - \frac{\boldsymbol{\mu}_{jm}^{\mathsf{acc}}\boldsymbol{\mu}_{jm}^{\mathsf{acc}T}}{\gamma_{jm}^2}. \qquad \blacksquare
$$

Task 3. <u>The whole process of Expectation Maximization Algorithm of HMM-GMM is as follows.</u> □

- Initialize θ with random values.
- Repeat the following two steps until certain criteria are reached.

  (For example, the number of iterations is large enough, or the change of value of parameters are within a significantly small range).
- **Expectation Step.**

  We use $\hat{\theta}$ of the former iteration in Expectation Step.

  For the first iteration, use the initial value as $\hat{\theta}$.

  * Calculate $\alpha_j(t), \beta_j(t)$ for each r with $\hat{\theta}$.

    (Definition is given in Task 0 Page 2).
  * Calculate $\gamma_{(i,j)}^{(r)}(t), \gamma_{jm}^{(r)}(t)$ and then compute $\gamma_{(i,j)}$ and γ_{jm}.

    (Definition is given in Task 0 Page 2, Task 2 Page 3 and Conclusion Page 6.)

* Calculate $\boldsymbol{\mu}_{jm}^{\texttt{acc}}$ and $\boldsymbol{\Sigma}_{jm}^{\texttt{acc}}$.

  (Definition is given in Conclusion Page 6.)

- **Maximization Step.**

  Update the parameters $a_{ij}, c_{jm}, \boldsymbol{\mu}_{jm}, \boldsymbol{\Sigma}_{jm}$ with calculated probabilities and expectations in Expectation Step.

$$a_{ij}^* = \frac{\gamma_{(i,j)}}{\sum_{j=1}^{N} \gamma_{(i,j)}}, \quad c_{jm}^* = \frac{\gamma_{jm}}{\sum_{m=1}^{M_j} \gamma_{jm}}, \quad \boldsymbol{\mu}_{jm}^* = \frac{\boldsymbol{\mu}_{jm}^{\texttt{acc}}}{\gamma_{jm}}, \quad \boldsymbol{\Sigma}_{jm}^* = \frac{\boldsymbol{\Sigma}_{jm}^{\texttt{acc}}}{\gamma_{jm}} - \frac{\boldsymbol{\mu}_{jm}^{\texttt{acc}} \boldsymbol{\mu}_{jm}^{\texttt{acc}T}}{\gamma_{jm}^2}.$$

  $a_{ij}^*, c_{jm}^*, \boldsymbol{\mu}_{jm}^*, \boldsymbol{\Sigma}_{jm}^*$ are the updated values of parameters.

End of Solution and Proof. ■