Chapter 17. Markov and hidden Markov models

17.1.

$$l_c(\boldsymbol{\theta}) = \sum_{i=1}^{N} \log p(\mathbf{x}_i, \mathbf{z}_i|\boldsymbol{\theta}) = \sum_{i=1}^{N}[\log p(z_{i1}) + \sum_{t=2}^{T_i} \log p(z_{it}|z_{i,t-1}) + \sum_{t=1}^{T_i} \log p(\mathbf{x}_{it}|z_{it})]$$

$$Q(\boldsymbol{\theta}, \boldsymbol{\theta}^{\text{old}}) = \mathbb{E}[l_c(\boldsymbol{\theta})|\mathcal{D}, \boldsymbol{\theta}^{\text{old}}] = \sum_{i=1}^{N} \mathbb{E}[\log p(z_{i1})]$$

$$+ \sum_{i=1}^{N}\sum_{t=2}^{T_i} \mathbb{E}[\log p(z_{it}|z_{i,t-1})] + \sum_{i=1}^{N}\sum_{t=1}^{T_i} \mathbb{E}[\log p(\mathbf{x}_{it}|z_{it})].$$

$$\sum_{i=1}^{N} \mathbb{E}[\log p(z_{i1})] = \sum_{i=1}^{N}\sum_{k=1}^{K} p(z_{i1} = k|\mathbf{x}_i, \boldsymbol{\theta}^{\text{old}}) \log p(z_{i1} = k) = \sum_{k=1}^{K} \mathbb{E}[N_k^1] \log \pi_k.$$

$$\sum_{i=1}^{N}\sum_{t=2}^{T_i} \mathbb{E}[\log p(z_{it}|z_{i,t-1})] = \sum_{i=1}^{N}\sum_{t=2}^{T_i}\sum_{j=1}^{K}\sum_{k=1}^{K} p(z_{it} = k, z_{i,t-1} = j|\mathbf{x}_i, \boldsymbol{\theta}^{\text{old}}) \log p(z_{it} = k|z_{i,t-1} = j)$$

$$= \sum_{j=1}^{K}\sum_{k=1}^{K} \mathbb{E}[N_{jk}] \log A_{jk}.$$

$$\sum_{i=1}^{N}\sum_{t=1}^{T_i} \mathbb{E}[\log p(\mathbf{x}_{it}|z_{it})] = \sum_{i=1}^{N}\sum_{t=1}^{T_i}\sum_{k=1}^{K} p(z_{it} = k|\mathbf{x}_i, \boldsymbol{\theta}^{\text{old}}) \log p(\mathbf{x}_{it}|z_i = k).$$

Summing those three term gives:

$$Q(\boldsymbol{\theta}, \boldsymbol{\theta}^{\text{old}}) = \sum_{k=1}^{K} \mathbb{E}[N_k^1] \log \pi_k + \sum_{j=1}^{K}\sum_{k=1}^{K} \mathbb{E}[N_{jk}] \log A_{jk}$$

$$+\sum_{i=1}^{N}\sum_{t=1}^{T_i}\sum_{k=1}^{K} p(z_{it}=k|\mathbf{x}_i,\boldsymbol{\theta}^{\text{old}})\log p(\mathbf{x}_{it}|\boldsymbol{\phi}_k).$$

17.2.

$$r_t(i) = p(S_t = i|\mathbf{x}_{t+1:T}) = \sum_j p(S_t = i, S_{t+1} = j|\mathbf{x}_{t+1:T})$$
$$= \sum_j p(S_{t+1} = j|\mathbf{x}_{t+1:T})p(S_t = i|S_{t+1} = j).$$
$$p(S_{t+1} = j|\mathbf{x}_{t+1:T}) = p(S_{t+1} = j|\mathbf{x}_{t+1}, \mathbf{x}_{t+2:T})$$
$$\propto p(\mathbf{x}_{t+1}|S_{t+1} = j)p(S_{t+1} = j|\mathbf{x}_{t+2:T}) \propto \phi_{t+1}(j)r_{t+1}(j).$$
$$\Rightarrow \mathbf{r}_t \propto \boldsymbol{\phi}_t \odot (\Xi^T \mathbf{r}_{t+1}),$$

where Ξ is the time-reversed transition matrix with $\xi_{ji} = p(S_t = i|S_{t+1} = j)$ with initial state given by $\Pi_T(i)$.
Computing $\gamma_t(i)$ can be done as follows:

$$\gamma_t(i) = p(S_t = i|\mathbf{x}_{1:T}) = p(S_t = i)p(\mathbf{x}_{1:T}|S_t = i)$$
$$= p(S_t = i)p(\mathbf{x}_{1:t}|S_t = i)p(\mathbf{x}_{t+1:T}|S_t = i) = \frac{1}{p(S_t = i)}p(S_t = i|\mathbf{x}_{1:t})p(S_t = i|\mathbf{x}_{t+1:T})$$
$$= \frac{\alpha_t(i)r_t(i)}{\Pi_t(i)}.$$

17.3.
Recall the result from 17.1:

$$Q(\boldsymbol{\theta},\boldsymbol{\theta}^{\text{old}}) = \sum_{k=1}^{K}\mathbb{E}[N_k^1]\log\pi_k + \sum_{j=1}^{K}\sum_{k=1}^{K}\mathbb{E}[N_{jk}]\log A_{jk}$$
$$+\sum_{i=1}^{N}\sum_{t=1}^{T_i}\sum_{k=1}^{K} p(z_{it}=k|\mathbf{x}_i,\boldsymbol{\theta}^{\text{old}})\log p(\mathbf{x}_{it}|\boldsymbol{\phi}_k).$$

Assume that each of the emission distributions

$$p(\mathbf{x}_t|z_t = k,\boldsymbol{\theta}) = \sum_{m=1}^{M_k} w_{km}\mathcal{N}(\mathbf{x}_t|\boldsymbol{\mu}_{km},\boldsymbol{\Sigma}_{km})$$

is itself a mixture of M_k Gaussians. Now we introduce a second latent variable y_{km}, which refers to the component m with state k. The latent variable y_{km} is itself a Markov chain with the transition matrix $\mathbf{\Omega} = [\omega_{(km),(k'm')}]$, where $\omega_{(km),(k'm')} = A_{kk'}w_{k'm'}$. (Therefore the transition between km and $k'm'$ only depends on the hidden state z at the previous time.)
Let $\gamma_{i,t}(k)$, $\xi_{i,t}(j,k)$ be smoothed node and edge marginals in Section 17.5.2.1., and $\delta_{i,t}(k,m) = p(y_t = km|z_t = k, \mathbf{x}_{i,1:T_i}, \boldsymbol{\theta})$ be marginals for mixture component transition. Then the last term becomes:

$$
\begin{aligned}
&\sum_{i=1}^{N}\sum_{t=1}^{T_i}\sum_{k=1}^{K} p(z_{it} = k|\mathbf{x}_i, \boldsymbol{\theta}^{\text{old}}) \log p(\mathbf{x}_{it}|\boldsymbol{\phi}_k) \\
&= \sum_{i=1}^{N}\sum_{t=1}^{T_i}\sum_{k=1}^{K}\sum_{m=1}^{M_k} \gamma_{i,t}(k)\delta_{i,t}(k,m) \log w_{km} \\
&+ \sum_{i=1}^{N}\sum_{t=1}^{T_i}\sum_{k=1}^{K}\sum_{m=1}^{M_k} \gamma_{i,t}(k)\delta_{i,t}(k,m) \log \mathcal{N}(\mathbf{x}_{it}|\boldsymbol{\mu}_{km}, \mathbf{\Sigma}_{km}).
\end{aligned}
$$

E-step:
The smoothed node and edge marginals can be estimated in the same manner in Section 17.5.2.1, running the forwards-backwards algorithm on each sequence.
The marginals for mixture component transition is computed as:

$$
\hat{\delta}_{i,t}(k,m) = \frac{w_{km}\mathcal{N}(\mathbf{x}_{it}|\boldsymbol{\mu}_{km}, \mathbf{\Sigma}_{km})}{\sum_{m'=1}^{M_k} w_{km'}\mathcal{N}(\mathbf{x}_{it}|\boldsymbol{\mu}_{km'}, \mathbf{\Sigma}_{km'})}
$$

M-step:
The M step for $\mathbf{A}$ and $\boldsymbol{\pi}$ is already discussed in Section 17.5.2.2. For other variables, we get:

$$
\hat{w}_{km} = \frac{\sum_{i=1}^{N}\sum_{t=1}^{T_i}\sum_{k=1}^{K} \hat{\gamma}_{i,t}(k)\hat{\delta}_{i,t}(k,m)}{\sum_{i=1}^{N}\sum_{t=1}^{T_i}\sum_{k=1}^{K} \hat{\gamma}_{i,t}(k)}.
$$

$$
\hat{\boldsymbol{\mu}}_{km} = \frac{\sum_{i=1}^{N}\sum_{t=1}^{T_i}\sum_{k=1}^{K} \hat{\gamma}_{i,t}(k)\hat{\delta}_{i,t}(k,m)\mathbf{x}_{it}}{\sum_{i=1}^{N}\sum_{t=1}^{T_i}\sum_{k=1}^{K} \hat{\gamma}_{i,t}(k)\hat{\delta}_{i,t}(k,m)}.
$$

$$
\hat{\mathbf{\Sigma}}_{km} = \frac{\sum_{i=1}^{N}\sum_{t=1}^{T_i}\sum_{k=1}^{K} \hat{\gamma}_{i,t}(k)\hat{\delta}_{i,t}(k,m)(\mathbf{x}_{it} - \hat{\boldsymbol{\mu}}_{km})(\mathbf{x}_{it} - \hat{\boldsymbol{\mu}}_{km})^T}{\sum_{i=1}^{N}\sum_{t=1}^{T_i}\sum_{k=1}^{K} \hat{\gamma}_{i,t}(k)\hat{\delta}_{i,t}(k,m)}.
$$

17.4.
This time each of the emission distributions becomes:

$$p(\mathbf{x}_t|z_t = k, \boldsymbol{\theta}) = \sum_{m=1}^{M} w_m \mathcal{N}(\mathbf{x}_t|\boldsymbol{\mu}_m, \boldsymbol{\Sigma}_m).$$

The expected complete data log likelihood is:

$$\begin{aligned} Q(\boldsymbol{\theta}, \boldsymbol{\theta}^{\text{old}}) = \sum_{k=1}^{K} \mathbb{E}[N_k^1] \log \pi_k + \sum_{j=1}^{K} \sum_{k=1}^{K} \mathbb{E}[N_{jk}] \log A_{jk} \\ + \sum_{i=1}^{N} \sum_{t=1}^{T_i} \sum_{k=1}^{K} \sum_{m=1}^{M_k} \gamma_{i,t}(k) \delta_{i,t}(m) \log w_m \\ + \sum_{i=1}^{N} \sum_{t=1}^{T_i} \sum_{k=1}^{K} \sum_{m=1}^{M_k} \gamma_{i,t}(k) \delta_{i,t}(m) \log \mathcal{N}(\mathbf{x}_{it}|\boldsymbol{\mu}_m, \boldsymbol{\Sigma}_m). \end{aligned}$$

E-step:

$$\hat{\delta}_{i,t}(m) = \frac{w_m \mathcal{N}(\mathbf{x}_{it}|\boldsymbol{\mu}_m, \boldsymbol{\Sigma}_m)}{\sum_{m'=1}^{M} w_{m'} \mathcal{N}(\mathbf{x}_{it}|\boldsymbol{\mu}_{m'}, \boldsymbol{\Sigma}_{m'})}.$$

M-step:

$$\hat{w}_m = \frac{\sum_{i=1}^{N} \sum_{t=1}^{T_i} \hat{\delta}_{i,t}(m)}{\sum_{i=1}^{N} \sum_{t=1}^{T_i} 1}.$$

$$\hat{\boldsymbol{\mu}}_m = \frac{\sum_{i=1}^{N} \sum_{t=1}^{T_i} \hat{\delta}_{i,t}(m) \mathbf{x}_{it}}{\sum_{i=1}^{N} \sum_{t=1}^{T_i} \hat{\delta}_{i,t}(m)}.$$

$$\hat{\boldsymbol{\Sigma}}_m = \frac{\sum_{i=1}^{N} \sum_{t=1}^{T_i} \hat{\delta}_{i,t}(m) (\mathbf{x}_{it} - \hat{\boldsymbol{\mu}}_m)(\mathbf{x}_{it} - \hat{\boldsymbol{\mu}}_m)^T}{\sum_{i=1}^{N} \sum_{t=1}^{T_i} \hat{\delta}_{i,t}(m)}.$$