Yannic Neuhaus, Sabri Aidi

1. Write down the expression for the log joint distribution of the latent and observed variables, i.e.,

$$\log p(\{\mathbf{x}_n, z_n\}_{n=1}^N \mid \boldsymbol{\Theta}).$$

$$\log p(\{\mathbf{x}_n, z_n\}_{n=1}^N \mid \boldsymbol{\Theta})$$

$$= \log \prod_{i=1}^{N} p(\mathbf{x}_n, z_n \mid \boldsymbol{\Theta})$$

$$= \sum_{n=1}^{N} \log \left(\prod_{k=1}^{K} p(\mathbf{x}_n, z_n = k \mid \boldsymbol{\Theta})^{[z_n=k]} \right)$$

$$= \sum_{n=1}^{N} \log \left(\prod_{k=1}^{K} [p(\mathbf{x}_n \mid \boldsymbol{\theta}_k) p(z_n = k \mid \boldsymbol{\Theta})]^{[z_n=k]} \right) \qquad , p(z_n = k \mid \boldsymbol{\Theta}) = \pi_k$$

$$= \sum_{n=1}^{N} \sum_{k=1}^{K} \log \left([\pi_k p(\mathbf{x}_n \mid \boldsymbol{\theta}_k)]^{[z_n=k]} \right) \qquad , p(\mathbf{x}_n \mid \boldsymbol{\theta}_k) = \prod_{j=1}^{W_n} \mathrm{Cat}(x_{nj} \mid \boldsymbol{\theta}_k)$$

$$= \sum_{n=1}^{N} \sum_{k=1}^{K} [z_n = k] \log \left(\pi_k \prod_{j=1}^{W_n} \mathrm{Cat}(x_{nj} \mid \boldsymbol{\theta}_k) \right)$$

$$= \sum_{n=1}^{N} \sum_{k=1}^{K} [z_n = k] \left(\log \pi_k + \log \left[\prod_{j=1}^{W_n} \prod_{m=1}^{|I|} \theta_{km}^{[x_{nj}=m]} \right] \right)$$

$$= \sum_{n=1}^{N} \sum_{k=1}^{K} [z_n = k] \left(\log \pi_k + \sum_{j=1}^{W_n} \sum_{m=1}^{|I|} [x_{nj} = m] \log \theta_{km} \right)$$

$$= \sum_{n=1}^{N} \left(\log \pi_{z_n} + \sum_{j=1}^{W_n} \sum_{m=1}^{|I|} [x_{nj} = m] \log \theta_{z_n m} \right)$$

$$= \sum_{n=1}^{N} \left(\log \pi_{z_n} + \sum_{j=1}^{W_n} \log \theta_{z_n x_{nj}} \right)$$

2. Compute the closed-form expression for the E-step, i.e.,

$$Q(\boldsymbol{\Theta}, \boldsymbol{\Theta}^{\text{old}})$$

Hint:

$$Q(\boldsymbol{\Theta}, \boldsymbol{\Theta}^{\text{old}}) = \sum_{n=1}^{N} \sum_{k=1}^{K} p(z_n = k \mid \mathbf{x_n}, \boldsymbol{\Theta}^{\text{old}}) \log p(\mathbf{x_n}, z_n = k \mid \boldsymbol{\Theta}) \approx \log p(\{\mathbf{x}_n\}_{n=1}^{N} \mid \boldsymbol{\Theta})$$

$$\begin{aligned}
p(z_n = k \mid \mathbf{x}_n, \boldsymbol{\Theta}^{\text{old}}) &= \frac{p(\mathbf{x}_n \mid z_n = k, \boldsymbol{\Theta}^{\text{old}})}{p(\mathbf{x}_n \mid \boldsymbol{\Theta}^{\text{old}})} \\
&= \frac{\pi_k p(\mathbf{x}_n \mid \boldsymbol{\theta}_k^{\text{old}})}{\sum_{l=1}^{K} \pi_l p(\mathbf{x}_n \mid \boldsymbol{\theta}_l^{\text{old}})} \\
&= \frac{\pi_k^{\text{old}} \prod_{i=1}^{W_n} \text{Cat}(x_{ni} \mid \boldsymbol{\theta}_k^{\text{old}})}{\sum_{l=1}^{K} \pi_l^{\text{old}} \prod_{j=1}^{W_n} \text{Cat}(x_{nj} \mid \boldsymbol{\theta}_l^{\text{old}})} \\
&= \frac{\pi_k^{\text{old}} \prod_{i=1}^{W_n} \prod_{m=1}^{|I|} \theta_{km}^{\text{old}[x_{ni}=m]}}{\sum_{l=1}^{K} \pi_l^{\text{old}} \prod_{j=1}^{W_n} \prod_{m'=1}^{|I|} \theta_{lm'}^{\text{old}[x_{nj}=m']}} \\
&= \frac{\pi_k^{\text{old}} \prod_{i=1}^{W_n} \theta_{kx_{ni}}^{\text{old}}}{\sum_{l=1}^{K} \pi_l^{\text{old}} \prod_{j=1}^{W_n} \theta_{lx_{nj}}^{\text{old}}}
\end{aligned}$$

$$\begin{aligned}
\log p(\mathbf{x}_n, z_n = k \mid \boldsymbol{\Theta}) &= \log \left[\pi_k p(\mathbf{x}_n \mid \boldsymbol{\theta}_k)\right] \\
&= \log \pi_k + \log p(\mathbf{x}_n \mid \boldsymbol{\theta}_k) \\
&= \log \pi_k + \sum_{j=1}^{W_n} \log \theta_{kx_{nj}}
\end{aligned}$$

$$\begin{aligned}
\Rightarrow Q(\boldsymbol{\Theta}, \boldsymbol{\Theta}^{\text{old}}) &= \sum_{n=1}^{N} \sum_{k=1}^{K} p(z_n = k \mid \mathbf{x}_n, \boldsymbol{\Theta}^{\text{old}}) \log p(\mathbf{x}_n, z_n = k \mid \boldsymbol{\Theta}) \\
&= \sum_{n=1}^{N} \sum_{k=1}^{K} \frac{\pi_k^{\text{old}} \prod_{i=1}^{W_n} \theta_{kx_{ni}}^{\text{old}}}{\sum_{l=1}^{K} \pi_l^{\text{old}} \prod_{j=1}^{W_n} \theta_{lx_{nj}}^{\text{old}}} \left(\log \pi_k + \sum_{j=1}^{W_n} \log \theta_{kx_{nj}} \right)
\end{aligned}$$

3. M step: Derive the expression of the MLE for the model parameter $\mathbf{\Theta} = \{\pi, \{\theta_k\}\}$

$\underline{\pi_k^{ML}:}$

Lagrangian and derivative:

$$Q_\pi(\mathbf{\Theta}, \mathbf{\Theta}^{\text{old}}) := Q(\mathbf{\Theta}, \mathbf{\Theta}^{\text{old}}) + \lambda \left(\sum_{k=1}^{K} \pi_k - 1 \right)$$

$$\begin{aligned}
\frac{\partial Q(\mathbf{\Theta}, \mathbf{\Theta}^{\text{old}})}{\partial \pi_k} &= \sum_{n=1}^{N} \frac{\partial}{\partial \pi_k} \sum_{l=1}^{K} \big(p(z_n = l \mid \mathbf{x}_n, \mathbf{\Theta}^{\text{old}}) \log \pi_l \big) \\
&= \sum_{n=1}^{N} \frac{\partial}{\partial \pi_k} \big(p(z_n = k \mid \mathbf{x}_n, \mathbf{\Theta}^{\text{old}}) \log \pi_k \big) \\
&= \sum_{n=1}^{N} p(z_n = k \mid \mathbf{x}_n, \mathbf{\Theta}^{\text{old}}) \frac{\partial}{\partial \pi_k} \log \pi_k \\
&= \sum_{n=1}^{N} p(z_n = k \mid \mathbf{x}_n, \mathbf{\Theta}^{\text{old}}) \frac{1}{\pi_k} \\
&= \frac{N_k}{\pi_k} \qquad \text{, where } N_k = \sum_{n=1}^{N} p(z_n = k \mid \mathbf{x}_n, \mathbf{\Theta}^{\text{old}})
\end{aligned}$$

$$\Rightarrow \frac{\partial Q_\pi(\mathbf{\Theta}, \mathbf{\Theta}^{\text{old}})}{\partial \pi_k} = \frac{N_k}{\pi_k} + \lambda$$

Critical points:

$$\frac{\partial Q_\pi(\mathbf{\Theta}, \mathbf{\Theta}^{\text{old}})}{\partial \pi_k} = \frac{N_k}{\pi_k} + \lambda \stackrel{!}{=} 0$$

$$\Rightarrow \frac{N_k}{\pi_k} + \lambda = 0 \quad \Rightarrow \quad N_k + \lambda \pi_k = 0 \quad \Rightarrow \quad \lambda \pi_k = -N_k \quad \Rightarrow \quad \lambda \sum_{k=1}^{K} \pi_k = -\sum_{k=1}^{K} N_k$$

$$\Rightarrow \lambda = -N \quad \text{, since } \sum_{k=1}^{K} \pi_k = 1 \text{ and } \sum_{k=1}^{K} N_k = \sum_{k=1}^{K} \sum_{n=1}^{N} p(z_n = k \mid \mathbf{x}_n, \mathbf{\Theta}^{\text{old}}) = \sum_{n=1}^{N} 1 = N$$

$$\frac{N_k}{\pi_k} - N = 0 \quad \Rightarrow \quad \frac{N_k}{\pi_k} = N \quad \Rightarrow \quad \pi_k N = N_k \quad \Rightarrow \quad \pi_k = \frac{N_k}{N}$$

$$\Rightarrow \pi_k^{ML} = \frac{N_k}{N}$$

$\underline{\theta_{km}^{ML}:}$

Lagrangian and derivative:

$$Q_\theta(\boldsymbol{\Theta}, \boldsymbol{\Theta}^{\text{old}}) := Q(\boldsymbol{\Theta}, \boldsymbol{\Theta}^{\text{old}}) + \lambda \left(\sum_{m=1}^{|I|} \theta_{km} - 1 \right)$$

$$\frac{\partial Q(\boldsymbol{\Theta}, \boldsymbol{\Theta}^{\text{old}})}{\partial \theta_{km}} = \sum_{n=1}^{N} \frac{\partial}{\partial \theta_{km}} \left(\sum_{k'=1}^{K} p(z_n = k' \mid \mathbf{x}_n, \boldsymbol{\Theta}^{\text{old}}) \sum_{i=1}^{W_n} \log \theta_{k' x_{ni}} \right)$$

$$= \sum_{n=1}^{N} p(z_n = k \mid \mathbf{x}_n, \boldsymbol{\Theta}^{\text{old}}) \frac{\partial}{\partial \theta_{km}} \left(\sum_{i=1}^{W_n} \log \theta_{k x_{ni}} \right)$$

$$= \sum_{n=1}^{N} p(z_n = k \mid \mathbf{x}_n, \boldsymbol{\Theta}^{\text{old}}) \sum_{i=1}^{W_n} [x_{ni} = m] \frac{1}{\theta_{k x_{ni}}}$$

$$\text{, because for } x_{ni} \neq m : \frac{\partial}{\partial \theta_{km}} \left(\sum_{i=1}^{W_n} \log \theta_{k x_{ni}} \right) = 0$$

$$= \sum_{n=1}^{N} \frac{p(z_n = k \mid \mathbf{x}_n, \boldsymbol{\Theta}^{\text{old}}) c_{nm}}{\theta_{km}} \qquad \text{, where } c_{nm} = \sum_{i=1}^{W_n} [x_{in} = m]$$

$$= \frac{1}{\theta_{km}} \sum_{n=1}^{N} p(z_n = k \mid \mathbf{x}_n, \boldsymbol{\Theta}^{\text{old}}) c_{nm}$$

$$= \frac{\tilde{c}_{km}}{\theta_{km}} \qquad \text{, where } \tilde{c}_{km} = \sum_{n=1}^{N} p(z_n = k \mid \mathbf{x}_n, \boldsymbol{\Theta}^{\text{old}}) c_{nm}$$

Critical points:

$$\frac{\partial Q_\theta(\mathbf{\Theta}, \mathbf{\Theta}^{\text{old}})}{\partial \theta_{km}} = \frac{\tilde{c}_{km}}{\theta_{km}} + \lambda \stackrel{!}{=} 0$$

$$\begin{aligned}
\Rightarrow \frac{\tilde{c}_{km}}{\theta km} + \lambda = 0 \quad &\Rightarrow \quad \tilde{c}_{km} + \lambda\theta_{km} = 0 \quad \Rightarrow \quad \lambda\theta_{km} = -\tilde{c}_{km} \\
\Rightarrow \lambda \sum_{m=1}^{|I|} \theta_{km} &= -\sum_{m=1}^{|I|} \tilde{c}_{km} \\
\Rightarrow \lambda &= -\sum_{m=1}^{|I|} \tilde{c}_{km} = -\sum_{m=1}^{|I|} \sum_{n=1}^{N} p(z_n = k \mid \mathbf{x}_n, \mathbf{\Theta}^{\text{old}}) \sum_{i=1}^{W_n} [x_{ni} = m] \\
&= -\sum_{n=1}^{N} p(z_n = k \mid \mathbf{x}_n, \mathbf{\Theta}^{\text{old}}) \sum_{i=1}^{W_n} \sum_{m=1}^{|I|} [x_{ni} = m] \\
&= -\sum_{n=1}^{N} p(z_n = k \mid \mathbf{x}_n, \mathbf{\Theta}^{\text{old}}) W_n \quad \text{, since } \sum_{m=1}^{|I|} [x_{ni} = m] = 1 \\
&= -\tilde{c}_k \qquad \text{, where } \tilde{c}_k = \sum_{n=1}^{N} p(z_n = k \mid \mathbf{x}_n, \mathbf{\Theta}^{\text{old}}) W_n
\end{aligned}$$

$$\frac{\tilde{c}_{km}}{\theta_{km}} - \tilde{c}_k = 0 \quad \Rightarrow \quad \theta_{km}\tilde{c}_k = \tilde{c}_{km} \quad \Rightarrow \quad \theta_{km} = \frac{\tilde{c}_{km}}{\tilde{c}_k}$$