ECE 537 Fundamentals of Speech Processing
Problem Set 6

UNIVERSITY OF ILLINOIS
Department of Electrical and Computer Engineering

Assigned: Wednesday, 10/12/2022; Due: Wednesday, 10/19/2022
Reading: Rabiner, "A tutorial on hidden Markov models and selected applications in speech recognition," 1989

1. Consider the problem of learning an HMM with mixture Gaussian observation probabilities, i.e., the type of probability density function shown in Eq. (49) of the article. Plug the article's Eq. (49) into Eq. (41), use Eq. (50) to devise an appropriate Lagrangian, and show that Eq. (52) is the result. Note: if you use Eq. (41) as the definition of Baum's auxiliary function, you will wind up with the following definition of $\gamma_t(j,k)$:
$$\gamma_t(j,k) = \gamma_t(j)\left[\frac{\bar{c}_{jk}\mathcal{N}(o_t,\bar{\mu}_{jk},\bar{U}_{jk})}{\sum_{m=1}^{M}\bar{c}_{jm}\mathcal{N}(o_t,\bar{\mu}_{jm},\bar{U}_{jm})}\right] \tag{1}$$
The article assumes, for this section only, a slightly different definition of Baum's auxiliary function, which is described in more detail in reference [35] of the article. That slightly different definition of Baum's auxiliary gives the definition of $\gamma_t(j,k)$ shown in the article:
$$\gamma_t(j,k) = \gamma_t(j)\left[\frac{c_{jk}\mathcal{N}(o_t,\mu_{jk},U_{jk})}{\sum_{m=1}^{M}c_{jm}\mathcal{N}(o_t,\mu_{jm},U_{jm})}\right] \tag{2}$$
Eq. (1) is the one you should find in your solution to this problem, but Eq. (2) is actually easier to use in practice.

Solution:
$$\begin{aligned}
Q(\lambda,\bar{\lambda}) &= \sum_{Q} P(Q|O,\lambda)\log P(O,Q|\bar{\lambda}) \\
&= \sum_{t=1}^{T}\sum_{j=1}^{N} P(q_t=j|O,\lambda)\log \bar{b}_j(o_t) + \text{other terms} \\
&= \sum_{t=1}^{T}\sum_{j=1}^{N} \gamma_t(j)\log \sum_{m=1}^{M} c_{jm}\mathcal{N}\left[o_t,\mu_{jm},U_{jm}\right] + \text{other terms}
\end{aligned}$$
We can therefore devise the Lagrangian as
$$\begin{aligned}
J(\lambda,\bar{\lambda}) = \sum_{t=1}^{T}\sum_{j=1}^{N} \gamma_t(j)\log \sum_{m=1}^{M} \bar{c}_{jm}\mathcal{N}\left[o_t,\bar{\mu}_{jm},\bar{U}_{jm}\right] \\
+ \sum_{j}\nu_j\left(1-\sum_{m=1}^{M}\bar{c}_{jm}\right) + \text{other terms}
\end{aligned}$$

Differentiating w.r.t. $\bar{c}_{jm}$, we get

$$\frac{\partial J(\lambda,\bar{\lambda})}{\partial \bar{c}_{jm}} = \sum_{t=1}^{T} \gamma_t(j) \left(\frac{\mathcal{N}\left[o_t, \bar{\mu}_{jm}, \bar{U}_{jm}\right]}{\sum_{k=1}^{M} \bar{c}_{jk} \mathcal{N}\left[o_t, \bar{\mu}_{jk}, \bar{U}_{jk}\right]} \right) - \nu_j$$

Setting that equal to zero, we find that

$$\begin{aligned}
\nu_j &= \sum_{t=1}^{T} \gamma_t(j) \left(\frac{\mathcal{N}\left[o_t, \bar{\mu}_{jm}, \bar{U}_{jm}\right]}{\sum_{k=1}^{M} \bar{c}_{jk} \mathcal{N}\left[o_t, \bar{\mu}_{jk}, \bar{U}_{jk}\right]} \right) \\
1 &= \frac{1}{\nu_j} \sum_{t=1}^{T} \gamma_t(j) \left(\frac{\mathcal{N}\left[o_t, \bar{\mu}_{jm}, \bar{U}_{jm}\right]}{\sum_{k=1}^{M} \bar{c}_{jk} \mathcal{N}\left[o_t, \bar{\mu}_{jk}, \bar{U}_{jk}\right]} \right) \\
\bar{c}_{jm} &= \frac{1}{\nu_j} \sum_{t=1}^{T} \gamma_t(j) \left(\bar{c}_{jm} \frac{\mathcal{N}\left[o_t, \bar{\mu}_{jm}, \bar{U}_{jm}\right]}{\sum_{k=1}^{M} \bar{c}_{jk} \mathcal{N}\left[o_t, \bar{\mu}_{jk}, \bar{U}_{jk}\right]} \right) \\
&= \frac{\gamma_t(j,m)}{\nu_j}
\end{aligned}$$

All that remains is to choose ν_j in order to satisfy the constraint, giving

$$\bar{c}_{jm} = \frac{\gamma_t(j,m)}{\sum_{k=1}^{M} \gamma_t(j,k)}$$

2. Suppose that Eqs. (96) and (97) in the article are true, but suppose that c_s is not defined as in Eq. (91); instead, suppose that c_s is just some arbitrary constant that depends on the time index, s. Prove that, even under this relaxed assumption, Eq. (95) is still equal to Eq. (40b).

Solution:

$$\begin{aligned}
\bar{a}_{ij} &= \frac{\sum_{t=1}^{T-1} \hat{\alpha}_t(i) a_{ij} b_j(o_{t+1}) \hat{\beta}_{t+1}(j)}{\sum_{t=1}^{T-1} \sum_{j=1}^{N} \hat{\alpha}_t(i) a_{ij} b_j(o_{t+1}) \hat{\beta}_{t+1}(j)} \\
&= \frac{\sum_{t=1}^{T-1} C_t \alpha_t(i) a_{ij} b_j(o_{t+1}) D_{t+1} \beta_{t+1}(j)}{\sum_{t=1}^{T-1} \sum_{j=1}^{N} C_t \alpha_t(i) a_{ij} b_j(o_{t+1}) D_{t+1} \beta_{t+1}(j)} \\
&= \frac{\left(\prod_{s=t}^{T} c_s\right) \sum_{t=1}^{T-1} \alpha_t(i) a_{ij} b_j(o_{t+1}) \beta_{t+1}(j)}{\left(\prod_{s=1}^{T} c_s\right) \sum_{t=1}^{T-1} \sum_{j=1}^{N} \alpha_t(i) a_{ij} b_j(o_{t+1}) \beta_{t+1}(j)} \\
&= \frac{\sum_{t=1}^{T-1} \alpha_t(i) a_{ij} b_j(o_{t+1}) \beta_{t+1}(j)}{\sum_{t=1}^{T-1} \sum_{j=1}^{N} \alpha_t(i) a_{ij} b_j(o_{t+1}) \beta_{t+1}(j)} \\
&= \frac{\sum_{t=1}^{T-1} \xi_t(i,j)}{\sum_{t=1}^{T-1} \sum_{j=1}^{N} \xi_t(i,j)} \\
&= \frac{\sum_{t=1}^{T-1} \xi_t(i,j)}{\sum_{t=1}^{T-1} \gamma_t(i)}
\end{aligned}$$