אוניברסיטת ת"א, ביה"ס להנדסת חשמל, למידת מכונה סטטיסטית

פתרון תרגיל בית 8

שאלה 1

אין שינוי בהגדרת ה- complete data לעומת האלגוריתם שלמדנו בכיתה. נסמן ב- M את מספר המצבים השונים במודל. אזי,

$$\log P\left(\mathbf{x}_1,\mathbf{S}_1,\mathbf{x}_2,\mathbf{S}_2,\ldots,\mathbf{x}_L,\mathbf{S}_L\left|\boldsymbol{\theta}\right.\right)=\sum_{i=1}^{L}\log P\left(\mathbf{x}_i,\mathbf{S}_i\left|\boldsymbol{\theta}\right.\right)=\sum_{i=1}^{L}\sum_{j=1}^{M}I_{\left\{S_{i,1}=j\right\}}\log\lambda_j+$$

$$\sum_{i=1}^{L}\sum_{t=2}^{T_i}\sum_{j=1}^{M}\sum_{k=1}^{M}I_{\left\{S_{i,t-1}=k,S_{i,t}=j\right\}}\log A\left(k,j\right)+\sum_{i=1}^{L}\sum_{t=1}^{T_i}\sum_{j=1}^{M}I_{\left\{S_{i,t}=j\right\}}\log\left[q_j^{x_{i,t}}\left(1-q_j\right)^{1-x_{i,t}}\right]$$

Hence,

$$Q\left(\boldsymbol{\theta}\middle|\boldsymbol{\theta}_0\right)=\sum_{i=1}^{L}E\left\{\log P\left(\mathbf{x}_i,\mathbf{S}_i\left|\boldsymbol{\theta}\right.\right)\middle|\mathbf{x}_i,\boldsymbol{\theta}_0\right\}=\sum_{i=1}^{L}\sum_{j=1}^{M}w_i\left(j,1\right)\log\lambda_j+$$

$$\sum_{i=1}^{L}\sum_{t=2}^{T_i}\sum_{j=1}^{M}\sum_{k=1}^{M}u_i\left(k,j,t\right)\log A\left(k,j\right)+\sum_{i=1}^{L}\sum_{t=1}^{T_i}\sum_{j=1}^{M}\left[x_{i,t}\log q_j+\left(1-x_{i,t}\right)\log\left(1-q_j\right)\right]w_i\left(j,t\right)$$

כאשר

$$w_i\left(j,t\right)=P\left(S_{i,t}=j\middle|\mathbf{x}_i,\boldsymbol{\theta}_0\right)$$
$$u_i\left(k,j,t\right)=P\left(S_{i,t-1}=k,S_{i,t}=j\middle|\mathbf{x}_i,\boldsymbol{\theta}_0\right)$$

לכן, כמו באלגוריתם שפתחנו בכיתה,

$$\hat{\lambda}_j=\frac{\sum_{i=1}^{L}w_i\left(j,1\right)}{L}$$

$$\hat{A}\left(k,j\right)=\frac{\sum_{i=1}^{L}\sum_{t=2}^{T_i}u_i\left(k,j,t\right)}{\sum_{i=1}^{L}\sum_{t=2}^{T_i}w_i\left(k,t-1\right)}$$

ואילו עבור q_j נקבל באופן דומה,

$$\hat{q}_j=\arg\max_{q_j}\left\{\sum_{i=1}^{L}\sum_{t=1}^{T_i}\left[x_{i,t}\log q_j+\left(1-x_{i,t}\right)\log\left(1-q_j\right)\right]w_i\left(j,t\right)\right\}$$

$$=\arg\max_{q_j}\left\{\log q_j\cdot\sum_{i=1}^{L}\sum_{t=1}^{T_i}x_{i,t}w_i\left(j,t\right)+\log\left(1-q_j\right)\cdot\sum_{i=1}^{L}\sum_{t=1}^{T_i}\left(1-x_{i,t}\right)w_i\left(j,t\right)\right\}$$

$$=\frac{\sum_{i=1}^{L}\sum_{t=1}^{T_i}x_{i,t}w_i\left(j,t\right)}{\sum_{i=1}^{L}\sum_{t=1}^{T_i}x_{i,t}w_i\left(j,t\right)+\sum_{i=1}^{L}\sum_{t=1}^{T_i}\left(1-x_{i,t}\right)w_i\left(j,t\right)}=\frac{\sum_{i=1}^{L}\sum_{t=1}^{T_i}x_{i,t}w_i\left(j,t\right)}{\sum_{i=1}^{L}\sum_{t=1}^{T_i}w_i\left(j,t\right)}$$

חישוב $w_i(j,t)$ ו- $u_i(k,j,t)$ יעשה באמצעות אלגוריתם forward-backward שלמדנו בכיתה, פרט לעובדה שיש להשתמש בפילוג המוצא הבדיד במקום פילוג המוצא הגאוסי הרציף שבו השתמשנו בכיתה.

<u>שאלה 2</u>

כאמור ברמז, אפשר לראות את בעית השערוך כאילו לפנינו שתי מילים מחוברות: האחת היא "one two" שממנה יש לנו L_1 הגיות והשניה היא "two one" שממנה יש לנו L_2 הגיות. כמו כן יש לנו אילוצי שיתוף (tying) בין הפרמטרים של שתי המילים המחוברות הללו כך שיש לנו רק שישה פילוגים שיש לשערך, וכ"א מהם מיוצג ע"י וקטור תוחלת ומטריצת שונות. המצבים של המילה הראשונה הם: "S","O","W","T","E","N","O","S". המצבים של המילה השניה הם "S","E","N","O","O","W","T","S". נסמן ב- j את אינדקס המצב במילה. אזי, עבור שתי המילים המחוברות, נסמן ב- $C_1(j)$ ($C_2(j)$, בהתאמה), $j=1,2,\dots,8$, את המיפוי מאינדקס המצב למצב במילה המחוברת הראשונה (שנייה), כך ש-

$$
\begin{aligned}
&C_1(1)="S",\ C_1(2)="O",C_1(3)="N",C_1(4)="E",\\
&\qquad C_1(5)="T",C_1(6)="W",C_1(7)="O",C_1(8)="S"\\
&C_2(1)="S",\ C_2(2)="T",C_2(3)="W",C_2(4)="O",\\
&\qquad C_2(5)="O",C_2(6)="N",C_2(7)="E",C_2(8)="S"
\end{aligned}
$$

נסמן את ההגיות הנתונות מהמילה המחוברת הראשונה (שניה, בהתאמה) ב- $\left\{\mathbf{x}_1^1,\mathbf{x}_2^1,\dots,\mathbf{x}_{L_1}^1\right\}$ ($\left\{\mathbf{x}_1^2,\mathbf{x}_2^2,\dots,\mathbf{x}_{L_2}^2\right\}$) עבור $\mathbf{x}_i^m=\left(\mathbf{x}_{i,1}^m,\dots,\mathbf{x}_{i,T_i^m}^m\right),\quad m\in\{1,2\},\quad i\in\{1,\dots,L_m\}$. סדרת המצבים המתאימה ל- $\mathbf{x}_i^m$ היא $\mathbf{s}_i^m=\left(s_{i,1}^m,\dots,s_{i,T_i^m}^m\right)$ כאשר $s_{i,t}^m\in\{1,2,\dots,8\}$ מציין את אינדקס המצב במילה המחוברת $m\in\{1,2\}$. ה- log-likelihood של ה- complete data נתון ע"י

$$
\begin{aligned}
&\log P\left(\mathbf{x}_1^1,\mathbf{x}_2^1,\dots,\mathbf{x}_{L_1}^1,S_1^1,S_2^1,\dots,S_{L_1}^1,\mathbf{x}_1^2,\mathbf{x}_2^2,\dots,\mathbf{x}_{L_2}^2,S_1^2,S_2^2,\dots,S_{L_2}^2\right)=\\
&\quad-\frac{1}{2}\sum_{i=1}^{L_1}\sum_{t=1}^{T_i^1}\sum_{j=1}^{8}1_{\left\{s_{i,t}^1=j\right\}}\left[\left(\mathbf{x}_{i,t}^1-\boldsymbol{\mu}_{C_1(j)}\right)^T\Sigma_{C_1(j)}^{-1}\left(\mathbf{x}_{i,t}^1-\boldsymbol{\mu}_{C_1(j)}\right)+\log\left|\Sigma_{C_1(j)}\right|\right]\\
&\quad-\frac{1}{2}\sum_{i=1}^{L_2}\sum_{t=1}^{T_i^2}\sum_{j=1}^{8}1_{\left\{s_{i,t}^2=j\right\}}\left[\left(\mathbf{x}_{i,t}^2-\boldsymbol{\mu}_{C_2(j)}\right)^T\Sigma_{C_2(j)}^{-1}\left(\mathbf{x}_{i,t}^2-\boldsymbol{\mu}_{C_2(j)}\right)+\log\left|\Sigma_{C_2(j)}\right|\right]+\text{Const}
\end{aligned}
$$

כאשר Const לא תלוי בפרמטרים הנעלמים, $\boldsymbol{\mu}_l,\Sigma_l$ עבור $l\in\{"S","O","N","E","T","W"\}$. מכאן:

$$
\begin{aligned}
\mathrm{Q}\left(\boldsymbol{\theta}\middle|\boldsymbol{\theta}_0\right)=&-\frac{1}{2}\sum_{i=1}^{L_1}\sum_{t=1}^{T_i^1}\sum_{j=1}^{8}w_i^1(j,t)\left[\left(\mathbf{x}_{i,t}^1-\boldsymbol{\mu}_{C_1(j)}\right)^T\Sigma_{C_1(j)}^{-1}\left(\mathbf{x}_{i,t}^1-\boldsymbol{\mu}_{C_1(j)}\right)+\log\left|\Sigma_{C_1(j)}\right|\right]\\
&-\frac{1}{2}\sum_{i=1}^{L_2}\sum_{t=1}^{T_i^2}\sum_{j=1}^{8}w_i^2(j,t)\left[\left(\mathbf{x}_{i,t}^2-\boldsymbol{\mu}_{C_2(j)}\right)^T\Sigma_{C_2(j)}^{-1}\left(\mathbf{x}_{i,t}^2-\boldsymbol{\mu}_{C_2(j)}\right)+\log\left|\Sigma_{C_2(j)}\right|\right]+\text{Const}
\end{aligned}
$$

כאשר $\boldsymbol{\theta}_0$ הוא וקטור הפרמטרים לפני האיטרציה ו- $\boldsymbol{\theta}=\left\{\boldsymbol{\mu}_l,\Sigma_l\right\}$ וקטור הפרמטרים החדש שיש לשערך. כמו כן, $w_i^m\left(j,t\right)=P\left(S_{i,t}^m=j\middle|\mathbf{x}_i,\boldsymbol{\theta}_0\right)$. עכשיו, כמו בפיתוח הרגיל, נוכל לחשב את $\boldsymbol{\mu}_l,\Sigma_l$ מתוך הדרישה שהנגזרות של $\mathrm{Q}\left(\boldsymbol{\theta}\middle|\boldsymbol{\theta}_0\right)$ ביחס לפרמטרים הללו יתאפסו. נקבל:

$$\hat{\boldsymbol{\mu}}_l=\frac{\sum_{i=1}^{L_1}\sum_{t=1}^{T_i^1}\sum_{j\,:\,C_1(j)=l}w_i^1\left(j,t\right)\mathbf{x}_{i,t}^1+\sum_{i=1}^{L_2}\sum_{t=1}^{T_i^2}\sum_{j\,:\,C_2(j)=l}w_i^2\left(j,t\right)\mathbf{x}_{i,t}^2}{\sum_{i=1}^{L_1}\sum_{t=1}^{T_i^1}\sum_{j\,:\,C_1(j)=l}w_i^1\left(j,t\right)+\sum_{i=1}^{L_2}\sum_{t=1}^{T_i^2}\sum_{j\,:\,C_2(j)=l}w_i^2\left(j,t\right)}$$

$$\hat{\Sigma}_l=\frac{\sum_{i=1}^{L_1}\sum_{t=1}^{T_i^1}\sum_{j\,:\,C_1(j)=l}w_i^1\left(j,t\right)\left(\mathbf{x}_{i,t}^1-\hat{\boldsymbol{\mu}}_l\right)\left(\mathbf{x}_{i,t}^1-\hat{\boldsymbol{\mu}}_l\right)^T+\sum_{i=1}^{L_2}\sum_{t=1}^{T_i^2}\sum_{j\,:\,C_2(j)=l}w_i^2\left(j,t\right)\left(\mathbf{x}_{i,t}^2-\hat{\boldsymbol{\mu}}_l\right)\left(\mathbf{x}_{i,t}^2-\hat{\boldsymbol{\mu}}_l\right)^T}{\sum_{i=1}^{L_1}\sum_{t=1}^{T_i^1}\sum_{j\,:\,C_1(j)=l}w_i^1\left(j,t\right)+\sum_{i=1}^{L_2}\sum_{t=1}^{T_i^2}\sum_{j\,:\,C_2(j)=l}w_i^2\left(j,t\right)}$$

את $w_i^m\left(j,t\right)$ נחשב תוך שימוש באלגוריתם forward-backward בדומה למקרה הרגיל שלמדנו בכיתה. ההבדל היחיד הוא שעכשיו אנחנו יודעים שכל הגייה חייבת להסתיים במצב שקט. לכן ההתחול של ה- β-ות יהיה כזה: $\beta_T(j)=1$ עבור j מצב השקט בסוף מילה, ו- $\beta_T(k)=0$ עבור כל מצב k אחר.

<u>שאלה 3</u>

נסמן $c_t=1/\sum_{l=1}^{M}\tilde{\alpha}_t\left(l\right)$. אזי קל לראות באינדוקציה שמתקיים $\hat{\alpha}_t\left(j\right)=\left(\prod_{l=1}^{t}c_l\right)\alpha_t\left(j\right)$ עבור $t=1,\ldots,T$ ו- $j=1,\ldots,M$. באופן דומה, קל לראות באינדוקציה שמתקיים $\hat{\beta}_t\left(j\right)=\left(\prod_{l=t}^{T}c_l\right)\beta_t\left(j\right)$ עבור $t=1,\ldots,T$ ו- $j=1,\ldots,M$. לכן $\hat{\alpha}_{t-1}\left(k\right)\hat{\beta}_t\left(j\right)=\alpha_{t-1}\left(k\right)\beta_t\left(j\right)\left(\prod_{l=1}^{T}c_l\right)$. מכאן,

$$\hat{A}\left(k,j\right)=\frac{\sum_{t=2}^{T}\alpha_{t-1}\left(k\right)\beta_t\left(j\right)A_{\boldsymbol{\theta}_0}\left(k,j\right)f\left(x_t\middle|S_t=j,\boldsymbol{\theta}_0\right)}{\sum_{l=1}^{M}\sum_{t=2}^{T}\alpha_{t-1}\left(k\right)\beta_t\left(l\right)A_{\boldsymbol{\theta}_0}\left(k,l\right)f\left(x_t\middle|S_t=l,\boldsymbol{\theta}_0\right)}$$
$$=\frac{\sum_{t=2}^{T}\hat{\alpha}_{t-1}\left(k\right)\hat{\beta}_t\left(j\right)A_{\boldsymbol{\theta}_0}\left(k,j\right)f\left(x_t\middle|S_t=j,\boldsymbol{\theta}_0\right)}{\sum_{l=1}^{M}\sum_{t=2}^{T}\hat{\alpha}_{t-1}\left(k\right)\hat{\beta}_t\left(l\right)A_{\boldsymbol{\theta}_0}\left(k,l\right)f\left(x_t\middle|S_t=l,\boldsymbol{\theta}_0\right)}$$

באופן דומה ניתן לנרמל את שאר נוסחאות העדכון של אלגוריתם ה- EM.