Chapter 18

stevenjin8

June 13, 2021

Exercises

Exercise 1

We first find the E step. The complete data log likelihood is given by

$$
\begin{aligned}
Q(\boldsymbol{\theta}, \boldsymbol{\theta}^{\text{old}}) &= \mathbb{E}\left[\sum_i \sum_t \log p(\mathbf{y}_{i,t}, \mathbf{z}_{i,t} | \mathbf{u}_{i,t}, \mathbf{z}_{i,t-1}) \middle| \mathbf{y}_i\right] \\
&= \sum_i \sum_t \mathbb{E}[\log p(\mathbf{y}_{i,t} | \mathbf{z}_{i,t}, \mathbf{u}_{i,t}) + \log p(\mathbf{z}_{i,t} | \mathbf{z}_{i,t-1}, \mathbf{u}_{i,t})] \\
&= \sum_i \sum_t \mathbb{E}[\log \mathcal{N}(\mathbf{y}_{i,t} | \mathbf{C}\mathbf{z}_{i,t} + \mathbf{D}\mathbf{u}_{i,t}, \mathbf{R}) + \log p(\mathbf{z}_{i,t} | \mathbf{A}\mathbf{z}_{i,t-1} + \mathbf{B}\mathbf{u}_{i,t}, \mathbf{Q})].
\end{aligned}
$$

For the M step, we can optimize with respect to each parameter. We will do so alphabetically:

$$
\begin{aligned}
\partial Q &= \sum_i \sum_t \mathbb{E}[\partial \log \mathcal{N}(\mathbf{z}_{i,t} | \mathbf{A}\mathbf{z}_{i,t-1} + \mathbf{B}\mathbf{u}_{i,t}, \mathbf{Q})] \\
&= \sum_i \sum_t \mathbb{E}\left[-\frac{1}{2}\partial\left((\mathbf{z}_{i,t} - \mathbf{A}\mathbf{z}_{i,t-1} - \mathbf{B}\mathbf{u}_{i,t})^T \mathbf{Q}^{-1} (\mathbf{z}_{i,t} - \mathbf{A}\mathbf{z}_{i,t-1} - \mathbf{B}\mathbf{u}_{i,t})\right)\right] \\
&= \sum_i \sum_t \mathbb{E}\left[-(\mathbf{z}_{i,t} - \mathbf{A}\mathbf{z}_{i,t-1} - \mathbf{B}\mathbf{u}_{i,t})^T \mathbf{Q}^{-1} \partial \mathbf{A}\mathbf{z}_{i,t-1}\right] \\
&= \sum_i \sum_t \operatorname{tr}\left[\mathbb{E}\left[-\mathbf{z}_{i,t-1}(\mathbf{z}_{i,t} - \mathbf{A}\mathbf{z}_{i,t-1} - \mathbf{B}\mathbf{u}_{i,t})^T\right] \mathbf{Q}^{-1} \partial \mathbf{A}\right] \\
\frac{\partial Q}{\partial \mathbf{A}} &= \sum_i \sum_t -\mathbf{Q}^{-1} \mathbb{E}\left[(\mathbf{z}_{i,t} - \mathbf{A}\mathbf{z}_{i,t-1} - \mathbf{B}\mathbf{u}_{i,t}) \mathbf{z}_{i,t-1}^T\right] \\
\mathbf{0} &= \sum_i \sum_t \mathbb{E}\left[\mathbf{z}_{i,t}\mathbf{z}_{i,t-1}^T\right] - \mathbf{A}\mathbb{E}\left[\mathbf{z}_{i,t-1}\mathbf{z}_{i,t-1}^T\right] - \mathbf{B}\mathbf{u}_{i,t}\mathbb{E}[\mathbf{z}_{i,t-1}^T] \\
\mathbf{A} &= \left(\sum_i \sum_t \mathbb{E}\left[\mathbf{z}_{i,t}\mathbf{z}_{i,t-1}^T\right] - \mathbf{B}\mathbf{u}_{i,t}\mathbb{E}\left[\mathbf{z}_{i,t-1}^T\right]\right)\left(\sum_i \sum_t \mathbb{E}\left[\mathbf{z}_{i,t-1}\mathbf{z}_{i,t-1}^T\right]\right)^{-1}.
\end{aligned}
$$

Doing the same thing for $\mathbf{B}$ gives

$$\begin{aligned}
\frac{\partial Q}{\partial \mathbf{B}} &= \sum_i \sum_t -\mathbf{Q}^{-1}\mathbb{E}\left[\left(\mathbf{z}_{i,t} - \mathbf{A}\mathbf{z}_{i,t-1} - \mathbf{B}\mathbf{u}_{i,t}\right)\mathbf{u}_{i,t}^T\right] \\
\mathbf{0} &= \sum_i \sum_t \mathbb{E}\left[\mathbf{z}_{i,t}\mathbf{u}_{i,t}\right] - \mathbf{A}\mathbb{E}[\mathbf{z}_{i,t-1}]\mathbf{u}_{i,t}^T - \mathbf{B}\mathbf{u}_{i,t}\mathbf{u}_{i,t}^T \\
\mathbf{B} &= \left(\sum_i \sum_t \mathbb{E}\left[\mathbf{z}_{i,t}\mathbf{u}_{i,t}\right] - \mathbf{A}\mathbb{E}[\mathbf{z}_{i,t-1}]\mathbf{u}_{i,t}^T\right)\left(\sum_i \sum_t \mathbf{u}_{i,t}\mathbf{u}_{i,t}^T\right)^{-1}.
\end{aligned}$$

For $\mathbf{C}$, we have

$$\begin{aligned}
\partial Q &= \sum_i \sum_t \mathbb{E}\left[\partial \log \mathcal{N}(\mathbf{y}_{i,t}|\mathbf{C}\mathbf{z}_{i,t} + \mathbf{D}\mathbf{u}_{i,t}, \mathbf{R})\right] \\
&= \sum_i \sum_t \operatorname{tr}\left[\mathbb{E}\left[-\mathbf{z}_{i,t}\left(\mathbf{y}_{i,t} - \mathbf{C}\mathbf{z}_{i,t} - \mathbf{D}\mathbf{u}_{i,t}\right)^T\right]\mathbf{R}^{-1}\partial\mathbf{C}\right] \\
\frac{\partial Q}{\partial \mathbf{C}} &= \sum_i \sum_t -\mathbf{R}^{-1}\mathbb{E}\left[\left(\mathbf{y}_{i,t} - \mathbf{C}\mathbf{z}_{i,t} - \mathbf{D}\mathbf{u}_{i,t}\right)\mathbf{z}_{i,t}^T\right] \\
\mathbf{0} &= \sum_i \sum_t \mathbf{y}_{i,t}\mathbb{E}\left[\mathbf{z}_{i,t}^T\right] - \mathbf{C}\mathbb{E}\left[\mathbf{z}_{i,t}\mathbf{z}_{i,t}^T\right] - \mathbf{D}\mathbf{u}_{i,t}\mathbb{E}\left[\mathbf{z}_{i,t}^T\right] \\
\mathbf{C} &= \left(\sum_i \sum_t \mathbf{y}_{i,t}\mathbb{E}\left[\mathbf{z}_{i,t}^T\right] - \mathbf{D}\mathbf{u}_{i,t}\mathbb{E}\left[\mathbf{z}_{i,t}^T\right]\right)\left(\sum_i \sum_t \mathbb{E}\left[\mathbf{z}_{i,t}\mathbf{z}_{i,t}^T\right]\right)^{-1}.
\end{aligned}$$

For $\mathbf{D}$, we have

$$\begin{aligned}
\frac{\partial Q}{\partial \mathbf{D}} &= \sum_i \sum_t -\mathbf{R}^{-1}\mathbb{E}\left[\left(\mathbf{y}_{i,t} - \mathbf{C}\mathbf{z}_{i,t} - \mathbf{D}\mathbf{u}_{i,t}\right)\mathbf{u}_{i,t}^T\right] \\
\mathbf{0} &= \sum_i \sum_t \mathbf{y}_{i,t}\mathbf{u}_{i,t}^T - \mathbf{C}\mathbb{E}\left[\mathbf{z}_{i,t}\right]\mathbf{u}_{i,t} - \mathbf{D}\mathbf{u}_{i,t}\mathbf{u}_{i,t}^T \\
\mathbf{D} &= \left(\sum_i \sum_t \mathbf{y}_{i,t}\mathbf{u}_{i,t}^T - \mathbf{C}\mathbb{E}\left[\mathbf{z}_{i,t}\right]\mathbf{u}_{i,t}\right)\left(\sum_i \sum_t \mathbf{u}_{i,t}\mathbf{u}_{i,t}^T\right)^{-1}.
\end{aligned}$$

The derivation of the covariance matrices $\mathbf{Q}$ and $\mathbf{R}$ is very similar to that of the MLE of a multivariate Gaussian. The trick is to differentiate with respect to $\mathbf{Q}^{-1}$ and $\mathbf{R}^{-1}$ respectively. Doing so gives us

$$\begin{aligned}
\mathbf{Q} &= \frac{1}{M}\sum_i \sum_t \mathbb{E}\left[\left(\mathbf{z}_{i,t} - \mathbf{A}\mathbf{z}_{i,t-1} - \mathbf{B}\mathbf{u}_{i,t}\right)\left(\mathbf{z}_{i,t} - \mathbf{A}\mathbf{z}_{i,t-1} - \mathbf{B}\mathbf{u}_{i,t}\right)^T\right] \\
\mathbf{R} &= \frac{1}{M}\sum_i \sum_t \mathbb{E}\left[\left(\mathbf{y}_{i,t} - \mathbf{C}\mathbf{z}_{i,t} - \mathbf{D}\mathbf{u}_{i,t}\right)\left(\mathbf{y}_{i,t} - \mathbf{C}\mathbf{z}_{i,t} - \mathbf{D}\mathbf{u}_{i,t}\right)^T\right],
\end{aligned}$$

where $M = \sum N_i$.

The above equations omit the base case $p(\mathbf{z}_{i,1}|\mathbf{u}_{i,1})$. We consider it now.

$$\begin{aligned}
\partial Q &= \sum_i \partial \mathbb{E}\left[\log \mathcal{N}\left(\mathbf{z}_{i,1}|\boldsymbol{\mu}_{1|0} + \mathbf{B}\mathbf{u}_{i,1}, \boldsymbol{\Sigma}_{1|0}\right)\right] \\
\boldsymbol{\mu}_{1|0} &= \frac{1}{N}\sum_i \mathbb{E}\left[\mathbf{z}_{i,1}\right] - \mathbf{B}\mathbf{u}_{i,1} \\
\boldsymbol{\Sigma}_{1|0} &= \frac{1}{N}\sum_i \mathbb{E}\left[\left(\mathbf{z}_{i,1} - \boldsymbol{\mu}_{1|0} - \mathbf{B}\mathbf{u}_{i,1}\right)\left(\mathbf{z}_{i,1} - \boldsymbol{\mu}_{1|0} - \mathbf{B}\mathbf{u}_{i,1}\right)^T\right].
\end{aligned}$$

Now we must compute the expectations the above formulas depend on:

$$\begin{aligned}
\mathbb{E}\left[\mathbf{z}_{i,t}|\mathbf{y}_{i,1:T}, \mathbf{u}_{i,1:T}\right] &= \boldsymbol{\mu}_{t|T} \\
\mathbb{E}\left[\mathbf{z}_{i,t}\mathbf{z}_{i,t}^T|\mathbf{y}_{i,1:T}, \mathbf{u}_{i,1:T}\right] &= \boldsymbol{\Sigma}_{t|T} + \boldsymbol{\mu}_{t|T}\boldsymbol{\mu}_{t|T}^T.
\end{aligned}$$

To compute the expected value of $\mathbf{z}_{i,t}\mathbf{z}_{i,t+1}$, we first see that the distribution is given by

$$\begin{aligned}
p(\mathbf{z}_{i,t}\mathbf{z}_{i,t+1}|\mathbf{y}_{i,1:T}, \mathbf{u}_{i,1:T}) &= \mathcal{N}(\mathbf{z}_{i,t}, \mathbf{z}_{i,t+1}|\boldsymbol{\mu}, \boldsymbol{\Sigma}) \\
&= p(\mathbf{z}_{i,t+1}|\mathbf{y}_{i,1:T}, \mathbf{u}_{i,1:T})p(\mathbf{z}_{i,t}|\mathbf{z}_{i,t+1}, \mathbf{y}_{i,1:t}, \mathbf{u}_{i,1:t}),
\end{aligned}$$

(see section 18.3.2 for the formulas in terms of the data and parameters) where

$$\begin{aligned}
\boldsymbol{\mu} &= \begin{bmatrix}\boldsymbol{\mu}_1 \\ \boldsymbol{\mu}_2\end{bmatrix} \\
\boldsymbol{\Sigma} &= \begin{bmatrix}\boldsymbol{\Sigma}_{1,1} & \boldsymbol{\Sigma}_{1,2} \\ \boldsymbol{\Sigma}_{2,1} & \boldsymbol{\Sigma}_{2,2}\end{bmatrix}.
\end{aligned}$$

It follows that

$$\mathbb{E}\left[\mathbf{x}_{i,t}\mathbf{x}_{i,t+1}^T|\mathbf{y}_{i,1:T}, \mathbf{u}_{i,1:T}\right] = \boldsymbol{\Sigma}_{1,2} + \boldsymbol{\mu}_1\boldsymbol{\mu}_2^T.$$