Supplementary Materials: Simultaneous Bayesian Clustering and Model Selection with Mixture of Robust Factor Analyzers

Shan Feng [1,2,*], Wenxian Xie [1] and Yufeng Nie [1,*]

1. Deriving the auxiliary posteriors of the latent variables

(i) $q(u_{nl}|\phi_{nl}, z_n)$: The evidence lower bound associated with the auxiliary posterior $q(u_{nl}|\phi_{nl} = 1, z_n = k)$ can be read as

$$\mathcal{L} = \left\langle \log p(y_{nl}, u_{nl}|\boldsymbol{x}_{nk}, \boldsymbol{r}_{nk}, \phi_{nl} = 1, z_n = k) - \log q(u_{nl}|\phi_{nl} = 1, z_n = k)\right\rangle_q. \tag{1}$$

According to the KKT condition, and using the Lagrange multiplier [1], we obtain the $q(u_{nl}|\phi_{nl} = 1, z_n = k)$ maximizing the lower bound as

$$\begin{aligned} q(u_{nl}|\phi_{nl} = 1, z_n = k) &\propto \exp\left[\langle \log p(y_{nl}|\boldsymbol{x}_{nk}, \boldsymbol{r}_{nk}, u_{nl}, \phi_{nl} = 1, z_n = k)\rangle + \langle \log p(u_{nl}|\phi_{nl} = 1, z_n = k)\rangle\right] \\ &\propto \exp\left[\frac{1}{2}\log u_{nl} - \frac{1}{2}\langle\sigma_{kl}\rangle\langle(\tilde{y}^k_{nl} - \mu_{kl})^2\rangle u_{nl} + \left(\frac{v_{kl}}{2} - 1\right)\log u_{nl} - \frac{v_{kl}}{2}u_{nl}\right], \end{aligned} \tag{2}$$

which gives

$$q(u_{nl}|\phi_{nl} = 1, z_n = k) = \mathcal{G}\left(u_{nl}|\hat{a}_{kl}, \hat{b}^k_{nl}\right), \tag{3}$$

as shown in the paper.

The evidence lower bound associated with $q(u_{nl}|\phi_{nl} = 0)$ can be read as

$$\mathcal{L} = \left\langle \sum_k \delta_{z_n,k} \log p(y_{nl}, u_{nl}|\boldsymbol{x}_{nk}, \boldsymbol{r}_{nk}, \phi_{nl} = 0, z_n = k) - \log q(u_{nl}|\phi_{nl} = 0)\right\rangle_q. \tag{4}$$

The posterior $q(u_{nl}|\phi_{nl} = 0)$ maximizing the lower bound is given by

$$\begin{aligned} q(u_{nl}|\phi_{nl} = 0) &\propto \exp\left[\sum_k \langle\delta_{z_n,k}\rangle\langle \log p(y_{nl}|\boldsymbol{x}_{nk}, \boldsymbol{r}_{nk}, u_{nl}, \phi_{nl} = 0, z_n = k)\rangle + \langle \log p(u_{nl}|\phi_{nl} = 0)\rangle\right] \\ &\propto \exp\left\{\frac{1}{2}\sum_k \langle\delta_{z_n,k}\rangle\left[\log u_{nl} - \langle\sigma_{0l}\rangle\langle(\tilde{y}^k_{nl} - \mu_{0l})^2\rangle u_{nl}\right] + \left(\frac{v_{0l}}{2} - 1\right)\log u_{nl} - \frac{v_{0l}}{2}u_{nl}\right\}, \end{aligned} \tag{5}$$

which gives

$$q(u_{nl}|\phi_{nl} = 0) = \mathcal{G}\left(u_{nl}|\hat{a}_{0l}, \hat{b}^0_{nl}\right), \tag{6}$$

as shown in the paper.

(ii) $q(\phi_{nl})$: The evidence lower bound associated with the auxiliary posterior $q(\phi_{nl})$ is

$$\begin{aligned} \mathcal{L} = \Bigg\langle \phi_{nl}\Bigg[&\sum_k \delta_{z_n,k} \log p(y_{nl}, u_{nl}|\boldsymbol{x}_{nk}, \boldsymbol{r}_{nk}, \phi_{nl} = 1, z_n = k) \\ &- \sum_k \delta_{z_n,k} \log q(u_{nl}|\phi_{nl} = 1, z_n = k) + \log\beta_l - \log q(\phi_{nl} = 1)\Bigg] \\ &+ (1 - \phi_{nl})\Bigg[\sum_k \delta_{z_n,k} \log p(y_{nl}, u_{nl}|\boldsymbol{x}_{nk}, \boldsymbol{r}_{nk}, \phi_{nl} = 0, z_n = k) \\ &- \log q(u_{nl}|\phi_{nl} = 0) + \log(1 - \beta_l) - \log q(\phi_{nl} = 0)\Bigg]\Bigg\rangle_q. \end{aligned} \tag{7}$$

Define

$$\overline{q}(\phi_{nl}=1) = \exp\bigg\{ \sum_k \langle \delta_{z_n,k} \rangle \Big\langle \log p(y_{nl}|\boldsymbol{x}_{nk}, \boldsymbol{r}_{nk}, u_{nl}, \phi_{nl}=1, z_n=k)$$

$$+ \log p(u_{nl}|\phi_{nl}=1, z_n=k) - \log q(u_{nl}|\phi_{nl}=1, z_n=k) \Big\rangle + \langle \log \beta_l \rangle \bigg\},$$

$$\overline{q}(\phi_{nl}=0) = \exp\bigg\{ \sum_k \langle \delta_{z_n,k} \rangle \langle \log p(y_{nl}|\boldsymbol{x}_{nk}, \boldsymbol{r}_{nk}, u_{nl}, \phi_{nl}=0, z_n=k) \rangle$$

$$+ \langle \log p(u_{nl}|\phi_{nl}=0) \rangle - \langle \log q(u_{nl}|\phi_{nl}=0) \rangle + \langle \log(1-\beta_l) \rangle \bigg\}. \quad (8)$$

Then, $q(\phi_{nl})$ can be obtained by

$$q(\phi_{nl}=1) = \frac{\overline{q}(\phi_{nl}=1)}{\overline{q}(\phi_{nl}=1) + \overline{q}(\phi_{nl}=0)}, \quad (9)$$

and $q(\phi_{nl}=0) = 1 - q(\phi_{nl}=1)$. Referring to the expectations calculated in Table A1 in the paper, we have

$$\overline{q}(\phi_{nl}=1) = \exp\bigg\{ \sum_k \langle \delta_{z_n,k} \rangle \bigg[\frac{1}{2} \langle \log \sigma_{kl} \rangle + \frac{v_{kl}}{2} \log \frac{v_{kl}}{2} - \log \Gamma\Big(\frac{v_{kl}}{2}\Big)$$

$$- \hat{a}_{kl} \log \hat{b}^k_{nl} + \log \Gamma(\hat{a}_{kl}) \bigg] + \langle \log \beta_l \rangle + \text{const.} \bigg\},$$

$$\overline{q}(\phi_{nl}=0) = \exp\bigg[\frac{1}{2} \langle \log \sigma_{0l} \rangle + \frac{v_{0l}}{2} \log \frac{v_{0l}}{2} - \log \Gamma\Big(\frac{v_{0l}}{2}\Big)$$

$$- \hat{a}_{0l} \log \hat{b}^0_{nl} + \log \Gamma(\hat{a}_{0l}) + \langle \log(1-\beta_l) \rangle + \text{const.} \bigg], \quad (10)$$

as shown in the paper.

(iii) $q(\boldsymbol{x}_{nk}|\boldsymbol{r}_{nk})$: Maximizing the evidence lower bound associated with $q(\boldsymbol{x}_{nk}|\boldsymbol{r}_{nk})$ gives

$$q(\boldsymbol{x}_{nk}|\boldsymbol{r}_{nk}) \propto \exp\bigg[\log \mathcal{N}(\boldsymbol{x}_{nk}|\mathbf{0}, \mathbf{I}_{p_k}) + \sum_l \langle \phi_{nl} \rangle \langle \log p(y_{nl}|\boldsymbol{x}_{nk}, \boldsymbol{r}_{nk}, u_{nl}, \phi_{nl}=1, z_n=k) \rangle$$

$$+ \sum_l \langle 1-\phi_{nl} \rangle \langle \log p(y_{nl}|\boldsymbol{x}_{nk}, \boldsymbol{r}_{nk}, u_{nl}, \phi_{nl}=0, z_n=k) \rangle \bigg]$$

$$\propto \exp\bigg[-\frac{1}{2} \boldsymbol{x}^T_{nk} \boldsymbol{x}_{nk} - \frac{1}{2} \sum_l \langle \phi_{nl} \rangle \langle \sigma_{kl} \rangle \langle u_{nl} \rangle^1_k \langle (y_{nl} - \boldsymbol{w}^T_{kl} \mathbf{R}_{nk} \boldsymbol{x}_{nk} - \mu_{kl})^2 \rangle$$

$$- \frac{1}{2} \sum_l \langle 1-\phi_{nl} \rangle \langle \sigma_{0l} \rangle \langle u_{nl} \rangle^0 \langle (y_{nl} - \boldsymbol{w}^T_{kl} \mathbf{R}_{nk} \boldsymbol{x}_{nk} - \mu_{0l})^2 \rangle \bigg]. \quad (11)$$

The expectations are taken by fixing on $\boldsymbol{x}_{nk}$ and conditioning on $\boldsymbol{r}_{nk}$. It can be obtained that

$$q(\boldsymbol{x}_{nk}|\boldsymbol{r}_{nk}) = \mathcal{N}\big(\boldsymbol{x}_{nk}|\hat{\boldsymbol{f}}^k_n(\boldsymbol{r}_{nk}), \hat{\mathbf{C}}^k_n(\boldsymbol{r}_{nk})\big), \quad (12)$$

as shown in the paper.

(iv) $q(\boldsymbol{r}_{nk})$: The evidence lower bound associated with $q(r_{nkj})$ is given by

$$\begin{aligned}\mathcal{L} = \Big\langle & \log \mathcal{N}(\boldsymbol{x}_{nk}|\mathbf{0}, \mathbf{I}_{p_k}) + \sum_l \phi_{nl} \log p(y_{nl}|\boldsymbol{x}_{nk}, \boldsymbol{r}_{nk}, u_{nl}, \phi_{nl} = 1, z_n = k) \\ & + \sum_l (1-\phi_{nl}) \log p(y_{nl}|\boldsymbol{x}_{nk}, \boldsymbol{r}_{nk}, u_{nl}, \phi_{nl} = 0, z_n = k) - \log q(\boldsymbol{x}_{nk}|\boldsymbol{r}_{nk}) \\ & + r_{nkj}[\log \rho_{kj} - \log q(r_{nkj} = 1)] + (1 - r_{nkj})[\log(1-\rho_{kj}) - \log q(r_{nkj} = 0)] \Big\rangle_q . \end{aligned} \tag{13}$$

If we define

$$\begin{aligned}\bar{q}(r_{nkj} = c) = \exp \Big[& \langle \log \mathcal{N}(\boldsymbol{x}_{nk}|\mathbf{0}, \mathbf{I}_{p_k}) \rangle + \sum_l \langle \phi_{nl} \rangle \langle \log p(y_{nl}|\boldsymbol{x}_{nk}, \boldsymbol{r}_{nk}, u_{nl}, \phi_{nl} = 1, z_n = k) \rangle \\ & + \sum_l \langle 1-\phi_{nl} \rangle \langle \log p(y_{nl}|\boldsymbol{x}_{nk}, \boldsymbol{r}_{nk}, u_{nl}, \phi_{nl} = 0, z_n = k) \rangle \\ & - \langle \log q(\boldsymbol{x}_{nk}|\boldsymbol{r}_{nk}) \rangle + c \langle \log \rho_{kj} \rangle + (1-c) \langle \log(1-\rho_{kj}) \rangle \Big], \end{aligned} \tag{14}$$

where the expectations are taken by fixing on $r_{nkj} = c$ ($c \in \{0, 1\}$), then, $q(r_{nkj})$ that maximizes the lower bound can be obtained as

$$q(r_{nkj} = 1) = \frac{\bar{q}(r_{nkj} = 1)}{\bar{q}(r_{nkj} = 1) + \bar{q}(r_{nkj} = 0)}, \tag{15}$$

and $q(r_{nkj} = 0) = 1 - q(r_{nkj} = 1)$. Mathematical manipulation gives

$$\begin{aligned}\bar{q}(r_{nkj} = c) = \exp \Big\{ & -\frac{1}{2} \langle \log |\mathbf{I} + \mathbf{A}_{nk} \odot \boldsymbol{r}_{nk} \boldsymbol{r}_{nk}^T| \rangle \\ & + \frac{1}{2} \mathrm{tr} \big[\langle (\mathbf{I} + \mathbf{A}_{nk} \odot \boldsymbol{r}_{nk} \boldsymbol{r}_{nk}^T)^{-1} (\mathbf{R}_{nk} \boldsymbol{t}_{nk}) (\mathbf{R}_{nk} \boldsymbol{t}_{nk})^T \rangle \big] \\ & + c \langle \log \rho_{kj} \rangle + (1-c) \langle \log(1-\rho_{kj}) \rangle + \text{const.} \Big\}, \end{aligned} \tag{16}$$

as shown in the paper.

(v) $q(z_n)$: For $q(z_n)$, we define the quantity

$$\begin{aligned}\bar{q}(z_n = k) = \exp \Big\{ & \langle \log \mathcal{N}(\boldsymbol{x}_{nk}|\mathbf{0}, \mathbf{I}_{p_k}) \rangle + \sum_l \langle \phi_{nl} \rangle \langle \log p(y_{nl}|\boldsymbol{x}_{nk}, \boldsymbol{r}_{nk}, u_{nl}, \phi_{nl} = 1, z_n = k) \rangle \\ & + \sum_l \langle 1-\phi_{nl} \rangle \langle \log p(y_{nl}|\boldsymbol{x}_{nk}, \boldsymbol{r}_{nk}, u_{nl}, \phi_{nl} = 0, z_n = k) \rangle \\ & + \sum_l \langle \phi_{nl} \rangle \langle \log p(u_{nl}|\phi_{nl} = 1, z_n = k) - \log q(u_{nl}|\phi_{nl} = 1, z_n = k) \rangle \\ & + \langle \log p(\boldsymbol{r}_{nk}) \rangle - \langle \log q(\boldsymbol{x}_{nk}|\boldsymbol{r}_{nk}) \rangle - \langle \log q(\boldsymbol{r}_{nk}) \rangle + \langle \log \pi_k \rangle \Big\}. \end{aligned} \tag{17}$$

Then, $q(z_n)$ that maximizes the evidence lower bound is given by

$$q(z_n = k) = \frac{\bar{q}(z_n = k)}{\sum_{k'} \bar{q}(z_n = k')}. \tag{18}$$

The expectations in equation (17) can be evaluated referring to Table A1. Mathematical manipulation gives the expression in the paper, where the expectations with respect to $q(\boldsymbol{r}_{nk})$ have been replaced by the random imputations.

2. Deriving the auxiliary posteriors of the parameters

The auxiliary posterior $q(\boldsymbol{\pi})$ that maximizes the evidence lower bound can be derived as

$$\begin{aligned} q(\boldsymbol{\pi}) &\propto \exp\left[\sum_{n,k}\langle\delta_{z_n,k}\rangle \log \pi_k + \log \mathcal{D}ir(\boldsymbol{\pi}|\boldsymbol{\alpha}_0)\right] \\ &\propto \exp\left[\sum_k \left(\alpha_{0k} + \sum_n \langle\delta_{z_n,k}\rangle - 1\right) \log \pi_k\right]. \end{aligned} \tag{19}$$

It follows that

$$q(\boldsymbol{\pi}) = \mathcal{D}ir(\boldsymbol{\pi}|\hat{\boldsymbol{\alpha}}), \tag{20}$$

as shown in the paper.

The auxiliary posterior $q(\beta_l)$ that maximizes the evidence lower bound is given by

$$\begin{aligned} q(\beta_l) &\propto \exp\left[\sum_n \langle\phi_{nl}\rangle \log \beta_l + \sum_n \langle 1-\phi_{nl}\rangle \log(1-\beta_l) + \log \mathcal{B}eta(\beta_l|\kappa_1,\kappa_2)\right] \\ &\propto \exp\left[\left(\kappa_1 + \sum_n \langle\phi_{nl}\rangle - 1\right)\log \beta_l + \left(\kappa_2 + \sum_n \langle 1-\phi_{nl}\rangle - 1\right)\log(1-\beta_l)\right]. \end{aligned} \tag{21}$$

It follows that

$$q(\beta_l) = \mathcal{B}eta(\beta_l|\hat{\kappa}_{1l}, \hat{\kappa}_{2l}), \tag{22}$$

as shown in the paper.

The auxiliary posterior $q(\rho_{kj})$ that maximizes the evidence lower bound is given as

$$\begin{aligned} q(\rho_{kj}) &\propto \exp\left[\sum_n \langle\delta_{z_n,k}\rangle\langle r_{nkj}\rangle \log \rho_{kj} + \sum_n \langle\delta_{z_n,k}\rangle\langle 1-r_{nkj}\rangle \log(1-\rho_{kj}) + \log \mathcal{B}eta(\rho_{kj}|\tau_1,\tau_2)\right] \\ &\propto \exp\left[\left(\tau_1 + \sum_n \langle\delta_{z_n,k}\rangle\langle r_{nkj}\rangle - 1\right)\log \rho_{kj} + \left(\tau_2 + \sum_n \langle\delta_{z_n,k}\rangle\langle 1-r_{nkj}\rangle - 1\right)\log(1-\rho_{kj})\right]. \end{aligned} \tag{23}$$

It follows that

$$q(\rho_{kj}) = \mathcal{B}eta(\rho_{kj}|\hat{\tau}_{1kj}, \hat{\tau}_{2kj}), \tag{24}$$

as in the paper.

The auxiliary posterior $q(\boldsymbol{w}_{kl})$ that maximizes the evidence lower bound can be obtained as

$$\begin{aligned} q(\boldsymbol{w}_{kl}) &\propto \exp\Bigg\{\sum_n \langle\delta_{z_n,k}\rangle\Big[\langle\phi_{nl}\rangle\langle \log p(y_{nl}|\boldsymbol{x}_{nk}, \boldsymbol{r}_{nk}, u_{nl}, \phi_{nl}=1, z_n=k)\rangle \\ &\qquad + \langle 1-\phi_{nl}\rangle\langle \log p(y_{nl}|\boldsymbol{x}_{nk}, \boldsymbol{r}_{nk}, u_{nl}, \phi_{nl}=0, z_n=k)\rangle\Big] + \log \mathcal{N}(\boldsymbol{w}_{kl}|\mathbf{0}, m_0\mathbf{I}_{p_k})\Bigg\} \\ &\propto \exp\Bigg\{-\frac{1}{2}\sum_n \langle\delta_{z_n,k}\rangle\Big[\langle\phi_{nl}\rangle\langle\sigma_{kl}\rangle\langle u_{nl}\rangle_k^1\langle(y_{nl} - \boldsymbol{w}_{kl}^T\mathbf{R}_{nk}\boldsymbol{x}_{nk} - \mu_{kl})^2\rangle \\ &\qquad + \langle 1-\phi_{nl}\rangle\langle\sigma_{0l}\rangle\langle u_{nl}\rangle^0\langle(y_{nl} - \boldsymbol{w}_{kl}^T\mathbf{R}_{nk}\boldsymbol{x}_{nk} - \mu_{0l})^2\rangle\Big] - \frac{m_0}{2}\boldsymbol{w}_{kl}^T\boldsymbol{w}_{kl}\Bigg\}. \end{aligned} \tag{25}$$

Mathematical manipulation gives

$$q(\boldsymbol{w}_{kl}) = \mathcal{N}(\boldsymbol{w}_{kl}|\hat{\boldsymbol{m}}_{kl}, \hat{\mathbf{M}}_{kl}), \tag{26}$$

as presented in the paper.

The auxiliary posterior $q(\mu_{kl})$ that maximizes the evidence lower bound can be obtained as

$$q(\mu_{kl}) \propto \exp\Bigg[\sum_n \langle\delta_{z_n,k}\rangle\langle\phi_{nl}\rangle\langle\log p(y_{nl}|\boldsymbol{x}_{nk},\boldsymbol{r}_{nk},u_{nl},\phi_{nl}=1,z_n=k)\rangle + \log\mathcal{N}(\mu_{kl}|s_l,\lambda_0)\Bigg]$$
$$\propto \exp\Bigg[-\frac{1}{2}\langle\sigma_{kl}\rangle\sum_n \langle\delta_{z_n,k}\rangle\langle\phi_{nl}\rangle\langle u_{nl}\rangle_k^1\langle(\tilde{y}_{nl}^k-\mu_{kl})^2\rangle - \frac{\lambda_0}{2}(\mu_{kl}-s_l)^2\Bigg], \tag{27}$$

which gives

$$q(\mu_{kl}) = \mathcal{N}(\mu_{kl}|\hat{s}_{kl},\hat{\lambda}_{kl}), \tag{28}$$

as shown in the paper.

The auxiliary posterior $q(\mu_{0l})$ that maximizes the evidence lower bound can be obtained as

$$q(\mu_{0l}) \propto \exp\Bigg[\sum_{n,k} \langle\delta_{z_n,k}\rangle\langle 1-\phi_{nl}\rangle\langle\log p(y_{nl}|\boldsymbol{x}_{nk},\boldsymbol{r}_{nk},u_{nl},\phi_{nl}=0,z_n=k)\rangle + \log\mathcal{N}(\mu_{0l}|s_l,\lambda_0)\Bigg]$$
$$\propto \exp\Bigg[-\frac{1}{2}\langle\sigma_{0l}\rangle\sum_{n,k} \langle\delta_{z_n,k}\rangle\langle 1-\phi_{nl}\rangle\langle u_{nl}\rangle^0\langle(\tilde{y}_{nl}^k-\mu_{0l})^2\rangle - \frac{\lambda_0}{2}(\mu_{0l}-s_l)^2\Bigg], \tag{29}$$

which gives

$$q(\mu_{0l}) = \mathcal{N}(\mu_{0l}|\hat{s}_{0l},\hat{\lambda}_{0l}), \tag{30}$$

as shown in the paper.

The auxiliary posterior $q(\sigma_{kl})$ that maximizes the evidence lower bound can be obtained as

$$q(\sigma_{kl}) \propto \exp\Bigg[\sum_n \langle\delta_{z_n,k}\rangle\langle\phi_{nl}\rangle\langle\log p(y_{nl}|\boldsymbol{x}_{nk},\boldsymbol{r}_{nk},u_{nl},\phi_{nl}=1,z_n=k)\rangle + \log\mathcal{G}(\sigma_{kl}|\frac{\eta_0}{2},\frac{\xi_0}{2})\Bigg]$$
$$\propto \exp\Bigg[\frac{1}{2}\sum_n \langle\delta_{z_n,k}\rangle\langle\phi_{nl}\rangle\log\sigma_{kl} - \frac{1}{2}\sigma_{kl}\sum_n \langle\delta_{z_n,k}\rangle\langle\phi_{nl}\rangle\langle u_{nl}\rangle_k^1\langle(\tilde{y}_{nl}^k-\mu_{kl})^2\rangle + \Big(\frac{\eta_0}{2}-1\Big)\log\sigma_{kl} - \frac{\xi_0}{2}\sigma_{kl}\Bigg], \tag{31}$$

which gives

$$q(\sigma_{kl}) = \mathcal{G}\left(\sigma_{kl}|\frac{\hat{\eta}_{kl}}{2},\frac{\hat{\xi}_{kl}}{2}\right), \tag{32}$$

as shown in the paper.

The auxiliary posterior $q(\sigma_{0l})$ that maximizes the evidence lower bound can be obtained as

$$q(\sigma_{0l}) \propto \exp\Bigg[\sum_{n,k} \langle\delta_{z_n,k}\rangle\langle 1-\phi_{nl}\rangle\langle\log p(y_{nl}|\boldsymbol{x}_{nk},\boldsymbol{r}_{nk},u_{nl},\phi_{nl}=0,z_n=k)\rangle + \log\mathcal{G}(\sigma_{0l}|\frac{\eta_0}{2},\frac{\xi_0}{2})\Bigg]$$
$$\propto \exp\Bigg[\frac{1}{2}\sum_n \langle 1-\phi_{nl}\rangle\log\sigma_{0l} - \frac{1}{2}\sigma_{0l}\sum_{n,k} \langle\delta_{z_n,k}\rangle\langle 1-\phi_{nl}\rangle\langle u_{nl}\rangle^0\langle(\tilde{y}_{nl}^k-\mu_{0l})^2\rangle + \Big(\frac{\eta_0}{2}-1\Big)\log\sigma_{0l} - \frac{\xi_0}{2}\sigma_{0l}\Bigg], \tag{33}$$

which gives

$$q(\sigma_{0l}) = \mathcal{G}\left(\sigma_{0l}|\frac{\hat{\eta}_{0l}}{2},\frac{\hat{\xi}_{0l}}{2}\right), \tag{34}$$

as in the paper.

To obtain v_{kl} that maximizes the evidence lower bound, we solve the equation

$$\frac{\partial}{\partial v_{kl}}\left\{\sum_n \langle\delta_{z_n,k}\rangle\langle\phi_{nl}\rangle\left[\left(\frac{v_{kl}}{2}-1\right)\langle\log u_{nl}\rangle_k^1-\frac{v_{kl}}{2}\langle u_{nl}\rangle_k^1+\frac{v_{kl}}{2}\log\frac{v_{kl}}{2}-\log\Gamma\left(\frac{v_{kl}}{2}\right)\right]\right\}=0, \tag{35}$$

which gives the equation in the paper.

To obtain v_{0l} we solve the equation

$$\frac{\partial}{\partial v_{0l}}\left\{\sum_n \langle 1-\phi_{nl}\rangle\left[\left(\frac{v_{0l}}{2}-1\right)\langle\log u_{nl}\rangle^0-\frac{v_{0l}}{2}\langle u_{nl}\rangle^0+\frac{v_{0l}}{2}\log\frac{v_{0l}}{2}-\log\Gamma\left(\frac{v_{0l}}{2}\right)\right]\right\}=0, \tag{36}$$

as given in the paper.

Author Contributions: Conceptualization, Y.N. and W.X.; methodology, S.F.; software, S.F.; validation, S.F.; formal analysis, S.F.; writing—original draft preparation, S.F.; writing—review and editing, W.X.; visualization, S.F.; supervision, Y.N.; funding acquisition, Y.N. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the National Key R&D Program of China grant number 2020YFA0713603 and the National Natural Science Foundation of China grant number 11971386.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Sun, J.; Zhou, A.; Keates, S.; Liao, S. Simultaneous Bayesian clustering and feature selection through student's t mixtures model. *IEEE Trans. Neural Netw. Learn. Syst.* **2018**, *29*, 1187–1199.