BAYES HW 2

Vayser Kirill

October 2021

0.1 M-step

For simplicity let us skip indices.

$$L = \log P(Y, X, U, Z) = \log P(Y|X, U, Z)\, P(X|U, Z)\, P(X|U, Z)\, P(U|Z)\, P(Z) =$$

$$= \log P(Y|X, U, Z) + \log P(X|U, Z) + \log P(X|U, Z) + \log P(U|Z) + \log P(Z) =$$

$$= \sum_{n=1}^{N} \sum_{k=1}^{K} z \left(\log \pi + \frac{v}{2} \log \frac{v}{2} - \log \Gamma\left(\frac{v}{2}\right) + \left(\frac{v}{2} - 1\right) \log u - \frac{v}{2} u + \right.$$

$$\left. + \left(\frac{D}{2} + \frac{d}{2}\right) \log \frac{u}{2\pi} - \frac{1}{2} u \left(\|y - Wx - \mu\|^2 + \|x\|^2\right) \right)$$

$$\nabla L_{\pi_k} = \sum_{n=1}^{N} \frac{z_{nk}}{\pi_k}$$

$$\nabla L_{v_k} = \sum_{n=1}^{N} \frac{1}{2} z_{nk} \left[\left(\log \frac{v}{2} + 1 \right) - \psi\left(\frac{v}{2}\right) + \log u - u \right]$$

$$\nabla L_{W_k} = \sum_{n=1}^{N} -z_{nk} u_{nk} \left(W_k x_{nk} x_{nk}^\top - (y_n - \mu_k) x_{nk}^\top \right)$$

$$\nabla L_{\mu_k} = \sum_{n=1}^{N} -z_{nk} u_{nk} \left(\mu_k + W_k x_{nk} - y_n \right)$$

So we need to calculate

$$P(z_{nk} = 1 | y_n)$$
$$P(u_{nk} | z_{nk} = 1, y_n)$$
$$P(x_{nk} | u_{nk}, y_n, z_{nk} = 1)$$

To solve following equations:

$$\mathbb{E}_z \nabla L_{\pi_k} + \lambda \left(1 - \sum_{k=1}^{K} \pi_k \right) = 0$$
$$\mathbb{E}_u \mathbb{E}_z \nabla L_{v_k} = 0$$
$$\mathbb{E}_x \mathbb{E}_u \mathbb{E}_z \nabla L_{W_k} = 0$$
$$\mathbb{E}_x \mathbb{E}_u \mathbb{E}_z \nabla L_{\mu_k} = 0$$

Let us calculate that:

$$P\left(x_{nk}|u_{nk}, y_n, z_{nk}=1\right) = \frac{P\left(y|x,u,z\right)P\left(x|u,z\right)P\left(u|z\right)P\left(z\right)}{P\left(u_{nk}, y_n, z_{nk}=1\right)} =$$
$$= AP\left(y|x,u,z\right)P\left(x|u,z\right) = Be^{-\frac{1}{2}u\left(\|y-Wx-\mu\|^2+\|x\|^2\right)} =$$
$$= Ce^{-\frac{1}{2}(x-m)^\top\Sigma^{-1}(x-m)},$$

where

$$\Sigma^{-1} = u\left(W^\top W + I\right) \quad \hat{\Sigma^{-1}} = \left(W^\top W + I\right)$$
$$m = \left(W^\top W + I\right)^{-1} W^T\left(y-\mu\right)$$

To see that, let us look at

$$u\left(\|y-Wx-\mu\|^2+\|x\|^2\right) = u\left[y^\top y - 2x^\top W^\top y + 2x^\top W^\top \mu - 2y^\top \mu + x^\top W^\top W x + \mu^\top \mu\right] =$$
$$= \left(x-m\right)^\top \Sigma^{-1}\left(x-m\right) + \left(y-\mu\right)^\top S\left(y-\mu\right)$$

where $S = u\left(I - W\left(W^\top W + I\right)^{-1}W^\top\right)$, $\hat{S} = \left(I - W\left(W^\top W+I\right)^{-1}W^\top\right)$.

So $P\left(x_{nk}|u_{nk}, y_n, z_{nk}=1\right)$ is $\mathcal{N}\left(x_{nk}|m, \Sigma\right)$

$$P\left(u_{nk}|y_n, z_{nk}=1\right) = \frac{\int P\left(y|x,u,z\right)P\left(x|u,z\right)P\left(u|z\right)P\left(z\right)dx}{P\left(y_n, z_{nk}=1\right)} = A\int P\left(y|x,u,z\right)P\left(x|u,z\right)P\left(u|z\right)dx =$$
$$= A\int\left(\frac{u}{2\pi}\right)^{\frac{D+d}{2}} e^{-\frac{1}{2}\left[(x-m)^\top\Sigma^{-1}(x-m)+(y-\mu)^\top S(y-\mu)\right]}\frac{\frac{v}{2}^{\frac{v}{2}}}{\Gamma\left(\frac{v}{2}\right)}u^{\frac{v}{2}-1}e^{-\frac{v}{2}u}dx =$$
$$= B\left(\frac{u}{2\pi}\right)^{\frac{D}{2}} u^{\frac{v}{2}-1}e^{-u\left[(y-\mu)^\top\hat{S}(y-\mu)+\frac{v}{2}\right]}$$

So $P\left(u_{nk}|y_n, z_{nk}=1\right)$ is $\Gamma\left(u_{nk}|\frac{D+v}{2}, \frac{(y-\mu)^\top\hat{S}(y-\mu)+v}{2}\right) = \Gamma\left(\alpha,\beta\right)$

$$P\left(z_{nk}=1|y_n\right) = \frac{\int P\left(y|x,u,z\right)P\left(x|u,z\right)P\left(u|z\right)P\left(z\right)dxdu}{P\left(y_n\right)} = \text{consider only numerator}$$
$$= \int\left(\frac{u}{2\pi}\right)^{\frac{D+d}{2}} e^{-\frac{1}{2}\left[(x-m)^\top\Sigma^{-1}(x-m)+(y-\mu)^\top S(y-\mu)\right]}\frac{\frac{v}{2}^{\frac{v}{2}}}{\Gamma\left(\frac{v}{2}\right)}u^{\frac{v}{2}-1}e^{-\frac{v}{2}u}\pi_k dxdu =$$
$$= \pi_k\frac{\frac{v}{2}^{\frac{v}{2}}}{\Gamma\left(\frac{v}{2}\right)}\frac{\det\left(W^\top W+I\right)^{\left(-\frac{d}{2}\right)}}{\left(2\pi\right)^{\frac{D}{2}}}\int u^{\frac{D+v}{2}-1}e^{-u\left[(y-\mu)^\top\hat{S}(y-\mu)+\frac{v}{2}\right]}du =$$
$$= \pi_k\frac{\frac{v}{2}^{\frac{v}{2}}}{\Gamma\left(\frac{v}{2}\right)}\frac{\det\left(W^\top W+I\right)^{\left(-\frac{d}{2}\right)}}{\left(2\pi\right)^{\frac{D}{2}}}\frac{\Gamma\left(\alpha\right)}{\beta^\alpha} = \hat{r_{nk}}$$
$$P\left(z_{nk}=1|y_n\right) = \frac{\hat{r_{nk}}}{\sum\limits_{k=1}^{K}\hat{r_{nk}}} = r_{nk}$$

So the expressions for parameters become

For π_k

$$\mathbb{E}_z \nabla L_{\pi_k} + \lambda\left(1 - \sum_{k=1}^{K} \pi_k\right) = \sum_{n=1}^{N} \frac{r_{nk}}{\pi_k} + \lambda\left(1 - \sum_{k=1}^{K} \pi_k\right) = 0$$

$$\Rightarrow \pi_k = \frac{1}{N} \sum_{n=1}^{N} r_{nk}$$

For v_k

$$\mathbb{E}_u \mathbb{E}_z \nabla L_{v_k} = \mathbb{E}_z\left[\mathbb{E}_u \nabla L_{v_k} | z\right] = \sum_{n=1}^{N} \frac{1}{2} r_{nk} \left[\left(\log \frac{v}{2} + 1\right) - \psi\left(\frac{v}{2}\right) + \mathbb{E}_{u_{nk}} \log u - \mathbb{E}_{u_{nk}} u\right] = 0$$

For W_k

$$\begin{aligned}
&\mathbb{E}_x \mathbb{E}_u \mathbb{E}_z \nabla L_{W_k} = \mathbb{E}_z\{\mathbb{E}_u\left[\mathbb{E}_x \nabla L_{W_k} | u, z\right] | z\} = \\
&= \mathbb{E}_z\{\mathbb{E}_u\left[\sum_{n=1}^{N} -z_{nk} u_{nk}\left(W_k \mathbb{E}_x\left[x_{nk}^\top x_{nk}\right] - (y_n - \mu_k)\mathbb{E}_x x_{nk}^\top\right) | u, z\right] | z\} = \\
&= \mathbb{E}_z\{\mathbb{E}_u \sum_{n=1}^{N} -z_{nk}\left(W_k\left[\hat{\Sigma} + u_{nk} m m^\top\right] - u_{nk} y_n m^\top\right) | z\} = \\
&= \sum_{n=1}^{N} -r_{nk}\left(W_k\left[\hat{\Sigma} + \mathbb{E}u_{nk} m m^\top\right] - \mathbb{E}u_{nk}(y_n - \mu_k) m^\top\right) = 0 \\
&\Rightarrow W_k = \sum_{n=1}^{N} r_{nk} \mathbb{E}u_{nk}(y_n - \mu_k) m^\top \left(\sum_{n=1}^{N} r_{nk}\left[\hat{\Sigma} + \mathbb{E}u_{nk} m m^\top\right]\right)^{-1}
\end{aligned}$$

For μ_k

$$\mathbb{E}_x \mathbb{E}_u \mathbb{E}_z \nabla L_{\mu_k} = \mathbb{E}_z\{\mathbb{E}_u\left[\mathbb{E}_x \nabla L_{\mu_k} | u, z\right] | z\} = \nabla L_{\mu_k} = \sum_{n=1}^{N} -r_{nk} \mathbb{E}u_{nk}\left(\mu_k + W_k m - y_n\right) = 0$$

$$\Rightarrow \mu_k = \frac{\sum_{n=1}^{N} r_{nk} \mathbb{E}u_{nk}(y_n - W_k m)}{\sum_{n=1}^{N} r_{nk} \mathbb{E}u_{nk}},$$

where

$$
\begin{aligned}
&\mathbb{E}_{u_{nk}} = \frac{\alpha}{\beta} \\
&\mathbb{E}_{u_{nk}} \log u_{nk} = \psi(\alpha) - \log \beta \\
&\alpha = \frac{D + v_k}{2} \\
&\beta = \frac{(y_n - \mu_k)^\top \left(I - W_k \left(W_k^\top W_k + I \right)^{-1} W_k^\top \right) (y_n - \mu_k) + v_k}{2} = \\
&= \frac{(y_n - \mu_k)^\top \left(\left(W_k W_k^\top + I \right)^{-1} \right) (y_n - \mu_k) + v_k}{2} \\
&\hat{\Sigma} = \left(W_k^\top W_k + I \right)^{-1} \\
&m = \left(W_k^\top W_k + I \right)^{-1} W_k^T (y_n - \mu_k) \\
&r_{nk} = \frac{\hat{r_{nk}}}{\sum\limits_{k=1}^{K} \hat{r_{nk}}} \\
&\hat{r_{nk}} = \pi_k \frac{\frac{v_k}{2}^{\frac{v_k}{2}}}{\Gamma\left(\frac{v_k}{2}\right)} \frac{\det \left(W_k^\top W_k + I \right)^{\left(-\frac{d}{2}\right)}}{(2\pi)^{\frac{D}{2}}} \frac{\Gamma(\alpha)}{\beta^\alpha}
\end{aligned}
$$