4) $\vec{x_i} = W\vec{z_i} + \vec{\mu} + \vec{\varepsilon_i}$

$x_i \in \mathbb{R}^n$

$z_i \in \mathbb{R}^d$

$\mu \in \mathbb{R}^n$

$\varepsilon_i \in \mathbb{R}^n$

$p(z_i) = \mathcal{N}(z_i | 0, I)$

$p(\varepsilon_i) = \mathcal{N}(\varepsilon_i | 0, \sigma^2 I)$

а) $p(X, Z | W, \mu, \sigma)$ — ?

$p = \prod_{i=1}^{m} \mathcal{N}(x_i | Wz_i + \mu, \sigma^2 I) \cdot \mathcal{N}(z_i | 0, I)$

б) $p(X | W, \mu, \sigma) \sim \sum_{i=1}^{m} -\frac{1}{2} \log\det(\sigma^2 I + WW^T) -$

$- \frac{1}{2}(\cancel{x_i} - \mu)^T (\sigma^2 I + WW^T)^{-1} (x_i - \mu)$

в) $p(X | W, \mu, \sigma^2) \to \max$

$F(q, \theta) = \log p(X | W, \mu, \sigma) - D_{KL}(q(z) \| p(Z | X, W, \mu, \sigma))$

E-шаг

$F(q, \theta) \to \max \iff q(z) = p(Z | X, W, \mu, \sigma)$

$\theta = (W, \mu, \sigma)$

$p(Z | X, W, \mu, \sigma) = \prod_{i=1}^{n} \mathcal{N}(z_i | z_i^0, \Sigma_{i0}^{-1})$ — апостериорное распределение $p(z_i | x_i, W, \sigma, \mu)$

$\Sigma_{i0} = I + \sigma^{-2} W^T W$

$z_i^0 = \sigma^{-2} \Sigma_{i0}^{-1} W^T x_i$

M-шаг

$\mathbb{E}_{q(z)} \log p(X, Z | \theta) = \int \prod \mathcal{N}(z_i | z_i^0, \Sigma_{i0}^{-2}) \cdot \log p(X, Z | \theta) dZ =$

$\mathbb{E}_{q(z)} \log p(X, Z | \theta) = \sum_i \mathbb{E}_{q(z)} \log p(x_i, z_i | W, \mu, \sigma) =$ (делим)

$= \sum \mathbb{E}_{q(z)} \log \mathcal{N}(x_i | Wz_i + \mu, \sigma^2 I) + \log(z_i | 0, I) =$

$= \sum_i \{ \frac{d}{2} \log \sigma^{-2} - \frac{\sigma^{-2}}{2} \mathbb{E}_{q(z)} \| x_i - Wz_i - \mu \|^2$ ~~$- \frac{1}{2}\mathbb{E}\|z_i\|^2$~~

$\frac{\partial F}{\partial \sigma} = -\frac{nm}{\sigma} + \sigma^{-3} \sum_i \mathbb{E} \| x_i - Wz_i - \mu \|^2$

$\sigma^2 = \frac{1}{nm} \sum_i \mathbb{E} \| x_i - Wz_i - \mu \|^2$

$\frac{\partial B}{\partial \mu}$ $\mathbb{E}(x-a)^2 = Dx + (\mu - a)^2$ $\Rightarrow$

$W \mathbb{E} z_i + \mu - x = 0$

$\Rightarrow \sigma^2 = \frac{1}{nm} \sum_{i,j}^{m,n} D\, w_j^T z_i \Rightarrow$

$\sigma^2 = \frac{1}{nm} \sum_{i,j}^{m,n} w_j^T \Sigma_{z_i} w_j$

$\mu, W - ?$ $\sum_i \mathbb{E}_{q(z)} \| x_i - W z_i - \mu \|^2 =$

$= \sum_{i,j} \mathbb{E}_{q(z)} | x_{ij} - \mu_j - w_j^T z_i |^2 = ?$

$j:$ $\sum_i \mathbb{E}_{q(z)} | x_{ij} - \mu_j - w_j^T z_{ij} |^2 =$

$= \sum_i \mathbb{E} | w_j^T z_{ij} - w_j^T \mu_z |^2 + \mathbb{E} | w_j^T \mu_z - \mu_j - x_{ij} |^2 =$

$= \sum_i D(w_j^T z_{ij}) + | w_j^T \mu_z - \mu_j - x_{ij} |^2 =$

$= \sum_i w_j^T \Sigma_{z_i}^{-1} w_j + | w_j^T \mu_{z_i} - \mu_j - x_{ij} |^2$ — это

обычная линейная регрессия. для μ и w

$\cancel{w_j = \ldots}$

$\cancel{\ldots} (Z^T Z - \Sigma_{z_i})^{-1}$

$w_j = ((\mathbb{E} Z)^T \mathbb{E} Z - \Sigma_z)^{-1} (\mathbb{E} Z)^T \cdot x_j$

$\frac{\partial R}{\partial \mu} = -\frac{\sigma^{-2}}{2} \mathbb{E} \frac{\partial \| x_i - W z_i - \mu \|^2}{\partial \mu} =$

$-\frac{\sigma^{-2}}{2} \mathbb{E}(\mu + W z_i - x_i) = -\frac{\sigma^{-2}}{2} (\mu + \mathbb{E} W z_i - x_i) = 0 \Rightarrow$

$\Rightarrow \mu_{new} = x_i - W_{new} \mathbb{E} z_i$

Чтобы учесть пропуски данных нужно добавить в модель скрытую переменную, отвечающую за пропуск. $z_{np} \sim Be(p)$, $z_{np} \in \{0, 1\}$

g) $\log q(W) = \mathbb{E}_{q(z)} \log p(W, z, x \mid \alpha, \sigma, \mu)$

$\log q(z) = \mathbb{E}_{q(W)} \log p(W, z, x \mid \alpha, \sigma, \mu)$, $\theta = (\alpha, \sigma, \mu)$

$\mathbb{E} \log p(W, z, x \mid \alpha, \sigma, \mu) = \cancel{\log p(x}\ \mathbb{E} \log p(x \mid W, z, \theta) +$

$+ \mathbb{E} \log p(W \mid \theta) + \log p(z \mid \theta)$

$p(x \mid W, z, \theta) = \left(\frac{1}{\sqrt{2\pi\sigma^2}}\right)^{mn} \prod_{i=1}^{n} \exp\left(-\frac{1}{2\sigma^2} \cdot (x_i - Wz_i - \mu)^T \cdot\right.$

$\left.\cdot (x_i - Wz_i + \mu)\right)$

$\log p(W \mid \theta) = \log \prod_{j=1}^{n} \left(\sqrt{\frac{\alpha_i}{2\pi}}\right)^n \exp\left(-\frac{\alpha_j}{2} w_j^T w_j\right)$

$\log (z \mid \theta) = \log \prod_{i=1}^{n} \left(\frac{1}{\sqrt{2\pi}}\right)^{d} \exp\left(-\frac{1}{2} z_i^T z_i\right)$

$\log p(x \mid W, z, \theta) = mn \ln \sigma^{-2} - \frac{1}{2\sigma^2} \| x_i - Wz_i + \mu \|_2^2$

$\log p(W \mid \theta) = \frac{n}{2}^2 \ln \alpha_j - \frac{\alpha_j}{2} w_j^T w_j$

$\log (z \mid \theta) = -\frac{1}{2} z_i^T z_i$

$\log p(x \mid W, z, \theta) = mn \ln \sigma^{-2} - \frac{1}{2\sigma^2} \| x_i - Wz_i + \mu \|_2^2 + \frac{n^2}{2} -$

$- \frac{\alpha_j}{2} w_j^T w_j$

$\mathbb{E}_{q(z_i)} p(x_i \mid W, z_i, \theta) = -\frac{1}{2\sigma^2} \| x_i - Wz_i + \mu \|_2^2 + \frac{n^2}{2} - \frac{\alpha_j}{2} w_j^T w_j$

$= -\frac{1}{2\sigma^2} \sum_{i,j}^{m,n} | x_{ij} - w_j^T z_i + \mu_j |^2 - \sum_{j}^{n} \frac{\alpha_j}{2} w_j^T w_j + mn \ln \sigma^{-2} +$

$+ \cancel{\frac{n^2}{2} \ln \alpha_j}$

$\mathbb{E}_{q(z_i)} | x_{ij} - w_j^T z_i + \mu_j |^2 = \mathbb{E} | x_{ij} - w_j^T \mathbb{E} z_i + \mu_{ij} |^2 +$

$+ \mathbb{D}(w_j^T z_i) = \mathbb{E} | x_{ij} - w_j^T \mu_{z_i} + \mu_j |^2 + w_j^T \Sigma_{z_i} w_j^T$

$\log Q(W) = -\frac{1}{2\sigma^2} \sum_{i,j}^{m,n} | x_{ij} - w_j^T \mu_{z_i} + \mu_j |^2 + \sum_{i}^{n} \mu_i w_j^T \Sigma_{z_i} w_j -$

$- \sum_{j}^{n} \frac{\alpha_j}{2} w_j^T w_j + \cancel{mn \ln \sigma^{-2} + \frac{n^2}{2} \ln \alpha_j}$ const

$\log Q(W) = -\frac{1}{2\sigma^2} \sum_{i,j}^{m,n} w_j^T \mu_{z_i} \mu_{z_i}^T w_j + 2(\mu_j + x_{ij}) \mu_{z_i}^T w_j -$

$- \sum_{j}^{n} w_j^T \left(\frac{n}{2\sigma^2} \Sigma_{z_i} + \frac{\alpha_j}{2} I\right) w_j =$

$$= -\sum_{i,j}^{n,m} w_j^T \left\{ \frac{\mu_{zi}\mu_{zi}^T}{2\sigma^2} + \Sigma_{zi} + \frac{\alpha_j}{2} I \right\} w_j -$$

$$- 2(\mu_j + x_{ij})\mu_{zi}^T w_j \Rightarrow$$

$$\Rightarrow w_j \sim \mathcal{N}(w_j \mid w_j^*, \Sigma_{w_j})$$

$$\Sigma_{w_j}^{-1} = \sum_i^m \left[\mu_{zi}\mu_{zi}^T + \Sigma_{zi} + \alpha_j I \right] \frac{1}{\sigma^2}$$

$$w_j^* = \Sigma_{w_j} \left(m \cdot \mu_j + \sum_i^m x_{ij} \right)$$

$$\log q(z_i) = -\sum_j^n \mathbb{E}_{q(w)} |x_{ij} - w_j^T z_i + \mu_j|^2 =$$

$$= -\sum_j^n |x_{ij} - z_i^T \mathbb{E} w_j + \mu_j|^2 - z_i^T \Sigma_{w_j} z_i$$

$$\log q(z_i) = -\sum_j^n (\mathbb{E} w_j^T z_i - (\mu_j + x_{ij})) - z_i^T \Sigma_{w_j} z_i$$

$$\Sigma_{zi}^{-1} = \sum_j^n (\mathbb{E} w_j \mathbb{E} w_j^T + \Sigma_{w_j}) = \sum_j^m (w_j^* w_j^{*T} + \Sigma_{w_j})$$

$$\mu_{zi} = \Sigma_{zi} \sum_j^n \mathbb{E} w_j^* (\mu_j + x_{ij}) = \Sigma_{zi} \sum_j^n w_j^* (\mu_j + x_{ij})$$

$$z_i^* = \Sigma_{zi} \cdot \sum_j^n \mathbb{E} w_j^* (\mu_j + x_{ij})$$

M-шаг

То же самое как и в пункте в, только вместо W используем $\mathbb{E}W$, где

$$w_j^* = \Sigma_{w_j}^{-1} (m \mu_j + \hat{x}_j)$$

$$\Sigma_{w_j}^{-1} = m \frac{1}{\sigma^2} [\mu_{zi}\mu_{zi}^T + \Sigma_{zi} + \alpha_j I]$$

И соответственно мы решаем задачу не по W, а по α_j

Поэтому происходит отбор главных компонент, т.к. мы сильно сужаем степени свободы.