Section 3

- $\mu_K \sim N(0, \sigma^2)$ $\quad k = 1, \dots, K$
- indicator vector $c_i \sim \text{categorical}(\frac{1}{K}, \dots, \frac{1}{K})$ $\quad i = 1, \dots, n$ $\Big\}$ latent
- $x_i \mid c_i, \underset{\sim}{\mu} \sim N(c_i^T \underset{\sim}{\mu}, 1)$ $\quad i = 1, \dots, n$ $\leftarrow$ observed

Joint of latent & observed :

- $$p(\underset{\sim}{\mu}, \underset{\sim}{c}, \underset{\sim}{x}) = p(\underset{\sim}{\mu}) \prod_{i=1}^{n} p(c_i)\, p(x_i \mid c_i, \underset{\sim}{\mu}) \qquad (8)$$

M-F family

K-vector assignment probabilities

- $$q(\underset{\sim}{\mu}, \underset{\sim}{c}) = \prod_{k=1}^{K} q(\underbrace{\mu_k ; m_k, s_k^2}_{\text{gaussian}}) \prod_{i=1}^{n} q(\underbrace{c_i ; \varphi_i}_{\text{categorical}}) \qquad (16)$$

generic

- $$\text{ELBO}(q) = \mathbb{E}_q[\log p(\underset{\sim}{z}, \underset{\sim}{x})] - \mathbb{E}_q[\log q(\underset{\sim}{z})] \qquad (13)$$

$$\Rightarrow \text{ELBO}(\underset{\sim}{m}, \underset{\sim}{s}^2, \underset{\sim}{\varphi}) = \mathbb{E}_q\Big[\log p(\underset{\sim}{\mu}) + \sum_{i=1}^{n}\big(\log p(c_i) + \log p(x_i \mid c_i, \underset{\sim}{\mu})\big)\Big]$$
$$- \mathbb{E}_q\Big[\sum_{k=1}^{K} \log q(\mu_k) - \sum_{i=1}^{n} \log q(c_i)\Big]$$

$$= \sum_{k=1}^{K} \mathbb{E}[\log \overset{m_k, s_k^2}{p(\mu_k)}] + \sum_{i=1}^{n}\Big(\mathbb{E}[\log \overset{\varphi_i}{p(c_i)}] + \mathbb{E}[\log \overset{\varphi_i, \underset{\sim}{m}, \underset{\sim}{s}^2}{p(x_i \mid c_i, \underset{\sim}{\mu})}]\Big)$$
$$- \sum_{i=1}^{n} \mathbb{E}[\log \overset{\varphi_i}{q(c_i)}] - \sum_{k=1}^{K} \mathbb{E}[\log \overset{m_k, s_k^2}{q(\mu_k)}]$$

Section 3.1 : update for cluster assignment c_i

(eq 18). $\bullet$ $q_j^*(z_j) \propto \exp\{\mathbb{E}_{-j}[\log p(z_j, \underset{\sim}{z}_{-j}, \underset{\sim}{x})]\}$

$$\Rightarrow q^*(c_i ; \varphi_i) \propto \exp\{\mathbb{E}_{-c_i}[\log p(\underset{\sim}{c}, \underset{\sim}{\mu}, \underset{\sim}{x})]\}$$

$$= \exp\{\mathbb{E}_{-c_i}[\log[p(x_i \mid c_i, \underset{\sim}{\mu})) \cdot p(c_i)\, \cancel{p(\underset{\sim}{c}_{-i})}\, \cancel{p(\underset{\sim}{\mu})}]]$$

$$\propto \exp\{\underbrace{\log p(c_i)} + \mathbb{E}_{-c_i}[\underbrace{\log p(x_i \mid c_i, \underset{\sim}{\mu})}]]\}$$

$\log p(c_i) = \log \frac{1}{K} = -\log K$

c_i'th normal density

$c_{ik} \in \{0, 1\}$

$$\rightarrow p(x_i \mid c_i, \underset{\sim}{\mu}) = \prod_{k=1}^{K} p(x_i \mid \mu_k)^{c_{ik}}$$

②: $\mathbb{E}_{-c_i}[\log p(x_i \mid c_i, \mu)]$ $\qquad$ ($\log p(x_i \mid c_i, \mu) = \sum_{k=1}^{K} \log p(x_i \mid \mu_k)^{c_{ik}}$)

$= \mathbb{E}_{-c_i}\left[\sum_{k=1}^{K} c_{ik} \log p(x_i \mid \mu_k)\right]$

$= \sum_{k=1}^{K} c_{ik}\, \mathbb{E}_{-c_i}[\log \underbrace{p(x_i \mid \mu_k)}_{\mathcal{N}(\mu_k, 1)}]$ $\qquad$ $\left(\frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\frac{(x-\mu)^2}{\sigma^2}} \right)$

$= \sum_k c_{ik}\, \mathbb{E}_{-c_i}\left[-\frac{1}{2}\cdot \underbrace{(x_i - \mu_k)^2}_{x_i^2 - 2x_i\mu_k + \mu_k^2}\right] + C_1$

$= \sum_k c_{ik} \cdot \left(\mathbb{E}[\mu_k] x_i + \mathbb{E}[\mu_k^2]\right) + C_2$

$\Rightarrow q(c_i) \propto \exp\left\{ \sum_k c_{ik} \left(\mathbb{E}_{\mu}[\mu_k] x_i - \frac{1}{2}\mathbb{E}_{\mu}[\mu_k^2] \right) \right\}$ $\qquad$ categorical dist is of ~~the form~~ $\exp\{\sum_{k=1}^{K} x_k \log p_k\}$

$\rightarrow \varphi_{ik} \propto \exp\left\{ \mathbb{E}_{-c_i}[\mu_k] x_i - \frac{1}{2}\mathbb{E}_{-c_i}[\mu_k^2] \right\}$

Section 3.2 : $q(\mu_k)$

- $q_j^*(z_j) \propto \exp\{\mathbb{E}_{-j}[\log p(z_j, z_{-j}, x)]\}$ $\quad$ $p(z, x)$

- $q^*(\mu_k) \propto \exp\{\mathbb{E}_{-\mu_k}[\log[p(x \mid c, \mu)\, p(\mu)\, \cancel{p(c)}]]\}$

$\qquad \propto \exp\left\{ \log p(\mu_k) + \sum_{i=1}^{n} \mathbb{E}_{-\mu_k}[\log p(x_i \mid c_i, \mu); \varphi_i, m_{-k}, s_{-k}^2] \right\}$

note $\varphi_{ik} = \mathbb{E}[c_{ik}; \varphi_i]$

- $\log q^*(\mu_k) = \log p(\mu_k) + \sum_i \mathbb{E}[\log p(x_i \mid c_i, \mu); \varphi_i, m_{-k}, s_{-k}^2] + C$

$= \log \underbrace{p(\mu_k)}_{\text{normal}} + \sum_i \mathbb{E}_{-\mu_k}[c_{ik} \log p(x_i \mid \mu_k)] + C$ $\qquad$ (using $p(x_i \mid c_i, \mu) = \prod_{k=1}^{K} p(x_i \mid \mu_k)^{c_{ik}}$ as before)

$= -\frac{\mu_k^2}{2\sigma^2} + \sum_i \log p(x_i \mid \mu_k)\, \mathbb{E}[c_{ik}] + C$

$= \frac{-\mu_k^2}{2\sigma^2} + \sum_i \varphi_{ik}\left[-\frac{1}{2}\underbrace{(x_i - \mu_k)^2}_{-\frac{1}{2}x_i^2 + x_i\mu_k - \frac{1}{2}\mu_k^2} \right] + C$

$= \frac{-\mu_k^2}{2\sigma^2} + \sum_i \left(\varphi_{ik} x_i \mu_k - \varphi_{ik}\frac{\mu_k^2}{2} \right) + C$

$= \left(\sum_i \varphi_{ik} x_i \right) \mu_k - \frac{1}{2}\left(\frac{1}{\sigma^2} + \sum_i \varphi_{ik} \right) \mu_k^2 + C$

$= \mathcal{N}(m_k^*, s_k^{2*})$ where $m_k^* = \frac{\sum_i \varphi_i x_i}{\sigma^2 + \sum_i \varphi_{ik}}$, $\quad s_k^{2*} = \frac{1}{\frac{1}{\sigma^2} + \sum_i \varphi_{ik}}$