① $Z_i \overset{iid}{\sim} \mathcal{M}(1, \pi), \quad \forall i \in \{1, \dots, n\}$

$$\begin{pmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix} \updownarrow K \qquad \begin{pmatrix} 0 \\ 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix} \qquad \dots \qquad \begin{pmatrix} 0 \\ \vdots \\ 0 \\ 1 \end{pmatrix}$$

one-hot encoding.

cluster 1   cluster 2   ...   cluster K

$$\mathbb{P}(\underbrace{Z_{ik} = 1}_{\text{1 in position } k \text{ for vector } Z_i}) = \pi_k \Rightarrow p(z_i \mid \pi) = \prod_{k=1}^{K} \pi_k^{z_{ik}}$$

$\Leftrightarrow$ observation i in cluster k

② $$X_i \mid Z_{ik} = 1 \sim \mathcal{N}(\mu_k, \Sigma_k) \Rightarrow p(x_i \mid z_i, \theta) = \prod_{k=1}^{K} \mathcal{N}(x_i; \mu_k, \Sigma_k)^{z_{ik}}$$

$$p(x_i, z_i \mid \pi, \theta) = p(x_i \mid z_i, \theta)\, p(z_i \mid \pi)$$
$$= \prod_{k=1}^{K} \left(\pi_k \mathcal{N}(x_i; \mu_k, \Sigma_k)\right)^{z_{ik}}$$

$$\Rightarrow p(x_i \mid \pi, \theta) = \sum_{z_i} p(x_i, z_i \mid \pi, \theta)$$

$$p(x_i \mid \pi, \theta) = \sum_{z_i} \left\{ \prod_{k=1}^{K} \left(\pi_k \mathcal{N}(x_i ; \mu_k, \Sigma_k)\right)^{z_{ik}} \right\}$$

$$= \pi_1 \mathcal{N}(x_i ; \mu_1, \Sigma_1) + \pi_2 \mathcal{N}(x_i ; \mu_2, \Sigma_2) + \ldots + \pi_K \mathcal{N}(x_i ; \mu_K, \Sigma_K)$$

$$= \sum_{k=1}^{K} \pi_k \mathcal{N}(x_i ; \mu_k, \Sigma_k)$$

$$L_{(x_i, z_i)_i}(\pi, \theta) = \sum_{k=1}^{K} \sum_{x_i \in C_k} \log\left[\pi_k \mathcal{N}(x_i ; \mu_k, \Sigma_k)\right]$$

assumptions: $\pi_k = \dfrac{1}{K}, \forall k$

$$\Sigma_k = I_d, \forall k$$

$$\Rightarrow L_{(x_i, z_i)}(\pi, \theta) = \sum_{k=1}^{K} \sum_{x_i \in C_k} \log(\overset{\frac{1}{K}}{\pi_k}) + \sum_{k=1}^{K} \sum_{x_i \in C_k} \log \mathcal{N}(x_i ; \mu_k, \Sigma_k)$$

$$= \underbrace{n \log\left(\frac{1}{K}\right)}_{} + \sum_{k=1}^{K} \sum_{x_i \in C_k} \log\left\{ \frac{1}{\sqrt{(2\pi)^d |I_d|}} \exp\left(-\frac{1}{2}(x_i - \mu_k)^T I_d^{-1} (x_i - \mu_k)\right) \right\}$$

$$= \sum_{k=1}^{K} \sum_{x_i \in C_k} \left(-\frac{1}{2}\right) \|x_i - \mu_k\|^2 + ct$$

$$= -2 W$$

$$P(z_i \mid x_i, \pi, \theta) = \frac{P(x_i, z_i \mid \pi, \theta)}{P(x_i \mid \pi, \theta)}$$

$$= \frac{\prod_{k=1}^{K} \left(\pi_k \mathcal{N}(x_i ; \mu_k, \Sigma_k)\right)^{z_{ik}}}{\sum_{l=1}^{K} \pi_l \mathcal{N}(x_i ; \mu_l, \Sigma_l)}$$

$$= \prod_{k=1}^{K} \left(\underbrace{\frac{\pi_k \mathcal{N}(x_i ; \mu_k, \Sigma_k)}{\sum_{l=1}^{K} \pi_l \mathcal{N}(x_i ; \mu_l, \Sigma_l)}}_{= \tau_{ik}} \right)^{z_{ik}}$$

$$= \prod_{k=1}^{K} \tau_{ik}^{z_{ik}}$$

$$= \mathcal{M}(z_i ; 1, \tau_i) \qquad \tau_i = \begin{pmatrix} \tau_{i1} \\ \tau_{i2} \\ \vdots \\ \tau_{iK} \end{pmatrix}$$

$$0 \leq \tau_{ik} \leq 1 \qquad \sum_{k=1}^{K} \tau_{ik} = 1 .$$