

①

$$z^{new} \leftarrow E(z)$$

$$\therefore E \text{ step: } q^{t+1}(z) \leftarrow P(z|x, \theta^t)$$

$$M \text{-step: } \theta^{t+1} \leftarrow \arg \max_{\theta} E_{z|x, \theta^t} [L(\theta)]$$

① obs: $X_{n,d} = [\tilde{x}_{n,d}, \tilde{x}_{n,d}, \dots, \tilde{x}_{n,d}]^T$ $\tilde{x}_{n,d} = [0_{n,d}, 0_{n,d}, \dots, 0_{n,d}]$ \rightarrow pixel

$N \times 3$ obs $D \times 3$ dimension $n \in [1, N]$ $d \in [1, D]$ $x_{n,d} \in \{0, 1\}$

$x_i \perp x_j$

Bernoulli Distribution:

$$P(X=1) = p, P(X=0) = 1-p$$

$$f(x|p) = \begin{cases} p & x=1 \\ 1-p & x=0 \end{cases}$$

$$\downarrow$$

$$f(x|p) = p^x \cdot (1-p)^{1-x} \quad x \in \{0, 1\}$$

$$X = \begin{bmatrix} 0_{1,d} & \dots & 0_{N,d} \\ \vdots & & \vdots \\ x_{1,d} & \dots & x_{N,d} \end{bmatrix}$$

$N \times D$

② probability $z_i \in [0, 1]$ $X_{n,d}$ follows $Ber(\mu_{k,d})$ $P(X_{n,d} | \mu_{k,d}) = \mu_{k,d}^{x_{n,d}} \cdot (1-\mu_{k,d})^{1-x_{n,d}}$

$$\sum_{i=1}^K z_i = 1 \quad z = [z_1, z_2, \dots, z_K]^T \quad \sum_{i=1}^K z_i = 1$$

$i \in 1 \dots K$ \rightarrow 对 X_n 聚类, 落在 cluster 1, ... cluster K 的概率 $P(\tilde{x} | z) = \sum_{i=1}^K P(\tilde{x} | z_i)$

③ Cluster:

z : the proportion of images in each cluster.

$$P(X_n \in k) = z_k = \frac{\text{\# of images in cluster } k}{\sum \text{\# of images}}$$

\uparrow mixing coefficient.

$$\begin{bmatrix} \frac{N_1}{N} & \frac{N_2}{N} & \dots & \frac{N_K}{N} \end{bmatrix}$$

\rightarrow parameter $\mu_{k,d} \in [0, 1]$ $1 \times D \times K$

$$\mu_{k,d} = \begin{bmatrix} \mu_{1,d} & \mu_{2,d} & \dots & \mu_{K,d} \end{bmatrix}$$

$$\mu_{k,d} = \begin{bmatrix} 0.1 & 0.2 & 0.5 & 0.7 & \dots & \dots \end{bmatrix}$$

$3 \times D \times K$

μ_i : 0, 1 的概率 \rightarrow pixel distribution

$$P(X_n | \mu, z) = \sum_i z_i \cdot P(X_n | \mu_i)$$

Marginal Distribution

④ Assignment: C_n

$1 \times 3 \times N$ 有 $3 \times C_n$ $\tilde{C}_n = [C_1^m, C_2^m, \dots, C_K^m]^T$ $C_k^m \in \{0, 1\}$

每个 X_n 属于 cluster? binary

$$\sum_{k=1}^K C_k^m = 1$$

$$C_{n,k} = \begin{bmatrix} k_1 & k_2 & k_3 & \dots & k_K \end{bmatrix}^T$$

1 of K

$$C = [\tilde{C}_1, \tilde{C}_2, \dots, \tilde{C}_N]$$

$$C_{n,i} = \begin{bmatrix} x_1 & x_2 & \dots & x_D \end{bmatrix}$$

$$C_{n,i} = \begin{bmatrix} 0.5 & 0.1 & \dots & \dots \\ 0.2 & 0.7 & \dots & \dots \\ 0.3 & 0.2 & \dots & \dots \end{bmatrix}$$

$3 \times N$

\downarrow col sum = 1

$$P(\tilde{x}_n | \mu_i) = \prod_{d=1}^D \mu_{k,d}^{x_{n,d}} (1-\mu_{k,d})^{1-x_{n,d}}$$

\uparrow Bernoulli Dst

对 x_n 中 $x_{n,d}=1$ 的项求 $p^x (1-p)^{1-x}$

if $P(C_{n,i}=1) = z_i$

$$P(\tilde{C} | z) = \prod_{i=1}^K z_i^{C_{n,i}}$$

if $P(\tilde{x}_n | \tilde{C}_{n,i}=1) = P(\tilde{x}_n | \mu_i)$

$C_i \sim \text{Bern}(\mu_i)$ 对 $j \in 1, 2, \dots, K$, $C_j = 1$ 的概率是 μ_j

$$P(\tilde{x}_n | \tilde{C}_{n,i}, \mu, z) = \prod_{i=1}^K P(\tilde{x}_n | \mu_i)^{C_{n,i}}$$

$$P(C | z) = \prod_{n=1}^N P(\tilde{C}_n | z) = \prod_{n=1}^N \prod_{i=1}^K z_i^{C_{n,i}}$$

$$P(X, C | \mu, z) = \prod_{n=1}^N P(X_n | C_n, \mu, z) = \prod_{n=1}^N \prod_{i=1}^K P(X_n | \mu_i)^{C_{n,i}}$$

$$= \prod_{n=1}^N \prod_{i=1}^K \left(\prod_{d=1}^D \mu_{k,d}^{x_{n,d}} (1-\mu_{k,d})^{1-x_{n,d}} \right)^{C_{n,i}}$$

$$P(X, C | \mu, z) = P(X | C, \mu, z) \cdot P(C | \mu, z)$$

$$= \prod_{n=1}^N \prod_{i=1}^K \left(z_i \prod_{d=1}^D \mu_{k,d}^{x_{n,d}} (1-\mu_{k,d})^{1-x_{n,d}} \right)^{C_{n,i}}$$

⑤ $\therefore z_i = \frac{N_i}{N}$ $N_i = \sum_{n=1}^N C_{n,i}$

$$= \frac{\sum_{n=1}^N C_{n,i}}{N}$$

Mid = $\frac{\text{\# of } (x_{n,d}=1)}{N_i}$

$$\text{Mid} = \frac{\sum x_{n,d} \cdot C_{n,i}}{N_i}$$

Bernoulli Distribution:

$$p(x=1) = p, \quad p(x=0) = 1-p$$

$$f(x|p) = \begin{cases} p & x=1 \\ 1-p & x=0 \end{cases}$$

$$f(x|p) = p^x \cdot (1-p)^{1-x} \quad x \in \{0, 1\}$$

①

$$z^{new} \leftarrow E(z)$$

$$\therefore E \text{ step: } q^{t+1}(z) \leftarrow P(z|X, \theta^t)$$

$$M \text{-step: } \theta^{t+1} \leftarrow \arg \max_{\theta} E_{q^{t+1}} [L(\theta)] \quad \text{pixel}$$

① obs: $X_{n,d} = [\vec{x}_{n,1}, \vec{x}_{n,2}, \dots, \vec{x}_{n,D}]^T$ $\vec{x}_{n,d} = [0_{n,1}, 0_{n,2}, \dots, 0_{n,D}]$

$N \times D$ obs D dimension $n \in \{1, \dots, N\}$ $d \in \{1, \dots, D\}$ $x_{n,d} \in \{0, 1\}$

$X_i \perp X_j$

$0_{n,1}$	$0_{n,2}$	$0_{n,3}$	$0_{n,4}$	$0_{n,5}$	$0_{n,6}$	$0_{n,7}$	$0_{n,8}$	$0_{n,9}$	$0_{n,10}$

② probability $\lambda_i \in [0, 1]$

$$X_{n,d} \text{ follows } \text{Ber}(\mu_{k,d}) \quad P(X_{n,d} | \mu_{k,d}) = \mu_{k,d}^{x_{n,d}} \cdot (1 - \mu_{k,d})^{1-x_{n,d}}$$

$$\sum_{i=1}^K \lambda_i = 1 \quad \lambda = [\lambda_1, \lambda_2, \dots, \lambda_K]^T \quad \sum_{i=1}^K \lambda_i = 1$$

$$i \in 1 \dots K \quad \rightarrow \text{对 } X_n \text{ 聚类, 落在 cluster 1, \dots, cluster K 的概率} \quad P(\vec{X} | \lambda) = \sum_{i=1}^K P(\vec{X} | \lambda_i)$$

③ Cluster:

λ : the proportion of images in each cluster.

$$P(X_{n,d} = k) = \lambda_k = \frac{\text{\# of images in cluster } k}{\sum_{i=1}^K \text{\# of images}}$$

↑ mixing coefficient.

λ_1	λ_2	λ_3	λ_4	λ_5	λ_6	λ_7	λ_8	λ_9	λ_{10}

Parameter $\mu_{k,d} \in [0, 1] \quad 1 \times D \times K$

$\mu_{k,d}$: $\begin{matrix} \mu_1 & \mu_2 & \mu_3 & \mu_4 & \mu_5 & \mu_6 & \mu_7 & \mu_8 & \mu_9 & \mu_{10} \end{matrix}$

μ_1	0.3	0.2	0.5	0.7					
μ_2									
μ_3									

μ_i : 0_i 为 1 的概率

(pixel distribution)

$$P(X_n | \mu, \lambda) = \sum_{i=1}^K \lambda_i \cdot P(X_n | \mu_i) \quad \text{Marginal Distribution}$$

④ Assumption: C_n

$$1 \text{ of } X_n \text{ 有 } \vec{C}_n, \vec{C}_n = [C_1^m, C_2^m, \dots, C_K^m]^T \quad C_k^m \in \{0, 1\}$$

每个 X_n 有 cluster? binary

$$\sum_{k=1}^K C_k^m = 1$$

$$\text{if } P(C_{ni} = 1) = \lambda_i$$

$$C_{n,K} = \begin{bmatrix} k_1 & k_2 & k_3 \\ 1 & 0 & 0 \end{bmatrix}^T \quad 1 \text{ of } K$$

$$P(\vec{C} | \lambda) = \prod_{i=1}^K \lambda_i^{C_{ni}}$$

$C_i \sim \text{Bern}(\mu_i)$
对 i 个 X_i , $C_i = 1$ 的概率是 μ_i

$$\text{if } P(\vec{X}_n | \vec{C}_{ni} = 1) = P(\vec{X}_n | \mu_i)$$

$$P(\vec{X}_n | \vec{C}_n, \mu, \lambda) = \prod_{i=1}^K P(\vec{X}_n | \mu_i)^{C_{ni}}$$

$$P(C | \lambda) = \prod_{n=1}^N P(\vec{C}_n | \lambda) = \prod_{n=1}^N \prod_{i=1}^K \lambda_i^{C_{ni}}$$

$$P(X, C | \mu, \lambda) = \prod_{n=1}^N P(X_n | C_n, \mu, \lambda) = \prod_{n=1}^N \prod_{i=1}^K P(X_n | \mu_i)^{C_{ni}}$$

$$= \prod_{n=1}^N \prod_{i=1}^K \left(\prod_{d=1}^D \mu_{i,d}^{x_{n,d}} (1 - \mu_{i,d})^{1-x_{n,d}} \right)^{C_{ni}}$$

②

$$L(\theta) = \log P(X, C | \theta) = \log P(X, C | \mu, \lambda)$$

$$= \sum_{n=1}^N \sum_{i=1}^K C_{ni} \left(\log \lambda_i + \sum_{d=1}^D x_{n,d} \log \mu_{i,d} + (1 - x_{n,d}) \log (1 - \mu_{i,d}) \right)$$

$$\therefore P(X, C | \mu, \lambda) = P(X | C, \mu, \lambda) \cdot P(C | \mu, \lambda)$$

$$= \prod_{n=1}^N \prod_{i=1}^K \left(\lambda_i \prod_{d=1}^D \mu_{i,d}^{x_{n,d}} (1 - \mu_{i,d})^{1-x_{n,d}} \right)^{C_{ni}}$$

③

$$\lambda_i = \frac{N_i}{N} \quad N_i = \sum_{n=1}^N C_{ni}$$

$$= \frac{\sum_{n=1}^N C_{ni}}{N} \quad \text{Mid} = \frac{\text{\# of } (X_{n,d} = 1)}{N_i}$$

$$\text{Mid} = \frac{\sum_{n,d} X_{n,d} \cdot C_{ni}}{N_i}$$