Bernoulli Distribution:

$p(X=1) = p \quad p(X=0) = 1-p$

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

↓

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

$X =$

	θ_1	...	θ_{784}
x_1			
⋮			
x_N			

$N \times 784$

①

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

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

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

① obs: $X_{n,d} = [\vec{x}_{1d}, \vec{x}_{2d} \dots \vec{x}_{Nd}]^T \quad \vec{x}_{nd} = [\theta_{n1}, \theta_{n2} \dots \theta_{nD}]$ pixel

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

$x_i \perp\!\!\!\perp x_j$

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

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

$i \in 1 \dots K$. →对于 x_n 来说，落在 [cluster 1, … cluster k] 的概率 $P(\vec{x}|\lambda) = \sum_{i=1}^{K} P(\vec{x}|\lambda_i)$

③ Cluster.

λ: the proportion of images in each cluster.

$p(x_i \in k_i) = \lambda_i = \frac{\#\text{ of images in i cluster}}{\sum \#\text{ of images}}$

↑ mixing coefficient

k_1	k_2	...			k_K
$\frac{N_1}{N}$					

→ parameter $\mu_{k,d} \in [0,1]$ 1×784

μ矩阵:

	θ_1				...	θ_{784}
μ_1	0.3	0.2	0.5	0.7		
μ_2						
μ_3						

3×784.

μ_i: θ_i 为1 的概率

$\boxed{\text{pixel distribution}}$

$\therefore P(x_n | \mu, \lambda) = \sum_i \lambda_i \cdot P(x_{n,d}|\mu_i)$ marginal Distribution

$\because P(\vec{x}_n | \mu_i) = \prod_{d=1}^{D} \mu_{k,d}^{x_{n,d}} (1-\mu_{k,d})^{1-x_{nd}}$ 相当于把 $x_{n,d}=1$ 的概率乘 $p^x (1-p)^{1-x}$.

↑ Bernoulli Dist

④ Assignment: c_n

假设 x_n 有一个 $\vec{c}_n$, $\vec{c}_n = [c_1^{(n)}, c_2^{(n)} \dots c_K^{(n)}]^T$ $c_k^{(n)} = \{0,1\}$.

属于哪个 cluster? binary $\sum_{k=1}^{K} c_k^{(n)} = 1$

$c_{n,k} =$

k_1	k_2	k_3
1	0	0

^{T} 1 of K

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

$P(\vec{c}|\lambda) = \prod_{i=1}^{K} \lambda_i^{c_{n,i}}$

if $P(\vec{x}_n | \vec{c}_{ni}=1) = P(\vec{x}_n|\mu_i)$

$P(\vec{x}_n | \vec{c}_n, \mu, \lambda) = \prod_{i=1}^{K} P(\vec{x}_n|\mu_i)^{c_{n,i}}$

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

$c_i \sim Bern(\mu_i)$ 对于 obs x_i, $c_i=1$ 的概率就是 μ_i

$C = [\vec{c}_1\ \vec{c}_2 \dots \vec{c}_N] =$

	x_1	x_2		x_N
k_1	0.5	0.1		
k_2	0.2	0.7		
k_3	0.3	0.2		

$3 \times N$.

col sum = 1

$c_{n,i}$

$\therefore P(X|C,\mu,\lambda) = \prod_{n=1}^{N} P(x_n|c_n,\mu,\lambda) = \prod_{n=1}^{N} \prod_{i=1}^{K} P(x_n|\mu_i)^{c_{n,i}}$

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

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

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

②

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

$= \sum_{n=1}^{N} \sum_{i=1}^{K} c_{n,i} \left(\lg \lambda_i + \sum_{d=1}^{D} x_{nd} \cdot \lg \mu_{id} + (1-x_{n,d}) \cdot \lg(1-\mu_{id}) \right)$

③ $\therefore \lambda_i = \frac{N_i}{N} \quad N_i = \sum_{n=1}^{N} c_{ni}$

$= \frac{\sum_{n=1}^{N} c_{ni}}{N} \quad \mu_{id} = \frac{\#\text{ of }(x_{nd}=1)}{N_i}$

$\mu_{id} = \frac{\sum x_{nd} \cdot c_{ni}}{N_i}$

Bernoulli Distribution:

$p(X=1)=p \quad p(X=0)=1-p$

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

$\downarrow$

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

①

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

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

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

① obs: $X_{n,d}=[\vec{x}_{1d}, \vec{x}_{2d} \ldots \vec{x}_{Nd}]^T$ $\quad \vec{x}_{nd}=[\theta_{n1}, \theta_{n2} \ldots \theta_{nD}]$ (pixel)

N个 obs $\quad$ D个 dimension $\quad n\in[1,N] \quad d\in[1,D] \quad X_{n,d}\in\{0,1\}$

$x_i \perp\!\!\!\perp x_j$

$X=$

	θ_1	…	θ_{784}
x_1			
⋮			
x_N			

$N\times 784$

② probability $\lambda_i\in[0,1]$ $\quad \sum_{i=1}^{K}\lambda_i=1$

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

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

$i\in 1\ldots K$. $\quad \to$ 对于 X_n 来说, 落在[cluster 1, … 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.

$P(x_n\in k_i)=\lambda_i=\dfrac{\#\text{ of images in } i \text{ cluster}}{\sum \#\text{ of images}}$

$\uparrow$ mixing coefficient

k_1	k_2				k_K
$\frac{N_1}{N}$					

$\to$ parameter $\mu_{k,d}\in[0,1]$ $\quad 1\times 784$

M_{k}:

	θ_1	…			θ_{784}
μ_1	0.3	0.2	0.5	0.7	
μ_2					
μ_3					

3×784.

μ_i: θ_i 为1 的概率

[pixel distribution]

$\therefore P(X_n|\mu,\lambda_i)=\sum_{i}\lambda_i\cdot P(X_{n,d}|\mu_i)$ ☆ marginal Distribution

☆ $P(\vec{x}_n|\mu_i)=\prod_{d=1}^{D}\mu_{k,d}^{\,x_{n,d}}(1-\mu_{k,d})^{1-x_{nd}}$ $\quad$ 相当于把 $x_{nd}=1$ 的概率求出来 $\quad p^{x}(1-p)^{1-x}$.

$\uparrow$ Bernoulli Dist

④ Assignment: c_n

每个 x_n 有一个 $\vec{c}_n$. $\quad \vec{c}_{n,k}=[c_1^{(n)}, c_2^{(n)} \ldots c_K^{(n)}]^T \quad c_k^{(n)}=\{0,1\}$.

属于哪个 cluster? binary $\quad \sum_{k=1}^{K}c_k^{(n)}=1$

$c_{n,k}=$

k_1	k_2	k_3
1	0	0

^{T} $\quad$ 1 of K

if $P(c_{n,i}=1)=\lambda_i$

$P(\vec{c}|\lambda)=\prod_{i=1}^{K}\lambda_i^{\,c_{n,i}}$

$c_i\sim Bern(\mu_i)$ 对于 obs x_i, $c_i=1$ 的概率 就是 μ_i

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

$P(\vec{x}_n|\vec{c}_n,\mu,\lambda)=\prod_{i=1}^{K}P(\vec{x}_n|\mu_i)^{c_{n,i}}$

$C=[\vec{c}_1\ \vec{c}_2 \ldots \vec{c}_N]=$

	x_1	x_2		x_N
k_1	0.5	0.1		
k_2	0.2	0.7		
k_3	0.3	0.2		

$3\times N$.

col sum = 1

$c_{n,i}$

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

②

$\therefore P(X|C,\mu,\lambda)=\prod_{n=1}^{N}P(x_n|c_n,\mu,\lambda)=\prod_{n=1}^{N}\prod_{i=1}^{K}P(x_n|\mu_i)^{c_{n,i}}$

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

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

$=\sum_{n=1}^{N}\sum_{i=1}^{K}c_{n,i}\left(\lg\lambda_i+\sum_{d=1}^{D}x_{nd}\cdot\lg\mu_{id}+(1-x_{n,d})\cdot\lg(1-\mu_{id})\right)$

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

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

③ $\therefore \lambda_i=\dfrac{N_i}{N} \quad N_i=\sum_{n=1}^{N}c_{ni}$

$=\dfrac{\sum_{n=1}^{N}c_{ni}}{N} \quad \mu_{id}=\dfrac{\#\text{ of }(x_{nd}=1)}{N_i}$

$\mu_{id}=\dfrac{\sum x_{nd}\cdot c_{ni}}{N_i}$