GMM $\sum_{i=1}^{K} \pi_i N(X|\mu_K, \Sigma_K)$

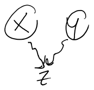
$X \sim N(\mu_1, \Sigma_1)$ $\Leftrightarrow$ $N(X|\mu_1, \Sigma_1)$

$Y \sim N(\mu_2, \Sigma_2)$

2 RANDOM VARIABLES !

$Z = X + Y$ IS GAUSSIAN

$Z \sim N(\dots, \dots)$



P1) $J(X, \underline{Z, \mu}) = \sum_{i=1}^{N} \sum_{j=1}^{K} Z_{ij} \cdot \|X_i - \mu_j\|_1 \to \min$

a) UPDATE OF Z

μ_3 μ_1

μ_2 x

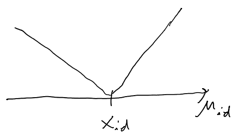
$$Z_{ij} = \begin{cases} 1 & \text{IF } j = \arg\min_k \|X_i - \mu_k\|_1 \\ 0 & \text{ELSE} \end{cases}$$

b) UPDATE OF μ

SOLVE FOR μ_1: $\sum_{\substack{i=1 \\ Z_{i1}=1}}^{N} \|X_i - \mu_1\|_1$

$$= \sum_{\substack{i=1 \\ Z_{i1}=1}}^{N} \sum_{d=1}^{D} |X_{id} - \mu_{1d}| \to \min$$

$$= \sum_{d=1}^{D} \underbrace{\sum_{i, Z_{i1}=1} |X_{id} - \mu_{1d}|}_{\text{MIN THIS ! } =(1)}$$



$$\frac{\partial}{\partial \mu_{jd}} |X_{id} - \mu_{jd}| = \begin{cases} 1 & \text{IF } \mu_{jd} > X_{id} \\ -1 & \text{IF } \mu_{jd} < X_{id} \\ 0 & \text{IF } \mu_{jd} = X_{id} \end{cases}$$ ATTENTION !!!

$\frac{\partial (1)}{\partial \mu_{jd}} \overset{!}{=} 0$ IN THE OPTIMAL POINT

5 6 7 10

x_{1d} x_{2d} x_{3d} x_{4d}

6.5

5 6 7 10 11

$\mu - 5$
$+ \mu - 6$
$+ 7 - \mu$
$+ 10 - \mu$

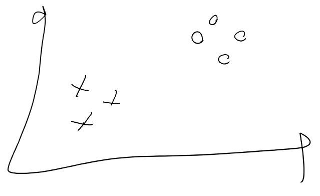
$\mu \in [6, 7]$

TODO !

$\Downarrow$

$0 \cdot \mu - 5 - 6 + 7 + 10 =$ CONST

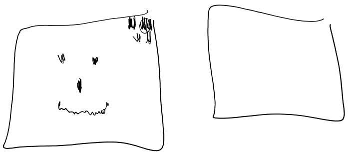
$\mu_{jd} = \underline{\text{MEDIAN}} \{X_{id} \mid i \text{ WHERE } Z_{ij} = 1\}$



$$|x| = \begin{cases} -x & x < 0 \\ x & x \geq 0 \end{cases}$$

P4) "MULTIVARIATE" - BERNOULLI - MIXTURE MODEL

GOAL: CLUSTER BINARY DATA



Z $\quad Z \sim CAT(\pi)$

$\downarrow$

X $\quad x_i \sim \mathcal{N}(\mu_{z_i}, \Sigma_{z_i})$

$\leftarrow$ CONTINUOUS DATA

$$\begin{array}{ccc} x_1 & x_2 & x_3 \\ \begin{bmatrix} 0 \\ 0 \\ 1 \\ 1 \end{bmatrix} & \begin{bmatrix} 1 \\ 1 \\ 0 \\ 0 \end{bmatrix} & \begin{bmatrix} 1 \\ 0 \\ 1 \\ 1 \end{bmatrix} \end{array}$$

INPUT DATA: $x_i \in \{0,1\}^D$

BINARY DATA

Z $\quad Z \sim CAT(\pi)$

$\downarrow$

X $\quad x_{id} \sim BER(\theta_{z_i, d})$

$$\theta_1 \in [0,1]^D \quad \sim \begin{bmatrix} 0.9 \\ 0.9 \\ 0.1 \\ 0.1 \end{bmatrix}$$

$$\theta_2 \in [0,1]^D \quad \sim \begin{bmatrix} 0.2 \\ 0.1 \\ 0.8 \\ 0.9 \end{bmatrix}$$

FOR x_2

$0.9 \cdot 0.9 \cdot 0.9 \cdot 0.9$

$0.2 \cdot 0.1 \cdot 0.2 \cdot 0.1$

$$p_\theta(x_i | z_i = k) = \prod_{d=1}^{D} \theta_{kd}^{x_{id}} \cdot (1 - \theta_{kd})^{(1 - x_{id})}$$

$$p(z = k | x, \theta) \Rightarrow r_{nk}(\theta) = \frac{p(z = k|\theta) \cdot p(x | z = k, \theta)}{\sum_{j=1}^{K} p(z = j|\theta) \cdot p(x | z = j, \theta)}$$

$Z \sim CAT(\frac{1}{K}, \frac{1}{K}, \frac{1}{K})$

$\Rightarrow p(z = k|\theta) = \text{CONST}$

E-STEP

$$\Rightarrow \frac{p(x | z = k, \theta)}{\sum_{j=1}^{K} p(x | z = j, \theta)} = r_{nk}(\theta)$$

$$\overset{1/K}{\overset{\shortparallel}{p(z|\theta)}} \cdot p(x|z, \theta)$$

$\uparrow$

M-STEP

$$\mathbb{E}_{z \sim p(z|x, \theta^{old})}\left[\log p(x, z | \theta)\right]$$

$$= \sum^{N} \sum^{K} r_{nk}(\theta^{old}) \cdot \ln\left(\frac{1}{K} \cdot \prod_{d=1}^{D} \theta_{kd}^{x_{id}} \cdot (1 - \theta_{kd})^{1 - x_{id}}\right)$$

$$= \sum_{n=1}^{N} \sum_{k=1}^{K} r_{nk}(\theta^{old}) \cdot \ln\left(\frac{1}{K} \cdot \prod_{d=1}^{D} \theta_{kd}^{x_{id}} \cdot (1-\theta_{kd})^{1-x_{id}}\right)$$

$$\ln\frac{1}{K} + \sum_{d=1}^{D} x_{id} \cdot \ln(\theta_{kd}) + (1-x_{id}) \cdot \ln(1-\theta_{kd})$$

$$= f(\theta)$$

$$\max_{\theta} f(\theta)$$

$$\frac{\partial f(\theta)}{\theta_{k'd'}} = \sum_{n=1}^{N} r_{nk'}(\theta^{old}) \cdot \frac{\partial}{\partial \theta_{k'd'}}\Big[x_{id'} \cdot \ln(\theta_{k'd'}) + (1-x_{id'}) \cdot \ln(1-\theta_{k'd'}) \Big]$$

$$= \sum_{n=1}^{N} r_{nk'}(\theta^{old}) \cdot \left[\frac{x_{id'}}{\theta_{k'd'}} - \frac{(1-x_{id'})}{(1-\theta_{k'd'})} \right]$$

$$\overset{!}{=} 0$$

$$\Rightarrow \theta_{k'd'} = \frac{\sum_{n=1}^{N} r_{nk}(\theta^{old}) \cdot x_{id'}}{\sum_{n=1}^{N} r_{nk}(\theta^{old})}$$

0 0 0 1 1 1

1 1 1

$\frac{1}{3}$ $\frac{2}{3}$ $\frac{\#1}{\#0 + \#1}$

$$\mathbb{E}_{Z \sim p(Z|\theta^{old})}\left[\log p(X, Z|\theta)\right]$$

$$= \sum_{Z} p(Z|\theta^{old}) \cdot \log p(X, Z|\theta)$$

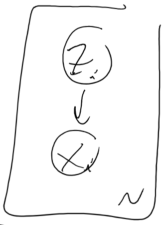
$$= \sum_{Z} p(Z|\theta^{old}) \cdot \log \prod_{i=1}^{N} p(x_i, Z|\theta)$$

$$= \sum_{Z} p(Z|\theta^{old}) \cdot \sum_{i} \log p(x_i, Z|\theta)$$

$$= \sum_{Z} \sum_{i=1}^{N} p(Z|\theta^{old}) \cdot \log p(x_i, Z|\theta)$$

$$= \sum_{i=1}^{N} \sum_{Z_1} \sum_{Z_2} \sum_{Z_3} \cdots \sum_{Z_N} \Big[p(Z_1|\theta^{old}) \cdot p(Z_2|\theta^{old}) \cdots p(Z_N|\theta^{old}) \cdot \log p(x_i, Z_i|\theta) \Big]$$

$= 1$



$$p(x_1|Z_1) = p(x_1|Z_1, Z_4, Z_3, Z_4)$$

$$p(x_i, Z|\theta) = p(x_i|Z, \theta) \cdot p(Z|\theta) = p(x_i|Z_i, \theta) \cdot \prod_i p(Z_i|\theta)$$

$$\log p(x_i, z_i \mid \theta) \Big] \quad = 1$$

$$= \sum_{i=1}^{N} \boxed{\sum_{z_1} \sum_{z_2} \cdots \sum_{z_{i-1}} \sum_{z_{i+1}} \cdots \sum_{z_N} \prod_{\substack{j=1 \\ j \neq i}}^{N} p(z_j \mid \theta^{old})} \quad \Big| \sum_{\substack{z_j \\ j \neq i}}$$

$$\sum_{z_i} p(z_i \mid \theta^{old}) \cdot \log p(x_i, z_i \mid \theta)$$

$$= \sum_{i=1}^{N} \sum_{z_i} p(z_i \mid \theta^{old}) \cdot \log p(x_i, z_i \mid \theta)$$

$$= \sum_{i=1}^{N} \sum_{k=1}^{K} \underbrace{p(z_i = k \mid \theta^{old})}_{r_{nk}(\theta^{old})} \cdot \log \ldots$$

$$\leftarrow p(z_i = k \mid \theta) \cdots$$

$z_1, \ldots, z_N$

$\uparrow$

domain
$\{1, 2, \ldots, K\}$

$$p(z_{ik} = 1 \mid \theta) \ldots$$

$\uparrow$

$z_{11}, z_{12}, \ldots, z_{1K}$ domain $\{0, 1\}$

$\vdots$

$z_{N1} \quad \ldots \quad z_{NK}$

$$X \sim \mathrm{Ber}(\theta)$$

$$\theta^{x} \cdot (1-\theta)^{(1-x)}$$

$$x \cdot \theta + (1-x) \cdot (1-\theta)$$

$$\begin{bmatrix} \theta_1 \\ \theta_2 \end{bmatrix} \qquad \theta_1^{x_1} \cdot \theta_2^{(1-x)}$$

s.t. $\theta_1 + \theta_2 = 1$

$\theta_i \geq 0$

$$\frac{\pi_j \cdot p(x_i \mid \mu_j, \Sigma_j)}{\sum_{j'} \pi_{j'} \cdot p(x_i \mid \mu_{j'}, \Sigma_{j'})} = p(z = j \mid x, \mu, \pi, \Sigma)$$

$$= \frac{\pi_z \cdot p(x_i \mid \mu_z, \Sigma_z)}{\dots}$$

$$z_i \sim \mathrm{CAT}(\pi)$$

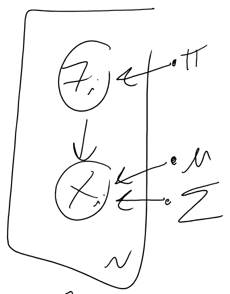
$$x_i \sim \mathcal{N}(\mu_{z_i}, \Sigma_{z_i})$$

M-STEP GMM

$$\mathbb{E}_{z \sim p(z \mid \theta^{old})} \left[\log \left[p(Z \mid \theta) \cdot p(X \mid Z, \theta) \right] \right]$$

$$= \sum_{i=1}^{N} \sum_{k=1}^{K} r_{ik}(\theta^{old}) \cdot \left[\ln p(z_i^{k} \mid \theta) + \ln p(x_i \mid z_i^{k}, \theta) \right]$$

$$= \sum_{i=1}^{N} \sum_{k=1}^{K} r_{ik}(\theta^{old}) \cdot \left[\ln \pi_k + \ln \mathcal{N}(x_i \mid \mu_k, \Sigma_k) \right]$$



$$\theta = \{\pi, \mu, \Sigma\}$$

$$= \sum_{i=1}^{N} \sum_{k=1}^{K} r_{ik}(\theta^{old}) \cdot \left[\ln \pi_k - \frac{1}{2} \ln \det(\Sigma_k^{-1}) - \frac{1}{2} (x_i - \mu_k)^T \cdot \Sigma_k^{-1} \cdot (x_i - \mu_k) + \mathrm{const} \right]$$

$$= Q(\pi, \mu, \Sigma)$$

GOAL: $\to$

$$\max_{\pi, \mu, \Sigma} Q(\pi, \mu, \Sigma)$$

$$\frac{\partial Q}{\partial \mu_{k'}} \left[\sum_{i=1}^{N} r_{ik'}(\theta^{old}) \cdot \left[-\frac{1}{2} (x_i - \mu_k)^T \Sigma_k^{-1} (x_i - \mu_k) \right] \right]$$

$$= \sum_{i=1}^{N} r_{ik'}(\theta^{old}) \cdot \Sigma_k^{-1} (x_i - \mu_k)$$

$$= \Sigma_k^{-1} \cdot \underbrace{\left[\sum_{i=1}^{N} r_{ik'}(\theta^{old}) \cdot (x_i - \mu_k) \right]}_{\overset{!}{=} 0}$$

$$\boxed{\begin{aligned} &\frac{\partial}{\partial s} (x-s)^T \cdot W (x-s) \\ &= -2 \cdot W \cdot (x-s) \end{aligned}}$$

$$\overset{!}{=} 0$$

$$\Leftrightarrow \mu_k = \frac{\sum_{i=1}^{N} r_{ik'}(\theta^{old}) \cdot x_i}{\sum_{i=1}^{N} r_{i,k'}(\theta^{old})}$$

$$\boxed{\begin{aligned} &A \cdot b + A \cdot c \\ &= A \cdot [b + c] \end{aligned}}$$

$$\frac{\partial Q}{\cdot} = \frac{\partial}{\cdot} \sum_{i=1}^{N} r_{ik'}(\theta^{old}) \cdot \left[-\frac{1}{2} \ln \det(\Sigma_k^{-1}) - \frac{1}{2} (x_i - \right.$$

$$\mu_k)^T \Sigma_{k'}^{-1} \cdot (x_i - \mu_{k'})\Big]$$

$$\frac{\partial Q}{\partial(\Sigma_{k'}^{-1})} = \frac{\partial}{\partial \Sigma_{k'}^{-1}} \sum_{i=1} r_{ik'}(\theta^{old}) \cdot \Big[-\frac{1}{2} \ln \det(\Sigma_k) - \frac{1}{2}(x_i -$$

$$= \sum_{i=1}^{N} r_{ik'}(\theta^{old}) \cdot \Big[-\frac{1}{2} \big(\underbrace{(\Sigma_k^{-1})^{-1}}_{\Sigma_k}\big)^T -$$

$$-\frac{1}{2} (x_i - \mu_{k'}) \cdot (x_i - \mu_{k'})^T \Big]$$

$$\stackrel{!}{=} 0$$

$$\Leftrightarrow \Sigma_k = \frac{\sum_{i=1}^{N} r_{ik'}(\theta^{old}) (x_i - \mu_{k'}) \cdot (x_i - \mu_{k'})^T}{\sum_{i=1}^{N} r_{ik'}(\theta^{old})}$$

$$\pi_k = \frac{\sum_{i=1}^{N} r_{ik}(\theta^{old})}{N}$$

$$\tilde{x}_i = ($$

$$\tilde{x}_i^T$$

$$= tr(\tilde{x}$$

$\mu_k) \Sigma_{k'} (x_i - \mu_{k'}))$

$$\frac{\partial}{\partial A} \log \det(A) = (A^{-1})^T$$

$$\frac{\partial}{\partial A} tr(B \cdot A) = B^T$$

$$tr(ABC) = tr(C \cdot A \cdot B)$$

$(x_i - \mu_k)$

$\Sigma_k^{-1} \cdot \tilde{x}_i =$

$^T \cdot \Sigma_k^{-1} \tilde{x}_i)$

$= \mathrm{tr}\,($

$$\frac{\partial f}{\partial \Sigma_\mu^{-1}}$$

$$\tilde{x}_i \cdot \tilde{x}^T \cdot \Sigma_k^{-1}] = 6$$

$$= [\tilde{x}_n \cdot \tilde{x}_i^T]^T = \tilde{x}_i^{T^T} \cdot \tilde{x}_n^T$$

$$= \tilde{x}_n \cdot \tilde{x}_i^T$$

$$= (x_i - \mu_k)^T \cdot (x_n - \mu_k)^T$$