E-step:

$$Q(\theta) = E_{z|x}\left[L(\theta) \right]$$

M-step:

$$\hat{\theta} = \arg\max_{\theta} Q(\theta)$$

$$P(x|\theta) = (1-\alpha)\underline{N(x|\mu_0, \Sigma_0)} + \alpha \underline{N(x|\mu_1, \Sigma_1)}$$

$$\theta = (\alpha, \mu_0, \Sigma_0, \mu_1, \Sigma_1) \qquad 0 \leq \alpha \leq 1$$

$$D = \{x_1, \dots, x_n\}$$

E-step:

$$\gamma_i^t = Pr(z_i = 1 | x_i, \theta^t) = E_{z_i|x_i}[z_i] \qquad 0 \leq \gamma_i^t \leq 1$$

M-step:

$$\hat{\mu}_0 = \frac{\sum_{i=1}^{n} x_i (1-\gamma_i^t)}{\sum_{i=1}^{n} (1-\gamma_i^t)}$$

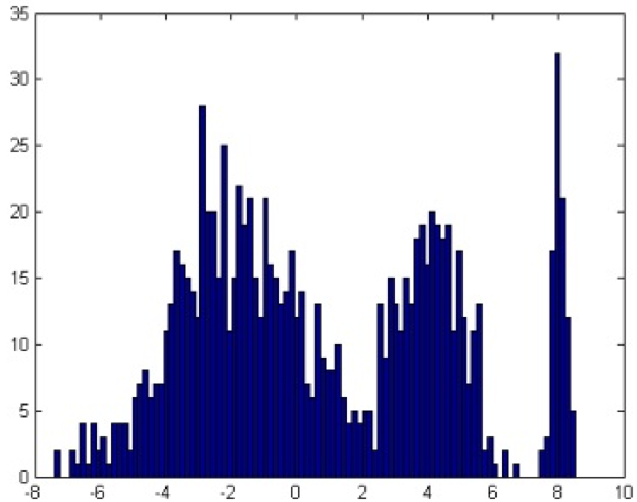
$$\boxed{\hat{\alpha} = \frac{\sum_{i=1}^{n} \gamma_i^t}{n}}$$

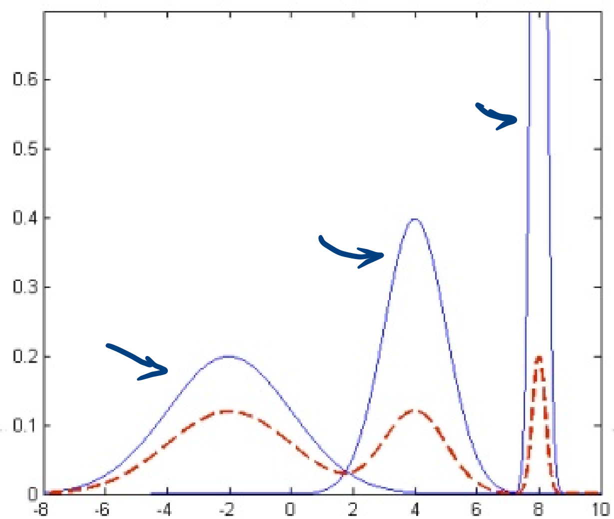
$$\hat{\Sigma}_1 = \frac{\sum_{i=1}^{n} \gamma_i^t (x_i - \mu_1)(x_i - \mu_1)^T}{\sum_{i=1}^{n} \gamma_i^t}$$

$x_i \qquad z_i$

?

GMM: 1-D Example



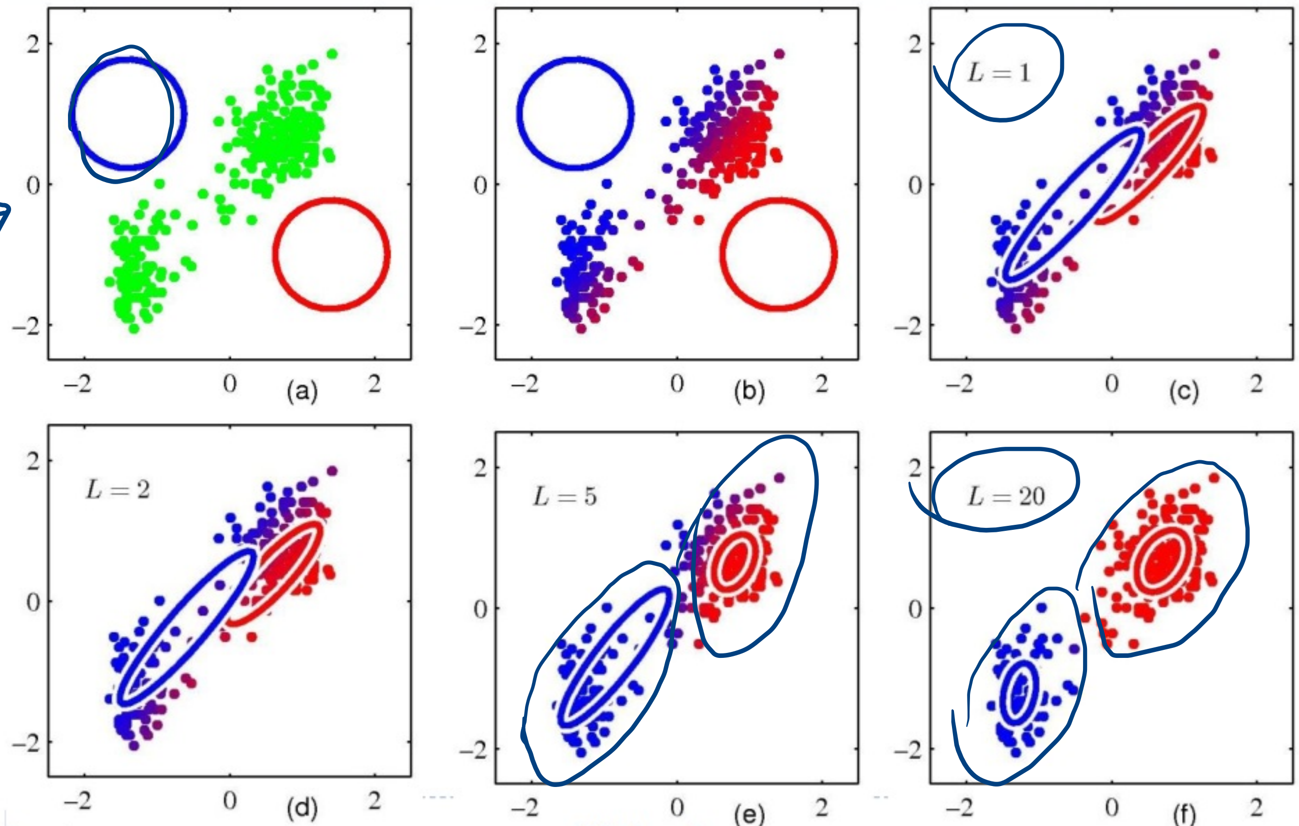


$\mu_1 = -2$
$\sigma_1 = 2$
$\pi_1 = 0.6$

$\mu_2 = 4$
$\sigma_2 = 1$
$\pi_2 = 0.3$

$\mu_3 = 8$
$\sigma_3 = 0.2$
$\pi_3 = 0.1$

EM & GMM: Example



[Bishop]

$\gamma_i^t \geq \frac{1}{2}$

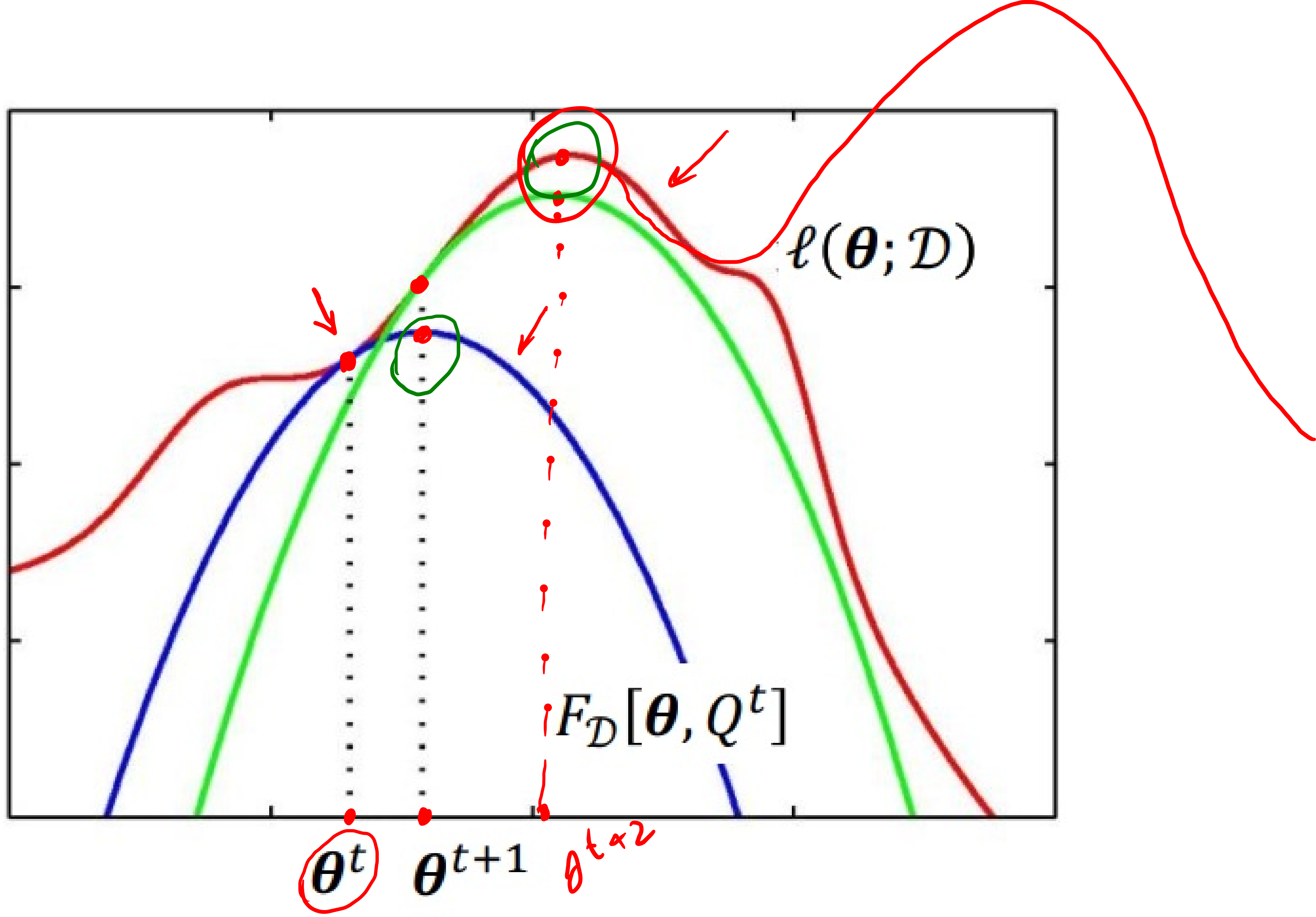
$$D = \{x_1, \dots, x_n\} \qquad H = \{z_1, \dots, z_n\}$$

$$\hat{\theta}_{ML} = \arg\max_{\theta} \underbrace{\log P(D|\theta)}_{\text{Likelihood}} = \arg\max_{\theta} \log P(x_1, \dots, x_n | \theta)$$

$$= \arg\max_{\theta} \sum_{i=1}^{n} \log P(x_i|\theta) = \arg\max_{\theta} \boxed{\sum_{i=1}^{n} \log \sum_{z_i} P(x_i, z_i|\theta)}$$

log likelihood

$$\theta^{t+1} = \theta^{t} + \eta \nabla_{\theta} L(\theta^{t})$$



$\ell(\boldsymbol{\theta}; \mathcal{D})$
$F_{\mathcal{D}}[\boldsymbol{\theta}, Q^t]$
$\boldsymbol{\theta}^t$
$\boldsymbol{\theta}^{t+1}$
θ^{t+2}

$$P(x|\theta) = \sum_{k=1}^{K} \alpha_k N(x|\mu_k, \Sigma_k) \qquad \sum_{k=1}^{K} \alpha_k = 1$$

$$\alpha_k \geq 0$$

$$D = \{x_1, x_2, x_3, \dots, x_n\}$$

$$H = \{z_1, z_2, z_3, \dots, z_n\}$$

$$\boxed{z_i = \begin{bmatrix} 0 \\ 0 \\ 1 \\ \vdots \\ 0 \end{bmatrix}_{K\times 1} \to k}$$

$$P(x_i, z_i|\theta) = \prod_{k=1}^{K} \left(\alpha_k N(x_i|\mu_k, \Sigma_k)\right)^{z_{ik}}$$

$$\cancel{= \sum_{k=1}^{K} z_k \alpha_k N(x|\mu_k, \Sigma_k)}$$

$$L(\theta) = \log P(D, H|\theta) = \log P(x_1, z_1, x_2, z_2, \dots, x_n, z_n|\theta) = \log \prod_{i=1}^{n} P(x_i, z_i|\theta) = \sum_{i=1}^{n} \log P(x_i, z_i|\theta)$$

$$= \sum_{i=1}^{n} \sum_{k=1}^{K} z_{ik}\left(\log \alpha_k + \log N(x_i|\mu_k, \Sigma_k)\right)$$

$$Q(\theta) = E_{z_i}\left[L(\theta)\right] = \sum_{i=1}^{n} \sum_{k=1}^{K} \underbrace{E[z_{ik}]}_{\gamma_{ik}^{t}} \left(\log \alpha_k + \log N(x_i|\mu_k, \Sigma_k)\right)$$

$$\underset{z_i|x_i}{E}[z_{ik}] = Pr(z_{ik}=1 \mid x_i, \theta^t)$$

$$= \frac{p(x_i \mid z_{ik}=1, \theta^t)\, p(z_{ik}=1 \mid \theta^t)}{p(x_i \mid \theta^t)}$$

$$= \frac{N(x_i \mid \mu_k^t, \Sigma_k^t)\ \alpha_k^t}{\sum_{k=1}^{K} \alpha_k N(x_i \mid \mu_k^t, \Sigma_k^t)} = \gamma_{ik}^t$$

$$\frac{\partial Q(\theta)}{\partial \alpha_k} = 0$$

$$\boxed{\hat{\alpha}_k = \frac{\sum_{i=1}^{n} \gamma_{ik}^t}{n}}$$

$$\boxed{\hat{\mu}_k = \frac{\sum_{i=1}^{n} \gamma_{ik}^t x_i}{\sum_{i=1}^{n} \gamma_{ik}^t}}$$

$$Q(\theta) = \underbrace{\sum_{i=1}^{n}} \sum_{k=1}^{K} \gamma_{ik}^t \left(\log \alpha_k + \underbrace{\log N(x_i \mid \mu_k, \Sigma_k)}\right) - \lambda\left(\underbrace{\sum_{k=1}^{K} \alpha_k} - 1\right)$$

$$\frac{\partial Q(\theta)}{\partial \alpha_j} = \sum_{i=1}^{n} \gamma_{ij}^t \frac{1}{\alpha_j} - \lambda = 0 \implies \boxed{\alpha_j = \frac{\sum_{i=1}^{n} \gamma_{ij}^t}{\lambda}} \quad j = 1, \ldots, K$$

$$1 = \sum_{j=1}^{K} \alpha_j = \frac{\sum_{j=1}^{K} \sum_{i=1}^{n} \gamma_{ij}^t}{\lambda}$$

$$\alpha_j = \frac{\sum_{i=1}^{n} \gamma_{ij}^t}{\sum_{j=1}^{K} \sum_{i=1}^{n} \gamma_{ij}^t} = \frac{\sum_{i=1}^{n} \gamma_{ij}^t}{n}$$