Suppose we observe $X_1, \dots, X_n \overset{\text{iid}}{\sim} \pi_1 N(\mu_1, \sigma_1^2) + \pi_2 N(\mu_2, \sigma_2^2)$ where $\pi_2 = 1 - \pi_1$ (1)

Goal: estimate $\theta := (\pi_1, \mu_1, \mu_2, \sigma_1^2, \sigma_2^2)$ via MLE, ie

$$\hat{\theta} = \underset{\theta}{\operatorname{argmax}}\ p_\theta(x)$$

This is difficult so let's be clever and introduce latent variables Z : (1) equiv. to

$$Z_i \overset{\text{iid}}{\sim} \text{Bern}(1-\pi_1) + 1$$
$$X_i \mid Z_i = 1 \overset{\text{iid}}{\sim} N(\mu_1, \sigma_1^2)$$
$$X_i \mid Z_i = 2 \overset{\text{iid}}{\sim} N(\mu_2, \sigma_2^2)$$

Observed data : X

Latent/unobserved : Z

Likelihood:

$$p_\theta(x_i, z_i) = p_\theta(x_i \mid z_i)\, p_\theta(z_i)$$
$$= \begin{cases} \frac{1}{\sqrt{2\pi_1 \sigma_1^2}} e^{-\frac{(x-\mu_1)^2}{2\sigma_1^2}} \pi_1, & \text{if } z_i = 1 \\ \frac{1}{\sqrt{2\pi_2 \sigma_2^2}} e^{-\frac{(x-\mu_2)^2}{2\sigma_2^2}} \pi_2, & \text{if } z_i = 2 \end{cases}$$

$$\Rightarrow \log p_\theta(x_i, z_i) = \begin{cases} -\frac{1}{2}\log(2\pi_1) - \log\sigma_1 - \frac{(x_i-\mu_1)^2}{2\sigma_1^2} + \log\pi_1, & \text{if } z_i = 1 \\ -\frac{1}{2}\log(2\pi_2) - \log\sigma_2 - \frac{(x_i-\mu_2)^2}{2\sigma_2^2} + \log\pi_2, & \text{if } z_i = 1 \end{cases}$$

1) E-Step: Compute $Q(\theta \mid \theta^{(t)}) = \mathbb{E}\{\log L(\theta; X, Z) \mid \theta^{(t)}, X\}$

$$Q(\theta \mid \theta^{(t)}) = \mathbb{E}\{\log p_\theta(x, z) \mid \theta^{(t)}, X\}$$
$$= \sum_{i=1}^{n} \mathbb{E}\{\log p_\theta(x_i, z_i) \mid \theta^{(t)}, X\} \quad \text{using independence of } X_i\text{'s}$$
$$= \sum_{i=1}^{n} \Big[\mathbb{E}\{\log p_\theta(x_i, z_i) \mid \theta^{(t)}, X, Z_i = 1\}\, \mathbb{P}(Z_i = 1 \mid \theta^{(t)}, X)$$
$$+ \mathbb{E}\{\log p_\theta(x_i, z_i) \mid \theta^{(t)}, X, Z_i = 2\}\, \mathbb{P}(Z_i = 2 \mid \theta^{(t)}, X) \Big] \quad \text{by total law of expectation}$$
$$= \sum_{i=1}^{n} \Big[\Big(\log\pi_1 - \tfrac{1}{2}\log(2\pi_1) - \log\sigma_1 - \frac{(x_i-\mu_1)^2}{2\sigma_1^2}\Big) \cdot Z_{i,1}^{(t)}$$
$$+ \Big(\log\pi_2 - \tfrac{1}{2}\log(2\pi_2) - \log\sigma_2 - \frac{(x_i-\mu_2)^2}{2\sigma_2^2}\Big) \cdot Z_{i,2}^{(t)} \Big]$$

where
$$Z_{i,j}^{(t)} = \mathbb{P}(Z_i = j \mid \theta^{(t)}, X)$$
$$= \frac{\pi_j^{(t)} f(x_i, \mu_j^{(t)}, \sigma_j^{(t)})}{\sum_{k=1}^{2} \pi_k^{(t)} f(x_i, \mu_k^{(t)}, \sigma_k^{(t)})} \quad \text{for } j = 1, 2,$$

$$f(x, \mu, \sigma) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \quad [\text{ie pdf of } N(\mu, \sigma^2)]$$

M step: maximize $Q(\theta | \theta^{(t)})$ wrt $\theta = (\pi_1, \mu_1, \sigma_1, \mu_2, \sigma_2)$

(i) Take derivative of $Q(\theta|\theta^{(t)})$ wrt π_1 and set to 0

$$\frac{\partial Q}{\partial \pi_1} = \frac{\partial}{\partial \pi_1}\left[\sum_{i=1}^{n}\left(\log \pi_1 \cdot z_{i1}^{(t)} + \log(1-\pi_1)\cdot z_{i2}^{(t)}\right)\right]$$

$$= \sum_{i=1}^{n} z_{i1}^{(t)} \cdot \frac{1}{\pi_1} - \sum_{i=1}^{n} z_{i2}^{(t)} \cdot \frac{1}{1-\pi_1}$$

$$\overset{\text{set}}{=} 0$$

$$\Rightarrow \pi_1^{(t+1)} = \frac{\sum z_{i1}^{(t)}}{\sum (z_{i1}^{(t)} + z_{i2}^{(t)})} = \frac{1}{n}\sum_{i=1}^{n} z_{i1}^{(t)}$$

(ii) Take derivative of $Q(\theta|\theta^{(t)})$ wrt μ_1 and set to 0

$$\frac{\partial Q}{\partial \mu_1} = \frac{\partial}{\partial \mu_1}\left[\sum_{i=1}^{n}\left(-\frac{(x_i-\mu_1)^2}{2\sigma_1^2}\right) z_{i1}^{(t)}\right]$$

$$= \sum_{i=1}^{n} z_{i1}^{(t)} \cdot \frac{(x_i-\mu_1)}{\sigma_1^2}$$

$$\overset{\text{set}}{=} 0$$

$$\Rightarrow \mu_1^{(t+1)} = \frac{\sum_{i=1}^{n} z_{i1}^{(t)} x_i}{\sum_{i=1}^{n} z_{i1}^{(t)}}$$

(iii) Take derivative of $Q(\theta|\theta^{(t)})$ wrt σ_1 and set to 0

$$\frac{\partial Q}{\partial \sigma_1} = \frac{\partial}{\partial \sigma_1}\left[\sum_{i=1}^{n}\left(-\log \sigma_1 - \frac{(x_i-\mu_1)^2}{2\sigma_1^2}\right) z_{i1}^{(t)}\right]$$

$$= \sum_{i=1}^{n}\left(-\frac{1}{\sigma_1} + \frac{(x_i-\mu_1)^2}{\sigma_1^3}\right) z_{i1}^{(t)}$$

$$\overset{\text{set}}{=} 0$$

$$\Rightarrow \sum_{i=1}^{n} z_{i1}^{(t)}\left((x_i-\mu_1)^2 - \sigma_1^2\right) = 0 \quad (\text{assuming } \sigma_1 \neq 0)$$

$$\Rightarrow (\sigma_1^2)^{(t+1)} = \frac{\sum_{i=1}^{n} z_{i1}^{(t)} (x_i - \mu_1^{(t+1)})^2}{\sum_{i=1}^{n} z_{i1}^{(t)}}$$

(iv) Can get analogous update formulas for μ_2, σ_2.

Thus, EM Algorithm for mixture of 2 Gaussians :

Initialize $\theta^{(0)} = (\pi_1^{(0)}, \mu_1^{(0)}, \sigma_1^{(0)}, \mu_2^{(0)}, \sigma_2^{(0)})$

Repeat until convergence (or until reach max # of iterations) :

Compute $z_{i1}^{(t)}, z_{i2}^{(t)}$ for each $i = 1, \dots, n$

Update : $\pi_1^{(t+1)} = \frac{1}{n} \sum_{i=1}^{n} z_{i1}^{(t)}$

$$\left.\begin{aligned} \mu_j^{(t+1)} &= \frac{\sum_{i=1}^{n} z_{ij}^{(t)} x_i}{\sum_{i=1}^{n} z_{ij}^{(t)}} \\ (\sigma_j^2)^{(t+1)} &= \frac{\sum_{i=1}^{n} z_{ij}^{(t)} (x_i - \mu_j^{(t+1)})^2}{\sum_{i=1}^{n} z_{ij}^{(t)}} \end{aligned}\right\} \text{for } j = 1, 2$$

End

Output : $\hat{\pi}_1, \hat{\mu}_1, \hat{\sigma}_1^2, \hat{\mu}_2, \hat{\sigma}_2^2, \hat{z}_{i1}, \hat{z}_{i2}$