Homework 4 Yan Liu

1 a) $$Q(\theta', \theta) = \sum_{i=1}^{N} E_{\theta}[\log P(\hat{x}_i, z_i \mid \theta')]$$

$$= \sum_{i=1}^{N} [r_{i1} \log P(\hat{x}_i, z_i = 1 \mid \theta') + r_{i2} \log P(\hat{x}_i, z_2 = 2 \mid \theta')]$$

Plug in the pdfs of the normals, we have

$$Q(\theta', \theta) = \sum_{i=1}^{N} [r_{i1} \log (p_1 \cdot \frac{1}{\sqrt{2\pi\sigma_1^2}} e^{-(\hat{x}_i - \mu_1)^2/2\sigma_1^2}) + r_{i2} \log (p_2 \cdot \frac{1}{\sqrt{2\pi\sigma_2^2}} e^{-(\hat{x}_i - \mu_2)^2/2\sigma_2^2})]$$

$$r_{i1} = P(z_i = 1 \mid \hat{x}^{(i)}, \theta)$$

$$= \frac{P(z_i = 1, \hat{x}^{(i)} \mid \theta)}{P(\hat{x}^{(i)})}$$

$$= \frac{p_1 \cdot \frac{1}{\sqrt{2\pi\sigma_1^2}} e^{-(\hat{x}_i - \mu_1)^2/2\sigma_1^2}}{p_1 \cdot \frac{1}{\sqrt{2\pi\sigma_1^2}} e^{-(\hat{x}_i - \mu_1)^2/2\sigma_1^2} + p_2 \cdot \frac{1}{\sqrt{2\pi\sigma_2^2}} e^{-(\hat{x}_i - \mu_2)^2/2\sigma_2^2}}$$

Similarly, we have

$$r_{i2} = P(z_i = 2 \mid \hat{x}^{(i)}, \theta)$$

$$= \frac{P(z_i = 2, \hat{x}^{(i)} \mid \theta)}{P(\hat{x}^{(i)})}$$

$$= \frac{p_2 \cdot \frac{1}{\sqrt{2\pi\sigma_2^2}} e^{-(\hat{x}_i - \mu_2)^2/2\sigma_2^2}}{p_1 \cdot \frac{1}{\sqrt{2\pi\sigma_1^2}} e^{-(\hat{x}_i - \mu_1)^2/2\sigma_1^2} + p_2 \cdot \frac{1}{\sqrt{2\pi\sigma_2^2}} e^{-(\hat{x}_i - \mu_2)^2/2\sigma_2^2}}$$

b) Follow the expression derived in a), we have

$$Q(\theta';\theta) = \sum_{i=1}^{N} \left(r_{i1} \log \left(p_1 \cdot \frac{1}{\sqrt{2\pi\sigma_1^2}} e^{-(\hat{x}_i-\mu_1)^2/2\sigma_1^2} \right) + r_{i2} \log \left(p_2 \cdot \frac{1}{\sqrt{2\pi\sigma_2^2}} e^{-(\hat{x}_i-\mu_2)^2/2\sigma_2^2} \right) \right)$$

Take the log, we have

$$Q(\theta';\theta) = \sum_{i=1}^{N} \Big(r_{i1} \log p_1 + r_{i1} \log \left(\frac{1}{\sqrt{2\pi\sigma_1^2}}\right) - r_{i1}(\hat{x}_i-\mu_1)^2/2\sigma_1^2 + r_{i2} \log p_2 + r_{i2} \log \left(\frac{1}{\sqrt{2\pi\sigma_2^2}}\right) - r_{i2}(\hat{x}_i-\mu_2)^2/2\sigma_2^2 \Big)$$

Next, we solve $\nabla_{\theta'} Q(\theta';\theta) = 0$

where $\nabla_{\theta'} Q(\theta',\theta)$ is the gradient. $\theta' = (\mu_1, \mu_2, \sigma_1, \sigma_2, p_1, p_2)$.

① Solve for μ_1, μ_2

$$\frac{\partial Q(\theta';\theta)}{\partial \mu_1} = 0 \Rightarrow \sum_{i=1}^{N} r_{i1} \cdot 2(\hat{x}_i - \mu_1)/2\sigma_1^2 = 0$$

$$\Rightarrow \sum_{i=1}^{N} r_{i1} \cdot (\hat{x}_i - \mu_1) = 0$$

$$\Rightarrow \mu_1 = \frac{\sum_{i=1}^{N} r_{i1} \hat{x}_i}{\sum_{i=1}^{N} r_{i1}}$$

Similarly, $\frac{\partial Q(\theta';\theta)}{\partial \mu_2} = 0 \Rightarrow \mu_2 = \frac{\sum_{i=1}^{N} r_{i2} \hat{x}_i}{\sum_{i=1}^{N} r_{i2}}$

② Solve for σ_1, σ_2

Let $v_1 = \sigma_1^2$, we have

$$\frac{\partial Q(\theta', \theta)}{\partial v_1} = 0$$

$$\Rightarrow \sum_{i=1}^{N} \left(r_{i1} \log\left(\frac{1}{v_1}\right)' - r_{i1} \cdot \frac{(\hat{x}_i - \mu_1)^2}{2} \left(\frac{1}{v_1}\right)' \right) = 0$$

$$\Rightarrow \sum_{i=1}^{N} \left(r_{i1} \cdot \left(-\frac{1}{v_1} + \frac{(\hat{x}_i - \mu_1)^2}{2} \cdot \frac{1}{v_1^2} \right) \right) = 0$$

Solve for v_1, we have $v_1 = \dfrac{\sum_{i=1}^{N} r_{i1} (\hat{x}_i - \hat{\mu}_1)^2}{\sum_{i=1}^{N} r_{i1}} = \sigma_1^2$

Similarly for σ_2, we have $v_2 = \dfrac{\sum_{i=1}^{N} r_{i2} (\hat{x}_i - \mu_2)^2}{\sum_{i=1}^{N} r_{i2}} = \sigma_2^2$

③ Solve for p_1, p_2 $(p_1 + p_2 = 1)$

$$\frac{\partial Q(\theta', \theta)}{\partial p_1} = \sum_{i=1}^{N} \left(r_{i1} \frac{1}{p_1} - r_{i2} \frac{1}{1-p_1} \right) = 0$$

$$\Rightarrow \frac{\sum_{i=1}^{N} r_{i1}}{p_1} = \frac{\sum_{i=1}^{N} (1 - r_{i1})}{1 - p_1}$$

$$\Rightarrow \sum_{i=1}^{N} r_{i1} (1 - p_1) = \left(N - \sum_{i=1}^{N} r_{i1} \right) p_1$$

$$\Rightarrow N p_1 = \sum_{i=1}^{N} r_{i1}$$

$$\Rightarrow p_1 = \frac{\sum_{i=1}^{N} r_{i1}}{N}$$

Since $p_1 + p_2 = 1$, $p_2 = 1 - p_1 = \dfrac{N - \sum_{i=1}^{N} r_{i1}}{N} = \dfrac{\sum_{i=1}^{N} r_{i2}}{N}$

c) d) see the other pdf file.

2. a) Write down the EM iteration for the mixture model.

Solution:

Let $z_i = \begin{cases} 1, & \text{probability } r_{i1} \\ 2, & \text{probability } r_{i2} \end{cases}$

where $z_i=1$ represents that x_i is from mixture 1.

$z_i=2$ represents that x_i is from mixture 2.

r_{i1} indicates that r_{i1} portion of x_i is from mixture 1

r_{i2} indicates that r_{i2} portion of x_i is from mixture 2

$r_{i1} + r_{i2} = 1$ In other words,

Let $R_{i1} = p_1 \prod_{j=1}^{10} (\mu_j^{(1)})^{\hat{x}_j^{(i)}} (1-\mu_j^{(1)})^{\hat{x}_j^{(i)}}$, $R_{i2} = p_2 \prod_{i=1}^{10} (\mu_j^{(2)})^{\hat{x}_j^{(i)}} (1-\mu_j^{(2)})^{\hat{x}_j^{(i)}}$

Then $r_{i1} = P(z_i = 1 \mid \hat{x}_i, \theta) = R_{i1} / (R_{i1} + R_{i2})$

$r_{i2} = P(z_i = 2 \mid \hat{x}_i, \theta) = R_{i2} / (R_{i1} + R_{i2})$

Step 1: Pick $\theta^{(t)} = (\mu^{(1)}, \mu^{(2)}, p_1, p_2)$

Step 2: Expectation Step:

$$Q(\theta^{(t)}, \theta) = \sum_{i=1}^{N} E_\theta [\log P(\hat{x}_i, z_i) \mid \theta^{(t)}]$$

Maximization Step:

Update $\theta^{(t)}$ by the formul below:

$$\theta^{(t+1)} := \arg\max_\theta Q(\theta^{(t)}, \theta)$$

Step 3: Iterate over step 2 until

$$E_{\theta^{(t+1)}} [\log P(\hat{x}_i, z_i) \mid \theta^{(t+1)}] \leqslant E_{\theta^{(t)}} [\log P(\hat{x}_i, z_i) \mid \theta^{(t)}] + \varepsilon$$

where ε is some preset threshold.

Specifically in $Q(\theta^{(t)}, \theta) = \sum_{i=1}^{N} E_\theta [\log P(\hat{x}_i, z_i) \mid \theta^{(t)}]$

since each mixture follows a binomial distribution, we have:

$$Q(\theta^{(t)}, \theta) = \sum_{i=1}^{N} \Big(r_{i1} \log P(\hat{x}_i, z_i = 1 \mid \theta^{(t)}) + r_{i2} \log P(\hat{x}_i, z_i = 2 \mid \theta^{(t)}) \Big)$$

$$= \sum_{i=1}^{N} \Big(r_{i1} \log \big(p_1 \cdot \prod_{j=1}^{10} (\mu_j^{(1)})^{\hat{x}_j^{(i)}} \cdot (1-\mu_j^{(1)})^{1-\hat{x}_j^{(i)}} \big) + r_{i2} \log \big(p_2 \cdot \prod_{j=1}^{10} (\mu_j^{(2)})^{\hat{x}_j^{(i)}} \cdot (1-\mu_j^{(2)})^{1-\hat{x}_j^{(i)}} \big) \Big)$$

$$= \sum_{i=1}^{N} \Big(r_{i1} \log p_1 + r_{i2} \log p_2 + \sum_{j=1}^{10} \big(r_{i1} (\hat{x}_j^{(i)} (\mu_j^{(1)}) + (1-\hat{x}_j^{(i)})(1-\mu_j^{(1)})) + r_{i2} (\hat{x}_j^{(i)} (\mu_j^{(2)}) + (1-\hat{x}_j^{(i)})(1-\mu_j^{(2)})) \big) \Big)$$

Let $\nabla Q(\theta^{(t)}, \theta) = 0$ where $\theta^{(t)} = (\mu^{(1)}, \mu^{(2)}, p_1, p_2)$

$$\frac{\partial Q}{\partial \mu_k^{(1)}} = \sum_{i=1}^{N} r_{i1} \frac{\hat{X}_k^{(i)}}{\mu_k^{(1)}} - \frac{1-\hat{X}_k^{(i)}}{1-\mu_k^{(1)}} = 0$$

Solve for $\mu_k^{(1)}$, we have $\mu_k^{(1)} = \dfrac{\sum_{i=1}^{N} r_{i1} X_k^{(i)}}{\sum_{i=1}^{N} r_{i1}}$

Similarly, we have $\mu_k^{(2)} = \dfrac{\sum_{i=1}^{N} r_{i2} X_k^{(i)}}{\sum_{i=1}^{N} r_{i2}}$

As for p_1, p_2, we have $\dfrac{\partial Q}{\partial p_1} = \dfrac{\partial}{\partial p_1} \Big(\sum_{i=1}^{N} (r_{i1} \log p_1 + r_{i2} \log p_2) \Big) = 0$

Similarly as in 1b) ③, solve for p_1, p_2 we have

$$p_1 = \frac{\sum_{i=1}^{N} r_{i1}}{N}, \quad p_2 = \frac{\sum_{i=1}^{N} r_{i2}}{N}.$$