University of Waterloo
CS 480
r2knowle: 2023-11-13

Exercise # 3

Q3a) For this question we are going to derive both the expectation step and the maximization step independently:

Expectation Step: To begin we are given that S_k is diagonal, which gives us the following properties:

$$|S_k| = \sigma_1^2 \times \sigma_2^2 \times ... \times \sigma_n^2 = \prod_i^n \sigma_i^2$$

$$S_k^{-1} = \begin{bmatrix} \frac{1}{\sigma_1^2} & ... & 0 \\ \vdots & & \vdots \\ 0 & & \frac{1}{\sigma_n^2} \end{bmatrix}$$

Continuing from the slides the we get that:

$$\begin{aligned} r_{ik} &= q_i(Z_i = k) \\ &= p_\theta(Z_i = k|x_i) \\ &= \frac{p_\theta(Z_i = j, x_i)}{p_\theta(x_i)} \\ &= \frac{\pi_k N(\mu_k, S_k, x_i)}{\sum_{l=1}^k \pi_k N(\mu_l, S_l, x_i))} \end{aligned}$$

Note that the denominator is calculated outside of the red step, and the red step is only the numerator. Therefore any optimization for a diagonal matrix will have to occur within $p_\theta(Z_i = j, x_i)$, thus we get the following expansion:

$$\begin{aligned} \pi_k N(\mu_k, S_k, X_i) &= \frac{1}{\sqrt{|2\pi S_k|}} \exp\left(-\frac{1}{2}(x_i - \mu_k)^T S_k^{-1}(x_i - \mu_k)\right) \\ &= \frac{1}{\sqrt{2\pi^k |S_k|}} \exp\left(-\frac{1}{2}\left(\frac{(x_{i1} - \mu_{k1})^2}{\sigma_1^2} + \frac{(x_{i2} - \mu_{k2})^2}{\sigma_2^2} + ... + \frac{(x_{in} - \mu_{kn})^2}{\sigma_n^2}\right)\right) \\ &= \frac{\exp\left(-\frac{1}{2}\left(\frac{(x_{i1}-\mu_{k1})^2}{\sigma_1^2} + \frac{(x_{i2}-\mu_{k2})^2}{\sigma_2^2} + ... + \frac{(x_{in}-\mu_{kn})^2}{\sigma_n^2}\right)\right)}{\sqrt{2\pi^k \sigma_1^2 \times \sigma_2^2 \times ... \times \sigma_n^2}} \\ &= \frac{\exp\left(-\frac{1}{2}\frac{(x_{i1}-\mu_{k1})^2}{\sigma_1^2}\right)}{\sqrt{2\pi\sigma_1^2}} \times \frac{\exp\left(-\frac{1}{2}\frac{(x_{i2}-\mu_{k2})^2}{\sigma_2^2}\right)}{\sqrt{2\pi\sigma_2^2}} \times ... \times \frac{\exp\left(-\frac{1}{2}\frac{(x_{in}-\mu_{kn})^2}{\sigma_n^2}\right)}{\sqrt{2\pi\sigma_n^2}} \end{aligned}$$

Maximization Step As given in the slides are our is to maximize the following:

$$p_\theta(x) = \sum_i^k \pi_k N(\mu_k, S_k, x_i)$$

Which from the slides can we know can be rewritten as:

$$\begin{aligned}
&= \text{argmax}_\theta \sum_{i=1}^n \sum_{j=1}^k q_i(Z_i = j) \log p_\theta(x_i, Z_i = j) \\
&= \text{argmax}_\theta \sum_{i=1}^n \sum_{j=1}^k q_i(Z_i = j) \log \left[\frac{\pi_k}{\sqrt{|2\pi S_k|}} \exp\left(-\frac{1}{2}(x_i - \mu_k)^T S_k^{-1} (x_i - \mu_k) \right) \right] \\
&= \text{argmax}_\theta \sum_{i=1}^n \sum_{j=1}^k q_i(Z_i = j) \left[\log(\pi_k) - \frac{k}{2} \log(2\pi) - \frac{1}{2} \log(|S_k|) + \left(-\frac{1}{2}(x_i - \mu_k)^T S_k^{-1} (x_i - \mu_k) \right) \right] \\
&= \text{argmax}_\theta \sum_{i=1}^n \sum_{j=1}^k r_{ik} \left[\log(\pi_k) - \frac{k}{2} \log(2\pi) - \frac{1}{2} \log(|S_k|) + \left(-\frac{1}{2}(x_i - \mu_k)^T S_k^{-1} (x_i - \mu_k) \right) \right]
\end{aligned}$$

We can then take the derivative w.r.t to S_k and set to zero, to find where this is concave function is maximized:

$$\begin{aligned}
\sum_{i=1}^n \sum_{j=1}^k r_{ik} \left[\log(\pi_k)' - \frac{k}{2} \log(2\pi)' - \frac{1}{2} \log(|S_k|)' + \left(-\frac{1}{2}(x_i - \mu_k)^T S_k^{-1} (x_i - \mu_k) \right)' \right] &= 0 \\
\sum_{i=1}^n \sum_{j=1}^k r_{ik} \left[0 - 0 - \frac{1}{2} \log\left(\prod_i^n \sigma_i^2 \right)' + \left(-\frac{1}{2}(x_i - \mu_k)^T (x_i - \mu_k) \right) \right] &= 0 \\
\sum_{i=1}^n \sum_{j=1}^k r_{ik} \left[0 - 0 - \frac{1}{\sigma_{ik}^2} + \left(-\frac{1}{2}(x_i - \mu_k)^T (x_i - \mu_k) \right) \right] &= 0
\end{aligned}$$

We can rearrange to get:

$$\begin{aligned}
\sum_{i=1}^n \frac{r_{ik}}{\sigma_{ik}^2} &= \sum_{i=1}^n \frac{1}{2}(x_i - \mu_k)^T (x_i - \mu_k) \\
S_k &= \sum_{i=1}^n \frac{\frac{1}{2}(x_i - \mu_k)^T (x_i - \mu_k)}{r_{ik}}
\end{aligned}$$

On the next page is my implementation of the algorithm.