EM Extra excercises

(2)

a.) $r_{\mu k} = p(k|x^\mu, \theta_t)$

$$= \frac{\pi_k \mathcal{N}(x^\mu | a_k, \sigma_k^2)}{\sum_{k'} \pi_{k'} \mathcal{N}(x^\mu | a_{k'}, \sigma_{k'}^2)}$$

b.) $$Q(\theta, q^*) = \sum_{\mu, k} r_{\mu k} \left(\log(\pi_k) + \log \mathcal{N}(x^\mu | a_k, \sigma_k^2)\right)$$

$$= \sum_{\mu, k} r_{\mu k} \left(\log(\pi_k) - \log(\sqrt{2\pi}) - \log(\sigma_k) - \frac{1}{2\sigma_k^2}(x^\mu - a_k)^2\right)$$

c.) For the next steps we will have to use the partial derivatives of the above expression.

*) $$\frac{\partial Q}{\partial a_k} = \sum_{\mu} r_{\mu k} \frac{1}{\sigma_k^2}(x^\mu - a_k) \overset{!}{=} 0 \quad (1)$$

*) $$\frac{\partial Q}{\partial \sigma_k} = \sum_{\mu} r_{\mu k}\left(-\frac{1}{\sigma_k} + \frac{\cancel{2}(x^\mu - a_k)^2}{\cancel{2}\sigma_k^3}\right) \overset{!}{=} 0 \quad (2)$$

For the last one we note that we have

*) a condition that needs to be specified, namely: $\sum_k \pi_k = 1$, thus we add this, together with a Lagrange multiplier to the expression that we are maximizing. Thus we get:

$$\frac{\partial Q}{\partial \pi_k} = \sum_{\mu} r_{\mu k} \frac{1}{\pi_k} \Longrightarrow \frac{\sum_{\mu} r_{\mu k}}{\pi_k} + \lambda = 0 \quad (3)$$

Now we are going to solve these equations:

① : given that $\sigma_k^2 \neq 0$ we find:

$$\boxed{a_k = \frac{\sum_\mu r_{\mu k}\, x^\mu}{\sum_\mu r_{\mu k}}}$$

② $\sum_\mu r_{\mu k} \left(\frac{-1}{\sigma_k} + \frac{(x^\mu - a_k)^2}{\sigma_k^3} \right) = 0$

$$\to \sum_\mu r_{\mu k} \frac{(x^\mu - a_k)^2}{\sigma_k^2} = \sum_\mu r_{\mu k}$$

$$\to \quad \boxed{\sigma_k^2 = \frac{\sum_\mu r_{\mu k} (x^\mu - a_k)^2}{\sum_\mu r_{\mu k}}}$$

③ we find: $\pi_k = -\lambda \sum_\mu r_{\mu k}$

But we know:

$\sum_k \pi_k \stackrel{!}{=} 1 \quad \to \quad \sum_k \left(-\lambda \sum_\mu r_{\mu k}\right) = -\lambda \sum_{\mu,k} r_{\mu k}$

Now we know:

$\sum_k r_{\mu k} = \sum_k p(k|x^\mu, \theta^t)$, which since it is a probability must equal 1

$$\to \sum_{\mu,k} r_{\mu k} = \sum_\mu 1 = N$$

$\Rightarrow \quad -\lambda N = 1 \quad \to \quad \lambda = \frac{-1}{N}$

And thus we have: $\boxed{\pi_k = \frac{\sum_\mu r_{\mu k}}{N}}$.

If we check these expressions with those found on slide 204, we see that they are (very) equivalent. Where we do note that for σ_k^2 the expr. between $(..)^2$ is the expression filled into the expectation value for the def. of σ_k, as is the same for $(\Sigma_k)_{ij}$ on slide 204!