Machine Learning Ex. 3

Aufg. 2) Free Energy is defined as:

$$F = \sum_n \sum_c \left[q^{(n)}(c) \log(p(\vec{x}^{(n)}, c|\theta)) - q^{(n)}(c) \log(q^{(n)}(c)) \right]$$

It is $p(\vec{x}|c,\theta) = \frac{1}{\sqrt{\det(2\pi\Sigma_c)}} \exp(-\frac{1}{2}(\vec{x}-\vec{\mu}_c)^T \Sigma_c^{-1} (\vec{x}-\vec{\mu}_c))$

and $p(c|\theta) = \pi_c$ fulfilling $\sum_c \pi_c = 1$

With the product rule we know that $p(\vec{x}, c|\theta) = p(\vec{x}|c,\theta)\, p(c|\theta)$.

a) Maximize with respect to $\vec{\mu}$.

$$\frac{\partial F}{\partial \vec{\mu}_c} = 0 \iff 0 = \frac{\partial}{\partial \mu_c} \left[\sum_n \sum_c \left[q^{(n)}(c) \log\left(\frac{\pi_c}{\sqrt{\det(2\pi\Sigma_c)}} \exp(-\frac{1}{2}(\vec{x}-\vec{\mu}_c)^T \Sigma_c^{-1} (\vec{x}-\vec{\mu}_c)) \right) - q^{(n)}(c) \log(q^{(n)}(c)) \right] \right]$$

Using $\frac{\partial}{\partial \vec{\mu}} q^{(n)}(c) = 0$ and since Σ_c is the diagonal matrix of the variances it is $\frac{\partial}{\partial \vec{\mu}} \Sigma_c = 0$ and $\frac{\partial}{\partial \vec{\mu}} \Sigma_c^{-1} = 0$ as a consequence we have not only $\frac{\partial}{\partial \vec{\mu}} \log(\det(\Sigma_c)) = tr(\Sigma_c^{-1} \frac{\partial \Sigma_c}{\partial \vec{\mu}}) = 0$

but can also use $\log(a \cdot b) = \log(a) + \log(b)$ together with the linearity of the derivative and get:

$$\frac{\partial F}{\partial \vec{\mu}_c} = \sum_n \sum_{c'} \left(\left[-\frac{1}{2} \cdot \frac{1}{\sqrt{2\pi}} \frac{1}{\det(\Sigma_c)^{3/2}}\, tr(\underbrace{\Sigma_c^{-1} \frac{\partial \Sigma_c}{\partial \vec{\mu}}}_{=0}) + (-\tfrac{1}{2})(-2)\Sigma_c^{-1}(\vec{x}^{(n)} - \vec{\mu}_c) + (-\tfrac{1}{2})(\vec{x}-\vec{\mu})^T \overbrace{\frac{\partial \Sigma_c^{-1}}{\partial \vec{\mu}}}^{=0} (\vec{x}-\vec{\mu}) \right] q^{(n)}(c) \cdot \delta_{cc'} \right) = 0$$

$$\Rightarrow 0 = +\sum_n q^{(n)}(c)(\vec{x}^{(n)} - \vec{\mu}_c) \Rightarrow \sum_n q^{(n)}(c)\, \vec{x}^{(n)} = \sum_n q^{(n)}(c)\, \vec{\mu}_c$$

$$\Rightarrow \vec{\mu}_c = \frac{\sum_n q^{(n)}(c)\, \vec{x}^{(n)}}{\sum_n q^{(n)}(c)}$$

b) Maximize π_c with the constraint $\sum_c \pi_c = 1$

$$\frac{\partial F}{\partial \pi_c} + \lambda \frac{\partial}{\partial \pi_c}\left(\sum_c \pi_c - 1\right) = 0 \quad \forall c$$

$$\Rightarrow \frac{\partial}{\partial \pi_c}\left(\sum_n \sum_c q^{(n)} \log(\pi_c)\right) + \lambda \left(\underbrace{\sum_c \delta_{cc'}}_{=1}\right) = 0$$

$$\Rightarrow \sum_n q^{(n)} \frac{1}{\pi_c} + \lambda = 0 \quad (*)$$

From this follows $\lambda \pi_c = -\sum_n q^{(n)} \quad \forall c$

and summing over c gets

$$\lambda \underbrace{\sum_c \pi_c}_{=1} = -\sum_n \underbrace{\sum_c q^{(n)}}_{=1} = -\sum_n 1 = -N$$

So with $\lambda = -N$ $(*)$ results in:

$$\pi_c = \frac{1}{N} \sum_n q^{(n)}$$