1 Exercise 1

We have N samples (x_i, y_i) $\boldsymbol{\pi} = (\pi_1, ..., \pi_M), \boldsymbol{\theta} = (\theta_{1,1}, ...\theta_{M,K})$
$a_m = \mid \{i \mid z_i = m, \forall i \in [1, N]\} \mid$, $b_{m,k} = \mid \{i \mid z_i = m$ and $x_i = k, \forall i \in [1, N]\} \mid$

$$\begin{aligned} l(\boldsymbol{\pi}, \boldsymbol{\theta}) &= \sum_{i=0}^{n} log(p(x_i, z_i)) \\ &= \sum_{i=0}^{n} log(p(x_i \mid z_i)p(z_i)) \\ &= \sum_{i=0}^{n} (log(\theta_{z_i,x_i}) + log(\pi_{x_i})) \end{aligned}$$

$l(\boldsymbol{\pi}, \boldsymbol{\theta})$ is concave. We want to minimize $-l(\boldsymbol{\pi}, \boldsymbol{\theta})$ subjected to $\sum_{k=1}^{K} \pi_k = 1$ and $\sum_{k=1}^{K} \sum_{m=1}^{M} \theta_{m,k} = 1$ Lets introduce the langrangian.

$$L(\boldsymbol{\pi}, \boldsymbol{\theta}, \lambda_1, \lambda_2) = -(\sum_{i=0}^{n} (log(\theta_{z_i,x_i}) + log(\pi_{x_i}))) + \lambda_1(\sum_{k=1}^{K} \pi_k - 1) + \lambda_2(\sum_{k=1}^{K} \sum_{m=1}^{M} \theta_{m,k} - 1)$$

The Slaters constraint qualification are trivialy verified and therefore the problem has strong duality property. Therefore we have

$$\min_{\boldsymbol{\pi}, \boldsymbol{\theta}} -l(\boldsymbol{\pi}, \boldsymbol{\theta}) = \max_{\lambda_1, \lambda_2} L(\boldsymbol{\pi}, \boldsymbol{\theta}, \lambda_1, \lambda_2)$$

Moreover the lagrangian is convex with respect to $\boldsymbol{\pi}$ and $\boldsymbol{\theta}$

$$\frac{\partial L}{\partial \pi_m} = 0 \Rightarrow \pi_m = \frac{a_m}{\lambda_1}$$

$$\frac{\partial L}{\partial \theta_{m,k}} = 0 \Rightarrow \theta_{m,k} = \frac{b_{m,k}}{\lambda_2}$$

Using the constrains we can calculate λ_1, λ_2.

$$\begin{aligned} \pi_m &= \frac{a_m}{N} \\ \theta_{m,k} &= \frac{b_{m,k}}{N} \end{aligned}$$

2 Exercicse 2

2.1 Generative model LDA

We have N samples and $n =\mid \{i, y_i = 0, \forall i \in [1, N]\} \mid$

$$\begin{aligned}
l(\omega, \Sigma, \mu_1, \mu_2) &= \sum_{i=1}^{N} log(p(x_i, y_i)) \\
&= \sum_{i=1}^{N} log(p(x_i \mid y_i)p(y_i)) \\
&= \sum_{\substack{i=1,\\ y_i=0}}^{N} log(p(x_i \mid y_i = 0)) + nlog(\omega) + \sum_{\substack{i=1,\\ y_i=0}}^{N} log(p(x_i \mid y_i = 1)) + (N-n)log(1-\omega) \\
&= -\frac{Nd}{2}log(2\pi) + \frac{N}{2}log(\mid \Sigma^{-1} \mid) - \sum_{\substack{i=1,\\ y_i=0}}^{N} \frac{1}{2}(x_i-\mu_1)^T\Sigma^{-1}(x_i-\mu_1) \\
&- \sum_{\substack{i=1,\\ y_i=1}}^{N} \frac{1}{2}(x_i-\mu_2)^T\Sigma^{-1}(x_i-\mu_2) + nlog(\omega) + (N-n)log(1-\omega)
\end{aligned}$$

This log likelyhood is not concave in $(\omega, \Sigma, \mu_1, \mu_2)$. It is concave in (ω, μ_1, μ_2) with Σ fixed.

$$\nabla_\omega l = \frac{n}{\omega} - \frac{N-n}{1-\omega}$$

$\nabla_\omega l = 0$ gives us :

$$\omega = \frac{n}{N}$$

Calculating the gradient in μ_1, μ_2 and equalating it to 0 gives us.

$$\mu_1 = \frac{1}{n} \sum_{\substack{i=1,\\ y_i=0}}^{N} x_i$$

$$\mu_2 = \frac{1}{N-n} \sum_{\substack{i=1,\\ y_i=1}}^{N} x_i$$

Let us now differentiate l w.r.t. Σ^{-1}.
Let $A = \Sigma^{-1}$, $\Sigma_1 = \frac{1}{n}\sum_{\substack{i=1,\\ y_i=0}}^{N}(x_i-\mu_1)^T(x_i-\mu_1)$,
$\Sigma_2 = \frac{1}{N-n}\sum_{\substack{i=1,\\ y_i=1}}^{N}(x_i-\mu_2)^T(x_i-\mu_2)$

We have :

$$\begin{aligned}
l(\omega, \Sigma, \mu_1, \mu_2) =& -\frac{Nd}{2} log(2\pi) + \frac{N}{2} log(\mid \Sigma^{-1} \mid) + \frac{1}{2}\text{Trace}(A(n\Sigma_1 + (N-n)\Sigma_2) \\
& + nlog(\omega) + (N-n)log(1-\omega) \\
\nabla_A l =& \frac{N}{2} A^{-1} + \frac{1}{2}(n\Sigma_1 + (N-n)\Sigma_2)
\end{aligned}$$