TP - 2

Exercice 1

1 - On utilise la technique de la fonction de répartition inverse

$X = \{x_i\}_{i \leq n}$, on suppose X ordonné

$$F_X : X \longrightarrow [0,1]$$
$$x_i \longmapsto \mathbb{P}(X \leq x_i) = \sum_{k=1}^{i} p_k$$

On introduit la fonction inverse généralisée :

$$F^{\leftarrow}(u) = \inf\{x \in E, F(x) \geq u\}$$

Dans notre cas : $F(u) = x_k$ si $u \in \left[\sum_{i=1}^{k-1} p_i , \sum_{i=1}^{k} p_i\right[$

Exercice 2

1] $\Theta = \{\alpha_j, \mu_j, \Sigma_j\}$. $(\alpha_j)_j \in$ Simplexe : $\sum \alpha_j = 1, \alpha_j \geq 0$

. $(\mu_j) \in (\mathbb{R}^d)_j$

. $\Sigma_j \in S_n^{++}(d)$: ouvert

$$p_\theta(x) = \int_z p_\theta(x, z) dz$$

$$= \int_z p_\theta(z) p_\theta(x|z) dz = \sum_{j=1}^{p} p_\theta(z_j) \, p_\theta(x|z=j)$$

$$p_\theta(x|z=j) = \frac{1}{(2\pi)^{d/2}\sqrt{\det \Sigma}} e^{-\frac{1}{2}(x-\mu_j)^T \Sigma^{-1}(x-\mu_j)}$$

. $\log(p_\theta(x_i)_i) = \sum_{i=1}^{N} \log\left(\sum_{j=1}^{p} \alpha_j p_\theta(x_i|z=j)\right)$

2- EM algorithm.

On optimise la loi jointe $\mathbb{E}_{z \sim q}[\log p_\theta(x, z)]$,

avec $q = p_{\theta t}(\cdot|x)$

$$\mathbb{E}_{z \sim p_{\theta t}(\cdot|x)}[\log p_\theta(x, z)] = \int \log p_\theta(x, z) p_{\theta t}(z|x) dz$$

$$= \sum_{i=1}^{m} \sum_{j=1}^{p} \mathbb{P}_{\theta t}(z_i = j | x_i) \log p_\theta(x_i, z_i)$$

$$= \sum_{i=1}^{m} \sum_{j=1}^{p} \tau_{ij} [\underbrace{\log p_\theta(x_i|z_i = j)}_{\text{cste} - \frac{1}{2}\log\det(\Sigma_j) - \frac{1}{2}(x-\mu_j)^T\Sigma_j^{-1}(x-\mu_j)} + \log \alpha_j]$$

$$\boxed{\tau_{ij} = \frac{\alpha_j^t \, p_{\theta t}(x_i | z_i = j)}{\sum_{h=1}^{p} \alpha_h^t p_{\theta t}(x_i | z_i = h)}}$$

Le problème d'optimisation s'écrit donc :

$$\boxed{\begin{array}{l} \displaystyle\max_{(\alpha_j),(\mu_j),(\Sigma_j)} \sum_{i=1}^{m}\sum_{j=1}^{p} \tau_{ij}\left[\text{cste} + \frac{1}{2}\log(\det(\Sigma^{-1})) - \frac{1}{2}(x-\mu_j)^T\Sigma_j^{-1}(x-\mu_j) + \log\alpha_j\right] \overset{"}{=} F(\theta) \\ \text{s.t.} \quad \displaystyle\sum_{j=1}^{p}\alpha_j = 1 \end{array}}$$

Lagrangien :

$$\mathcal{L}(\theta,\lambda) = F(\theta) + \lambda\left(1 - \sum_{j=1}^{m}\alpha_j\right)$$

KKT conditions (problème convexe et faisable) → concave donc convexe

$$\frac{\partial\mathcal{L}(\theta,\lambda)}{\partial\alpha_k} = 0 \iff \sum_{i=1}^{m}\sum_{j=1}^{p}\tau_{ij}\frac{\partial\log\alpha_j}{\partial\alpha_k} - \lambda = 0$$

$$\iff \sum_{i=1}^{m}\frac{\tau_{ij}}{\alpha_k} - \lambda = 0$$

$$\iff \alpha_k = \frac{1}{\lambda}\sum_{i=1}^{m}\tau_{ik}$$

$$\sum_{k=1}^{p}\alpha_k = 1 \Rightarrow \lambda = \sum_{i=1}^{m}\sum_{j=1}^{m}\tau_{ik} \ (= n)$$

$$\Rightarrow \boxed{\alpha_k = \frac{\sum_{i=1}^{m}\tau_{ik}}{\sum_{i=1}^{m}\sum_{j=1}^{p}\tau_{ik}}}$$

$$\frac{\partial\mathcal{L}(\theta,\lambda)}{\partial\mu_k} = \sum_{i=1}^{m}\sum_{j=1}^{p} -\tau_{ij}\frac{1}{2}\frac{\partial}{\partial\mu_k}(x_i-\mu_j)^T A_j (x_i-\mu_j) \quad \Sigma$$

$$\frac{\partial}{\partial\mu_k}(x_i-\mu_j)^T A_j (x_i-\mu_j) = -\delta_{jk}\, 2A_j(x_i-\mu_j)$$

$$\Rightarrow \frac{\partial \mathcal{L}(\Theta, \alpha)}{\partial \mu_k} = \sum_{i=1}^{m} \frac{\tau_{ik}}{2} 2 A_k (x_i - \mu_k) = 0$$

$$\Leftrightarrow \sum_{i=1}^{m} \tau_{ik} x_i = \sum_{i=1}^{m} \tau_{ik} \mu_k$$

$$\Leftrightarrow \boxed{\mu_k = \frac{\sum_{i=1}^{m} \tau_{ik} x_i}{\sum_{i=1}^{m} \tau_{ik}}}$$

on pose $A = \Sigma^{-1}$

$$\frac{\partial \mathcal{L}(\Theta, \alpha)}{\partial A_k} = \sum_{i=1}^{m} \sum_{j=1}^{p} \tau_{ij} \frac{1}{2} \frac{\partial}{\partial A_k} \log \det(A_j) - \frac{\partial}{\partial A_k} \frac{1}{2} (x_i - \mu_j)^T A_j (x_i - \mu_j)$$

- $(x - \mu_j)^T A_j (x - \mu_j) = \mathrm{Tr}[(x - \mu_j)(x - \mu_j)^T A_j]$

$$\Rightarrow \frac{\partial}{\partial A_k} (x - \mu_j)^T A_j (x - \mu_j) = \delta_{jk} (x - \mu_k)(x - \mu_k)^T$$

$$\frac{\partial}{\partial A_k} (\log \det(A_j)) = \delta_{jk} A_j^{-1}$$

Hence: $\sum_{i=1}^{m} \tau_{ij} \frac{A_k^{-1}}{2} - \frac{1}{2} (x_i - \mu_k)(x_i - \mu_k)^T \tau_{ij} = 0$

$$\Rightarrow \boxed{A_k^{-1} = \frac{\sum_{i=1}^{m} \tau_{ij} (x_i - \mu_k)(x_i - \mu_k)^T}{\sum_{i=1}^{m} \tau_{ij}} = \Sigma}$$

$\Rightarrow$

Exercise 3

4 - in (iii), we want to optimize:

$$\sum_{i=1}^{n} \tilde{w}_i^{(0)} \log q_\theta(X_i^{(0)})$$

It is the log likelihood, but with additionnal weights. We can use the same proof as in the original EM:

$$\cancel{\tilde{w}_i(0)\log}\ q_\theta(X_i^{(0)}) = \int \cancel{\tilde{w}_i(0)\log}\ q_\theta(X_i^{(0)}, z_i)\, dz_i$$

$$= \int \cancel{\tilde{w}_i(0)\log}\ q_\theta(X_i^{(0)}, z_i) \frac{p(z_i)}{p(z_i)} dz_i$$

$$= \mathbb{E}_{z \sim p(\cdot)} \left[\frac{\cancel{\tilde{w}_i(0)\log}\ q_\theta(X_i^{(0)}, z)}{p(z)} \right]$$

=> Jensen: ~~log (w̃_i (0)~~

$$\log q_\theta(X_i^{(0)}) \geq \mathbb{E}_{z_i \sim p(\cdot)} \left[\log \frac{q_\theta(X_i, z_i)}{p(z_i)}\right]$$

$$\Rightarrow \sum_{i=1}^{n} \tilde{w}_i(0) \log q_\theta(X_i^{(0)}) \geq \sum_{i=1}^{n} \tilde{w}_i(0)\, \mathbb{E}_{z_i \sim p(\cdot)} \left[\log \frac{q_\theta(X_i, z_i)}{p(z_i)}\right]$$

$$= \cancel{\sum_{i=1}^{n} \sum_{j=1}^{p}} \tilde{w}_i(0)\, \tau_{ij}$$

The $(\tilde{w}_i(0))$ do not change the optimization over β, hence when optimizing over θ we have the ELBO:

$$\sum_{i=1}^{n} \sum_{j=1}^{p} \tilde{w}_i(0)\, \tau_{ij} \left[\mathbb{E}_{z_i \sim q_{\theta^{(0)}}(\cdot \mid X_i)} [\log q_\theta(X_i, z_i)] \right]$$

We recognize the same pb as before,
Hence: $\tau_{ij}' \leftarrow \tau_{ij}\tilde{w}_i(0)$

$$\alpha_k = \frac{\sum_{i=1}^{n} \tau_{ik}\tilde{w}_i(0)}{\sum_{i\leq n}\sum_{j\leq p} \tau_{ik}\tilde{w}_i(0)}$$

$$\mu_k = \frac{\sum_{i=1}^{n} \tau_{ik}\tilde{w}_i(0)\, x_i}{\sum_{i\leq n} \tau_{ik}\tilde{w}_i(0)}$$

$$\Sigma_k = \frac{\sum_{i\leq n} \tau_{ij}\tilde{w}_i(0)\,(x_i-\mu_k)(x_i-\mu_k)^T}{\sum_{i\leq n} \tau_{ij}\tilde{w}_i(0)}$$