TAYEBJEE
Meïlame
MVA 23-24

TD 2 - Computational Statistics

Exercice 1

1) On considère les $\{x_1 \dots x_n\}$ triés dans l'ordre croissant.
On cherche à appliquer la méthode de la transformée inverse, et pour cela on s'intéresse à l'inverse généralisée de la fonction de répartition de la loi cible qui est :

$$F_X(x) = \sum_{k=1}^{i} p_k$$

Son inverse généralisée $F^{-1}(u) = \inf_{\{x_1 \dots x_n\}} \{x \mid F(x) \geq u\}$ se définit par :

$$F^{-1}(u) = x_i \iff u \in \left[\sum_{k=1}^{i-1} p_k \, ; \, \sum_{k=1}^{i} p_k \right]$$

Si $U \hookrightarrow \mathcal{U}_{[0,1]}$, $F^{-1}(U)$ suit la loi cible

en effet : $\mathbb{P}(F^{-1}(U) = x_i) = \mathbb{P}\left(U \in \left[\sum_{k=1}^{i-1} p_k \, ; \, \sum_{k=1}^{i} p_k \right]\right) = p_i$

On a alors le pseudo-code suivant :

- Input : $p = (p_1 \dots p_n)$, $x = (x_1 \dots x_n)$ $x_1 < \dots < x_n$
- Compute $\mathcal{P} = (p_1, p_1 + p_2, \dots, \sum_{k=1}^{n} p_k)$
- Générer $U \hookrightarrow \mathcal{U}_{[0,1]}$
- Trouver l'unique i_0 / $\mathcal{P}_{i_0 - 1} < U < \mathcal{P}_{i_0}$
- Return x_{i_0} (qui est $F^{-1}(U)$)

$\boxed{\text{Exercice 2}}$

1/ i/ Les paramètres sont ici :

$$\theta = (\alpha_j, \mu_j, \Sigma_j)_{1 \leq j \leq p}$$

où
$$\begin{cases} (\alpha_j)_j \in \{\alpha \in \mathbb{R}^p \mid \sum_{j=1}^{p} \alpha_j = 1, \alpha_j > 0 \ j = 1 \dots p\} \\ (\mu_j)_j \in \mathbb{R}^d \\ \Sigma_j \in S_n^{++}(d) \ \underline{\text{ouvert}} \end{cases}$$

ii/ Pour une observation x donnée :

$$p_\theta(x) = \int_Z p_\theta(x, z)\, dz = \int_Z p_\theta(z)\, p_\theta(x|z)\, dz$$

$$= \sum_{j=1}^{p} \underbrace{p_\theta(z_j)}_{\alpha_j} \underbrace{p_\theta(x \mid z = j)}_{(*)}$$

où on a : $(*) = p_\theta(x \mid z = j) := \mathcal{N}(x ; \mu_j, \Sigma_j)$

$$= \frac{1}{(2\pi)^{d/2}} \frac{1}{\sqrt{\det \Sigma_j}} \exp\left\{-\frac{1}{2}(x - \mu_j)^T \Sigma_j^{-1} (x - \mu_j)\right\}$$

Il vient alors pour la vraisemblance des $(X_1 - X_n)$ iid :

$$\boxed{p_\theta(x_1 - x_n) = \prod_{i=1}^{n} p_\theta(x_i) = \prod_{i=1}^{n} \sum_{j=1}^{p} \frac{\alpha_j}{(2\pi)^{d/2}} \frac{1}{\sqrt{|\Sigma_j|}} \exp\left\{-\frac{1}{2}(x_i - \mu_j)^T \Sigma_j^{-1} (x_i - \mu_j)\right\}}$$

2/ NOTEBOOK.

3/ On s'intéresse donc ici à l'algorithme EM pour le GMM.
À l'étape $(M)_t$, on cherche à résoudre le programme, à $\theta^{(t)} = (\alpha^{(t)}, \mu^{(t)}, \Sigma^{(t)})$ donné

$$\theta^{t+1} \in \underset{\substack{\theta \\ \alpha_j > 0}}{\operatorname{argmax}} \ \mathbb{E}_{Z \sim P_{\theta^t}(\cdot | (x_i)_{1 \le i \le n})} \left[\log P_\theta \left((x_i)_i , (z_i)_i \right) \right]$$

$$= \underset{\substack{\theta \\ \alpha_j > 0}}{\operatorname{argmax}} \ \sum_{i=1}^{N} \underbrace{\mathbb{E}_{Z_i \sim P_{\theta^t}(\cdot | x_i)} \left[\log \left(P_\theta (x_i, z_i) \right) \right]}_{:= F_i(\theta)}$$

$$\underline{\text{s.c}} \quad \sum \alpha_j = 1$$

Car $P_\theta \left((x_i)_i (z_i)_i \right) = \prod_{i=1}^{N} P_\theta (x_i, z_i)$ par indépendance des $\left((x_i, z_i) \right)_{1 \le i \le N}$

On a :

$$F_i(\theta) = \sum_{j=1}^{p} \underbrace{\log \left(P_\theta (x_i, z_i = j) \right)} \underbrace{P_{\theta^t}(Z_i = j \mid x_i)}_{\tau_{ij}}$$

- On calcule la loi a posteriori de Z_i :

$$P_{\theta^t}(Z_i = j \mid x_i) = \frac{P_{\theta^t}(x_i \mid Z_i = j) \overbrace{P_{\theta^t}(Z_i = j)}^{= \alpha_j^{(t)}}}{\sum_{k=1}^{p} P_{\theta^t}(x_i \mid Z_i = k) P_{\theta^t}(Z_i = k)} \quad \text{(Bayes)}$$

$$= \boxed{\frac{\alpha_j^{(t)} \mathcal{N}(x_i ; \mu_j^{(t)}, \Sigma_j^{(t)})}{\sum_{k=1}^{p} \alpha_k^{(t)} \mathcal{N}(x_i ; \mu_k^{(t)}, \Sigma_k^{(t)})} \equiv \tau_{ij}}$$

où $\mathcal{N}(x_i ; \mu_j, \Sigma_j)$ est la densité $\mathcal{N}(\mu_j, \Sigma_j)$ en x_i

$\hookrightarrow$ indépendant de θ / (dépend uniquement de θ^t)

- On a :

$$\log\left(P_\theta(x_i, z_i = j)\right) = \log\left(P_\theta(x_i \mid z_i = j)\, P_\theta(z_i = j)\right)$$

$$= \log\left(\mathcal{N}(x_i ; \mu_j, \Sigma_j)\right) + \log(\alpha_j)$$

$$= \text{constante} - \frac{1}{2}\log\det\Sigma_j - \frac{1}{2}(x_i-\mu_j)^T\Sigma_j^{-1}(x_i-\mu_j) + \log(\alpha_j)$$

Si bien qu'in fine, le programme se réduit à :

$$\theta^{t+1} \in \underset{\substack{\theta \\ \alpha_j > 0}}{\operatorname{argmax}} \sum_{i=1}^{N}\sum_{j=1}^{P} \tau_{ij}\left[\log(\alpha_j) - \frac{1}{2}\log\det\Sigma_j - \frac{1}{2}(x_i-\mu_j)^T\Sigma_j^{-1}(x_i-\mu_j)\right]$$

$$= \sum_{j=1}^{P}\left\{-\frac{1}{2}\log\det\Sigma_j\, T_j + T_j\log(\alpha_j) - \frac{1}{2}\sum_{i=1}^{n}\tau_{ij}(x_i-\mu_j)^T\Sigma_j^{-1}(x_i-\mu_j)\right\}$$

s.c $\sum_{j=1}^{P}\alpha_j = 1$ où $\boxed{T_j = \sum_{i=1}^{n}\tau_{ij}}$

La fonction objectif est concave en tant que somme finie des $(F_i)_i$, elles-mêmes étant concaves par combinaison convexe ($\tau_{ij} \geq 0$ $\forall i \forall j$, $\boxed{\forall i : \sum_{j=1}^{P}\tau_{ij} = 1}$) des fonctions concaves $\left(\theta \mapsto \log(P_\theta(x_i, z_i=j))\right)_{j\in[\![1,P]\!]}$ pour $i \in [\![1,n]\!]$.

La contrainte est linéaire.

Donc tout point critique du Lagrangien est un maximum global ($\nabla_\theta \mathcal{L} = 0$)

Le Lagrangien est, pour $\lambda \in \mathbb{R}$:

$$\mathcal{L}(x,\theta,\lambda) = \sum_{j=1}^{p} \left\{ -\frac{1}{2} \log \det \Sigma_j \, T_j + T_j \log(\alpha_j) - \frac{1}{2} \sum_{i=1}^{n} \tau_{ij} (x_i - \mu_j)^T \Sigma_j^{-1} (x_i - \mu_j) \right\}$$

$$+ \lambda \left\{ 1 - \sum_{j=1}^{p} \alpha_j \right\}.$$

Déterminons les points critiques :

- $\nabla_{\mu_j} \mathcal{L} = 0 \iff 2 \sum_{i=1}^{N} \tau_{ij} \Sigma_j^{-1} (x_i - \mu_j) = 0$

$$\iff \boxed{\mu_j^{(t+1)} = \frac{\sum_{i=1}^{N} \tau_{ij} x_i}{T_j} \quad \forall j}$$

- On utilise que $\nabla_{\Sigma_j^{-1}} \log \det \Sigma_j = - \nabla_{\Sigma_j^{-1}} \log \det \Sigma_j^{-1} = -\Sigma_j$

et que $(x_i - \mu_j)^T \Sigma_j^{-1} (x_i - \mu_j)$

$= \mathrm{Tr} \left[\Sigma_j^{-1} (x_i - \mu_j)(x_i - \mu_j)^T \right]$

où $x_i - \mu_j$ est une colonne

et que $\nabla_{\Sigma_j^{-1}} \mathrm{Tr} \left(\Sigma_j^{-1} (x_i - \mu_j)(x_i - \mu_j)^T \right)$

$= (x_i - \mu_j)(x_i - \mu_j)^T$

Pour obtenir

$$\nabla_{\Sigma_j^{-1}} \mathcal{L} = 0 \iff \frac{T_j}{2} \Sigma_j - \frac{1}{2} \sum_i \tau_{ij} (x_i - \mu_j)(x_i - \mu_j)^T = 0$$

$$\iff \boxed{\Sigma_j^{(t+1)} = \frac{1}{T_j} \sum_i \tau_{ij} (x_i - \mu_j)(x_i - \mu_j)^T \quad \forall j}$$

$\bullet \quad \nabla_{\alpha_j} = 0 \iff \dfrac{\tau_j}{\alpha_j} = \lambda$

$$\iff \sum_{j=1}^{P} \tau_j = \lambda \sum_{j=1}^{P} \alpha_j = \lambda .$$

or $\displaystyle\sum_{j=1}^{P} \tau_j = \sum_{j=1}^{P} \sum_{i=1}^{N} \tau_{ij} = \sum_{i=1}^{N} \sum_{j=1}^{P} \tau_{ij} = \sum_{i=1}^{N} \underbrace{\sum_{j=1}^{P} P_\theta(z_j | x_i)}_{=1} = N$

D'où $\dfrac{\tau_j}{\alpha_j} = N \iff \boxed{\alpha_j^{(t+1)} = \dfrac{\tau_j}{N} \quad \forall j}$

$\bullet$ À l'étape $(E)_{t+1}$ on cherche à évaluer la loi a posteriori $P_{\theta^{(t+1)}}(z_i \mid x_i) \ \forall i$, autrement dit on "met à jour" les τ_{ij} (les "responsabilités") avec le nouveau $\theta^{(t+1)}$

Si bien que le pseudo-code de l'EM devient dans le cas GMM :

$$\boxed{\begin{array}{ll}
\underline{\text{Init}} & \theta = (\alpha, \Sigma, \mu) \\[2ex]
\underline{\text{Repeat}} & (E) : \quad \tau_{ij} \leftarrow \dfrac{\alpha_j \, \mathcal{N}(x_i ; \mu_j, \Sigma_j)}{\sum_{k=1}^{P} \alpha_k \, \mathcal{N}(x_i ; \mu_k, \Sigma_k)} \quad \forall i, \forall j \\[2ex]
\underline{\text{Until}} \\ \underline{\text{convergence}} & (M) : \quad \mu_j \leftarrow 1/\tau_j \sum_{i=1}^{N} \tau_{ij} x_i \\[2ex]
& \qquad \Sigma_j \leftarrow \dfrac{1}{\tau_j} \sum_{i=1}^{N} \tau_{ij} (x_i - \mu_j)(x_i - \mu_j)^T \quad \forall j \\[2ex]
& \qquad \alpha_j \leftarrow \tau_j / N \quad \forall j
\end{array}}$$

Exercice 3

4/ Ici, $q_{\theta^{(0)}}(X_i^{(0)}) = p_\theta(x_i)$ en reprenant les notations du GMM (exercice 2) $\left(\begin{array}{l} x_i \sim \mathcal{N}(\mu_j, \Sigma_j) \\ x|z_i = j \quad , P(z_i = j) = \alpha_j \end{array} \right)$

On garde, comme dans le cadre du GMM, la borne inférieure sur la log-vraisemblance :

$$\forall i, \quad w_i^{(0)} \log\left(q_\theta(X_i^{(0)})\right) \geq w_i^{(0)} Q(\theta, \theta^{(0)})$$

car $w_i^{(0)} \propto \nu(X_i^{(0)}) / q_{\theta^{(0)}}(X_i^{(0)}) \geq 0$

avec $Q(\theta, \theta^{(0)}) = \mathbb{E}_{Z_i \sim p(\cdot|x,\theta^{(0)})}\left[\log p_\theta(x_i, z_i)\right]$

En reprenant le fil de Exercice 2 q.3/, on peut donc appliquer un algorithme EM :

Étape M Le programme devient :

$$\theta^{(1)} \in \underset{\theta}{\operatorname{argmax}} \sum_{i=1}^{N} \underline{w_i^{(0)}} \underbrace{\mathbb{E}_{Z_i \sim p_\theta(\cdot|x_i)}\left[\log\left(p_\theta(x_i, z_i)\right)\right]}_{:= F_i(\theta)}$$

$$\underline{\text{s.c}} \quad \sum \alpha_j = 1$$

Le calcul des F_i ne change pas. On a par suite :

$$\theta^{(1)} \in \underset{\substack{\theta \\ \alpha_j > 0}}{\operatorname{argmax}} \sum_{i=1}^{N} w_i^{(0)} \sum_{j=1}^{p} \tau_{ij} \Big[\log(\alpha_j) - \frac{1}{2} \log\det \Sigma_j - \frac{1}{2} (x_i - \mu_j)^T \Sigma_j^{-1} (x_i - \mu_j) \Big]$$

$$= \sum_{j=1}^{p} \Big\{ \log(\alpha_j)\,(w\tau)_j - \frac{1}{2} (w\tau)_j \log\det\Sigma_j - \frac{1}{2} \sum_{i=1}^{N} w_i^{(0)} \tau_{ij} (x_i - \mu_j)^T \Sigma_j^{-1} (x_i - \mu_j) \Big\}$$

<u>s.c</u> $\sum_{j=1}^{p} \alpha_j = 1$

où $\boxed{(w\tau)_j = \sum_{i=1}^{N} w_i^{(0)} \tau_{ij}}$

Le problème est toujours concave ($w_i^{(0)} \geq 0$) :

on calcule toujours un point critique du Lagrangien :

- $\nabla_{\mu_j} \mathcal{L} = 0 \iff 2 \sum_{i=1}^{N} \tau_{ij} w_i^{(0)} \Sigma_j^{-1} (x_i - \mu_j) = 0$

$$\iff \boxed{\mu_j^{(1)} = \frac{\sum_{i=1}^{N} w_i^{(0)} \tau_{ij}\, x_i}{(w\tau)_j} \quad \forall j}$$

- $\nabla_{\Sigma_j^{-1}} \mathcal{L} = 0 \iff \frac{(w\tau)_j}{2} \Sigma_j - \frac{1}{2} \sum_i w_i^{(0)} \tau_{ij} (x_i - \mu_j)(x_i - \mu_j)^T = 0$

$$\iff \boxed{\Sigma_j^{(1)} = \frac{1}{(w\tau)_j} \sum_i w_i^{(0)} \tau_{ij} (x_i - \mu_j)(x_i - \mu_j)^T \quad \forall j}$$

- $\nabla_{\alpha_j} = 0 \iff \dfrac{(w\tau)_j}{\alpha_j} = \lambda$

$$\Rightarrow \sum_{j=1}^{p} (w\tau)_j = \lambda \sum \alpha_j = \lambda$$

or $$\sum_{j=1}^{p} (w\tau)_j = \sum_{j=1}^{p} \sum_{i=1}^{N} w_i^{(0)} \tau_{ij} = \sum_{i=1}^{N} w_i^{(0)} \underbrace{\left(\sum_{j=1}^{p} \tau_{ij} \right)}_{= \sum_{j=1}^{p} p_{\theta^{(0)}}(z_i = j \mid x_i) = 1}$$

$$\Rightarrow \lambda = W^{(0)} := \sum_{i=1}^{N} w_i^{(0)}$$

D'où $$\boxed{\alpha_j = \frac{(w\tau)_j}{W^{(0)}}}$$

En notant que si les poids sont normalisés, $W^{(0)} = n$

- L'étape (E) reste la même, étant donné $\theta^{(1)}$, on met à jour la posterior de z ie :

$$\tau_{ij} \longleftarrow \frac{\alpha_j^{(1)} \mathcal{N}(x_i ; \mu_j^{(1)}, \Sigma_j^{(1)})}{\sum_{k=1}^{p} \alpha_k^{(1)} \mathcal{N}(x_i ; \mu_k^{(1)}, \Sigma_k^{(1)})}$$