TP2 : Algorithme EM

de Charrin Théotime
10/11/2023

Exercice 1

1. On utilisera la méthode de la transformée inverse pour cette distribution discrète.
Soit notre de valeurs E ordonnées : $E = \{x_1, x_2, \dots x_n\}$, $x_1 < x_2 < \dots < x_n$.
La fonction de répartition $F : \begin{cases} E \mapsto [0,1] \\ x_k \mapsto P(x \leq x_k) = \sum_1^k \end{cases}$

Soit u a $\mathcal{U}([0,1])$
L'inverse généralisée est $F^{-1} : \begin{cases} [0,1] \mapsto E \\ u \mapsto \inf \{x_k \in E, F(x_k) \geq u\} \end{cases}$

Soit $\{p_1, p_2, \dots p_n\}$ les probabilités associées à E.

On a bien si $u \in [0, p_1)$, $F^{-1}(u) = x_1$
De manière générale si $u \in \left[\sum_1^{k-1} p_i ; \sum_1^k p_i\right]$, on a $F^{-1}(u) = x_k$

2. On calcule la distribution cumulative et on utilise la transformée inverse

Exercice 2

(1) Les paramètres θ du modèle GMM sont les probabilités de chaque gaussienne, les moyennes et les variances.
Soit pour un modèle à p gaussiennes :

$$\theta = \left\{ \{\alpha_j\}_{j \in [\![1,p]\!]}, \{\mu_j\}_{j \in [\![1,p]\!]}, \{\Sigma_j\}_{j \in [\![1,p]\!]} \right\}$$

On considère les x_i iid donc la vraisemblance est :

$$\mathcal{L}(x_1, x_2, \dots x_n ; \theta) = p_\theta(x_1, x_2 \dots, x_n) = \prod_{i=1}^n p_\theta(x_i)$$

On peut écrire $p_\theta(x)$ comme la marginale entre x et z.

D'où $p_\theta(x) = \int p_\theta(x, z)\, dz$

Comme la variable Z est une mesure de comptage à valeurs discrètes :

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

Or $p_\theta(z_j) = \mathbb{P}(Z = z_j) = \alpha_j$

D'où $p_\theta(x) = \sum_{j=1}^{p} \alpha_j\, p_\theta(x|z_j)$

$p_\theta(x|z_j)$ est une densité gaussienne de paramètres (μ_j, Σ_j).

On a $\mathcal{L}(x_1, x_2, \dots, x_n) = \prod_{i=1}^{n} \sum_{j=1}^{p} \alpha_j\, p_\theta(x_i|z_j)$.

3. L'algorithme EM cherche à maximiser la log-vraisemblance

$$\mathcal{L} = \sum_1^n \log p_\theta(x_i) = \sum_1^n \log\left(\sum \alpha_j\, p_\theta(x_i|z_j)\right)$$

Plus précisément on a une Empirical Lower Bound (ELBO) :

$$\log p_\theta(x) \geq \mathbb{E}_{q(z)}\left[\log \frac{p_\theta(x, z)}{q(z)}\right]$$

$$\geq \mathbb{E}_{q(z)}\left[\log p_\theta(x, z)\right] - \int \log q(z)\, dz$$

$$\geq \underbrace{-KL\left(q(z) \,\|\, p_\theta(z|x)\right)}_{\text{gap à minimiser}} + \log p_\theta(x)$$

On cherche donc à calculer à chaque itération t :

$$\theta^{t+1} = \underset{\theta}{\operatorname{argmax}}\ ELBO = \underset{\theta}{\operatorname{argmax}}\ \mathbb{E}_{q(z)}\left[\log p_\theta(x,z)\right]$$

$$q^{t+1} = p_{\theta^{t+1}}(z|x)$$

En combinant les deux :

$$\theta^{t+1} = \underset{\theta}{\operatorname{argmax}}\ \mathbb{E}_{z\sim p_{\theta^t}(\cdot|x)}\left[\log p_\theta(x,z)\right]$$

On pose $f_x(\theta) = \mathbb{E}_{z\sim p_{\theta^t}(\cdot|x)}\left[\log p_\theta(x,z)\right]$ qu'on veut maximiser.

$$= \sum_{j=1}^{p} \log p_\theta(x,z_j)\ p_{\theta^t}(z_j|x)$$

$$= \sum_{j=1}^{p} \Big[\underbrace{\log p_\theta(x|z_j)}_{\sim\mathcal{N}(\mu_j,\Sigma_j)} + \underbrace{\log p_\theta(z_j)}_{\log \alpha_j}\Big]\ \underbrace{p_{\theta^t}(z_j|x)}_{\text{on ne connaît pas, on note } \tau_j}$$

$$\text{D'où } f_x(\theta) \propto \sum_{j=1}^{p} \tau_j \left(\log\left[\frac{1}{|\det \Sigma_j|^{1/2}} \exp\left(-\frac{1}{2}(x-\mu_j)^t \Sigma_j^{-1}(x-\mu_j)\right)\right] + \log \alpha_j\right)$$

$$f_x(\theta) \propto \sum_{j=1}^{p} \tau_j \left(-\frac{1}{2}\log|\det \Sigma_j| - \frac{1}{2}(x-\mu_j)^t \Sigma_j^{-1}(x-\mu_j) + \log \alpha_j\right)$$

On sait que $|\det A^{-1}| = -|\det A|$

$$f_x(\theta) \propto \sum_{j=1}^{p} \tau_j \left(\frac{1}{2}\log|\det \Sigma_j^{-1}| - \frac{1}{2}(x-\mu_j)^t \Sigma_j^{-1}(x-\mu_j) + \log \alpha_j\right)$$

On a le problème d'optimisation suivant :

Maximiser $f_x(\theta)$ sous contraintes : $\sum_{j=1}^{p} \alpha_j = 1 \quad \alpha_j > 0\ \forall j \in [\![1;p]\!]$

Σ_j matrice de covariance inversible.

On a le Lagrangien suivant :

$$\mathcal{L}(\theta, \lambda) = F_x(\theta) + \lambda \left(1 - \sum_1^p \alpha_j\right)$$

F_x ~~est~~ concave en θ (log).

On cherche θ^* maximiseur global, ie $\begin{cases} \nabla_\theta \mathcal{L}(\theta, \lambda)\big|_{\theta^*} = 0 \\ 1 - \sum \alpha_j = 0 \end{cases}$

• On a $\dfrac{\partial \mathcal{L}}{\partial \mu_j} = \displaystyle\sum_{j=1}^p \tau_j \left(-\frac{1}{2} \times \left(-2 \Sigma_j^{-1} (x - \mu_j) \right) \right)$

$$= \sum_{j=1}^p \tau_j \Sigma_j^{-1} (x - \mu_j)$$

On veut $\dfrac{\partial \mathcal{L}}{\partial \mu_j} = 0 \iff \displaystyle\sum_1^p \tau_j \cancel{\Sigma_j^{-1}} x = \sum_1^p \tau_j \cancel{\Sigma_j^{-1}} \mu_j$

$$\iff \mu_j = \frac{\sum_1^p \tau_j x}{\sum_1^p \tau_j}$$

On obtient la moyenne empirique pondérée par les probabilités postérieures.

• $\dfrac{\partial \mathcal{L}}{\partial \alpha_j} = \displaystyle\sum_1^p \tau_j \frac{1}{\alpha_j} - \lambda$

$\dfrac{\partial \mathcal{L}}{\partial \alpha_j} = 0 \iff \displaystyle\sum_1^p \tau_j \frac{1}{\alpha_j} = \lambda$

$$\iff \alpha_j = \frac{\sum \tau_j}{\lambda}$$

Or on sait $\displaystyle\sum_1^n \alpha_j = 1$ et $\displaystyle\sum_1^p \tau_j = 1$

donc $\displaystyle\sum_1^p \alpha_j = \frac{p}{\lambda} = 1 \iff \lambda = n$

d'où $\alpha_j = \dfrac{\sum \tau_j}{n}$

- $\dfrac{\partial \mathcal{L}}{\partial \Sigma_j^{-1}} = \dfrac{\partial}{\partial \Sigma^{-1}} \left(\sum_1^p \tau_j \left(\frac{1}{2} \log |\det \Sigma_j^{-1}| - \frac{1}{2} (x-\mu_j)^t \Sigma_j^{-1} (x-\mu) \right) \right)$

$\dfrac{\partial}{\partial M} \log |\det M| = M^{-1}$ et $\dfrac{\partial}{\partial M} a^t M a = a a^t$

Donc $\dfrac{\partial \mathcal{L}}{\partial \Sigma_j^{-1}} = \sum_1^p \tau_j \times \frac{1}{2} \left((\Sigma_j^{-1})^{-1} - (x-\mu_j)(x-\mu_j)^t \right)$

$$\frac{\partial \mathcal{L}}{\partial \Sigma_j^{-1}} = 0 \iff \frac{1}{2} \sum_1^p \tau_j \left[\Sigma_j - (x-\mu_j)(x-\mu_j)^t \right] = 0$$

$$\iff \Sigma_j = \frac{\sum_1^p \tau_j (x-\mu_j)(x-\mu_j)^t}{\sum_1^p \tau_j}$$

On retrouve la covariance empirique pondérée par les posterior.

Comme on ne connaît pas les τ_j :

on approxime de manière empirique,

$$\tau_j = P_{\theta^t}(\mathfrak{z}_j | x) \approx \frac{P_{\theta^t}(x|\mathfrak{z}_j)\, P_{\theta^t}(\mathfrak{z}_j)}{\sum_1^p P_{\theta^t}(x|\mathfrak{z}_j)\, P_{\theta^t}(\mathfrak{z}_j)} = \frac{\tilde{\tau}_j}{\sum_1^p \tilde{\tau}_j}$$

On a donc :

Etape (E) : $\tau_j = \dfrac{\tilde{\tau}_j}{\sum_1^p \tilde{\tau}_j}$

Etape (M) : $\alpha_j = \dfrac{\sum_1^p \tau_j}{n}$ $\quad \mu_j = \dfrac{\sum_1^p \tau_j x}{\sum_1^p \tau_j}$

$$\Sigma_j = \frac{\sum_1^p \tau_j (x-\mu_j)(x-\mu_j)^t}{\sum_1^p \tau_j}$$

Exercice 3

4]. En posant q_θ la densité d'une famille de mixture de gaussiennes, le PBMC (iii) nous fait calculer

$$\theta^{t+1} = \underset{\theta}{\operatorname{argmax}} \sum_{i=1}^{n} \tilde{\omega}_i^t \log q_\theta(x_i^t)$$

Avec $\tilde{\omega}_i^t = \dfrac{\partial(x_i^t)}{q_{\theta^t}(x_i^t)}$ qui est donc connue.

C'est très similaire à notre problème initial d'EM : $\underset{\theta}{\operatorname{argmax}} \log p_\theta(x)$.

En réécrivant $p_\theta(x)$ comme au début de l'exercice 2 :

$$p_\theta(x) = p_\theta(x_1, x_2, \dots, x_n) = \prod_1^n p_\theta(x_i)$$

Avec φ une densité normale : $= \prod_1^n \sum_1^J \alpha_j \varphi(x_i ; \mu_j, \Sigma_j)$.

$$\log p_\theta(x) = \sum_1^n \log\left(\sum_1^J \alpha_j \varphi(x_i ; \mu_j, \Sigma_j)\right)$$

$$= \sum_1^n \log q_\theta(x_i)$$

On retrouve notre problème d'EM à une constante $\tilde{\omega}_i^t$ près qui est connue car approximée à chaque étape (E) par $q_{\theta^t}(x_i^t)$ et non q_θ.

Notre problème est donc équivalent à maximiser $F_x(\theta)$:

$$F_x(\theta) = \mathbb{E}_{q_{\theta^t}(z|x)}\left[\tilde{\omega}_i^t \log q_\theta(z, x)\right] \quad \text{sc } \sum \alpha_j = 1$$

La constante $\tilde{\omega}_i^t$ va être dérivée à chaque étape de résolution du lagrangien car n'étant dépendante ni de α_j, μ_j ou Σ_j.

En dérivant, il vient donc assez naturellement :

$$\alpha_j^{t+1} = \sum_{i=1}^{n} \tilde{\omega}_i \, \tau_{j,i}$$

$$\mu_j^{t+1} = \frac{\sum_{i=1}^{n} \tilde{\omega}_i^t \, \tau_{j,i} \, x_i}{\sum_{i=1}^{n} \tilde{\omega}_i^t \, \tau_{j,i}}$$

$$\Sigma_j^{t+1} = \frac{\sum_{i=1}^{n} \tilde{\omega}_i^t \, \tau_{j,i} \, (x - \mu_j)(x - \mu_j)^T}{\sum_{i=1}^{n} \tilde{\omega}_i^t \, \tau_{j,i}}$$