

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 ensemble de valeurs E ordonnées: $E = \{x_1, x_2, \dots, x_n\}$, $x_1 < x_2 < \dots < x_n$.
 La fonction de répartition F: $\begin{cases} E \mapsto [0,1] \\ x_k \mapsto P(X \leq x_k) = \sum_{i=1}^k p_i \end{cases}$

Soit $u \in U([0,1])$

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 $u \in \left[\sum_{i=1}^{k-1} p_i ; \sum_{i=1}^k p_i \right]$, 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:

$$\theta = \left\{ \begin{array}{l} \{ \alpha_j \}_{j \in [1, p]} \\ \{ N_j \}_{j \in [1, p]} \\ \{ \Sigma_j \}_{j \in [1, p]} \end{array} \right\}$$

On considère les x_i iid donc la vraisemblance est:

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

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

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

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

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

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

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

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

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

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

$$\mathcal{L} = \sum_{i=1}^n \log p_{\theta}(x_i) = \sum_{i=1}^n \log \left(\sum_{j=1}^p \alpha_j p_{\theta}(x_i|z_j) \right)$$

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

$$\begin{aligned} \log p_{\theta}(x) &\geq \mathbb{E}_{q(z)} \left[\log \frac{p_{\theta}(x, z)}{q(z)} \right] \\ &\geq \mathbb{E}_{q(z)} \left[\log p_{\theta}(x, z) \right] - \int \log q(z) dz \\ &\geq \underbrace{-KL(q(z) || p_{\theta}(z|x))}_{\text{gap à minimiser}} + \log p_{\theta}(x) \end{aligned}$$

On cherche donc à calculer à chaque itération t :

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

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

En combinant les deux:

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

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

$$= \sum_{j=1}^P \log p_{\theta}(x | \beta_j) p_{\theta^t}(\beta_j | x)$$

$$= \sum_{j=1}^P \underbrace{[\log p_{\theta}(x | \beta_j) + \log p_{\theta}(\beta_j)]}_{\sim \mathcal{N}(\mu_j, \Sigma_j)} p_{\theta^t}(\beta_j | x)$$

$\log \alpha_j$ on ne connaît pas,
on connait τ_j

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

$$f_x(\theta) \approx \sum_{j=1}^P \tau_j \left(-\frac{1}{2} \log |\det \Sigma_j| - \frac{1}{2} (\mu - \mu_j)^T \Sigma_j^{-1} (\mu - \mu_j) + \log \alpha_j \right)$$

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

$$f_x(\theta) \approx \sum_{j=1}^P \tau_j \left(\frac{1}{2} \log |\det \Sigma_j| - \frac{1}{2} (\mu - \mu_j)^T \Sigma_j^{-1} (\mu - \mu_j) + \log \alpha_j \right)$$

On a le problème d'optimisation suivant:

Maximiser $f_x(\theta)$ sous contraintes: $\sum_{j=1}^P \alpha_j = 1$ $\alpha_j > 0$ $\forall j \in [1; P]$
 Σ_j matrice de covariance inversible.

On a le Lagrangien suivant:

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

F_x est concave en θ (log).

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

- On a $\frac{\partial \mathcal{L}}{\partial \alpha_j} = \sum_{j=1}^P \tau_j \left(-\frac{1}{2} \times \left(-2 \sum_j^{-1} (\mu - \mu_j) \right) \right)$
- $= \sum_{j=1}^P \tau_j \sum_j^{-1} (\mu - \mu_j)$

On veut $\frac{\partial \mathcal{L}}{\partial \alpha_j} = 0 \Leftrightarrow \sum_1^P \tau_j \cancel{\sum_j^{-1} \mu} = \sum_1^P \tau_j \cancel{\sum_j^{-1} \mu_j}$

 $\Leftrightarrow \mu_j = \frac{\sum_1^P \tau_j \mu}{\sum_1^P \tau_j}$

On obtient la moyenne empirique pondérée par les probabilités postérieures.

- $\frac{\partial \mathcal{L}}{\partial \alpha_j} = \sum_j^P \tau_j \frac{1}{\alpha_j} - \lambda$
- $\frac{\partial \mathcal{L}}{\partial \alpha_j} = 0 \Leftrightarrow \sum_1^P \tau_j \frac{1}{\alpha_j} = \lambda$
- $\Leftrightarrow \alpha_j = \frac{\sum \tau_j}{\lambda}$

Or on sait $\sum_1^P \alpha_j = 1$ et $\frac{P}{1} \tau_j = 1$

donc $\sum_1^P \alpha_j = \frac{P}{\lambda} = 1 \Leftrightarrow \lambda = P$

d'où $\alpha_j = \frac{\sum \tau_j}{P}$

$$\frac{\partial \mathcal{L}}{\partial \Sigma^{-1}} = \frac{\partial}{\partial \Sigma^{-1}} \left(\sum_{j=1}^P \tau_j \left(\frac{1}{2} \log |\det \Sigma_j^{-1}| - \frac{1}{2} (\mu - \mu_j)^t \Sigma_j^{-1} (\mu - \mu_j) \right) \right)$$

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

$$\text{Donc } \frac{\partial \mathcal{L}}{\partial \Sigma^{-1}} = \sum_{j=1}^P \tau_j \times \frac{1}{2} \left((\Sigma_j^{-1})^{-1} - (\mu - \mu_j)(\mu - \mu_j)^t \right)$$

$$\begin{aligned} \frac{\partial \mathcal{L}}{\partial \Sigma^{-1}} = 0 &\Leftrightarrow \frac{1}{2} \sum_{j=1}^P \tau_j \left[\Sigma_j^{-1} - (\mu - \mu_j)(\mu - \mu_j)^t \right] = 0 \\ &\Leftrightarrow \Sigma = \frac{\sum_{j=1}^P \tau_j (\mu - \mu_j)(\mu - \mu_j)^t}{\sum_{j=1}^P \tau_j} \end{aligned}$$

On retrouve la covariance empirique pondérée par les postérieurs.

Comme on ne connaît pas les τ_j :

on approche de manière empirique,

$$\tilde{\tau}_j = P_{\Theta^t}(\beta_j | \mu) \approx \frac{P_{\Theta^t}(\mu | \beta_j) P_{\Theta^t}(\beta_j)}{\sum_1^P P_{\Theta^t}(\mu | \beta_j) P_{\Theta^t}(\beta_j)} = \frac{\tilde{\tau}_j}{\sum_1^P \tilde{\tau}_j}$$

On a donc:

$$\text{Etape (E)} : \tilde{\tau}_j = \frac{\tilde{\tau}_j}{\sum_1^P \tilde{\tau}_j}$$

$$\begin{aligned} \text{Etape (M)} : \alpha_j &= \frac{\sum_1^P \tilde{\tau}_j}{N_j} & N_j &= \frac{\sum_1^P \tilde{\tau}_j \cdot n}{\sum_1^P \tilde{\tau}_j} \\ \Sigma &= \frac{\sum_1^P \tilde{\tau}_j (\mu - \mu_j)(\mu - \mu_j)^t}{\sum_1^P \tilde{\tau}_j} \end{aligned}$$

Exercice 3

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

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

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

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

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

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

Avec φ une densité normale : $= \prod_{i=1}^n \sum_j \alpha_j \varphi(x_i; \mu_j, \Sigma_j)$.

$$\begin{aligned} \log p_\theta(\mathbf{x}) &= \sum_{i=1}^n \log \left(\sum_j \alpha_j \varphi(x_i; \mu_j, \Sigma_j) \right) \\ &= \sum_{i=1}^n \log q_\theta(x_i) \end{aligned}$$

On retrouve notre problème d'EM à une constante \tilde{w}_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 $F_x(\theta)$:

$$F_x(\theta) = \underset{q_{\theta^t}(\mathbf{z}|\mathbf{x})}{\mathbb{E}} \left[\sum_i \log q_\theta(z_i, x_i) \right] \text{ sc } \sum \alpha_j = 1$$

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

En dérivant, il vient donc assez naturellement :

$$\alpha_j^{t+1} = \sum_{i=1}^n \tilde{w}_i \tilde{\gamma}_{j,i}$$

$$\nu_j^{t+1} = \frac{\sum_{i=1}^n \tilde{w}_i^t \tilde{\gamma}_{j,i} x_i}{\sum_{i=1}^n \tilde{w}_i^t \tilde{\gamma}_{j,i}}$$

$$\Sigma_j^{t+1} = \frac{\sum_{i=1}^n \tilde{w}_i^t \tilde{\gamma}_{j,i} (x - \nu_j^t) (\mu - \nu_j^t)^T}{\sum_{i=1}^n \tilde{w}_i^t \tilde{\gamma}_{j,i}}$$