Bastien LE CHENADEC
bastien.le-chenadec@eleves.enpc.fr

TP 2
Computational statistics

23 novembre 2023

## Exercice 1 Discrete distributions

1. "Avec les mains" on divise le segment $[0,1]$ en $n$ intervalles de longueur $p_i$, et on regarde dans quel intervalle tombe $U$. Plus formellement, d'après le théorème d'inversion, on peut tirer $X \sim F^{-1}(U)$ avec $U \sim \mathcal{U}([0,1])$ :

$$x_i = \mathbb{1}_{[0,p_1]}(u_i)x_1 + \sum_{j=1}^{n-1} \mathbb{1}_{[\sum_{k=1}^{j} p_k, \sum_{k=1}^{j+1} p_k]}(u_i)x_{j+1}$$

2. et 3. Sur le notebook.

## Exercice 2 Gaussian mixture model and the EM algorithm

1. On cherche les paramètres $\theta = (\alpha_j, \Sigma_j, \mu_j)_{1\leq j\leq p}$. La vraisemblance s'écrit :

$$\begin{aligned}\mathcal{L}(x_1, \dots, x_n, \theta) &= \prod_{i=1}^{n} p_\theta(x_i) \\ &= \prod_{i=1}^{n} \sum_{j=1}^{p} p_\theta(x_i|z_i = j)p_\theta(z_i = j) \\ &= \prod_{i=1}^{n} \sum_{j=1}^{p} \alpha_j f_{\mu_j,\Sigma_j}(x_i)\end{aligned}$$

où $f$ est la densité gaussienne multivariée :

$$f_{\mu,\Sigma}(x) = \frac{1}{\sqrt{|2\pi\Sigma|}} \exp\left(-\frac{1}{2}(x-\mu)^T\Sigma^{-1}(x-\mu)\right)$$

2. Sur le notebook.

3. $Q(\theta|\theta_t) = \mathbb{E}[\log(p(Z,x))]$ avec $Z \sim P_{\theta_t}(Z|x)$ est une borne inférieure de la vraisemblance qu'on veut maximiser. L'algorithme EM est le suivant :

- Étape E : on calcule $Q(\theta|\theta_t)$.
- Étape M : on calcule $\theta_{t+1} = \arg\max Q(\theta|\theta_t)$.

Étape E :

On peut décomposer $Q$ :

$$Q(\theta|\theta_t) = \sum_{i=1}^{n} \mathbb{E}_{Z_i\sim p_\theta(Z_i|x_i)}[\log p(Z_i, x_i)]$$

De plus $\log p(Z_i = j, x_i) = \log p_\theta(x_i|Z_i = j) + \log p(Z_i = j)$. D'où :

$$Q(\theta|\theta_t) = \sum_{i=1}^{n} \sum_{j=1}^{p} [\log \alpha_j + \log f_{\mu_j,\Sigma_j}(x_i)] Z_{ij}(\theta_t)$$

où on a posé :

$$\begin{aligned} Z_{ij}(\theta_t) &= P_{\theta_t}(Z_i = j|x_i) \\ &= \frac{P(x_i|Z_i = j)P(Z_i = j)}{P(x_i)} \\ &= \frac{\alpha_j^{(t)} f_{\mu_j^{(t)},\Sigma_j^{(t)}}(x_i)}{\sum_{j=1}^{p} \alpha_j^{(t)} f_{\mu_j^{(t)},\Sigma_j^{(t)}}(x_i)} \end{aligned}$$

On peut calculer les $Z_{ij}$ à partir de $\theta_t$. L'étape E est donc explicite.

Étape M :

On a le problème d'optimisation sous contraintes :

$$\min_{\alpha_i,\mu_j,\Sigma_j} - \sum_{\substack{1\leq i\leq n \\ 1\leq j\leq p}} \left[\log \alpha_j + \log f_{\mu_j,\Sigma_j}(x_i)\right] Z_{ij}(\theta_t)$$
$$\text{s.c.} \sum_{j=1}^{p} \alpha_j = 1, \alpha_j \geq 0, \Sigma_j \in S_d^{++}(\mathbb{R})$$

On veut montrer que cet objectif est convexe. Il est clair que $u \mapsto -\log u$ est convexe, la difficulté réside dans la convexité de $\mu, \Sigma \mapsto -\log f_{\mu,\Sigma}(x)$.

On a :

$$\begin{aligned} -\log f_{\mu,\Sigma}(x) &= -\log\left(\frac{1}{\sqrt{|2\pi\Sigma|}} \exp\left(-\frac{1}{2}(x-\mu)^T\Sigma^{-1}(x-\mu)\right)\right) \\ &= \frac{1}{2}(x-\mu)^T\Sigma^{-1}(x-\mu) + \frac{1}{2}\log|2\pi\Sigma| \\ &= \frac{1}{2}(x-\mu)^T\Sigma^{-1}(x-\mu) + \frac{1}{2}\log|\Sigma| + \frac{d}{2}\log 2\pi \\ &= g(\mu,\Sigma) \end{aligned}$$

En dimension 1, en dérivant deux fois $\log f_{\mu,\Sigma}$ par rapport à $\mu$ et $\Sigma$, on montre aisément que $\mu, \Sigma \mapsto \log f_{\mu,\Sigma}$ est concave. Ce résultat peut se généraliser en dimension $d$ en utilisant la matrice hessienne de $g$. C'est un résultat classique : la densité gaussienne est log-concave. En notant à présent qu'on a une combinaison linéaire de fonctions convexes, on en déduit que l'objectif est convexe. On peut donc utiliser les conditions de Karush-Kuhn-Tucker pour trouver le minimum.

La contrainte $\Sigma_j \in S_d^{++}$ ne pose pas de problème car $S_d^{++}$ est un cône convexe. Il en va de même pour la positivité des $\alpha_j$.

La contrainte $\sum_{j=1}^{p} \alpha_j = 1$ est une contrainte d'égalité. On peut la traiter en introduisant un multiplicateur de Lagrange $\lambda$ et en minimisant le lagrangien :

$$L(\alpha,\mu,\Sigma) = -\sum_{\substack{1\leq i\leq n \\ 1\leq j\leq p}} \left[\log \alpha_j + \log f_{\mu_j,\Sigma_j}(x_i)\right] Z_{ij}(\theta_t) + \lambda\left(\sum_{j=1}^{p} \alpha_j - 1\right)$$

On a [1] :

$$\nabla_{\alpha_j} L = -\sum_{i=1}^{n} \frac{Z_{ij}(\theta_t)}{\alpha_j} + \lambda = 0 \qquad (A)$$

$$\nabla_{\mu_j} L = \sum_{i=1}^{n} Z_{ij}(\theta_t)\Sigma_j^{-1}(x_i - \mu_j) = 0 \qquad (B)$$

$$\nabla_{\Sigma_j} L = \sum_{i=1}^{n} Z_{ij}(\theta_t)\left(\frac{1}{2}\Sigma_j^{-1} - \frac{1}{2}\Sigma_j^{-1}(x_i - \mu_j)(x_i - \mu_j)^T\Sigma_j^{-1}\right) = 0 \qquad (C)$$

On en déduit, en utilisant la contrainte d'égalité sur les $\alpha_j$ avec $(A)$ :

$$\lambda = \sum_{i=1}^{n}\sum_{j=1}^{p} Z_{ij}(\theta_t)$$

Par conséquent,

$$\alpha_j^{(t+1)} = \frac{\sum_{i=1}^{n} Z_{ij}(\theta_t)}{\sum_{i=1}^{n}\sum_{j'=1}^{p} Z_{ij'}(\theta_t)}$$

En multipliant $(B)$ par $\Sigma_j$ de part et d'autre, on obtient :

$$\mu_j^{(t+1)} = \frac{\sum_{i=1}^{n} Z_{ij}(\theta_t)x_i}{\sum_{i=1}^{n} Z_{ij}(\theta_t)}$$

Finalement dans $(C)$ on peut multiplier chaque côté de l'égalité par $\Sigma_j$ à gauche et à droite pour obtenir :

$$\frac{1}{2}\sum_{i=1}^{n} Z_{ij}(\theta_t)\left(\Sigma_j - (x_i - \mu_j)(x_i - \mu_j)^T\right) = 0$$

Ce qui nous permet finalement d'obtenir :

$$\Sigma_j^{(t+1)} = \frac{\sum_{i=1}^{n} Z_{ij}(\theta_t)(x_i - \mu_j^{(t+1)})(x_i - \mu_j^{(t+1)})^T}{\sum_{i=1}^{n} Z_{ij}(\theta_t)}$$

4. , 5 et 6. Sur le notebook.

## Exercice 3 Importance sampling

1. Sur le notebook.
2. Sur le notebook.
3. Sur le notebook.

1. Je recommande le formidable Matrix Cookbook pour calculer des dérivées matricielles : https ://www2.imm.dtu.dk/pubdb/edoc/imm3274.pdf

4. Dans l'étape (iii) de Population Monte Carlo, on veut maximiser (en reprenant les notations de l'exercice précédent) :

$$\sum_{i=1}^{n} \tilde{\omega}_i \log q_\theta(x_i) = \sum_{i=1}^{n} \tilde{\omega}_i \log \left[ \sum_{i=1}^{K} \alpha_i f_{\mu_i, \Sigma_i}(x_i) \right]$$

Or dans le cas de l'algorithme EM avec un mélange gaussien, on souhaitait maximiser la log-vraissemblance :

$$\sum_{i=1}^{n} \sum_{j=1}^{p} [\log \alpha_j + \log f_{\mu_j, \Sigma_j}(x_i)]$$

Et on a choisi de maximiser $Q(\theta|\theta_t)$ qui est une borne inférieure de la log-vraisemblance. On peut donc s'inspirer de cette démarche et maximiser :

$$\sum_{i=1}^{n} \sum_{j=1}^{p} \tilde{\omega}_i Z_{ij}(\theta_t) [\log \alpha_j + \log f_{\mu_j, \Sigma_j}(x_i)]$$

où $Z_{ij}(\theta_t)$ est défini comme dans l'exercice précédent. On voit tout de suite que les calculs de gradients sont les mêmes que dans l'exercice précédent aux constantes $\tilde{\omega}_i^{(0)}$ près. On obtient donc les formules :

$$\alpha_j^{(t+1)} = \frac{\sum_{i=1}^{n} \tilde{\omega}_i Z_{ij}(\theta_t)}{\sum_{i=1}^{n} \sum_{j'=1}^{p} \tilde{\omega}_i Z_{ij'}(\theta_t)}$$

$$\mu_j^{(t+1)} = \frac{\sum_{i=1}^{n} \tilde{\omega}_i Z_{ij}(\theta_t) x_i}{\sum_{i=1}^{n} \tilde{\omega}_i Z_{ij}(\theta_t)}$$

$$\Sigma_j^{(t+1)} = \frac{\sum_{i=1}^{n} \tilde{\omega}_i Z_{ij}(\theta_t)(x_i - \mu_j^{(t+1)})(x_i - \mu_j^{(t+1)})^T}{\sum_{i=1}^{n} \tilde{\omega}_i Z_{ij}(\theta_t)}$$

5. Sur le notebook.