TP2 : Expectation-Maximisation algorithm – Importance sampling

Exercice 1 : Discrete distributions

1. On cherche à générer une variable aléatoire X telle que

$$\mathbb{P}(X = x_i) = p_i \qquad i = 1, \dots n.$$

On peut générer une variable aléatoire $U \sim \mathcal{U}([0,1])$.

$$\mathbb{P}(X \leq k) = \sum_{i=0}^{k} p_i = \mathbb{P}\left(U \leq \sum_{i=0}^{k} p_i\right)$$

En prenant $F = \max k$ sous contrainte $U \leq \sum_{i=1}^{k} p_i$ et $k \leq n$

on a $F(U) \sim X$.

Et on a ainsi généré X.

2.3. Voir le notebook.

Exercice 2 : Gaussian mixture model and the EM algorithm

1. Les paramètres dénotés par θ sont $(\alpha_j, \mu_j, \Sigma_j)_{j \in [\![1, m]\!]}$.

On note f_j la densité associée à la loi $\mathcal{N}(\mu_j, \Sigma_j)$.

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

Par définition, la likelihood s'écrit :

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

formule des probabilités totales

$$= \prod_{i=1}^{n} \left(\sum_{j=1}^{m} \underbrace{f_\theta(x_i \mid Z_i = j)}_{f_j(x_i)} \underbrace{\mathbb{P}_\theta(Z_i = j)}_{= \alpha_j} \right)$$

$$\boxed{\mathcal{L}(x_1, \dots x_n; \theta) = \prod_{i=1}^{n} \sum_{j=1}^{m} \alpha_j f_j(x_i)}$$

3. Algorithme EM :

La quantité $Q(\theta, \theta^t)$ est un minimiseur de la log-likelihood.
Maximiser $Q(\theta,\theta^t)$ revient à maximiser la loglikelihood.

$$Q(\theta,\theta^t) = \sum_{i=1}^{n} \mathbb{E}_{Z_i \sim P_{\theta^t}(\cdot|x_i)} \left[\ln\left(\frac{P_{\theta^t}(X_i = x_i, Z_i)}{P_{\theta^t}(Z_i | X_i = x_i)} \right) \right]$$

Or, $$\mathbb{P}_{\theta^t}(X_i = x_i, Z_i) = \sum_{j=1}^{m} \mathbb{P}_{\theta^t}(X_i = x_i \; Z_i = j)$$

$$= \sum_{j=1}^{m} \mathbb{P}_{\theta^t}(X_i = x | Z_i = j)\, \mathbb{P}_{\theta^t}(Z_i = j)$$

$$= \sum_{j=1}^{m} f_{\mu_j, \Sigma_j}(x_i)\, \alpha_j .$$

et $$\mathbb{P}_{\theta^t}(Z_i = j | X_i = x_i) = \frac{\mathbb{P}(Z_i = j, X_i = x_i)}{\mathbb{P}(X_i = x_i)}$$ (formule de Bayes)

$$= \frac{\alpha_j f_{\mu_j \Sigma_j}(x_i)}{\sum_{l=1}^{m} \alpha_l f_{\mu_l, \Sigma_l}(x_i)}$$

$$\triangleq T_{i,j}$$ (voir Notebook)

Ainsi : $$Q(\theta,\theta^t) = \sum_{i=1}^{n}\sum_{j=1}^{m} T_{ij} \left(\log(f_{\mu_j, \Sigma_j}(x_i)) + \log(\alpha_j) - \log(T_{ij}) \right)$$

Etape E : On calcule $Q(\theta,\theta^t)$ en calculant $T_{i,j}$ $i = 1, \dots, n$
$j = 1, \dots, m$

On veut ensuite prendre $\theta^{t+1} \in \underset{\theta}{\operatorname{argmax}}\, Q(\theta, \theta^t)$.

ce qui revient à : $$\min_{\theta} - \sum_{i=1}^{n}\sum_{j=1}^{m} T_{ij} \left(\log(f_{\mu_j, \Sigma_j}(x)) + \log \alpha_j \right)$$

(car $T_{ij}(-\log T_{ij})$ ne dépend que de θ^t) sous contrainte :

$$\sum_{j=1}^{m} \alpha_j = 1.$$

On reconnait un problème d'optimisation, dont le Lagrangien s'écrit :

$$\mathcal{L}(\theta, \lambda) = -\sum_{i=1}^{n}\sum_{j=1}^{m} T_{ij}\left(\log \alpha_j + \log f_{\mu_j, \Sigma_j}(x_i)\right) + \lambda\left(\sum_{j=1}^{m} \alpha_j - 1\right)$$

On cherche à vérifier l'annulation du gradient de $\mathcal{L}$:

$$\nabla_{\alpha_j} \mathcal{L}(\theta, \lambda) = 0 \Rightarrow \alpha_j^{t+1} = \frac{1}{n} \sum_{i=1}^{n} T_{ij}(\theta^t)$$

$$\nabla_{\mu_j} \mathcal{L}(\theta, \lambda) = 0 \Rightarrow \mu_j^{t+1} = \frac{1}{n\alpha_j^{t+1}} \sum_{i=1}^{n} x_i T_{ij}(\theta^t)$$

$$\nabla_{\Sigma_j} \mathcal{L}(\theta, \lambda) = 0 \Rightarrow \Sigma_j^{t+1} = \frac{1}{n\alpha_j^{t+1}} \sum_{i=1}^{n} T_{ij}(\theta^t)(x_i - \mu_j^{t+1})(x_i - \mu_j^{t+1})^T$$

<u>Etape M :</u> Actualiser les valeurs α_j^{t+1}, μ_j^{t+1}, Σ_j^{t+1}.

<u>Exercice 3 - Importance sampling</u>

1-3. Voir le notebook.

4. A l'étape (iii) de l'algorithme Population Monte Carlo, on calcule :

$$\theta^{(l)} \in \underset{\theta}{\operatorname{argmax}} \sum_{i=1}^{n} \tilde{\omega}_i^{(l)} \log q_\theta(X_i^{(l)})$$

qui est une approximation de $\underset{\theta}{\operatorname{argmax}}\ \mathbb{E}_\nu\left[\log q_\theta(X_i^{(l)})\right]$

Comme dans l'exercice 2, on cherche à maximiser la log-likelihood $\log q_\theta$. On avait pour cela utilisé l'algorithme EM, et en gardant les mêmes notation :

$$Q(\theta, \theta^t) = \sum_{i=1}^{n} \tilde{\omega}_i^{(t)} \sum_{j=1}^{m} T_{ij}(\theta^t)\left(\log \alpha_j + \log f_{\mu_j, \Sigma_j}(x_i) - \log T_{ij}(\theta^t)\right)$$

De la même manière, par annulation du gradient du Lagrangien, on obtient les paramètres suivants :

$$\alpha_j^{t+1} = \frac{\sum_{i=1}^{n} \tilde{\omega}_i^{(t)} T_{ij}(\theta^t)}{\sum_{k=1}^{m} \sum_{i=1}^{n} \tilde{\omega}_i^{(t)} T_{ik}(\theta^t)} \qquad j = 1, \dots, m$$

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

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