①

Achari
Benada
Younef

TP 2: Expectation-Maximisation algorithm - Importance sampling.

Exercice 1, Discrete distributions,

$n \in \mathbb{N}^*$, $X = \{x_1, \dots, x_n\}$ n distincts real number

$$\forall i \in [\![1, n]\!] \quad p_i > 0, \qquad \sum_{i=1}^{n} p_i = 1.$$

1) $\forall i \in [\![1, n]\!] \quad P(X = x_i) = p_i$
with $(p_1, \dots, p_n) \in \mathbb{R}_+^n$ such that $\sum_{i=1}^{n} p_i = 1$

Algorithm
(i) Sample $x \sim \mathcal{U}([0,1])$
(ii) $\Rightarrow k \;/\; \sum_{i=1}^{k-1} p_i \leq x < \sum_{i=1}^{k} p_i$
(iii) Return $X(k)$.

2) 3) See Python Code.

2

Exercice 2, Gaussian mixture model and the EM Algorithm.

$$\begin{cases} P(Z_i = j) = \alpha_j \\ X_i \mid Z_i = j \sim N(\mu_j, \Sigma_j) \end{cases} \qquad \Theta = \{\alpha_i, \mu_i, \Sigma_i\}_{i=1,\dots,n}$$

1) $f(x, z \mid \theta) = f(x \mid z, \theta) \times f(z \mid \theta)$

$$= \prod_{i=1}^{n} \underbrace{f(x_i \mid z_i, \theta)}_{N(\mu_{z_i}, \Sigma_{z_i})} \times \underbrace{f(z_i \mid \theta)}_{\alpha_{z_i}}$$

$$= \prod_{i=1}^{n} \sum_{j=1}^{m} \underbrace{f(x_i \mid \{Z_i = j\}, \theta)}_{N(\mu_j, \Sigma_j)} \times \overset{P(Z_i=j)}{\alpha_j} \mathbb{1}_{\{Z_i = j\}} =$$

Pour le cas du mélange Gaussien :

$$\log f(x, z ; \theta) = \sum_{i=1}^{n} \sum_{j=1}^{m} \frac{1}{2} \log(2\pi |\Sigma_j|) \mathbb{1}_{\{Z_i = j\}}$$

$$- \sum_{i=1}^{n} \sum_{j=1}^{m} \frac{1}{2} (x_i - \mu_j) \Sigma_j^{-1} (x_i - \mu_j) \mathbb{1}_{\{Z_i = j\}}$$

$$+ \sum_{i=1}^{n} \sum_{j=1}^{M} \log\{\alpha_j\} \mathbb{1}_{\{Z_i = j\}}$$

3) The parameters of the EM algorithm are :

$$\Theta = \{\alpha_i, \{\mu_i, \Sigma_i\}\}_{i=1,\dots,M}$$

We note q such that

$$q(z) = p(z \mid x, \Theta)$$

EM Algorithm

- Initialization $\theta = \theta_0$.
- E M: ~~(1- Expectation:)~~

At step t, We note $q_{i,k}^{(t)} = P(z_k^{(i)} = 1)$ sample k belong to component (i).

1- Expectation step:

Update of q: $q_{i,k}^{(t+1)} \leftarrow \dfrac{\alpha_k^{t-1} N(x^{(i)}, \mu_k^{(t-1)}, \Sigma_k^{(t-1)})}{\sum_{j=1}^{m} \alpha_j^{t-1} N(x^{(i)}, \mu_j^{(t-1)}, \Sigma_j^{(t-1)})}$

2- Maximization Step.

Update of θ: $\mu_k^{(t+1)} = \dfrac{\sum_{i=1}^{m} x^{(i)} q_{i,k}^{(t)}}{\sum_i q_{i,k}^{(t)}}$, $\Sigma_k^{(t+1)} = \dfrac{\sum_{i=1}^{m} (x^{(i)} - \mu_k)(x^{(i)} - \mu_k)^T \times q_{i,k}^{(t)}}{\sum_{i=1}^{m} q_{i,k}^{(t)}}$

$$\alpha_k^{(t+1)} = \frac{\sum_i q_{i,k}^{(t)}}{\sum_{i,k'} q_{i,k'}^{(t)}}$$

Implementation: See in the Code.

5) Application The [redacted] data Crude Birth / Death Rate has a shape of a banana.
So it is not well adapted for the EM algorithm with Gaussian Mixtures.

4

Exercice 3: Importance Sampling:

3.A - Poor Importance Sampling

$$f(x) = 2\sin\left(\frac{\pi}{1.5}x\right)$$

$$p(x) = x^{(1.65)-1} e^{-\frac{x^2}{2}} \mathbb{1}_{\mathbb{R}^+}(x)$$

and $$q(x) = \frac{2}{\sqrt{2\pi(1.5)}} e^{-\frac{((0.8)-x)^2}{2\times(1.5)}}$$

3.B Adaptative Importance Sampling

$$Q: \quad q(x) = \sum_{i=1}^{M} \alpha_i \varphi(x; \mu_i, \Sigma_i)$$

4- The EM algorithm can be used to maximize the loglikelihood by finding the new parameters $\theta^{(t+1)} = (\alpha^{(t+1)}, \mu^{(t+1)}, \Sigma^{(t+1)})$ from $\theta^{(t)} = (\alpha^{(t)}, \mu^{(t)}, \Sigma^{(t)})$. What changes here is the weights $(\tilde{w}_i)$.

We can consider the given expression as the Expectation of a Gaussian Mixture model under the proba $q^{(t)}$ and $\alpha_j^{(t)} = E_{q^{(t)}}\left[\mathbb{1}_{\{z=j\}}\right]$.

So we can maximize by doing the following updates:

For j in $1..M$:

$$\mu_j^{(t+1)} \longleftarrow \frac{\sum_{i=1}^{n} \tilde{w}_i \, x^{(i)} q_{i,j}^{(t)}}{\sum_{i=1}^{n} \tilde{w}_i \, q_{i,j}^{(t)}}$$

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

$$\alpha_j^{(t+1)} \longleftarrow \frac{\sum_{i=1}^{n} \tilde{w}_i \, q_{i,j}^{(t)}}{\sum_{i=1}^{n} \sum_{j=1}^{M} \tilde{w}_i \, q_{i,j}^{(t)}}$$