Solutions: Assignment 4

31 March 2024 20:09

Mainak Biswas.
PhD student, IISc

1) Coding: Find the implementation in the jupyter notebook.

Theory

2) (Ignore $r_{ij} \geq 0$ constraint)

(i) $$\mathcal{L} = \sum_{i=1}^{N} \sum_{j=1}^{K} r_{ij}^2 \, \|x_i - u_j\|^2 + \sum_{i=1}^{N} \lambda_i \left[1 - \sum_{j=1}^{K} r_{ij} \right]$$

$$\frac{\partial L}{\partial u} = 0$$

(ii) $$\Rightarrow \frac{\partial L}{\partial u} = -2 \sum_{i=1}^{n} r_{ij}^2 (x_i - u_j) = 0$$

$$\Rightarrow u_j = \frac{\sum_{i=1}^{N} r_{ij}^2 x_i}{\sum_{i=1}^{N} r_{ij}^2}$$

$$\frac{\partial L}{\partial r_{ij}} = 2 r_{ij} \|x_i - u_j\|_2^2 - \lambda_i = 0$$

$$\lambda_i = 2 r_{ij} \|x_i - u_j\|^2$$

$$\therefore \sum_{j=1}^{K} r_{ij} = \sum_{j=1}^{K} \frac{\lambda_i}{2\|x_i - u_j\|^2} = 1$$

$$= \lambda_i = \frac{1}{\sum_{j=1}^{K} \frac{1}{2\|x_i - u_j\|^2}}$$

$$= \lambda_i = \sum_{j=1}^{k} \frac{1}{2\|x_i - u_j\|^2}$$

$$\therefore \gamma_{ij} = \frac{\frac{1}{\|x_i - u_j\|^2}}{\sum_{j=1}^{k} \frac{1}{\|x_i - u_j\|^2}} > 0 \quad \text{(positivity constraint satisfied)}$$

(iii) Set; $\gamma_{ij}^{(0)} = \frac{1}{K}$ and $u^{(0)}$ randomly.

Iterate until convergence:

$$\begin{cases} u_j^{\langle t+1 \rangle} = \dfrac{\sum_{i=1}^{N} \gamma_{ij}^{2 \langle t \rangle} x_i}{\sum_{i=1}^{N} \gamma_{ij}^{2 \langle t \rangle}} \\[2ex] \gamma_{ij}^{\langle t+1 \rangle} = \dfrac{\left(\frac{1}{\|x_i - u_j^{\langle t+1 \rangle}\|_2^2} \right)}{\sum_{j=1}^{K} \frac{1}{\|x_i - u_j^{\langle t+1 \rangle}\|_2^2}} \end{cases}$$

(d) At every step take the argmax $\gamma_{ij}^{\langle t \rangle} = \begin{cases} 1 & \text{if } \gamma_{ij} > \gamma_{ik} \ \forall k \neq j \\ 0 & \text{otherwise.} \end{cases}$

This would make the above algorithm equivalent to K means:-

3)

$$\log P_\theta(x) = \log \sum_{i=1}^{n} P_\theta(x \mid Z = z_i) \underbrace{P_\theta(Z = z_i)}$$

$$= \log \sum_{i=1}^{n} \lambda_i e^{-\lambda_i x} \cdot u_i$$

Consider the ELBO:

$$\log P_\theta(x) \geq F_\theta(q)$$

$$= \log P_\theta(x) - \mathbb{D}_{KL}\left[q(z|x) \,\|\, P_\theta(z|x)\right]$$

(i) Expectation step : (E-step)

$$q^{t+1}(z_j|x_i) = q_{ji}^{(t+1)} \propto P_\theta(x_i|z_j)\, P_\theta(z_j)$$

$$\propto \lambda_j e^{-\lambda_j x_i} \cdot u_j$$

we know,

$$\sum_{j=1}^{K} q_{ji}^{(t+1)} = 1 \quad \forall i$$

$$\therefore q_{ji}^{(t+1)} = \frac{u_j \lambda_j e^{-\lambda_j x_i}}{\sum_{j=1}^{K} u_j \lambda_j e^{-\lambda_j x_i}}$$

(ii) Maximization step:

$$\theta^{(t+1)} = \arg\max_\theta F_\theta(q^{(t+1)}) \quad \text{s.t}$$

$$= \mathbb{E}_{x_i \sim P_x} \mathbb{E}_{z_j \sim q(z|x_i)} \log \frac{P_\theta(x|z) \cdot P_\theta(z)}{\underbrace{q^{t+1}(z|x)}_{(\text{const})}}$$

$$= \frac{1}{N} \sum_{i=1}^{N} \sum_{j=1}^{K} q_{ji}^{(t+1)} \left[\log P_{\theta=\{u_j, \lambda_j\}}(x_i|z_j) + \log P_\theta(z_j)\right]$$

$$\text{ST} \quad \sum_{j=1}^{K} u_j = 1$$

$$\mathcal{L}(\lambda, u) = \frac{1}{N} \sum_{i=1}^{N} \sum_{j=1}^{K} q_{ji}^{(t+1)} \left[\log \lambda_j e^{-\lambda_j x_i} + \log u_j \right] + \lambda \left[1 - \sum_{j=1}^{K} u_j \right]$$

$$\frac{\partial L}{\partial u_j} = \frac{1}{N} \sum_{i=1}^{N} q_{ji}^{(t+1)} / u_j - \lambda = 0$$

$$\Rightarrow \sum_{j=1}^{k} (\lambda u_j) = \sum_{j=1}^{k} \frac{1}{N} \sum_{i=1}^{N} q_{ji}^{(t+1)}$$

$$\text{using } \sum_j u_j = 1$$

$$\Rightarrow \lambda = \frac{1}{N} \sum_{i=1}^{N} \underbrace{\sum_{j=1}^{k} q_{ji}^{(t+1)}}_{1}$$

$$= 1$$

$$\therefore \boxed{u_j^{(t+1)} = \frac{1}{N} \sum_{i=1}^{N} q_{ji}^{(t+1)}} \quad \forall j$$

$$\frac{\partial L}{\partial \lambda_j} = \frac{1}{N} \sum_{i=1}^{N} q_{ji}^{(t+1)} \left[\frac{\partial}{\partial \lambda_j} \log \lambda_j - \frac{\partial}{\partial \lambda_j} \lambda_j x_i \right] = 0$$

$$\Rightarrow \sum_{i=1}^{N} q_{ji}^{(t+1)} \left[\frac{1}{\lambda_j} - x_i \right] = 0$$

$$\Rightarrow \sum_{i=1}^{N} q_{ji}^{(t+1)} x_i = \frac{1}{\lambda_j} \sum_{i=1}^{N} q_{ji}^{(t+1)}$$

$$\lambda_j^{(t+1)} \quad \sum^{N} q_{ji}^{(t+1)}$$

$$\Rightarrow \quad \lambda_j^{(t+1)} = \frac{\sum_{i=1}^{N} q_{ji}^{(t+1)}}{\sum_{i=1}^{N} q_{ji}^{(t+1)} x_i} \longrightarrow$$

$$\Rightarrow \quad \boxed{\lambda_j^{(t+1)} = \frac{N\,\mu_j^{(t+1)}}{\sum_{i=1}^{N} q_{ji}^{(t+1)} x_i}}$$