Homework #5

Name: Changxin Miao
ID: 11853018

Q1

(a) ~~log~~ likelihood of the ~~Bernoulli function~~ Gaussian mixture model

$P(X, Z \mid \mu, \Sigma, \pi) = \prod_{n=1}^{N} \prod_{k=1}^{K} \pi_k^{z_{nk}} \mathcal{N}(x_n \mid \mu_k, \Sigma_k)^{z_{nk}}$

where $\pi_k = \frac{1}{N} \sum_{n=1}^{N} z_{nk}$

Posterior distribution: $p(Z \mid X, \mu, \Sigma, \pi) \propto \prod_{n=1}^{N} \prod_{k=1}^{K} [\pi_k \mathcal{N}(x_n \mid \mu_k, \Sigma_k)]^{z_{nk}}$

log: $\sum_{n=1}^{N} \sum_{k=1}^{K} \gamma(z_{nk}) \{\ln \pi_k + \ln \mathcal{N}(x_n \mid \mu_k, \Sigma_k)\}$

$$\gamma(z_{nk}) = \frac{\pi_k \mathcal{N}(x \mid \mu_k, \Sigma_k)}{\sum_{j=1}^{K} \pi_j \mathcal{N}(x \mid \mu_j, \Sigma_j)} \quad \text{(Bishop 9.13)}$$

$$-\sum_{n=1}^{N} \frac{\pi_k \mathcal{N}(x \mid \mu_k, \Sigma_k)}{\sum_{j=1}^{K} \pi_j \mathcal{N}(x \mid \mu_j, \Sigma_j)} \Sigma_k (x_n - \mu_k) = 0$$

$-\gamma(z_{nk}) \Sigma_k (x_n - \mu_k) = 0$ fixed $\gamma(z_{nk})$, update μ_k, π_k and Σ_k

$$\mu_k^{new} = \frac{1}{N_k} \sum_{n=1}^{N} \gamma(z_{nk}) x_n \qquad N_k = \sum_{n=1}^{N} \gamma(z_{nk})$$

$$\Sigma_k^{new} = \frac{1}{N_k} \sum_{n=1}^{N} \gamma(z_{nk}) (x_n - \mu_k)(x_n - \mu_k)^T$$

$\cancel{\pi_k^{(new)} =}$

$$\cancel{\frac{\partial L\ \text{posterior}}{\partial \pi_k} = \frac{\sum_{n=1}^{N} \gamma(z_{nk})}{\pi_k}}$$

When maximizing π_k, we need ~~no~~ to take in the constraint $\sum_{k=1}^{K} \pi_k = 1$

$$\hat{L} = \ln p(Z \mid X, \mu, \Sigma, \pi) + \lambda\left(\sum_{k=1}^{K} \pi_k - 1\right)$$

$$\frac{\partial \hat{L}}{\partial \pi_k} = \sum_{n=1}^{N} \frac{\mathcal{N}(x_n \mid \mu_k, \Sigma_k)}{\sum_j \pi_j \mathcal{N}(x_n \mid \mu_j, \Sigma_j)} + \lambda = 0 \quad \text{multiply both sides by } \sum_{k=1}^{K} \pi_k$$

$$\sum_{n=1}^{N} 1 + \sum_{k=1}^{K} \pi_k \lambda = 0 \qquad \lambda = -N$$

$$\pi_k^{new} = \frac{N_k}{N}$$

由 扫描全能王 扫描创建

(b) If we change the Σ_k into Σ.

$$\gamma(z_{nk})' = \frac{\pi_k \mathcal{N}(x|\mu_k, \Sigma)}{\sum_{j=1}^{k} \pi_j \mathcal{N}(x|\mu_j, \Sigma)} \qquad N_k' = \sum_{n=1}^{N} \gamma(z_{nk})'$$

similarly, for μ_k, π_k, the update rule will be similair to the previous question.

$$\mu_k^{new} = \frac{1}{N_k} \sum_{n=1}^{N} \gamma(z_{nk})' x_n$$

$$\pi_k^{new} = \frac{N_k'}{N}$$

Q2 We want to maximize the posterior distribution $p(\theta|X)$

$$p(\theta|X) = p(\theta, X) / p(X)$$

$$\ln p(\theta|X) = \ln p(\theta, X) - \ln p(X)$$

where $\ln p(\theta|X) = \mathcal{L}(q,\theta) + KL(q\|p) + \ln p(\theta) - \ln p(X)$ (Bishop 9.70)

$$\geq \mathcal{L}(q,\theta) + \cancel{\ln p(\theta)} \ln p(\theta) - \ln p(X) \qquad (1.1)$$

$$\mathcal{L}(q,\theta) = \sum_Z q(Z) \ln\left\{\frac{p(X, Z|\theta)}{q(Z)}\right\}$$ remains the same

The E-step maximize $\mathcal{L}(q,\theta)$ w.r.t q for the fixed θ

$$\mathcal{L}(q, \theta^{old}) = \sum_Z q(Z) \ln(p, Z|\theta)$$

The largest $\mathcal{L}$ will appear when $KL = 0$

$$q(Z) = p(Z|X, \theta^{old})$$

$$\mathcal{L}(q, \theta^{old}) = \sum_Z p(Z|X, \theta^{old}) \ln(p, Z|\theta)$$

For the M-step, Refer to the equation 1.1, the formaula becomes

$$\mathcal{L}(q,\theta) = \sum_Z p(Z|X, \theta^{old}) \ln(p, Z|\theta) + \ln p(\theta) - \ln p(X)$$

We now fixed q and try to maximize θ, as $\ln p(X)$ is a constant which is not related to θ, the quantity to be maximized become:

$$\sum p(Z|X, \theta^{old}) \ln(p, Z|\theta) + \ln p(\theta)$$

由 扫描全能王 扫描创建

Q3. After the E-step, we try to ~~fixed~~ fix z_n and maximize the log-likelihood w.r.t α, a_k, b_k.

Thus, the posterior distribution will be

$$p(\mu \mid \{x_n\}_{n=1}^{N}) = p(\pi/\alpha)\prod_{k=1}^{K} p(\mu_k/\alpha_k, b_k)\prod_{n=1}^{N} p(z_{nk}/\pi_k)\,p(x_n/\mu_k)^{z_{nk}}$$

where $p(z_{nk}/\pi_k) = \pi_k^{z_{nk}}$

Log-likelihood become >

$$\hat{L} = \ln p(\mu \mid \{x_n\}_{n=1}^{N}) = \ln p(\pi/\alpha) + \sum_{k=1}^{K}[\ln p(\mu_k/\alpha_k, \beta_k)] + \sum_{k=1}^{K}\sum_{n=1}^{N}(\ln p(\underbrace{z_{nk}}_{\pi_k^{z_{nk}}}/\pi_k) + z_{nk}\ln p(x_n/\mu_k))$$

$$= \ln p(\pi/\alpha) + \sum_{k=1}^{K}[\ln p(\mu_k/a_k, b_k)] + \sum_{k=1}^{K}\sum_{n=1}^{N} z_{nk}[\ln \pi_k + \ln p(x_n/\mu_k)]$$

$$= \ln p(\pi/\alpha) + \sum_{k=1}^{K}[\ln p(\mu_k/a_k, b_k)] + \sum_{k=1}^{K}\sum_{n=1}^{N} \gamma(z_{nk})[\ln \pi_k + \ln p(x_n/\mu_k)]$$

$$\gamma(z_{nk}) = \frac{\pi_k p(x_n/\mu_k)}{\sum_{j=1}^{K}\pi_j p(x_n/\mu_j)}$$

Further we replace ~~all~~ $p(\pi/\alpha)$, $p(\mu_k/\alpha_k, \beta_k)$ and $p(x_n/\mu_k)$ with specific distribution. Then take the corresponding logarithm derivative to each parameters.

$$p(\pi/\alpha) \propto \prod_{k=1}^{K}\pi_k^{\alpha_k-1} \qquad p(\mu_k/a_k, b_k) \propto \prod_{i=1}^{D}\mu_{ki}^{a_k-1}(1-\mu_{ki})^{b_k-1} \qquad p(x_n/\mu_k) = \prod_{i=1}^{D}\mu_{ki}^{x_{ni}}(1-\mu_{ki})^{1-x_{ni}}$$

$$\hat{L}' = \hat{L} + \ln\sum_{k=1}^{K}\pi_k^{\alpha_k-1} + \sum_{i=1}^{D}(a_k-1)\ln\mu_{ki} + \sum_{i=1}^{D}(b_k-1)\ln(1-\mu_{ki}) + \sum_{i=1}^{D}x_{ni}\,\mu_{ki} + \sum_{i=1}^{D}(1-x_{ni})(1-\mu_{ki})$$

$$\frac{\partial \hat{L}'}{\partial \mu_{ki}} = \frac{a_k-1}{\mu_{ki}} + (-1)\frac{b_k-1}{(1-\mu_{ki})} + \sum_{n=1}^{N}\gamma(z_{nk})\left[\frac{x_{ni}}{\mu_{ki}} - \frac{1-x_{ni}}{1-\mu_{ki}}\right] = 0$$

$$= \frac{(a_k-1)(1-\mu_{ki}) - \mu_{ki}(b_k-1)}{\mu_{ki}(1-\mu_{ki})} + \sum_{n=1}^{N}\gamma(z_{nk})\left[\frac{x_{ni}-\mu_{ki}}{\mu_{ki}(1-\mu_{ki})}\right]$$

$$= \frac{a_k-1-(a_k+b_k)\mu_{ki}+2\mu_{ki}}{\mu_{ki}(1-\mu_{ki})} + \sum_{n=1}^{N}\gamma(z_{nk})\left[\frac{x_{ni}-\mu_{ki}}{\mu_{ki}(1-\mu_{ki})}\right]$$

$$\mu_{ki} = \frac{a_k-1+N_k}{2-a_k-b_k-N_k}$$

To calculate π_k, we need to incorporate the constraint $\sum_{k=1}^{K}\pi_k = 1$.

$$\hat{L}'(\ldots\lambda) = \hat{L}' + \lambda\left(\sum_{k=1}^{K}\pi_k - 1\right)$$

$$\frac{\partial \hat{L}'(\ldots\lambda)}{\partial \pi_k} = \frac{a_k-1}{\pi_k} + \frac{1}{\pi_k}\sum_{n=1}^{N}\gamma(z_{nk}) + \lambda \qquad N_k = \sum_{n=1}^{N}\gamma(z_{nk})$$

$$= \frac{a_k-1}{\pi_k} + \frac{N_k}{\pi_k} + \lambda = 0 \qquad \lambda\pi_k = -\alpha_k + 1 - N_k$$

由 扫描全能王 扫描创建

$$\lambda \sum_{k=1}^{K} \pi_k = \sum_{k=1}^{K} (-\alpha_k + 1 - N_k)$$

$$\lambda = K - N - \sum_{k=1}^{K} \alpha_k$$

$$\pi_k = \frac{\alpha_k - 1 + N_k}{N - K + \sum_{k=1}^{K} \alpha_k}$$

由 扫描全能王 扫描创建