AI701 HW2 - 20194293 고경렬

(a) $$p(Z|X,\theta) = \frac{p(X,Z,\theta)}{p(X,\theta)} = \frac{p(X,Z|\theta)}{p(X|\theta)} = \prod_{i=1}^{n} \frac{\prod_{k=1}^{K}\left(\pi_k \mathcal{N}(x_i|\mu_k,\Sigma_k)\right)^{\mathbb{1}(z_i=k)}}{\sum_{k=1}^{K}\pi_k \mathcal{N}(x_i|\mu_k,\Sigma_k)}$$

$$p(Z\setminus\{z_i\}|X,\theta) = p(Z\setminus\{z_i\}|X\setminus\{x_i\},\theta) = \prod_{j\in\{1,\cdots,n\}\setminus\{i\}} \frac{\prod_{k=1}^{K}\left(\pi_k \mathcal{N}(x_i|\mu_k,\Sigma_k)\right)^{\mathbb{1}(z_j=k)}}{\sum_{k=1}^{K}\pi_k \mathcal{N}(x_i|\mu_k,\Sigma_k)}$$

$$\therefore\ p(z_i|Z\setminus\{z_i\},X,\theta) = \frac{p(Z|X,\theta)}{p(Z\setminus\{z_i\}|X,\theta)} = \frac{\prod_{k=1}^{K}\left(\pi_k \mathcal{N}(x_i|\mu_k,\Sigma_k)\right)^{\mathbb{1}(z_i=k)}}{\sum_{k=1}^{K}\pi_k \mathcal{N}(x_i|\mu_k,\Sigma_k)}$$

$$= \mathrm{Cat}\left(z_i \,\middle|\, \frac{\pi_1 \mathcal{N}(x_i|\mu_1,\Sigma_1)}{\sum_{k=1}^{K}\pi_k \mathcal{N}(x_i|\mu_k,\Sigma_k)}, \cdots, \frac{\pi_K \mathcal{N}(x_i|\mu_K,\Sigma_K)}{\sum_{k=1}^{K}\pi_k \mathcal{N}(x_i|\mu_k,\Sigma_k)}\right)$$

(b) $$p(\pi|X,Z,\theta\setminus\{\pi\}) \propto p(\pi)\cdot p(Z|X,\theta)$$

$$\propto \prod_{i=1}^{n}\prod_{k=1}^{K} a_{i,k}^{\mathbb{1}(z_i=k)} \quad \text{where } a_{i,k} = \frac{\pi_k \mathcal{N}(x_i|\mu_k,\Sigma_k)}{\sum_{k=1}^{K}\pi_k \mathcal{N}(x_i|\mu_k,\Sigma_k)}$$

$$\propto \prod_{k=1}^{K}\left(\prod_{i\in S_k} a_{i,k}\right) \quad \text{where } S_k = \{i \mid z_i = k \text{ for } i=1,\cdots,n\}$$

$$\propto \prod_{k=1}^{K}(\pi_k)^{1+|S_k|-1}$$

$$= \mathrm{Dirichlet}(\pi \mid 1+|S_1|,\cdots,1+|S_K|)$$

(c) $p(\mu_k', \lambda_k', v_k' \mid X, \pi, \mu, \lambda, v) \propto p(X \mid \pi, \{\mu, \lambda, v\}_{-k}, \mu_k', \lambda_k', v_k')\, p(\mu_k', \lambda_k', v_k')$

$q(\mu_k', \lambda_k', v_k' \mid \mu_k, \lambda_k, v_k) = \mathcal{N}(\mu_k' \mid \mu_k, \sigma_q^2 \cdot I_d) \cdot \log\mathcal{N}(\lambda_k' \mid \log \lambda_k, \sigma_q^2) \cdot \mathcal{N}(v_k' \mid v_k, \sigma_q^2 I)$

$$\Rightarrow \frac{p(\mu_k', \lambda_k', v_k' \mid X, \pi, \mu, \lambda, v)\, q(\mu_k, \lambda_k, v_k \mid \mu_k', \lambda_k', v_k')}{p(\mu_k, \lambda_k, v_k \mid X, \pi, \mu, \lambda, v)\, q(\mu_k', \lambda_k', v_k' \mid \mu_k, \lambda_k, v_k)}$$

$$= \overset{①}{\boxed{\frac{p(X \mid \pi, \{\mu, \lambda, v\}_{-k}, \mu_k', \lambda_k', v_k')}{p(X \mid \pi, \{\mu, \lambda, v\}_{-k}, \mu_k, \lambda_k, v_k)}}}\ \overset{②}{\boxed{\frac{p(\mu_k', \lambda_k', v_k')}{p(\mu_k, \lambda_k, v_k)}}}\ \overset{③}{\boxed{\frac{q(\mu_k, \lambda_k, v_k \mid \mu_k', \lambda_k', v_k')}{q(\mu_k', \lambda_k', v_k' \mid \mu_k, \lambda_k, v_k)}}}$$

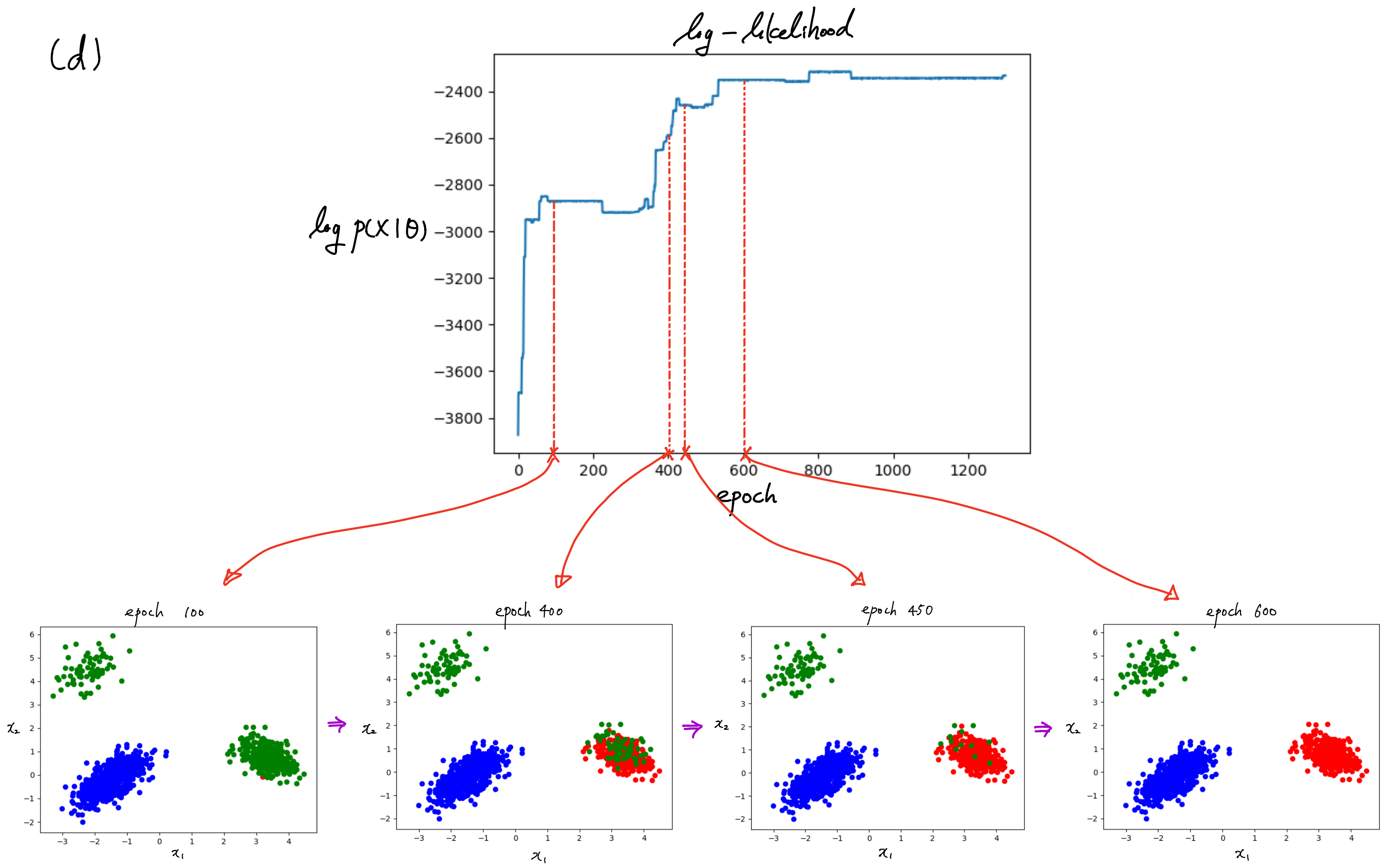
$$① = \prod_{i=1}^{n} \frac{\pi_k \mathcal{N}(x_i \mid \mu_k', \Sigma_k') - \pi_k \mathcal{N}(x_i \mid \mu_k, \Sigma_k) + \sum_{k=1}^{K} \pi_k \mathcal{N}(x_i \mid \mu_k, \Sigma_k)}{\sum_{k=1}^{K} \pi_k \mathcal{N}(x_i \mid \mu_k, \Sigma_k)}$$

$$② = \frac{\exp\left(-\frac{1}{10} \mu_k'^T \mu_k'\right) \cdot \cancel{\frac{1}{\lambda_k'}} \cdot \exp\left(-50 (\log \lambda_k' - 0.1)^2\right) \cdot \exp\left(-2 v_k'^T v_k'\right)}{\exp\left(-\frac{1}{10} \mu_k^T \mu_k\right) \cdot \cancel{\frac{1}{\lambda_k}} \cdot \exp\left(-50 (\log \lambda_k - 0.1)^2\right) \cdot \exp\left(-2 v_k^T v_k\right)}$$

$$③ = \frac{\cancel{\exp\left(-\frac{1}{2\sigma_q^2} (\mu_k - \mu_k')^T (\mu_k - \mu_k')\right)} \cdot \cancel{\frac{1}{\lambda_k}} \cdot \cancel{\exp\left(-\frac{1}{2\sigma_q^2} (\log \lambda_k - \log \lambda_k')^2\right)}}{\cancel{\exp\left(-\frac{1}{2\sigma_q^2} (\mu_k' - \mu_k)^T (\mu_k' - \mu_k)\right)} \cdot \cancel{\frac{1}{\lambda_k'}} \cdot \cancel{\exp\left(-\frac{1}{2\sigma_q^2} (\log \lambda_k' - \log \lambda_k)^2\right)}}$$
$$\cdot \frac{\cancel{\exp\left(-\frac{1}{2\sigma_q^2} (v_k - v_k')^T (v_k - v_k')\right)}}{\cancel{\exp\left(-\frac{1}{2\sigma_q^2} (v_k' - v_k)^T (v_k' - v_k)\right)}}$$

$$\therefore \text{Accept prob} = \min\left\{1,\ \prod_{i=1}^{n} \frac{\pi_k \mathcal{N}(x_i \mid \mu_k', \Sigma_k') - \pi_k \mathcal{N}(x_i \mid \mu_k, \Sigma_k) + \sum_{k=1}^{K} \pi_k \mathcal{N}(x_i \mid \mu_k, \Sigma_k)}{\sum_{k=1}^{K} \pi_k \mathcal{N}(x_i \mid \mu_k, \Sigma_k)} \right.$$
$$\left. \cdot \frac{\exp\left(-\frac{1}{10} \mu_k'^T \mu_k' - 50 (\log \lambda_k' - 0.1)^2 - 2 v_k'^T v_k'\right)}{\exp\left(-\frac{1}{10} \mu_k^T \mu_k - 50 (\log \lambda_k - 0.1)^2 - 2 v_k^T v_k\right)} \right\}$$

(d)



① In upper figure, the trace of log-likelihood ($\log p(X|\theta)$) is visualized.

This indicates that the parameters (π, μ, λ, ν) are trained to better explain the data assuming the proposed gaussian mixture model.

② In lower figures, the trace of the cluster assignment of each data are visualized.

(Different color indicates different cluster)

According to the upper plot, the epochs when the rapid increase of log-likelihood occurred were chosen.

It can be observed that the nearby data are grouped together.

③ The converged parameter values are summarized as follows

$$\pi : \begin{bmatrix} 0.3331688 \\ 0.0640807 \\ 0.60275049 \end{bmatrix} \quad \mu : \begin{bmatrix} 3.27665242, & 0.78549817 \\ -2.34010469, & 4.38982543 \\ -1.35254286, & -0.19921369 \end{bmatrix} \quad \lambda : \begin{bmatrix} 0.25735937 \\ 0.64322604 \\ 0.24108663 \end{bmatrix} \quad \nu : \begin{bmatrix} 0.1787624 & 0.19310355 \\ 0.17651964 & 0.08059265 \\ 0.47158767 & 0.36054678 \end{bmatrix}$$

$\Longrightarrow$ first row : red cluster second row : green cluster third row : blue cluster

(the value of z isn't written here due to space limit.)

(its value is included in "estimated_param.pickle" attached in submisson)

(e) $$p(X, Z, \phi) = \int p(X, Z, \theta)\, d\pi$$
$$= \int p(X, Z, \phi) \cdot p(\pi \mid X, Z, \phi)\, d\pi$$

$$\Rightarrow p(X, Z, \phi) = \frac{p(X, Z, \theta)}{p(\pi \mid X, Z, \phi)}$$

$$= \frac{p(\theta) \cdot \prod_{i=1}^{n} \prod_{k=1}^{K} \left(\pi_k \mathcal{N}(x_i \mid \mu_k, \Sigma_k)\right)^{\mathbb{1}(z_i = k)}}{\text{Dirichlet}\left(\pi \mid 1+|S_1|, \cdots, 1+|S_K|\right)}$$

$$\propto p(\theta) \cdot \prod_{k=1}^{K} |S_k|! \prod_{i=1}^{n} \prod_{k=1}^{K} \left(\mathcal{N}(x_i \mid \mu_k, \Sigma_k)\right)^{\mathbb{1}(z_i = k)}$$

(unnormalized form of $p(\phi, Z \mid X)$) $\Rightarrow$ use slice sampling