AI701 Homework 2

Sihwan Park (20204345)

Graduate School of AI, KAIST

November 3, 2020

Problem 1. *(a)*

$$\begin{aligned} P(z_i|Z \setminus \{z_i\}, X, \theta) &= P(z_i|x_i, \theta) \text{ since } z_i \text{ is dependent only to } x_i \\ &= \frac{P(z_i, x_i, \theta)}{P(x_i, \theta)} = \frac{P(z_i, \theta)P(x_i|z_i, \theta)}{\sum_{k=1}^{K} P(z_i = k, \theta)P(x_i|z_i = k, \theta)} \end{aligned} \tag{1}$$

Thus, we obtain

$$\begin{aligned} P(z_i = j|Z \setminus \{z_i\}, X, \theta) &= \frac{P(z_i = j, \theta)P(x_i|z_i = j, \theta)}{\sum_{k=1}^{K} P(z_i = k, \theta)P(x_i|z_i = k, \theta)} \\ &= \frac{\pi_j \mathcal{N}(x_i|\mu_j, \lambda_j I_d + v_j v_j^T)}{\sum_{k=1}^{K} \pi_k \mathcal{N}(x_i|\mu_k, \lambda_k I_d + v_k v_k^T)} \end{aligned} \tag{2}$$

Since we can easily obtain the conditional distributions, We can sample Z using Gibbs sampling.

(b) We can observe that

$$\begin{aligned} P(\pi|X, Z, \theta \setminus \{\pi\}) &= P(\pi|Z) \text{ by the conditional independency} \\ &\propto P(\pi)P(Z|\pi) \text{ by Bayes' rule} \\ &\propto (\pi_1^{\alpha_1 - 1} \cdots \pi_K^{\alpha_K - 1}) \times (\pi_1^{\alpha'_1} \cdots \pi_K^{\alpha'_K}) \\ &= \pi_1^{\alpha_1 + \alpha'_1 - 1} \cdots \pi_K^{\alpha_K + \alpha'_K - 1} \sim Dir(\alpha + \alpha') \end{aligned} \tag{3}$$

where $\alpha = (\alpha_1, \cdots, \alpha_K) = (1, 1, \cdots, 1), \alpha' = (\alpha'_1, \cdots, \alpha'_K), \alpha'_i =$ number of z_j's with $z_j = i$.

The third relation holds since $P(Z|\pi)$ is a categorical distribution with probability $(\pi_1, \cdots, \pi_K)$. (π_i will be multiplied α_i' times) And the last relation holds since the posterior of categorical likelihood with Dirichlet prior is also a Dirichlet distribution.

We can sample π by sampling the resulting Dirichlet distribution.

(c) Let $\phi = \{\mu_k, \lambda_k, v_k\}_{k=1}^K$.

The acceptance probability is

$$A(\phi'|\phi) = min\left\{1, \frac{P(\phi'|X,Z,\pi)q(\phi|\phi')}{P(\phi|X,Z,\pi)q(\phi'|\phi)}\right\} \tag{4}$$

Then let's find the $P(\phi|X,Z,\pi)$. We can observe that

$$\begin{aligned} P(\phi|X,Z,\pi) &= \frac{P(X,Z,\phi,\pi)}{P(X,Z,\pi)} \\ &= \frac{P(X|Z,\phi)P(Z|\pi)P(\pi)P(\phi)}{P(X,Z,\pi)} \end{aligned} \tag{5}$$

Then, we obtain

$$\begin{aligned} A(\phi'|\phi) &= min\left\{1, \frac{P(\phi'|X,Z,\pi)q(\phi|\phi')}{P(\phi|X,Z,\pi)q(\phi'|\phi)}\right\} \\ &= min\left\{1, \frac{\frac{P(X|Z,\phi')P(Z|\pi)P(\pi)P(\phi')}{P(X,Z,\pi)}q(\phi|\phi')}{\frac{P(X|Z,\phi)P(Z|\pi)P(\pi)P(\phi)}{P(X,Z,\pi)}q(\phi'|\phi)}\right\} \\ &= min\left\{1, \frac{P(X|Z,\phi')P(\phi')q(\phi|\phi')}{P(X|Z,\phi)P(\phi)q(\phi'|\phi)}\right\} \end{aligned} \tag{6}$$

Thus we can compute the acceptance probability since we have all the ingredients

$$P(X|Z,\phi) = \prod_{i=1}^{n}\prod_{k=1}^{K}\mathcal{N}(x_i|\mu_k, \lambda_k I_d + v_k v_k^T)^{\mathbf{1}_{\{z_i=k\}}} \tag{7}$$

$$P(\phi) = P(\{\mu_k,\lambda_k,v_k\}_{k=1}^K) = \prod_{k=1}^{K}\mathcal{N}\left(\mu_k; \mathbb{O}_d, 5.0\cdot I_d\right)\cdot \log\mathcal{N}\left(\lambda_k; 0.1, 0.1\right)\cdot\mathcal{N}\left(v_k; 0_d, 0.25\cdot I_d\right) \tag{8}$$

$$q(\phi'|\phi) = q(\{\mu_k',\lambda_k',v_k'\}_{k=1}^K|\{\mu_k,\lambda_k,v_k\}_{k=1}^K) = \prod_{k=1}^{K}\mathcal{N}\left(\mu_k'; \mu_k, \sigma_q^2 I_d\right)\log\mathcal{N}\left(\lambda_k'; \log\lambda_k, \sigma_q^2\right)\mathcal{N}\left(v_k'; v_k, \sigma_q^2 I\right) \tag{9}$$

(d) Before introduce the code implementation, we introduce our algorithm. We use the Gibbs sampling to sample Z, use Dirichlet distribution to sample π and use the Metropolis-Hastings algorithm to sample the ϕ. First, sample Z from the conditional distribution using Gibbs sampling. Second, using sampled Z, derive the conditional distribution $P(\pi|X, Z, \theta)$ and sample π. Third, using sampled Z and π, calculate the acceptance probability and sample ϕ by using the Metropolis-Hastings algorithm. And repeat these three steps until convergence.

Algorithm 1 $P(\theta, Z|X)$ sampler using Gibbs + Metropolis-Hastings

Input: Dataset $X = \{x_1, \cdots, x_n\}$
  Initialize $\pi^{(1)}, \phi^{(1)} \sim p(\theta)$ and $Z^{(1)} \sim P(Z|\pi^{(1)})$
  for all $t = 1, \cdots$ **do**
    Set $Z^{(t+1)} = Z^{(t)}$
    for all $j = 1, \cdots, n$ **do**
      Sample $z_j^{(t+1)} \sim P\left(z_j | \left\{z_k^{(t+1)}\right\}_{k \neq j}\right) = P(z_j|Z^{(t+1)} \setminus \{z_j^{(t+1)}\}, X, \theta)$
    end for
    for all $j = 1, \cdots, K$ **do**
      $\alpha_j^{(t+1)} =$ number of i with $z_i^{(t+1)} = j$
    end for
    Sample $\pi^{(t+1)} \sim Dir(\mathbf{1} + \alpha^{(t+1)})$ where $\mathbf{1} = (1, 1, \cdots, 1) \in \mathbb{Z}^K$
    Propose $\phi' \sim q(\phi'|\phi^{(t)})$
    Compute $A(\phi'|\phi) = min\left\{1, \frac{P(X|Z,\phi')P(\phi')q(\phi|\phi')}{P(X|Z,\phi)P(\phi)q(\phi'|\phi)}\right\}$
    Sample $u \sim$ Unif [0,1]
    if $u \leq A(\phi'|\phi)$ **then**
      $\phi^{(t+1)} = \phi'$
    else
      $\phi^{(t+1)} = \phi^{(t)}$
    end if
  end for

For more details on the code implementation, please refer to the implementation report.

(e) We observe that

$$P(X, Z, \theta) = P(X, Z, \phi, \pi) = P(X|Z, \phi)P(Z|\pi)p(\pi)P(\phi) \tag{10}$$

So

$$\begin{aligned}P(X,Z,\phi) &= \int P(X,Z,\theta)\mathrm{d}\pi \\ &= \int P(X|Z,\phi)P(Z|\pi)p(\pi)P(\phi) \\ &= P(X|Z,\phi)P(\phi)\int P(Z|\pi)p(\pi)\mathrm{d}\pi\end{aligned} \tag{11}$$

We know that

$$\begin{aligned}\int P(Z|\pi)p(\pi)\mathrm{d}\pi &= \int Cat(\pi)Dir(\pi|1,1,\cdots,1)\mathrm{d}\pi \\ &= \int\cdots\int(\pi_1^{n_1}\cdots\pi_K^{n_K})\frac{\Gamma(K)}{\Gamma(1)^K}\pi_1^{1-1}\cdots\pi_K^{1-1}\mathrm{d}\pi_1\cdots\mathrm{d}\pi_K \\ &= \Gamma(K)\int\cdots\int\pi_1^{n_1}\cdots\pi_K^{n_K}\mathrm{d}\pi_1\cdots\mathrm{d}\pi_K \\ &= \Gamma(K)\int_0^1\pi_1^{n_1}\mathrm{d}\pi_1\cdots\int_0^1\pi_K^{n_K}\mathrm{d}\pi_K \\ &= \Gamma(K)\prod_{i=1}^K\frac{1}{n_i+1} = (K-1)!\prod_{i=1}^K\frac{1}{n_i+1}\end{aligned} \tag{12}$$

where n_i is the number of j's with $z_j = i$.

Thus, we have

$$\begin{aligned}P(X,Z,\phi) &= P(X|Z,\phi)P(\phi)\int P(Z|\pi)p(\pi)\mathrm{d}\pi \\ &= P(X|Z,\phi)P(\phi)(K-1)!\prod_{i=1}^K\frac{1}{n_i+1} \\ &= \prod_{i=1}^n\prod_{k=1}^K\mathcal{N}(x_i|\mu_k,\lambda_k I_d+v_k v_k^T)^{\mathbf{1}_{\{z_i=k\}}} \\ &\times\prod_{k=1}^K\mathcal{N}\left(\mu_k;\mathbb{O}_d,5.0\cdot I_d\right)\cdot\log\mathcal{N}\left(\lambda_k;0.1,0.1\right)\cdot\mathcal{N}\left(v_k;0_d,0.25\cdot I_d\right) \\ &\times(K-1)!\prod_{i=1}^K\frac{1}{n_i+1}\end{aligned} \tag{13}$$

The last equality comes from eq (7), (8).

Next, let's consider $P(\phi, Z|X)$. Observe that

$$P(\phi, Z|X) = \frac{P(X, Z, \phi)}{P(X)} \tag{14}$$

and since we know the $P(X, Z, \phi)$, we can use the Gibbs + Metropolis-Hastings similar to (d).

So, similarly, we can propose an algorithm as follows.

Algorithm 2 $P(\phi, Z|X)$ sampler using Gibbs + Metropolis-Hastings

Input: Dataset $X = \{x_1, \cdots, x_n\}$
  Initialize $\phi^{(1)} \sim p(\phi)$ and $Z^{(1)}$ randomly.
  for all $t = 1, \cdots$ **do**
    Propose $\phi' \sim q(\phi'|\phi^{(t)})$
    Compute $A(\phi'|\phi) = min\left\{1, \frac{P(X|Z,\phi')P(\phi')q(\phi|\phi')}{P(X|Z,\phi)P(\phi)q(\phi'|\phi)}\right\}$
    Sample $u \sim$ Unif [0,1]
    if $u \leq A(\phi'|\phi)$ **then**
      $\phi^{(t+1)} = \phi'$
    else
      $\phi^{(t+1)} = \phi^{(t)}$
    end if
    Set $Z^{(t+1)} = Z^{(t)}$
    for all $j = 1, \cdots, n$ **do**
      Sample $z_j^{(t+1)} \sim P\left(z_j | \left\{z_k^{(t+1)}\right\}_{k \neq j}\right) = P(z_j | Z^{(t+1)} \setminus \{z_j^{(t+1)}\}, X, \phi)$
      where $P(z_i = j | Z \setminus \{z_i\}, X, \phi) = \frac{\mathcal{N}(x_i|\mu_j, \lambda_j I_d + v_j v_j^T)}{\sum_{k=1}^{K} \mathcal{N}(x_i|\mu_k, \lambda_k I_d + v_k v_k^T)}$
    end for
  end for