Haniyeh Ehsani Oskouie/ 306300374

No.

Date

Problem 1:

1.

$$f(x;\alpha) = \underbrace{\frac{\Gamma(\sum_{i=1}^{K} \alpha_i)}{\prod_{i=1}^{K} \Gamma(\alpha_i)}}_{\text{Constant}} \prod_{i=1}^{K} x_i^{\alpha_i - 1}$$

$$\frac{\Gamma'(z)}{\Gamma(z)} = \psi(z) = \frac{d}{dz} \log \Gamma(z)$$

$$\log f(x;\alpha) = \log \Gamma(\sum_{i=1}^{K} \alpha_i) - \sum_{i=1}^{K} \log \Gamma(\alpha_i) + \sum_{i=1}^{K} (\alpha_i - 1) \log x_i$$

$$\frac{d \overset{\rightarrow f}{\log} \overset{\rightarrow g}{\Gamma} (\overset{\rightarrow h}{\sum_{i=1}^{K} \alpha_i})}{d\alpha_i} = \frac{df}{dg} \times \frac{dg}{dh} \times \frac{dh}{d\alpha_i}$$

$$= \frac{g'}{g} \times [\psi(h)\Gamma(h)] \cdot 1$$

$$= \psi(h) \cdot \psi(h) \cdot \Gamma(h) = \psi^2(h) \cdot \Gamma(h)$$

$$\frac{d \log f(x;\alpha)}{d\alpha_i} = \psi^2(\sum_{i=1}^{K} \alpha_i) \Gamma(\sum_{i=1}^{K} \alpha_i) - \psi(\alpha_i) + \log x_i = 0$$

$$\Rightarrow \alpha_i = \psi^{-1}(\psi^2(\sum_{i=1}^{K} \alpha_i) \Gamma(\sum_{i=1}^{K} \alpha_i) + \log x_i)$$

2. Kernel of pdf: $\prod_{i=1}^{K} x_i^{\alpha_i - 1}$

$$P(\theta | x) \propto P(x;\theta)\, p(\theta)$$

$$= \exp(-\frac{1}{2} \sum_{i=1}^{N} (x_i - \mu)^T \Sigma^{-1} (x_i - \mu)) \exp(-\frac{1}{2} (\mu - 0)^T I^{-1} (\mu - 0))$$

$$= \exp(-\frac{1}{2} [(\sum_{i=1}^{N} x_i^T \Sigma^{-1} x_i) + N \mu^T \Sigma^{-1} \mu - 2N \bar{x} \Sigma^{-1} \mu + \mu^T \mu])$$

we remove parts unrelated to θ. We will have:

$$\bar{x} = \frac{\sum_{i=1}^{N} x_i}{N}$$

$= exp(-\frac{1}{2}[N\mu^T \Sigma^{-1}\mu - 2N\bar{x}\Sigma^{-1}\mu + \mu^T\mu])$

$= exp(-\frac{1}{2}[\mu^T(N\Sigma^{-1}+I)\mu - 2\mu(N\Sigma^{-1}\bar{x})])$

$M = N\Sigma^{-1} + I$

$b = N\Sigma^{-1}\bar{x}$

$p(\theta | x) = exp(-\frac{1}{2}[(\mu - M^{-1}b)^T M(\mu - M^{-1}b) - b^T M^{-1} b])$

$= \mathcal{N}_D(\mu | M^{-1}b, M^{-1})$

$= \mathcal{N}(\mu | (N\Sigma^{-1}+I)^{-1}\sum_{i=1}^{N} x_i, (N\Sigma^{-1}+I)^{-1})$

Problem 2:

1. Similarities: Both sample words (which are iid) from the same distribution (same likelihood function)

Differences: In MUM, we sample word by word but in MMM, we sample the whole document at once.

For Multinomial Mixture Model:

$$L = \prod_d \frac{(\sum_n x_{dn})!}{\prod_n x_{dn}!} \sum_k p(z=k) \prod_n \beta_{kn}^{x_{dn}}$$

For Mixture of Unigrams:

$$L = \prod_d \sum_k p(z=k) \prod_i \beta_{k w_{di}}$$

The likelihood functions are proven to be similar.

2. (a) $p(w, d, z; \theta, \beta)$

$= p(w|z,d)\, p(z|d)\, p(d)$

$= \beta_{zw}\, \theta_{dz}\, \pi_d$

(b) $p(z|w,d;\theta,\beta) = \dfrac{\overset{p(z,w|d)}{p(w|z,d)\, p(z|d)}}{\underbrace{\sum_{z'} p(w|z',d)\, p(z'|d)}_{p(w|d)}} = \dfrac{\beta_{zw}\theta_{dz}}{\sum_{z'} \beta_{z'w}\theta_{dz'}}$

(c) $p(w|z,d;\theta,\beta) = \beta_{zw}$

$$= \frac{\sum_d p(z|w,d)\, c(w,d)}{\sum_{w,d} p(z|w,d)\, c(w,d)}$$

Continue of (a) $\rightarrow p(w,d,z;\theta,\beta)$

$$= \frac{\sum_d p(z|w,d)\, c(w,d)}{\sum_{w,d} p(z|w,d)\, p(w,d)} \; \frac{\sum_w p(z|w,d)\, c(w,d)}{N_d} \; \pi_d$$

3.

Likelihood of corpus is: $\prod_{d=1} \pi_d \prod_{n=1}^{N_d} \sum_k \theta_{dk}\beta_{k w_n}$

$\underset{\theta,\beta}{\max} \log L = \sum_{dw} c(w,d) \left(\log \underbrace{\sum_z \theta_{dz}\beta_{zw}}_{p(w|d;\theta,\beta)} \right) \rightarrow$ log evidence

$p(w,z|d) = \beta_{zw}\,\theta_{dz}$

$KL(p(z|w,d) \,\|\, p(w|d)) \rightarrow$ KL divergence

$-\mathbb{E}[\log p(z|w,d)] + \mathbb{E}[\log p(z,w)] \rightarrow$ ELBO

$\log p(w|d;\theta,\beta) = KL(p(z|w,d;\theta,\beta) \,\|\, p(w|d;\theta,\beta))$
$\mathbb{E}[\log p(z,w;\theta,\beta)] - \mathbb{E}[\log p(z|w,d;\theta,\beta)]$