MOLTEN COMPONENT ANALYSIS

KO ABE

1. Normal distribution

1.1. Estimation. Now, consider following data generating process;

$$\begin{aligned} y_n &\sim \mathcal{N}\left(\sum_{l=1}^{L}\prod_{d=1}^{D}\mu_{dl}^{x_{nd}}, \lambda^{-1}\right) \\ \mu_{dl} &\sim \mathcal{N}(0, \tau^{-1}) \end{aligned}$$

where $\mathcal{N}(\mu, \sigma^2)$ is normal distribution with mean μ and variance σ^2.

After some manipulation, we get following result;

$$\begin{aligned}
&\sum_{n=1}^{N}\log p(\mu_{dl}|y, X, \mu_{dl}^{(c)}, \sigma^2) \\
&= \sum_{n=1}^{N}\left(-\frac{\lambda}{2}\left\{y_n - \sum_{l=1}^{L}\prod_{j=1}^{D}\mu_{j,l}^{x_{nj}}\right\}^2\right) - \tau\frac{\mu_{dl}^2}{2} + C \\
&= \sum_{n=1}^{N}\left(-\frac{\lambda}{2}\left\{y_n - \sum_{l'\neq l}\prod_{j=1}^{D}\mu_{jl'}^{x_{n,j}} - \prod_{j=1}^{D}\mu_{jl}^{x_{nj}}\right\}^2\right) - \tau\frac{\mu_{jl}^2}{2} + C \\
&= \sum_{n=1}^{N}\left(-\frac{\lambda}{2}\left\{\prod_{j=1}^{D}\mu_{jl}^{2x_{nj}} - 2\prod_{j=1}^{D}\mu_{jl}^{x_{nj}}\left(y_n - \sum_{l'\neq l}\prod_{j=1}^{D}\mu_{jl'}^{x_{n,j}}\right)\right\}\right) - \tau\frac{\mu_{dl}^2}{2} + C \\
&= -\frac{\lambda}{2}\left\{\sum_{n=1}^{N}\left(\mu_{dl}^{2x_{nd}}\prod_{d'\neq d}\mu_{d'l}^{2x_{nd'}} - 2\mu_{dl}^{x_{nd}}\prod_{d'\neq d}\mu_{d'l}^{x_{nd'}}\left(y_n - \sum_{l'\neq l}\prod_{j=1}^{D}\mu_{jl'}^{x_{nd}}\right)\right)\right\} - \tau\frac{\mu_{dl}^2}{2} + C \\
&= -\frac{\tau + \lambda\sum_{n=1}^{N}x_{nd}\prod_{d'\neq d}\mu_{d'l}^{2x_{nd'}}}{2} \\
&\quad\times\left(\mu_{dl}^2 - 2\mu_{dl}\lambda\frac{\sum_{n=1}^{N}x_{nd}\prod_{d'\neq d}\mu_{d'l}^{x_{nd'}}\left[y_n - \sum_{l'\neq l}\prod_{j=1}^{D}\mu_{jl'}^{x_{nj}}\right]}{\sum_{n=1}^{N}x_{nd}\prod_{d\neq d'}\mu_{d'l}^{2x_{nd'}}}\right) + C,
\end{aligned}$$

where C is constant term which not depends on μ_{dl}.

Using mean-field approximation, we get following variational posterior distribution;

$$q(\mu_{dl}) = \mathcal{N}(\hat{\mu}_{dl}, \hat{\sigma}_{dl}),$$

where $\hat{\mu}_{dl}$ and $\hat{\sigma}_{dl}$ are defined by

$$\hat{\mu}_{dl} = \frac{\sum_{n=1}^{N} x_{nd} \prod_{d' \neq d} \langle \mu_{d'l} \rangle^{x_{nd'}} \left[y_n - \sum_{l' \neq l} \prod_{j=1} \langle \mu_{jl'} \rangle^{x_{nj}} \right]}{\tau/\lambda + \left(\sum_{n=1}^{N} x_{nd} \prod_{d \neq d'} \langle \mu_{d'l}^2 \rangle^{x_{nd'}} \right)},$$

$$\hat{\sigma}^2 = \left(\tau + \lambda \left\{ \sum_{n=1}^{N} x_{nd} \prod_{d' \neq d} \langle \mu_{d'l}^2 \rangle^{x_{nd'}} \right\} \right)^{-1},$$

where $\langle x \rangle$ is expectation of x under the variational posterior distribution.

2. Poisson distribution

2.1. Estimation. Now, consider following data generating process;

$$y_n \sim \mathcal{P}\wr \left(\sum_{l=1}^{L} \prod_{d=1}^{D} \lambda_{dl}^{x_{nd}}, \right) \tag{1}$$

$$\lambda \sim \mathcal{G}(a, b),$$

where $\mathcal{P}\wr(\lambda)$ is Poisson distribution with mean λ and $\mathcal{G}(a, b)$ is gamma distribution with shape a, and scale b.

2.2. Variational inference. Equation **??** is equivalent to following:

$$y_n = \sum_{d=1}^{D} s_{n,d} + \sum_{l=1}^{L} u_{n,l},$$

$$u_{n,l} \sim \mathcal{P}\wr \left(\tau_n \prod_{d=1}^{D} \lambda_{d,l}^{x_{n,d}} \right) \tag{2}$$

Using the mean field assumption (Jordan *et al.*, 1999), we introduce closed-form variational Bayes inference updates for the proposal model. Let $q(\lambda_{k,l})$ and $q(\boldsymbol{h}_l)$ be the target approximate posterior distribution.

The updates for $u_{n,l}$ are given by $E[u_{n,l}] = y_n p_{n,l}$ where $p_{n,l}$ is defined as

$$p_{n,l} = \frac{\exp(x_{n,d} E[\log w_{d,l}])}{\exp(\sum_{d=1}^{D} x_{n,d} E[\log w_{d,l}]) + \sum_{l=1}^{L} \exp(x_{n,d} E[\log \lambda_{d,l}]) \exp(E[\log h_{l,k}])} \tag{3}$$

The updates for $s_{n,k}$ are given by $E[s_{n,k}] = y_n r_{n,d}$ where $r_{n,d}$ is defined as

$$r_{n,d} = \frac{\exp(E[\log w_d])}{\exp(\sum_{d=1}^{D} x_{n,d} E[\log w_{d,l}]) + \sum_{l=1}^{L} \exp(x_{n,d} E[\log \lambda_{d,l}]) \exp(E[\log h_{l,k}])} \tag{4}$$

The mean-field posterior $q(\lambda_{d,l})$ is gamma distribution with shape parameter $\hat{a}_{d,l}$ and rate parameter $\hat{b}_{d,l}$ where

$$\hat{a}_{d,l} = \sum_{n=1}^{N} \sum_{n=1}^{N} x_{n,d} u_{n,l} + a, \tag{5}$$

$$\hat{b}_{d,l} = \sum_{n=1}^{N} x_{n,d} \tau_n \left(\prod_{d' \neq d} \lambda_{d',l}^{x_{n,d'}} \right) + b. \tag{6}$$

The mean-field posterior $q(w_d)$ is gamma distribution with shape parameter $\hat{a}_d^{(w)}$ and rate parameter $\hat{b}_d^{(w)}$ where

$$\hat{a}_d^{(w)} = \sum_{n=1}^{N} s_{n,d} + a, \tag{7}$$

$$\hat{b}_d^{(w)} = \sum_{n=1}^{N} \tau_n + b. \tag{8}$$

Lower bound on marginal likelihood. Lower bound on marginal likelihood (evidence lower bound; ELBO) $\mathcal{L}(q)$ is used for model selection (Corduneanu and Bishop, 2001).

$$\mathcal{L}(q) = \int q(\theta) \log \frac{p(\boldsymbol{y}, \theta | X, a_w, b_w, a, b)}{q(\theta)} d\theta \tag{9}$$

$$= \int q(\theta) \log p(\boldsymbol{y}, \theta | X, a_w, b_w, a, b) d\theta - \int q(\theta) \log q(\theta) d\theta. \tag{10}$$

where θ is all the variational parameters.