MOLTEN COMPONENT ANALYSIS

KO ABE

1. Poisson distribution

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

$$y_n \sim \text{Poisson}\left(\sum_{l=1}^{L}\prod_{d=1}^{D}\lambda_{dl}^{x_{nd}},\right) \tag{1}$$
$$\lambda \sim \text{Gamma}(a, b),$$

where Poisson(λ) is Poisson distribution with mean λ and Gamma(a, b) is gamma distribution with shape a, and scale b.

1.2. Variational inference. Equation 1 is equivalent to following:

$$y_n = \sum_{l=1}^{L} u_{nl},$$
$$u_{nl} \sim \text{Poisson}\left(\prod_{d=1}^{D}\lambda_{dl}^{x_{nd}}\right) \tag{2}$$

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_{nl} are given by $E[u_{nl}] = y_n p_{nl}$ where p_{nl} is defined as

$$p_{nl} = \frac{\exp(x_{nd}E[\log \lambda_{dl}])}{\exp(\sum_{d=1}^{D} x_{nd}E[\log \lambda_{dl}])} \tag{3}$$

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

$$\hat{a}_{dl} = \sum_{n=1}^{N} x_{nd}u_{nl} + a, \tag{4}$$

$$\hat{b}_{dl} = \sum_{n=1}^{N} x_{nd}\left(\prod_{d' \neq d}\lambda_{d'l}^{x_{nd'}}\right) + b. \tag{5}$$

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

$$\hat{a}_{dl} = \sum_{n=1}^{N} x_{nd} u_{nd} + a, \tag{6}$$

$$\hat{b}_{dl} = \sum_{n=1}^{N} x_{nd} \left(\prod_{d' \neq d} \lambda_{d'l}^{x_{nd'}} \right) + b. \tag{7}$$

Thus, $E[\lambda_{dl}] = \hat{a}_{dl}/\hat{b}_{dl}$ and $E[\lambda_{dl}] = \psi(\hat{a}_{dl}) - \log(\hat{b}_{dl})$ where $\psi(\cdot)$ is digamma function.

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).

$$\mathcal{L}(q) = \int q(\theta) \log \frac{p(\boldsymbol{y}, \theta | X, a_w, b_w, a, b)}{q(\theta)} d\theta \tag{8}$$

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

where θ is all the variational parameters.