CS 698S

Homework - 3

Pratik Mishra
14493

Problem 1

Let $Z_{nm} = (u_n, v_m)$.

Thus, joint probability density:

$$p(z, X) = p(X \mid Z)\, p(Z)$$

$$= \left(\prod_n \prod_m p(X_{nm} \mid u_n, v_m)\right)\left(\prod_n p(u_n)\right)\left(\prod_m p(v_m)\right)$$

Log joint probability density:

$$\log(p(z,X)) = \sum_n \sum_m \log(p(X_{nm} \mid u_n, v_m)) + \sum_n \log(p(u_n))$$

$$+ \sum_m \log(p(v_m))$$

$$= \sum_n \sum_m -\frac{\beta}{2}(X_{nm} - u_n^T v_m)^2 - \sum_n \frac{\lambda_u}{2} u_n^T u_n$$

$$- \sum_m \frac{\lambda_v}{2} v_m^T v_m + \text{Constants}$$

Let $q(Z_{nm})$ be the proposed distributions with the mean field assumption, i.e, $q(u_n, v_m) = q_a(u_n)\, q_b(v_m)$.

Thus,

$$\log(q_a(u_n)) = E_{v_m}(\log(p(z,x)))$$

$$= E_{v_m}\Big(\sum_m \frac{\beta}{2}(x_{nm} - u_n^T v_m)^2 - \frac{\lambda_u}{2} u_n^T u_n - \frac{\lambda_v}{2} v_m^T v_m\Big) + \text{constants}$$

$$= \sum_m \frac{\cancel{\beta}\beta}{2}(u_n^T E(v_m))^2 - \beta X_{nm} u_n^T E(v_m) - \frac{\lambda_u}{2} u_n^T u_n + \text{constants}$$

$$= -\frac{1}{2}(u_n - \mu_{u_n})^T \Sigma_{u_n}^{-1}(u_n - \mu_{u_n})$$

where $\mu_{u_n} = -\Sigma_{u_n}\Big(\Sigma_m \frac{\beta}{2} X_{nm} E(v_m)\Big)$

and $\Sigma_{u_n} = \Big(\frac{\lambda_u}{2} I_K - \Sigma_m E(v_m) E(v_m)^T\Big)^{-1}$

Hence, $q_a(u_n)$ is normal distribution with mean μ_{u_n} and variance Σ_{u_n}. Similarly, $q_b(v_m)$ also follows a normal distribution with mean $\mu_{v_m} = -\Sigma_{v_m}\Big(\Sigma_n \frac{\beta}{2} X_{nm} E(u_n)\Big)$ and variance $\Sigma_{v_m} = \Big(\frac{\lambda_v}{2} I_K - \Sigma_n E(u_m) E(u_n)^T\Big)^{-1}$. Hence we get the desired updates for u_n & v_m.

Problem 2

Let $Z_{nm} = (u_n, v_m)$.

Then, log joint probability density:

$$\log(p(Z,X)) = \sum_n \sum_m \log(p(x_{nm} | u_n, v_m)) + \sum_n \log(p(u_n)) + \sum_m \log(p(v_m))$$

$$= \sum_n \sum_m (-u_n^T v_m + x_{nm} \log(u_n^T v_m)) + \sum_n \log\left(\prod_{k=1}^{K} u_{nk}^{a-1} e^{-b u_{nk}}\right) + \sum_m \log\left(\prod_{k=1}^{K} v_{mk}^{c-1} e^{-d v_{mk}}\right) + \text{constants}$$

$$= \sum_n \sum_m (-u_n^T v_m + x_{nm} \log(u_n^T v_m)) + \sum_n \sum_{k=1}^{K} ((a-1)\log(u_{nk}) - b u_{nk}) + \sum_n \sum_{k=1}^{K} ((c-1)\log(v_{mk}) - d v_{mk}) + \text{constants}$$

Let $q(Z_{nm})$ be the proposed distribution with the mean field assumption, i.e.

$$q(u_n, v_m) = q_u(u_n)\, q_v(v_m) = \prod_{k=1}^{K} q_u(u_{nk})\, q_v(v_{mk})$$

$$\log(q_u(u_n)) = E_{v_m}(\log(p(z,x)))$$

$$\therefore \log(q_u(u_{nk})) = E_{v_m, u_{nl}, l\neq k}(\log(p(z,x)))$$

$$= E_{v_m, u_{nl}, l\neq k}\Big[\sum_n \sum_m \Big(-u_n^T v_m + x_{nm}\log\Big(\sum_{l=1}^{k} u_{nl}v_{ml}\Big)$$

$$+ \sum_n ((a-1)\log(u_{nk}) - bu_{nk})\Big] + \text{c}$$

Consider the following,

$$\log\Big(\sum_{l=1}^{k} u_{nl}v_{ml}\Big) = \log\Big(\sum_{l=1}^{k} \phi_l \frac{u_{nl}v_{ml}}{\phi_l}\Big)$$

$$\geq \sum_{l=1}^{k} \phi_l \log\Big(\frac{u_{nl}v_{ml}}{\phi_l}\Big)$$

$$\Rightarrow E\Big[\log\Big(\sum_{l=1}^{k} u_{nl}v_{ml}\Big)\Big] \geq \sum_{l=1}^{k} \phi_l E\Big[\log\Big(\frac{u_{nl}v_{ml}}{\phi_l}\Big)\Big]$$

We use the value of ϕ_l for which R.H.S is maximum. Using the given approximation, we get:

$$\log(q_u(u_{nk})) = -u_{nk}\sum_n E_q[v_{mk}] + (a-1)\log(u_{nk})$$

$$- bu_{nk} + \sum_m \Big(x_{nm} E_q\Big[\log\Big(\sum_{l=1}^{k} u_{nl}v_{ml}\Big)$$

$$\geq -u_{nk}\sum_m E_q[v_{mk}] + (a-1)\log(u_{nk}) - bu_{nk}$$

$$+ \sum_m x_{nm}\Big(\sum_{l=1}^{k} \phi_l E\Big[\log\Big(\frac{u_{nl}v_{ml}}{\phi_l}$$

$$\Rightarrow -u_{nk} \sum_m E_q[v_{mk}] + (a-1)\log(u_{nk}) - b u_{nk}$$

$$+ \sum_m x_{nm} \left(\sum_{l=1}^{k} \frac{e^{\log(u_{nl}) + E_q[\log(v_{ml})]}}{\sum_{p \neq l} e^{\log(u_{np}) + E_q[\log(v_{mp})]}} \log\left(\sum_{p \neq l} e^{\log(u_{np}) + E_q[\log(v_{mp})]}\right)\right.$$

$$= (a-1)\log(u_{nk}) - u_{nk}\left(\sum_m \left(E_q[v_{mk}] - x_{nm}\, \alpha_{nm}^{k}\right) + b\right)$$

where $\alpha_{nm}^{k} = \left(\frac{e^{E_q[\log(v_{ml})]}}{\sum_{p \neq l} e^{\log(u_{np}) + E_q[\log(v_{mp})]}} \cdot \log\left(\sum_{p \neq l} e^{\log(u_{np}) + E_q(\log(v_{mp}))}\right)\right.$

Hence, $\underline{q_u(u_{nk}) \sim \text{Gamma}(a_{u_{nk}}, b_{u_{nk}})}$

where,

$a_{u_{nk}} = a$

and $b_{u_{nk}} = \sum_m \left(E_q(v_{mk}) - x_{nm}\, \alpha_{nm}^{k}\right) + b$

<u>Similarly</u>,

$q_v(v_{mk}) \sim \text{Gamma}(c_{v_{mk}}, d_{v_{mk}})$

where

$c_{v_{mk}} = c$ and $d_{v_{mk}} = \sum_m \left(E_q(u_{mk}) - x_{nm}\, \beta_{nm}^{k}\right)$

Problem 3

let $\epsilon_t \sim N(0, 2\eta_t)$.

Then SGLD update for μ_n:

$$u_n^{(t)} = u_n^{(t-1)} + \eta_t \nabla_{u_n}\left[\frac{N}{B}\sum_{t=1}^{B} \log(p(u_n, v_{m_t} | x_{nm_t}))\right] + \epsilon_t$$

$$= u_n^{(t-1)} + \eta_t \nabla_{u_n}\left[\frac{N}{B}\sum_{t=1}^{B} (\log(p(x_{nm_t} | u_n, v_{m_t})) + \log(p(u_n)) + \text{constants w.r.t } u_n)\right] + \epsilon_t$$

$$= u_n^{(t-1)} + \eta_t \nabla_{u_n}\left[\frac{N}{B}\sum_{t=1}^{B}\left(-\frac{\beta}{2}(x_{nm_t} - u_n^T v_{m_t})^2 - \frac{\lambda_u}{2} u_n^T u_n + \text{constants w.r.t } u_n\right)\right] + \epsilon_t$$

$$= u_n^{(t-1)} + \eta_t \left[\frac{N}{B}\sum_{t=1}^{B}(\beta(x_{nm_t} - u_n^T v_{m_t})v_{m_t} - \lambda_u u_n\right] + \epsilon_t$$

Hence, the update is:

$$u_n^{(t)} = u_n^{(t-1)} + \eta_t\left[\frac{N}{B}\sum_{t=1}^{B}(\beta(x_{nm_t} - v_{m_t}^T u_n^{(t-1)})v_{m_t} - \lambda_u u_n\right] + \epsilon_t$$

$$v_m^{(t)} = v_m^{(t-1)} + \eta_t\left[\frac{N}{B}\sum_{t=1}^{B}(\beta(x_{n_t m} - u_{n_t}^T v_m^{(t-1)})u_{n_t} - \lambda_v v_m\right] + \epsilon_t$$