Bayesian Model Machine Learning HW3

Ben Zhu
bjz2107

1) we have data set $\{(x_i, y_i)\}_{i=1}^N$ $x \in R^d$ $y \in R$

$y_i \overset{iid}{\sim} N(x_i^T w, \lambda^{-1} I)$

$w \sim \text{Normal}(0, \text{diag}(\alpha_1 \ldots \alpha_d)^{-1})$

$\alpha_k \overset{iid}{\sim} \text{Gamma}(a_0, b_0)$

$\lambda \overset{iid}{\sim} \text{Gamma}(e_0, f_0)$

use $\text{Gamma}(\eta | \tau_1, \tau_2) = \frac{\tau_2^{\tau_1}}{\Gamma(\tau_1)} \eta^{\tau_1 - 1} e^{-\tau_2 \eta}$

Variational Inference with

$q(w, \alpha_1 \ldots \alpha_d, \lambda) \approx p(w, \alpha_1 \ldots \alpha_d, \lambda | y, x)$

$q(w, \alpha_1 \ldots \alpha_d, \lambda) = q(w)\, q(\lambda) \prod_{k=1}^d q(\alpha_k)$

a) Using optimal form

$$q(\lambda) \propto \exp\left\{ E_{q(\alpha,w)}\left[\ln p(y, w, \alpha_1 \ldots \alpha_d, \lambda | x)\right]\right\}$$

not dependent on λ not dependent on λ

$$\propto \exp\left\{ E_{q(\alpha,w)}\left[\ln p(y | w, \alpha_1 \ldots \alpha_d, \lambda, x) + \ln p(w) + \ln p(\lambda) + \ln p(\alpha_1 \ldots \alpha_n)\right]\right\}$$

$$\propto \exp\left\{ E_{q(\alpha,w)}\left[\ln p(y | w, \alpha_1 \ldots \alpha_d, \lambda, x) + \ln p(\lambda)\right]\right\}$$

$$\propto \exp\left\{ E_{q(\alpha,w)}\left[\sum_{i=1}^N \frac{1}{2}\ln\frac{\lambda}{2\pi} - \frac{\lambda}{2}(y_i - x_i^T w)^2\right] + \ln p(\lambda)\right\}$$

$$\propto \exp\left\{ E_{q(\alpha,w)}\left[\sum_{i=1}^N \frac{1}{2}\ln\lambda - \frac{\lambda}{2}(y_i - x_i^T w)^2\right] + (e_0 - 1)\ln\lambda - f_0\lambda\right\}$$

$$\propto \lambda^{\frac{N}{2}} \exp\left\{-\frac{\lambda}{2} E_{q(\alpha,w)}\left[(y_i - x_i^T w)^2\right] + (e_0 - 1)\ln\lambda - f_0\lambda\right\}$$

$$\propto \lambda^{\frac{N}{2} + e_0 - 1} \exp\left\{-\frac{\lambda}{2} E_{q(\alpha,w)}\left[(y_i - x_i^T w)^2\right] - f_0\lambda\right\}$$

this is clearly a gamma

$q(\lambda) \sim \text{Gamma}(e', f')$

$$e' = e_0 + \frac{N}{2}$$

$$f' = f_0 + \frac{1}{2}\sum_{i=1}^N E_{q(\alpha,w)}\left[(y_i - x_i^T w)^2\right]$$

$$q(\alpha) \propto \exp\{E_{q(\lambda,w)}[\ln p(y,\alpha,w,\lambda|x)]\}$$

can be ignored

$$\propto \exp\{E_{q(\lambda,w)}[\underbrace{\ln p(y|\alpha,w,\lambda,x)}_{\text{can be ignored}} + \ln p(\lambda) + \ln p(w|\alpha) + \ln p(\alpha_1 \dots \alpha_d)]\}$$

$$\propto \exp\{E_{q(\lambda,w)}[\ln p(w|\alpha_1 \dots \alpha_d) + \sum_{k=1}^{d} \ln p(\alpha_1 \dots \alpha_d)]\}$$

$$\propto \alpha_k^{a_0-1} e^{-\alpha_k b_0} \exp\{E_{q(\lambda,w)}[\frac{1}{2}\sum_{k=1}^{d} \ln \alpha_k - \frac{1}{2} w^T \underbrace{\text{diagonal}(\alpha_1 \dots \alpha_d) w}_{\alpha_k w_k^T w_k}]$$

$$\propto \alpha_k^{a_0 - \frac{1}{2}} e^{-\alpha_k(b_0 + \frac{1}{2} w_k^2)}$$

gamma

$$\boxed{q(\alpha) = \prod_{i=1}^{d} \text{Gamma}(a', b') \quad \text{where} \quad a' = a_0 + \frac{1}{2}, \quad b' = b_0 + \frac{1}{2}(w_k^T w_k + \Sigma')}$$

$q(w) \propto \exp\{E_{q(\alpha,\lambda)}[\ln p(y, w, \alpha_1 \ldots \alpha_d, \lambda | x)]\}$

$\propto \exp\{E_{q(\alpha,\lambda)}[\ln p(y|w, \alpha_1 \ldots \alpha_d, \lambda, x) + \ln p(w) + \ln p(\lambda) + \ln p(\alpha_1 \ldots \alpha_k)]\}$

$\propto \exp\{E_{q(\alpha,\lambda)}[\ln p(y|w, \alpha_1 \ldots \alpha_d, \lambda, x) + \ln p(w|\alpha_1 \ldots \alpha_d)]\}$

$\propto \exp\{E_{q(\alpha,\lambda)}[\sum^N \frac{1}{2}\ln\lambda - \frac{\lambda}{2}(y_i - x_i^T w)^2] - \frac{1}{2} w^T \text{diagonal}(\alpha_1 \ldots \alpha_k) w \cancel{+ \frac{1}{2}\{\ln \alpha_w\}}\}$

$\propto \exp\{\sum_{i=1}^{N} \frac{E_q[\lambda]}{2} (y_i - x_i^T w)^2\} \exp(-\frac{1}{2} w^T \text{diagonal}(E_q[\alpha_1], \ldots E_q[\alpha_k]) w\}$

Normal distribution,

$q(w) \sim \text{Normal}(w | \mu', \Sigma')$

$$\Sigma' = [\text{diag}(E_q[\alpha_1], \ldots E_q[\alpha_k]) + E_{q(\lambda)}[\lambda] \sum_{i=1}^{N} x_i x_i^T)^{-1}$$

$$\mu' = \Sigma'(E_q[\lambda] \sum^N y_i x_i)$$

b) pseudocode

initialize all μ', Σ', a', b' and e', f'

for every iteration,

update $q(\lambda)$ so that $e_{t+1} = e_t + \frac{N}{2}$ $\quad f_{t+1} = f_t + \frac{1}{2}[\sum_{i=1}^{N} (y_i - x_i^T \mu_t) + x^T \Sigma' x_i]$

$q(\alpha)$ so that $a_{t+1} = a_t + \frac{1}{2}$ $\quad b_{t+1} = b_t + \frac{1}{2}[\mu_k \mu_k]$

$q(w)$ so that

$$\Sigma_{t+1} = \left(\frac{e_t}{f_t} \sum x_i^T x_i + \text{diag}\left(\frac{a_{\alpha_1 t}}{b_{\alpha_1 t}}, \ldots \frac{a_{\alpha_d t}}{b_{\alpha_d t}}\right)\right)^{-1}$$

$$\mu_{t+1} = \Sigma_{t+1}\left(\frac{e_t}{f_t} \sum^N y_i x_i\right)$$

evaluate the objective function

to see convergence

Variational objective function

c)

$$\mathcal{L}(a, b, e, f, \mu, \Sigma)$$

$$\ln p(y|x) \geq \mathcal{L}(q(w, \alpha, \lambda, x))$$

$$= \sum_{i=1}^{N} E[\ln p(y_i|x_i, w, \lambda)] + E[\ln p(w|\alpha)] + \sum_{k=1}^{J} E[\ln p(\alpha_k)] + E[\ln p(\lambda)]$$

$$- E[\ln q(w)] - \sum_{k=1}^{J} E[\ln q(\alpha_k)] - E[\ln q(\lambda)]$$

splitting it up

$$E[\ln p(y_i|x_i w, \lambda)] = -\frac{1}{2}\ln 2\pi + \frac{1}{2}E[\ln\lambda] - \frac{1}{2}E[\lambda(y_i - x_i^T w)^2]$$

$$= \frac{1}{2}(\psi(e') - \ln f') - \frac{1}{2}\ln 2\pi - \frac{1}{2}E[\lambda]E[(y_i - x_i^T w)^2]$$

$$= \frac{1}{2}(\psi(e') - \ln f') - \frac{1}{2}\ln 2\pi - \frac{1}{2}\frac{e'}{f'}[x_i^T \Sigma' x_i + (y_i - x_i^T \mu')^2]$$

$$E[\ln p(w|\alpha)] = \sum_{k=1}^{J} E[\ln p(w_k|\alpha)]$$

ignoring sum for now

$$= \frac{1}{2}E[\ln\alpha_k) - \frac{1}{2}\ln 2\pi - \frac{1}{2}E[\alpha_k]E[w_k]^2$$

$$= \frac{1}{2}(\psi(a_k') - \ln b_k') - \frac{1}{2}\ln 2\pi - \frac{1}{2}\frac{a_k'}{b_k'}(\Sigma_k' + \mu_k'^T \mu_k')$$

$$E[\ln p(\alpha_k)] = a_0 \ln b_0 - \ln\Gamma(a_0) + (a_0 - 1)E[\ln\alpha_k] - b_0 E[\alpha_k]$$

this one is weird

$$= a_0 \ln b_0 - \ln\Gamma(a_0) + (a_0 - 1)(\psi(a_k') - \ln b_k') - b_0 \frac{a_k'}{b_k'}$$

$$E[\ln p(\lambda)] = e_0 \ln f_0 - \ln\Gamma(e_0) + (e_0 - 1)(\psi(e') - \ln f') - f_0\frac{e'}{f'}$$

and we know the $q(w)$, $q(\alpha_k)$, $q(\lambda)$

$$\ln p(y|x) \geq \mathcal{L}(q(w, \alpha, \lambda, x)) =$$

$$= \sum_{i=1}^{N} \left[\frac{1}{2}\left(\psi(a_n') - \ln b_n'\right) - \frac{1}{2}\ln 2\pi - \frac{1}{2}\frac{a_k'}{b_k'}\left(\Sigma_t' + \mu_t'^T \mu_t\right) \right] + \sum_{k=1}^{d} \left[a_0 \ln b_0 - \ln\Gamma(a_0) + (a_0 - 1)\left(\psi(a_k') - \ln b_k'\right) - \frac{b_0 a_k'}{b_k'} \right]$$

$$+ \left[e_0 \ln f_0 - \ln\Gamma(e_0) + (e_0 - 1)\left(\psi(e') - \ln f'\right) - f_0 \frac{e'}{f'} \right] + \left[\frac{d}{2} + \frac{d}{2}\ln 2\pi + \frac{1}{2}\ln \Sigma' \right]$$

$$+ \left[e' - \ln f' + \ln\Gamma(e') + (1 - e')\psi(e') \right] + \sum_{k=1}^{d} \left(a_k' - \ln b_k' + \ln\Gamma(a_k') + (1 - a_k')\psi(a_k') \right)$$

For purposes of the next problem, evaluate this function

the only terms that matter are the variable parameters