Bayesian Model Machine Learning HW3

Ben Zhu
bjz2107

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

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

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

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

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

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

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

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

a) Using optimal form

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

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

($\ln p(w)$: not dependent on λ; $\ln p(\alpha_1 \dots \alpha_n)$: not dependent on λ)

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

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

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

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

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

this is clearly a gamma

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

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

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

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

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

($\ln p(\lambda)$: can be ignored; $\ln p(y|\alpha,w,\lambda,x)$: can be ignored)

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

$$\propto \alpha_k^{a_0-1} e^{-\alpha_k b_0} \exp\left\{E_{q(\alpha,w)}\left[\frac{1}{2}\sum_{k=1}^{d} \ln \alpha_k - \frac{1}{2} w^T \text{diagonal}(\alpha_1 \cdots \alpha_d) w\right]\right\}$$

($w^T \text{diagonal}(\alpha_1 \cdots \alpha_d) w \rightarrow \alpha_k w_k^T w_k$)

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

gamma

$$q(\alpha) = \prod_{i=1}^{d} \text{Gamma}(a', b')$$

where $a' = a_0 + \frac{1}{2}$

$b' = b_0 + \frac{1}{2} w_k^T w_k + \varepsilon'$

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

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

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

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

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

Normal distribution,

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

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

$$\mu' = \Sigma'\left(E_q[\lambda] \sum y_i x_i\right)$$

b) procedure

initialize all $\mu', \Sigma', a', b', e', f'$

for evaluation,

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

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

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

evaluate the objective function

to see convergence

c) Variational objective function

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

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

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

splitting it up

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

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

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

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

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

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

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

can we know the $q(w)$ $q(\alpha_k)$ $q(\lambda)$

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

$$= \sum_{i=1}^{N} \left[\frac{1}{2}\left(\psi(a'_n) - \ln b'_n\right) - \frac{1}{2}\ln 2\pi - \frac{1}{2}\frac{a'_n}{b'_k}\left(\xi'_t + M_t^T M_t\right) \right] + \sum_{k=1}^{d} \left[a_0 \ln b_0 - \ln\Gamma(a_0) + (a_0 - 1)\left(\psi(a'_k) - \ln b'_k\right) - \frac{b_0}{b'_k} \right.$$

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

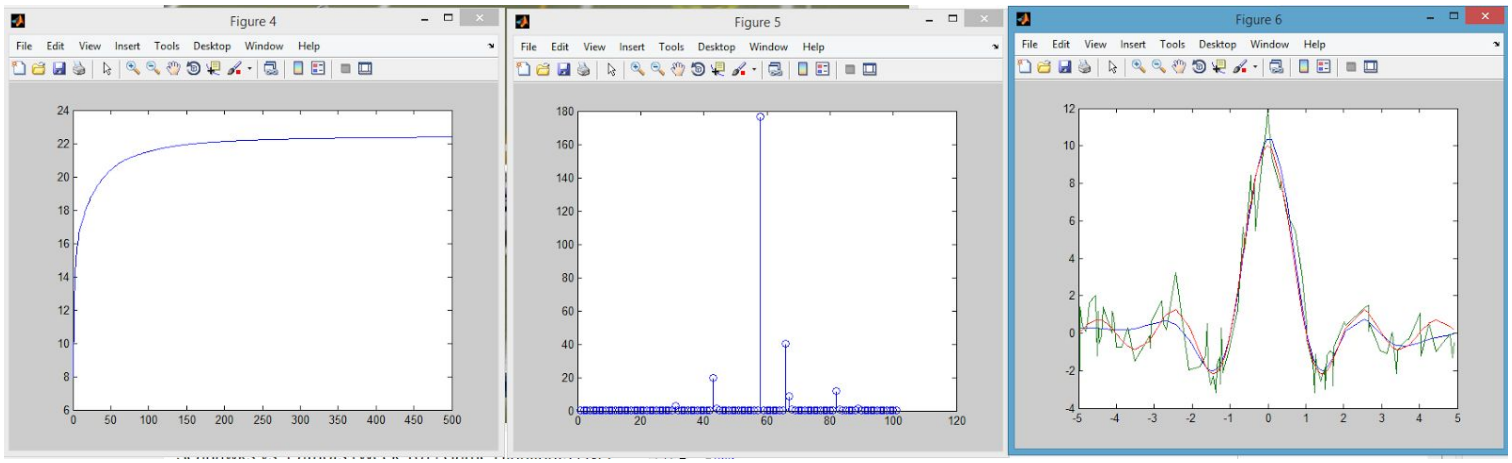
$$+ \left[e' - \ln f' + \ln\Gamma(e') + (1 - e')\psi(e') \right] + \sum_{k=1}^{d} \left(a'_k - \ln b'_k + \ln\Gamma(a'_k) + (1 - a'_k)\psi(a'_k) \right)$$

for purposes of the next problem, evaluate this function

the only terms that matter are the variable parameters

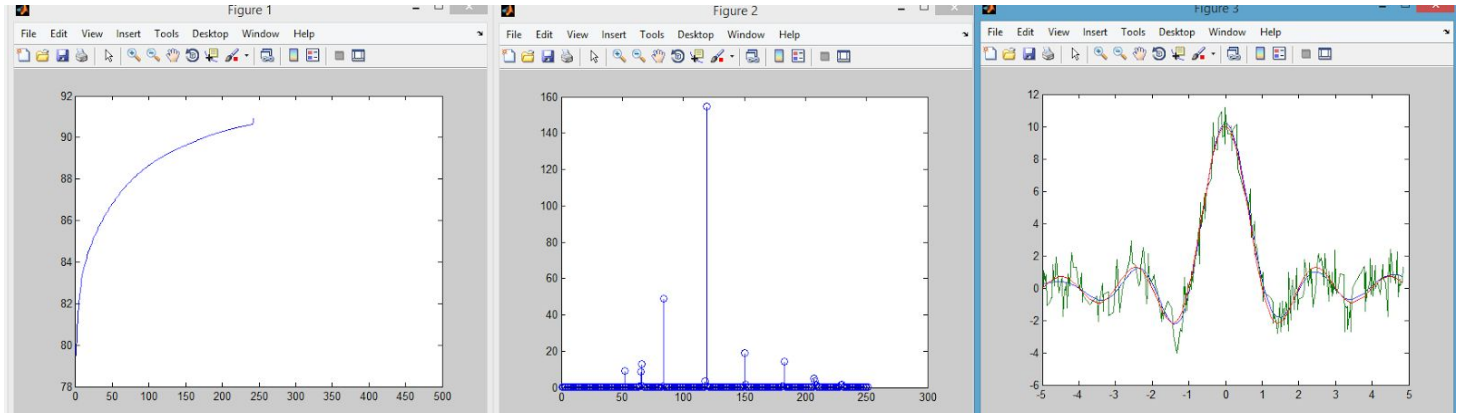
BMML Homework 3
Problem 2
DataSet 1

a) b) d)



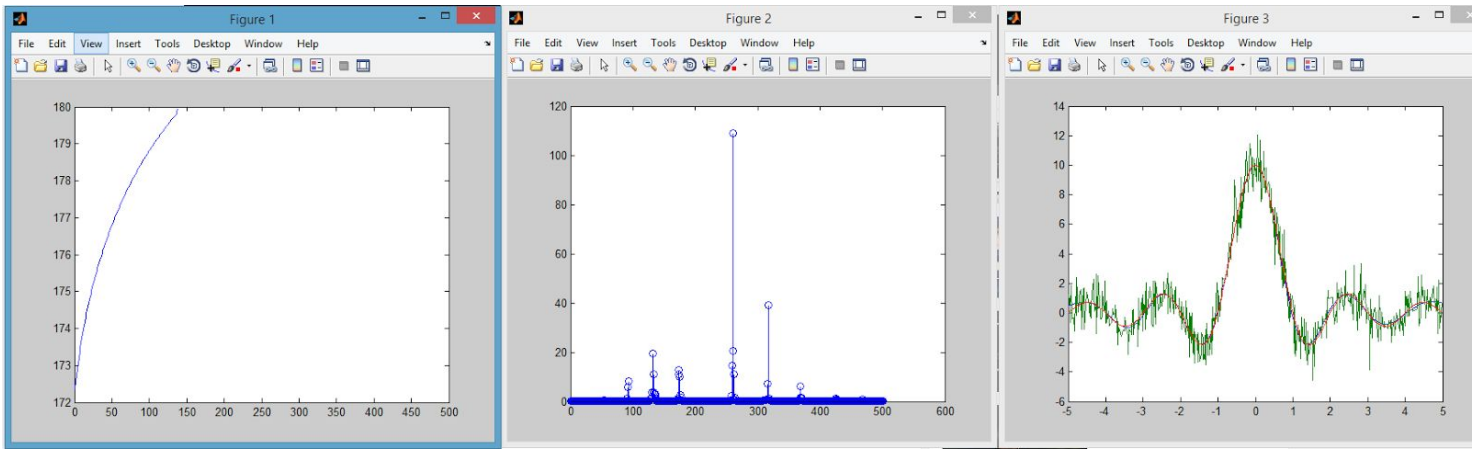
c) 1.0798

Data Set 2
a)b)d)



As you can see, it doesn't converge as t increases. I think modifications can be made to the objective function to fix this, as some values cause L to just go to infinity after a while
c) 0.8994

Data set 3
a)b)d)



c)
0.9781

This one doesn't converge either…

I used this as the main loop for my objective function

```
for k = 1:d
    b(k) = b0t(k) + 0.5*mumusigma(k,k);
    Elnpw = Elnpw + 0.5*(psi(a(k))-log(b(k))) - 0.5*a(k)./b(k)*mumusigma(k,k);
    Elnpalpha = Elnpalpha + (a0 - 1)*(psi(a(k))-log(b(k))) - b0*a(k)/b(k);
    Elnqalpha = Elnqalpha + log(gamma(a(k))) + (1 - a(k))*psi(a(k)) + a(k) - log(b(k));
    Eqalpha(k) = a(k)./b(k);
end


Elnpy = N/2*(psi(e)-log(f)) - 0.5*e/f*yxitmu;
Elnplambda = (e0 - 1)*(psi(e)-log(f)) - f0*e/f;
Elnqlambda = e - log(f) + (1 - e)*psi(e) +  gammaln(e);
Elnqw = 0.5*log(det(sigma));



L(i) = Elnpy + Elnpw + Elnpalpha + Elnplambda + Elnqw + Elnqalpha + Elnqlambda;
```