HW3

Tianbo Qiu tq2137

Problem 1

Model : $\{(x_i, y_i)\}_{i=1}^N$, $y_i \in R$, $x_i \in R^d$, $w \in R^d$

$y_i \overset{ind}{\sim} Normal(x_i^T w, \lambda^{-1})$. $p(y_i|x_i, w, \lambda) = \frac{\sqrt{\lambda}}{\sqrt{2\pi}} \exp\{-\frac{\lambda}{2}(y_i - x_i^T w)^2\}$, $\ln p(y_i|x_i, w, \lambda) = \frac{1}{2}\ln\lambda - \frac{1}{2}\ln 2\pi - \frac{\lambda}{2}(y_i - x_i^T w)^2$

$w \sim Normal(0, diag(\alpha_1, \cdots \alpha_d)^{-1})$. $p(w|\alpha_1, \cdots, \alpha_d) = \frac{\sqrt{\alpha_1 \cdots \alpha_d}}{\sqrt{2\pi^d}} \exp\{-\frac{1}{2} w^T \begin{bmatrix} \alpha_1 & & \\ & \alpha_2 & \\ & & \ddots \alpha_d \end{bmatrix} w\}$

$\ln p(w|\alpha_1, \cdots, \alpha_d) = \frac{1}{2}\sum_{i=1}^d \ln\alpha_i - \frac{d}{2}\ln 2\pi - \frac{1}{2} w^T \begin{bmatrix} \alpha_1 & & \\ & \ddots & \\ & & \alpha_d \end{bmatrix} w$

$\alpha_k \overset{iid}{\sim} Gamma(a_0, b_0)$. $p(\alpha_k) = \frac{b_0^{a_0}}{\Gamma(a_0)} \alpha_k^{a_0 - 1} e^{-b_0 \alpha_k}$. $\ln p(\alpha_k) = a_0 \ln b_0 - \ln\Gamma(a_0) + (a_0 - 1)\ln\alpha_k - b_0\alpha_k$

$\lambda \sim Gamma(e_0, f_0)$, $p(\lambda) = \frac{f_0^{e_0}}{\Gamma(e_0)} \lambda^{e_0 - 1} e^{-f_0 \lambda}$, $\ln p(\lambda) = e_0 \ln f_0 - \ln\Gamma(e_0) + (e_0 - 1)\ln\lambda - f_0 \ln\lambda$

$p(w, \alpha_1, \cdots, \alpha_d, \lambda | y, x) \approx q(w, \alpha_1, \cdots, \alpha_d, \lambda) = q(w) \prod_{i=1}^d q(\alpha_i) q(\lambda)$

Joint likelihood : $p(y, w, \lambda, \alpha_1, \cdots, \alpha_d | x) = p(y|x, w, \lambda) p(w|\alpha_1, \cdots, \alpha_d) \prod_{i=1}^d p(\alpha_i) p(\lambda)$

Log of Joint likelihood : $\ln p(y, w, \lambda, \alpha_1, \cdots \alpha_d | x) = \ln p(y|x, w, \lambda) + \ln p(w|\alpha_1, \cdots, \alpha_d) + \sum_{i=1}^d \ln p(\alpha_i) + \ln p(\lambda)$

$q(\lambda) \propto \exp\{\mathbb{E}_{q(w, \alpha_1, \cdots, \alpha_d)}[\ln p(y|x, \lambda, w) + \ln p(\lambda) + \ln p(w|\alpha_1, \cdots, \alpha_d) + \sum_{i=1}^d \ln p(\alpha_i)]\}$

$\propto \exp\{\mathbb{E}_{q(w)} \ln p(y|x, \lambda, w)\} p(\lambda) \propto \prod_{i=1}^N \exp\{\mathbb{E}_{q(w)}[\frac{1}{2}\ln\lambda - \frac{1}{2}\ln 2\pi - \frac{\lambda}{2}(y_i - x_i^T w)^2]\} p(\lambda)$

$\propto \{\prod_{i=1}^N \lambda^{\frac{1}{2}} \cdot e^{-\frac{\lambda}{2}\mathbb{E}_{q(w)}(y_i - x_i^T w)^2}\} \cdot \lambda^{e_0 - 1} \cdot e^{-f_0 \lambda}$ $\cancel{\propto \lambda^{\frac{N}{2} + e_0 - 1} \cdot e^{-\sum_{i=1}^n(\frac{1}{2}\mathbb{E}_{q(w)}[(y_i - x_i^T w)^2] + f_0)\lambda} \propto \lambda^{\frac{N}{2} + e_0 - 1} \cdot e}$

$= Gamma(e', f')$, where $e' = e_0 + \frac{N}{2}$, $f' = f_0 + \frac{1}{2}\sum_{i=1}^N \mathbb{E}_{q(w)}[(y_i - x_i^T w)^2]$

$q(\alpha_i) \propto \exp\{\mathbb{E}_{q(w, \alpha_{j \neq i}, \lambda)}[\ln p(y|x, \lambda, w) + \ln p(\lambda) + \ln p(w|\alpha_1, \cdots, \alpha_d) + \sum_{i=1}^d \ln p(\alpha_i)]\}$

$\propto \exp\{\mathbb{E}_{q(w, \alpha_{j \neq i})} \ln p(w|\alpha_1, \cdots \alpha_d) + \ln p(\alpha_i) + \mathbb{E}_{q(\alpha_{j \neq i})} \sum_{j \neq i} \ln \alpha_j\}$

$\propto \exp\{\mathbb{E}_{q(w, \alpha_{j \neq i})}[\frac{1}{2}\sum_{i=1}^d \ln\alpha_i - \frac{d}{2}\ln 2\pi - \frac{1}{2} w^T \begin{bmatrix} \alpha_1 & & \\ & \ddots & \\ & & \alpha_d \end{bmatrix} w\} p(\alpha_i) \propto \exp\{\frac{1}{2}\ln\alpha_i + \frac{1}{2}\sum_{j \neq i} \mathbb{E}_{q(\alpha_{j \neq i})} \ln\alpha_j - \frac{1}{2}\mathbb{E}_{q(w, \alpha_{j \neq i})} w^T \begin{bmatrix} \alpha_1 & & \\ & \ddots & \\ & & \alpha_d \end{bmatrix} w\} p(\alpha_i)$

$\mathbb{E}_{q(w)} w^T \begin{bmatrix} \alpha_1 & & \\ & \ddots & \\ & & \alpha_d \end{bmatrix} w = tr(\begin{bmatrix} \alpha_1 & & \\ & \ddots & \\ & & \alpha_d \end{bmatrix} \Sigma') + \mu'^T \begin{bmatrix} \alpha_1 & & \\ & \ddots & \\ & & \alpha_d \end{bmatrix} \mu' = \cancel{\sum_{i=1}^d \alpha_i(\Sigma'_{ii} + \mu'^2_i)} \sum_{i=1}^d \alpha_i(\Sigma'_{ii} + \mu'^2_i)$

Therefore. $q(\alpha_i) \propto \alpha_i^{\frac{1}{2}} \cdot \exp\{-\frac{1}{2}(\alpha_i(\Sigma'_{ii} + \mu'^2_i) + \sum_{j \neq i} \mathbb{E}_{q(\alpha_{j \neq i})} \alpha_j (\Sigma'_{jj} + \mu'_j)\} \alpha_i^{a_0 - 1} \cdot e^{-b_0 \alpha_i}$

$= Gamma(a', b')$, where . $a' = \frac{1}{2} + a_0$. $b' = b_0 + \frac{1}{2}(\Sigma'_{ii} + \mu'^2_i)$

$$q(w) \propto \exp\Big\{ \mathbb{E}_{q(\lambda,\alpha_1,\dots,\alpha_d)} \Big[\ln p(y|x,w,\lambda) + \ln p(\lambda) + \ln p(w|\alpha_1,\dots,\alpha_d) + \sum_{i=1}^{d} \ln p(\alpha_i) \Big] \Big\}$$

$$\propto \exp\Big\{ \sum_{i=1}^{N} \mathbb{E}_{q(\lambda)} \ln p(y_i|x_i,\lambda,w) + \mathbb{E}_{q(\alpha_1,\dots,\alpha_d)} p(w|\alpha_1,\dots,\alpha_d) \Big\}$$

$$\propto \cancel{ex} \prod_{i=1}^{N} \exp \mathbb{E}_{q(\lambda)} \Big[\frac{1}{2}\ln\lambda - \frac{1}{2}\ln 2\pi - \frac{\lambda}{2}(y_i - x_i^T w)^2 \Big] \cdot \exp\Big\{ \mathbb{E}_{q(\alpha_1,\dots,\alpha_d)} \Big[\frac{1}{2}\sum_{i=1}^{d}\ln\alpha_i - \frac{d}{2}\ln 2\pi - \frac{1}{2} w^T \begin{bmatrix} \alpha_1 & & \\ & \ddots & \\ & & \alpha_d \end{bmatrix} w \Big] \Big\}$$

$$\propto \prod_{i=1}^{N} \exp\Big\{ -\frac{\mathbb{E}_{q(\lambda)}[\lambda]}{2} (y_i - x_i^T w)^2 \Big\} \cdot \exp\Big\{ -\frac{1}{2} \mathbb{E}_{q(\alpha_1,\dots,\alpha_d)} w^T \begin{bmatrix} \alpha_1 & & \\ & \ddots & \\ & & \alpha_d \end{bmatrix} w \Big\}$$

$$\propto \exp\Big\{ -\frac{\mathbb{E}_{q(\lambda)}[\lambda]}{2} \sum_{i=1}^{N} (y_i - x_i^T w)^2 \Big\} \cdot \exp\Big\{ -\frac{1}{2} \begin{bmatrix} \mathbb{E}_{q(\alpha_1)}[\alpha_1] & & \\ & \ddots & \\ & & \mathbb{E}_{q(\alpha_d)}[\alpha_d] \end{bmatrix} w^T w \Big\}$$

$= \text{Normal}(\mu', \Sigma')$, where $\Sigma' = \left(\begin{bmatrix} \mathbb{E}_{q(\alpha_1)}[\alpha_1] & & \\ & \ddots & \\ & & \mathbb{E}_{q(\alpha_d)}[\alpha_d] \end{bmatrix} + \mathbb{E}_{q(\lambda)}[\lambda] \sum_{i=1}^{N} x_i x_i^T \right)^{-1}$, $\mu' = \Sigma' \left(\mathbb{E}_{q(\lambda)}[\lambda] \sum_{i=1}^{N} y_i x_i \right)$

Variational Inference Algorithm:

First, we define $q(\lambda) = \text{Gamma}(\lambda|e', f')$. $q(\alpha_i) =$ ~~Gamma (α~~ $\text{Gamma}(\alpha_i | a_i', b_i')$ $(i=1,2,\dots d)$

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

Then, we initialize $e', f', \mu', \Sigma', a_i', b_i'$ $(i=1,2,\dots,d)$

For every iteration, we update $q(\lambda)$ by setting $e'_t = e_0 + \frac{N}{2}$, $f'_t = f_0 + \frac{1}{2}\sum_{i=1}^{N} \mathbb{E}_{q(w)_{t-1}}[(y_i - x_i^T w)^2]$

update $q(\alpha_i)$ $i=1,2,\dots d$ by setting: $a'_{it} = \frac{1}{2} + a_0$

$b'_{it} = b_0 + \frac{1}{2}\left(\Sigma'_{(t-1)ii} + \mu'^{2}_{i(t-1)} \right)$

update $q(w)$ by setting $\Sigma'_t = \left(\begin{bmatrix} \mathbb{E}_{q(\alpha_1)(t)}[\alpha_1] & & \\ & \ddots & \\ & & \mathbb{E}_{q(\alpha_d)(t)}[\alpha_d] \end{bmatrix} + \mathbb{E}_{q(\lambda)(t)}[\lambda] \sum_{i=1}^{N} x_i x_i^T \right)^{-1}$

$\mu'_t = \Sigma'_t \left(\mathbb{E}_{q(\lambda)(t)}[\lambda] \sum_{i=1}^{N} y_i x_i \right)$

Then evaluate objective function $\mathcal{L}(a'_t, b'_t, \mu'_t, \Sigma'_t, e'_t, f'_t)$ to assess convergence.

b) VI algorithm:

Inputs: We define $q(\lambda) = \text{Gamma}(\lambda \mid e', f')$, $q(\alpha_i) = \text{Gamma}(\alpha_i \mid a_i', b_i')$, $i = 1, 2, \cdots, d$

$q(w) = \text{Normal}(w \mid \mu', \Sigma')$

Output: Values for $e', f', \mu', \Sigma', a_i', b_i'$ $(i = 1, 2, \cdots, d)$

1. Initialize $e_0', f_0', \mu_0', \Sigma_0', a_{0i}', b_{0i}'$ $(i = 1, 2, \cdots, d)$ in some way.

2. For iteration $t = 1, \cdots, T$

- Update $q(\lambda)$ by setting: $e_t' = e_0 + \frac{N}{2}$, $f_t' = f_0 + \frac{1}{2}\sum_{i=1}^{N}(y_i - x_i^T \mu'_{(t-1)})^2 + x_i^T \Sigma'_{(t-1)} x_i$

For $i = 1, \cdots, d$:

- Update $q(\alpha_i)$ by setting: $a_{it}' = a_0 + \frac{1}{2}$, $b_{it}' = b_0 + \frac{1}{2}(\Sigma'_{t-1,ii} + \mu'^2_{t-1,i})$

- Update $q(w)$ by setting: $\cancel{\mu_t'}$ $\Sigma_t' = \left(\begin{bmatrix} a_{1t}'/b_{1t}' & & \\ & \ddots & \\ & & a_{dt}'/b_{dt}' \end{bmatrix} + \frac{e_t'}{f_t'} \sum_{i=1}^{N} x_i x_i^T \right)^{-1}$

$$\mu_t' = \Sigma_t' \cdot \left(\frac{e_t'}{f_t'} \sum_{i=1}^{N} y_i x_i \right)$$

- Evaluate $\mathcal{L}(e_t', f_t', \mu_t', \Sigma_t', a_{1t}', b_{1t}', \cdots, a_{dt}', b_{dt}')$ to assess convergence.

c) $\mathcal{L}(e', f', a', b', \mu', \Sigma') = \int q(w) q(\lambda) \prod_{i=1}^{d} q(\alpha_i) \ln \frac{p(y, w, \lambda, \alpha_1, \cdots, \alpha_d \mid x)}{q(w) q(\lambda) \prod_{i=1}^{d} q(\alpha_i)} \, dw \, d\lambda \, d\alpha_1 \cdots d\alpha_d$

$$= \underbrace{\mathbb{E}_{q(w,\lambda)} \ln p(y \mid x, w, \lambda)}_{\mathcal{L}_1} + \underbrace{\mathbb{E}_{q(w, \alpha_1, \cdots, \alpha_d)} \ln p(w \mid \alpha_1, \cdots, \alpha_d)}_{\mathcal{L}_2} + \underbrace{\sum_{i=1}^{d} \mathbb{E}_{q(\alpha_i)} \ln p(\alpha_i)}_{\mathcal{L}_3} + \underbrace{\mathbb{E}_{q(\lambda)} \ln p(\lambda)}_{\mathcal{L}_4} \underbrace{- \int q(w) \ln q(w) dw - \int q(\lambda) \ln q(\lambda) d\lambda - \sum_{i=1}^{d} \int q(\alpha_i) \ln q(\alpha_i) d\alpha_i}_{\mathcal{L}_5}$$

$$\mathcal{L}_1 = \sum_{i=1}^{N} \mathbb{E}_{q(w,\lambda)} \left[\frac{1}{2} \ln \lambda - \frac{1}{2} \ln 2\pi - \frac{\lambda}{2} (y_i - x_i^T w)^2 \right] = \sum_{i=1}^{N} \left[\frac{1}{2} \mathbb{E}_{q(\lambda)} \ln \lambda - \frac{1}{2} \ln 2\pi - \frac{\mathbb{E}_{q(\lambda)}[\lambda]}{2} \mathbb{E}_{q(w)}[(y_i - x_i^T w)]^2 \right]$$

, where $\mathbb{E}_{q(\lambda)} \ln \lambda = \psi(e') - \ln(f')$, $\mathbb{E}_{q(w)}[(y_i - x_i^T w)^2] = (y_i - x_i^T \mu')^2 + x_i^T \Sigma' x_i$

$$\mathcal{L}_2 = \mathbb{E}_{q(w, \alpha_1, \cdots, \alpha_d)} \left[\frac{1}{2} \sum_{i=1}^{d} \ln \alpha_i - \frac{d}{2} \ln 2\pi - \frac{1}{2} w^T \begin{bmatrix} \alpha_1 & & \\ & \ddots & \\ & & \alpha_d \end{bmatrix} w \right] = \frac{1}{2} \sum_{i=1}^{d} \mathbb{E}_{q(\alpha_i)} \ln(\alpha_i) - \frac{d}{2} \ln 2\pi - \frac{1}{2} \sum_{i=1}^{d} \mathbb{E}_{q(\alpha_i)}(\alpha_i) (\Sigma'_{ii} + \mu_i'^2)$$

, where $\mathbb{E}_{q(\alpha_i)} \ln(\alpha_i) = \psi(a_i') - \ln(b_i')$, $\mathbb{E}_{q(\alpha_i)}(\alpha_i) = \frac{a_i'}{b_i'}$

$$\mathcal{L}_3 = \sum_{i=1}^{d} \mathbb{E}_{q(\alpha_i)} \ln p(\alpha_i) = \sum_{i=1}^{d} \mathbb{E}_{q(\alpha_i)} [a_0 \ln b_0 - \ln \Gamma(a_0) + (a_0 - 1) \ln \alpha_i - b_0 \alpha_i]$$

$$= \sum_{i=1}^{d} [a_0 \ln b_0 - \ln \Gamma(a_0) + (a_0 - 1) \mathbb{E}_{q(\alpha_i)} \ln \alpha_i - b_0 \mathbb{E}_{q(\alpha_i)}(\alpha_i)]$$

$$\mathcal{L}_4 = \mathbb{E}_{q(\lambda)} [e_0 \ln f_0 - \ln \Gamma(e_0) + (e_0 - 1) \ln \lambda - f_0 \lambda] = e_0 \ln f_0 - \ln \Gamma(e_0) + (e_0 - 1) \mathbb{E}_{q(\lambda)} \ln \lambda - f_0 \mathbb{E}_{q(\lambda)}[\lambda]$$

$$\mathcal{L}_2 = -\int q(w)\ln q(w)\,dw - \int q(\lambda)\ln q(\lambda)\,d\lambda - \sum_{i=1}^{d}\int q(\alpha_i)\ln q(\alpha_i)\,d\alpha_i$$

$$= \frac{1}{2}\ln\det(2\pi e\Sigma') + [e' - \ln f' + \ln\Gamma(e') + (1-e')\psi(e')] + \sum_{i=1}^{d}[a_i' - \ln b_i' + \ln\Gamma(a_i') + (1-a_i')\psi(a_i'$$

Therefore,

$$\mathcal{L}(e', f', a', b', \mu', \Sigma') = \sum_{i=1}^{N}\Big[\frac{1}{2}\mathop{\mathbb{E}}_{q(\lambda)}\ln\lambda - \frac{1}{2}\ln 2\pi - \frac{\mathbb{E}_{q(\lambda)}[\lambda]}{2}\mathop{\mathbb{E}}_{q(w)}[(y_i - x_i^T w)^2]\Big]$$

$$+ \frac{1}{2}\sum_{i=1}^{d}\mathop{\mathbb{E}}_{q(\alpha_i)}\ln(\alpha_i) - \frac{d}{2}\ln 2\pi - \frac{1}{2}\sum_{i=1}^{d}\mathop{\mathbb{E}}_{q(\alpha_i)}(\alpha_i)(\Sigma'_{ii} + \mu_i'^2)$$

$$+ \sum_{i=1}^{d}\Big[a_0\ln b_0 - \ln\Gamma(a_0) + (a_0 - 1)\mathop{\mathbb{E}}_{q(\alpha_i)}\ln\alpha_i - b_0\mathop{\mathbb{E}}_{q(\alpha_i)}[\alpha_i]\Big]$$

$$+ e_0\ln f_0 - \ln\Gamma(e_0) + (e_0 - 1)\mathop{\mathbb{E}}_{q(\lambda)}\ln\lambda - f_0\mathop{\mathbb{E}}_{q(\lambda)}[\lambda]$$

$$+ \frac{1}{2}\ln\det(2\pi e\Sigma') + [e' - \ln f' + \ln\Gamma(e') + (1-e')\psi(e')]$$

$$+ \sum_{i=1}^{d}[a_i' - \ln b_i' + \ln\Gamma(a_i') + (1-a_i')\psi(a_i')]$$

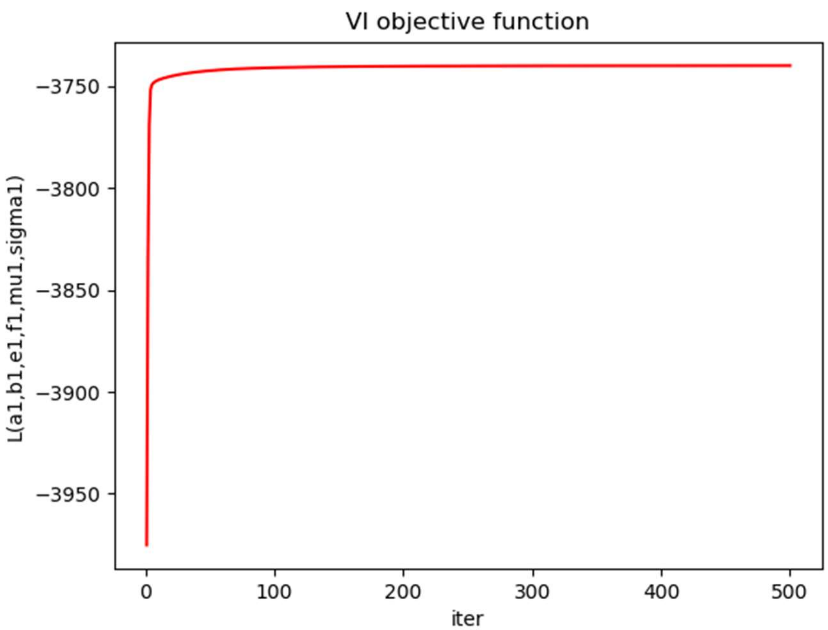
where $\mathop{\mathbb{E}}_{q(\lambda)}\ln\lambda = \psi(e') - \ln(f')$, $\mathop{\mathbb{E}}_{q(\lambda)}[\lambda] = \frac{e'}{f'}$, $\mathop{\mathbb{E}}_{q(w)}[(y_i - x_i^T w)^2] = (y_i - x_i^T\mu')^2 + x_i^T\Sigma' x_i$

$\mathop{\mathbb{E}}_{q(\alpha_i)}\ln(\alpha_i) = \psi(a_i') - \ln(b_i')$, $\mathop{\mathbb{E}}_{q(\alpha_i)}[\alpha_i] = \frac{a_i'}{f_i'}$

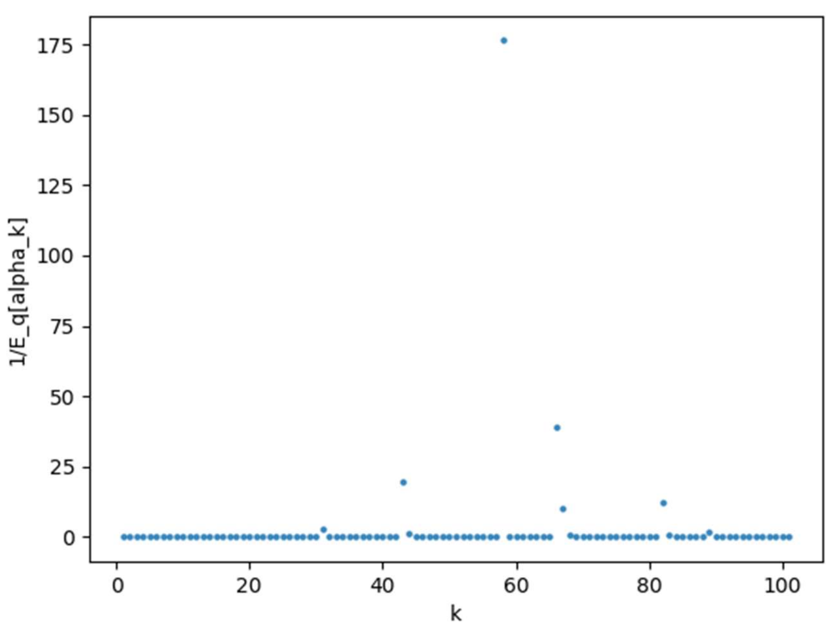
Problem 2

Dataset 1:

a)



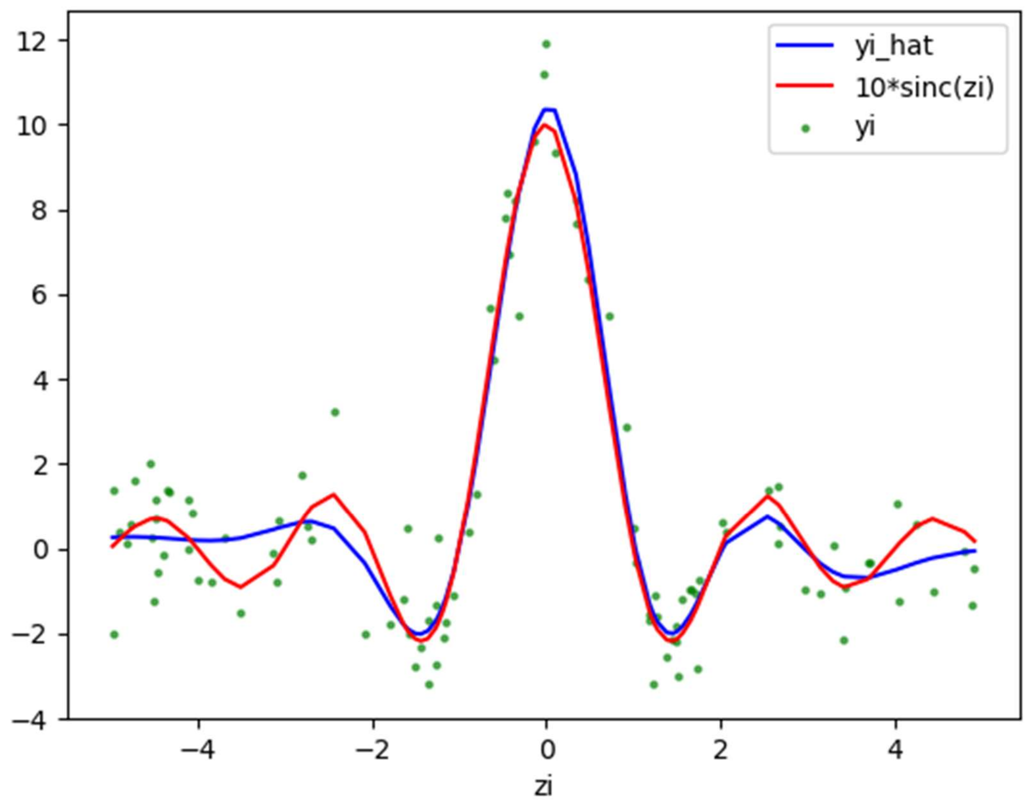
b)



c)

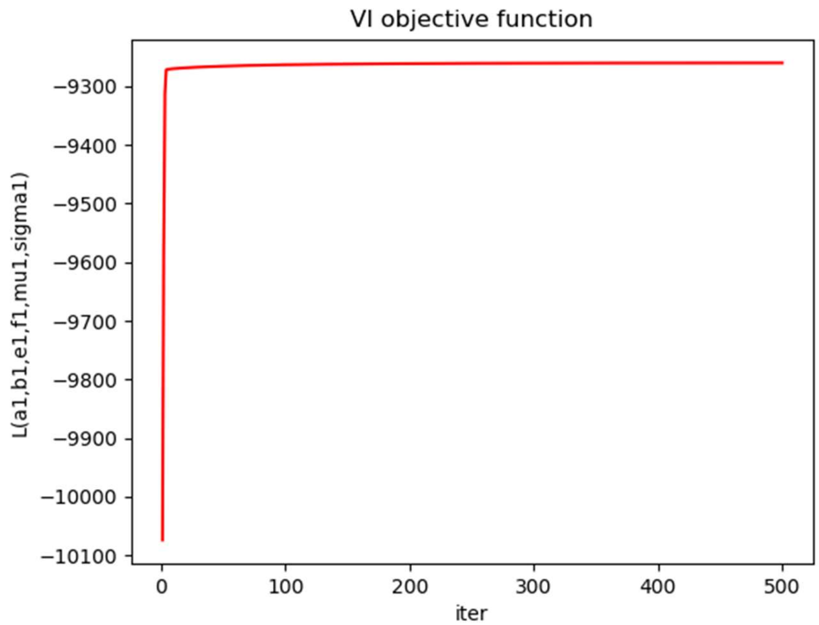
1 / $E_q[\lambda]$ = 1.07983307889

d)



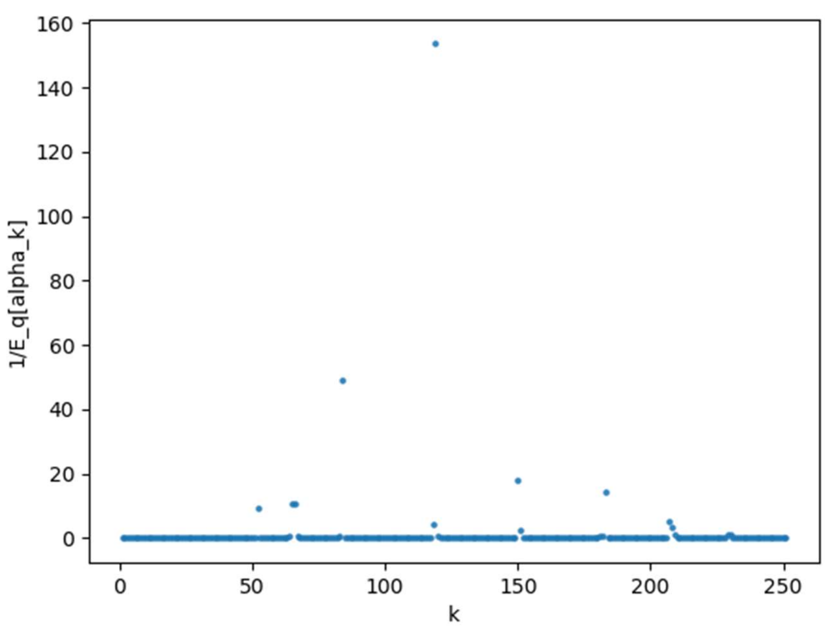
Dataset2:

a)



VI objective function
L(a1,b1,e1,f1,mu1,sigma1)
iter

b)

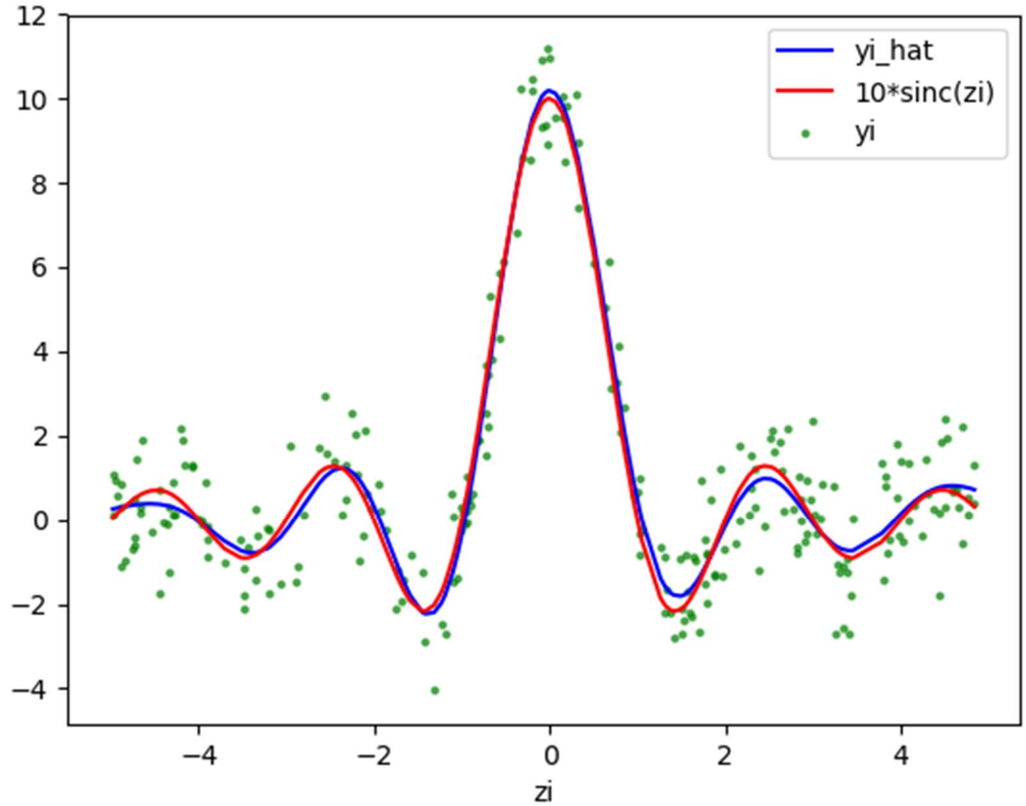


1/E_q[alpha_k]
k

c)

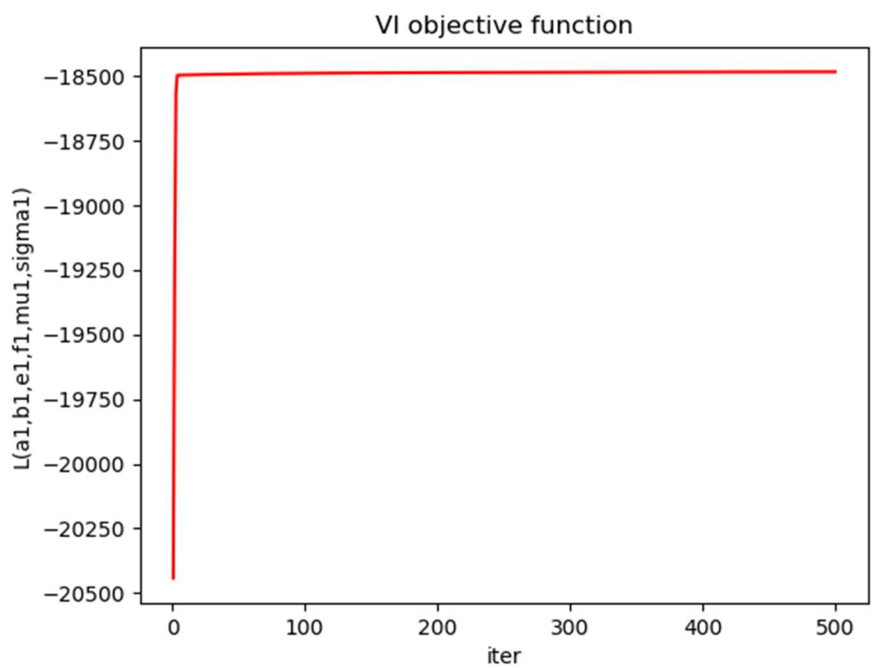
$1 / E_q[\lambda]$ =0.899465140248

d)

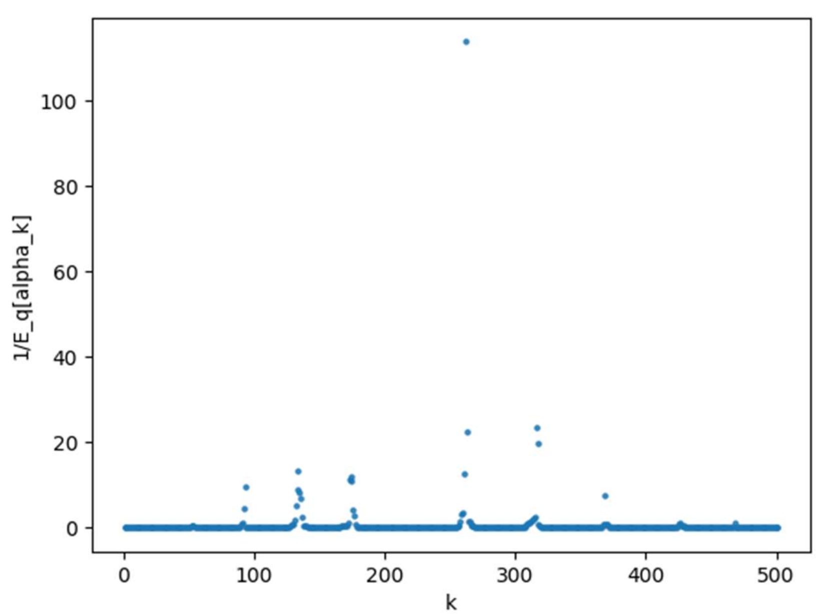


Dataset3:

a)


VI objective function
L(a1,b1,e1,f1,mu1,sigma1)
−18500
−18750
−19000
−19250
−19500
−19750
−20000
−20250
−20500
0
100
200
300
400
500
iter

b)


1/E_q[alpha_k]
100
80
60
40
20
0
0
100
200
300
400
500
k

c)

1 / $E_q[\lambda]$ =0.978142481386

d)