Problem 1

a.

To approach the optimal distribution of each q distribution,

$$q(\lambda) \propto \exp\left\{ E_{-q(\lambda)}[\ln p(Y|w, a_{1:n}, \lambda, X)] + \ln q(\lambda) \right\}$$

$$\propto \exp\left\{ E_{q(w)}\left[\frac{n}{2}\ln\lambda - \frac{\lambda}{2}(Y - X^T w)^T (Y - X^T w)\right] + (e_0 - 1)\ln\lambda - f_0 \lambda \right\}$$

$$\propto \exp\left\{ \left(\frac{n}{2} + e_0 - 1\right)\ln\lambda - \frac{1}{2}\left(Y^T Y - 2Y^T X^T E_{q(w)}[w] + tr(E_{q(w)}[ww^T]XX^T) + 2f_0\right)\lambda \right\}$$

$= \text{Gamma}(e_0', f_0')$ where $e_0' = \frac{n}{2} + e_0$, $f_0' = \frac{1}{2}(Y^T Y - 2Y^T X^T E_{q(w)}[w] + tr(E_{q(w)}[ww^T]XX^T) + 2f_0)$

$$q(a_k) \propto \exp\left\{ E_{-q(a_k)}[\ln p(w_k|a_k)] + \ln q(a_k) \right\}$$

$$\propto \exp\left\{ E_{q(w_k)}\left[\frac{1}{2}\ln a_k - \frac{a_k}{2}(w_k^2)\right] + (a_0 - 1)\ln a_k - b_0 a_k \right\}$$

$$\propto \exp\left\{ \left(\frac{1}{2} + a_0 - 1\right)\ln a_k - \frac{1}{2}\left(E_{q(w_k)}[w_k^2] + 2b_0\right) a_k \right\}$$

$= \text{Gamma}(a_0^{(k)'}, b_0^{(k)'})$ where

$$a_0^{(k)'} = \frac{1}{2} + a_0, \quad b_0^{(k)'} = \frac{1}{2}(E_{q(w_k)}[w_k^2] + 2b_0)$$

$$q(w) \propto \exp\left\{ E_{-q(w)}[\ln p(Y|w, a_1, \dots a_n, \lambda, X) + \ln q(w)] \right\}$$

$$\propto \exp\left\{ E_{q(\lambda)}\left[-\frac{\lambda}{2}(Y - X^T w)^T (Y - X^T w)\right] + E_{q(a_{1:d})}\left[-\frac{A}{2} w^T w\right] \right\}$$

Where $A = \text{diag}(a_1, \dots, a_d)$

$$\propto \exp\left\{ -\frac{1}{2} w^T (E_{q(\lambda)}[\lambda]XX^T + E_{q(a_{1:d})}[A]) w - 2w^T E_{q(\lambda)}[\lambda] XY \right\}$$

$$\propto \exp\left\{ -\frac{1}{2}\left(w^T M w - 2w^T M E_{q(\lambda)}[\lambda] M^{-1} XY + (E_{q(\lambda)}[\lambda] M^{-1} XY)^T M (E_{q(\lambda)}[\lambda] M^{-1} XY)\right) \right\}$$

$= N(E_{q(\lambda)}[\lambda] M^{-1} XY, M^{-1})$

Expections are defined as following:

$E_{q(\lambda)}[\lambda] = e_0' / f_0'$

$E_{q(a_{1:d})}[A] = \text{diag}(a_0^{(1)'} / b_0^{(1)'}, \dots, a_0^{(d)'} / b_0^{(d)'})$

$E_{q(w)}[w] = E_{q(\lambda)}[\lambda] M^{-1} X Y$

$E_{q(w_k)}[w_k^2] = (E_{q(w)}[w])_k^2 + M_{kk}^{-1}$

$E_{q(w)}[ww^T] = M^{-1} + E_{q(w)}[w] E_{q(w)}[w]^T$

b. Set a VI algorithm using pseudo-code:

First: Initialization for a_0', b_0', e_0', f_0', μ_0' and Σ_0'

Second:

For iteration $t = 1, \dots, T$:

1. Update $q(\lambda)$. Let

$$e_t' = \frac{n}{2} + e_0$$

$$f_t' = \frac{1}{2}(Y^TY - 2Y^TX^TE_{q(w)}[W_t] + tr(E_{q(w)}[W_tW_t^T]XX^T) + 2f_0)$$

Where $E_{q(w)}[W_t] = \mu_{t-1}$

$$E_{q(w)}[W_tW_t^T] = \Sigma_{t-1} + E_{q(w)}[W_t]E_{q(w)}[W_t]^T$$

2. Update each $q(\alpha_k)$. Let

$$a_t^{(k)'} = \frac{1}{2} + a_0$$

$$b_t^{(k)'} = \frac{1}{2}(E_{q(w_k)}[W_{kt}^2] + 2b_0)$$ where $E_{q(w_k)}[W_{kt}^2] = (E_{q(w)}[W_t])_k^2 + (\Sigma_{t-1})_{kk}$

3. Update $q(w)$. Let

$$M_t' = (E_{q(\lambda)}[\lambda_t]XX^T + E_{q(\alpha_{1:d})}[A_t])$$

$$\Sigma_t' = M_t^{-1}$$

$$\mu_t' = E_{q(\lambda)}[\lambda_t]M_t^{-1}XY$$ where $E_{q(\lambda)}[\lambda_t] = e_t'/f_t'$ and $E_{q(\alpha_{1:d})}[A_t] = diag(a_t^{(1)'}/b_t^{(1)'}, \dots, a_t^{(d)'}/b_t^{(d)'})$

4. Assess convergence while calculating

$$L(a_t', b_t', e_t', f_t', \mu_t', \Sigma_t')$$

c. $L((a_t^{(1)'}, b_t^{(1)'}), L((a_t^{(2)'}, b_t^{(2)'}), \dots (a_t^{(d)'}, a_t^{(d)'}), e_t', f_t', \mu_t', \Sigma_t')$

$$= \int_w\int_\lambda\int_{\alpha_1,\alpha_2\dots\alpha_d} q(w)q(\lambda)\prod_{k=1}^{d} q(\alpha_k)\ln\frac{p(w,\alpha_{1:d},\lambda \mid Y,X)}{q(w)q(\lambda)\prod_{k=1}^{d}q(\alpha_k)}\,d\alpha_{1:d}\,d\lambda\,dw$$

$$= \int_w\int_{\alpha_{1:d}} q(w)\ln p(w)\,d\alpha_{1:d}\,dw + \int_\lambda q(\lambda)\ln p(\lambda)\,d\lambda + \int_{\alpha_{1:d}}\prod_{k=1}^{d}q(\alpha_k)\ln p(\alpha_k)\,d\alpha_{1:d}$$

$$+ \int_w\int_\lambda q(w)q(\lambda)\ln p(Y\mid w,\lambda,X)\,d\lambda\,dw - \int_w q(w)\ln q(w)\,dw - \int_\lambda q(\lambda)\ln q(\lambda)\,d\lambda -$$

$$\int_{\alpha_{1:d}}\prod_{k=1}^{d}q(\alpha_k)\ln q(\alpha_k)\,d\alpha_{1:d}$$

For each term,

$$\int_\lambda q(\lambda)\ln p(\lambda)\,d\lambda = e_0\ln f_0 - \ln\Gamma(e_0) + (e_0-1)E_{q(\lambda)}[\ln\lambda] - f_0 E_{q(\lambda)}[\lambda]$$
$$= e_0\ln f_0 - \ln\Gamma(e_0) + (e_0-1)(\psi(e_t') - \ln f_t') - f_0\frac{e_t'}{f_t'}$$

$$\int_w\int_{a_{1:d}} q(w)\ln p(w)\,dw = -\frac{d}{2}\ln 2\pi + \frac{1}{2}\sum_{i=k}^{d} E_{a_{1:d}}[\ln a_k] - \frac{1}{2}E_{q(w),q(a_{1:d})}[w_t^T A w_t]$$
$$= -\frac{d}{2}\ln 2\pi + \frac{1}{2}\sum_{i=k}^{d}(\psi(a_t^{(k)'}) - \ln b_t^{(k)'}) - \frac{1}{2}tr((\mu_t'\mu_t'^T + \Sigma_t)E_{q(a_{1:d})}A)$$

$$\int_{a_{1:d}}\prod_{k=1}^{d} q(a_k)\ln p(a_k)\,da_{1:d}$$
$$= \sum_{k=1}^{d}\int_{a_k}(a_k)\ln p(a_k)\,da_k$$
$$= \sum_{k=1}^{d} a_0\ln b_0 - \ln\Gamma(a_0) + (a_0-1)(\psi(a_t^{(k)'}) - \ln b_t^{(k)'}) - b_0\frac{a_t^{(k)'}}{b_t^{(k)'}}$$

$$\int_w q(w)\ln q(w)\,dw = \frac{1}{2}\ln((2\pi e)^n|\Sigma_t|)$$

$$\int_\lambda q(\lambda)\ln q(\lambda)\,d\lambda = e_t'\ln f_t' - \ln\Gamma(e_t') + (e_t'-1)(\psi(e_t') - \ln f_t') - f_t'\frac{e_t'}{f_t'}$$
$$= \ln f_t' - \ln\Gamma(e_t') + (e_t'-1)\psi(e_t') - e_t'$$

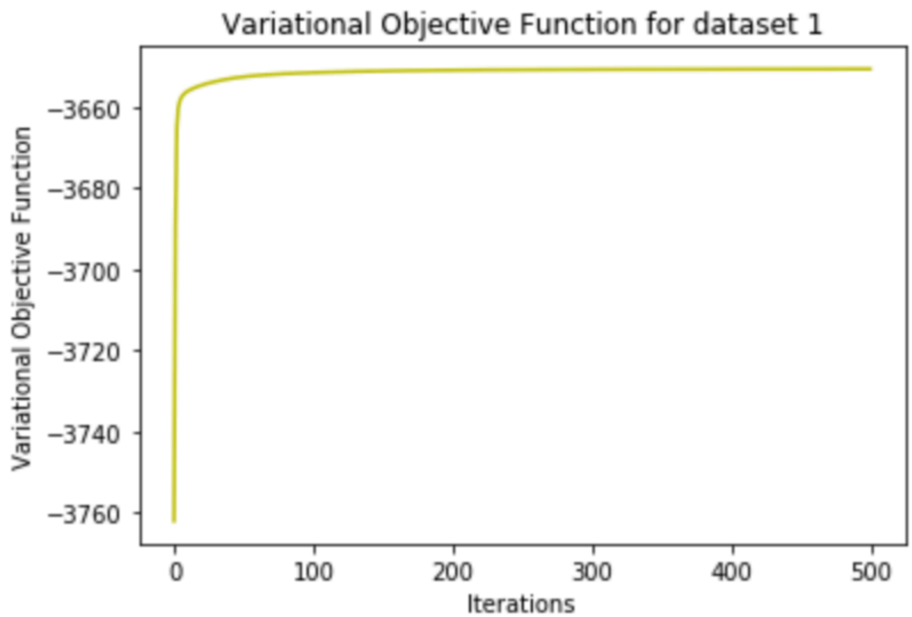
$$\int_{a_{1:d}}\prod_{k=1}^{d} q(a_k)\ln q(a_k)\,da_{1:d} = \sum_{k=1}^{d}\ln b_t^{(k)'} - \ln\Gamma(a_t^{(k)'}) + (a_t^{(k)'}-1)\psi(a_t^{(k)'}) - a_t^{(k)'}$$

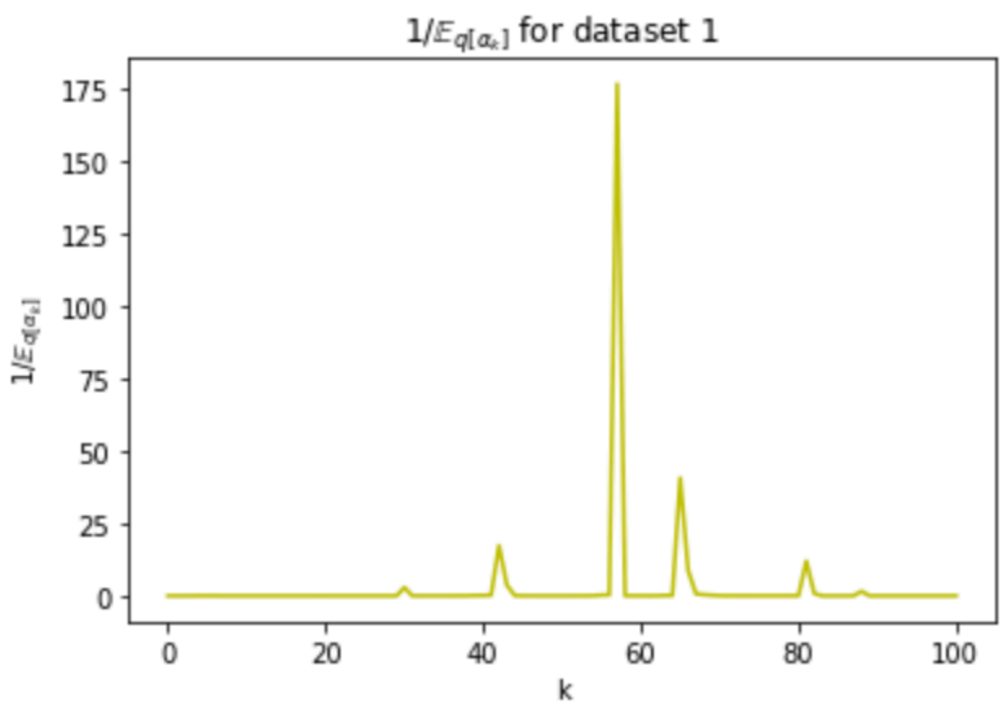
As a result,

$$\int_w\int_\lambda q(w)q(\lambda)\ln p(Y|w,\lambda,X)\,d\lambda\,dw.$$
$$= -\frac{n}{2}\ln 2\pi + \frac{n}{2}E_{q(\lambda)}[\ln\lambda] - \frac{E_{q(\lambda)}[\lambda]}{2}E_{q(w)}[Y - X^Tw)^T(Y - X^Tw)]$$
$$= -\frac{n}{2}\ln 2\pi + \frac{n}{2}(\psi(e_t') - \ln f_t') - \frac{e_t'}{2f_t'}(Y^TY - 2Y^TX^T\mu_t' - tr((\Sigma_t' + \mu_t'\mu_t'^T)XX^T))$$

Problem 2:

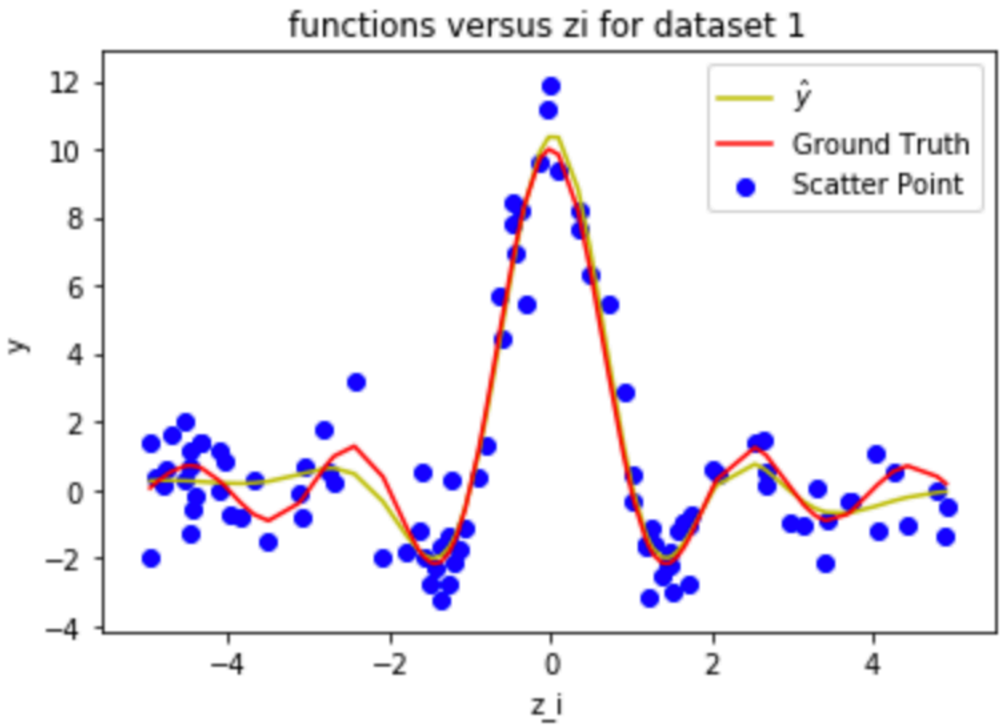
For dataset 1:



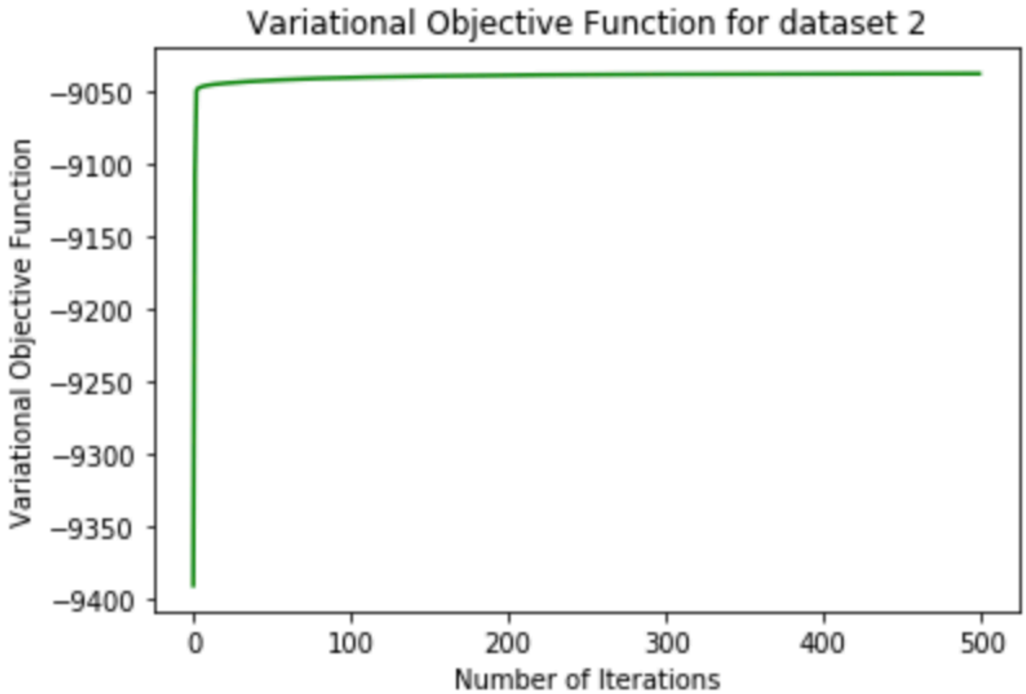


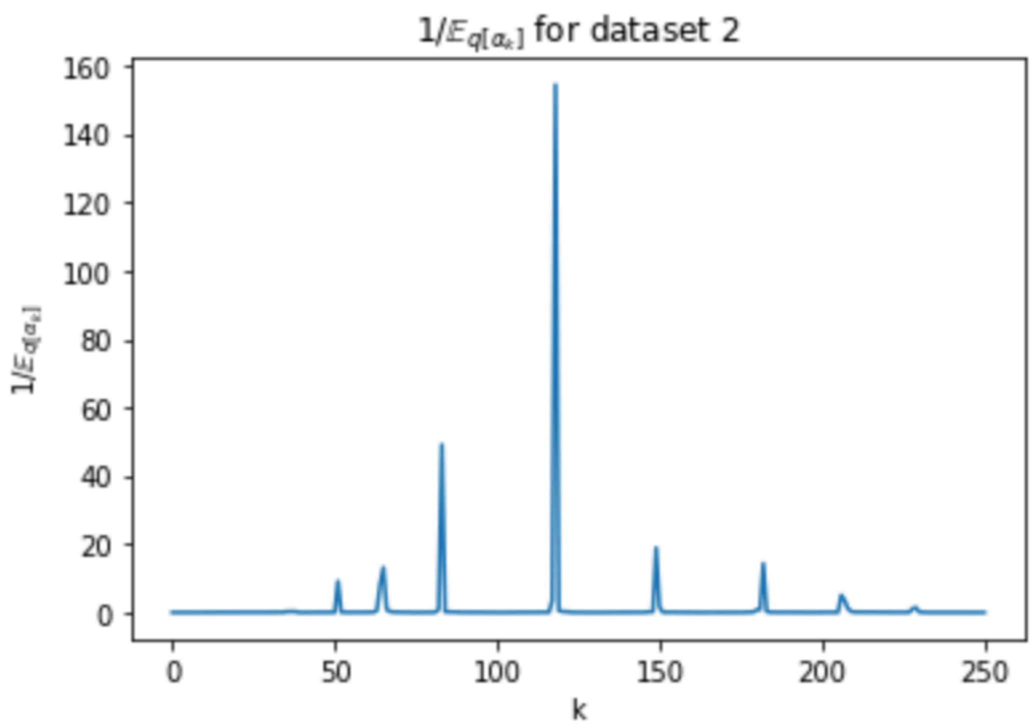
```
f_1/e_1
```

```
1.0800490653683776
```



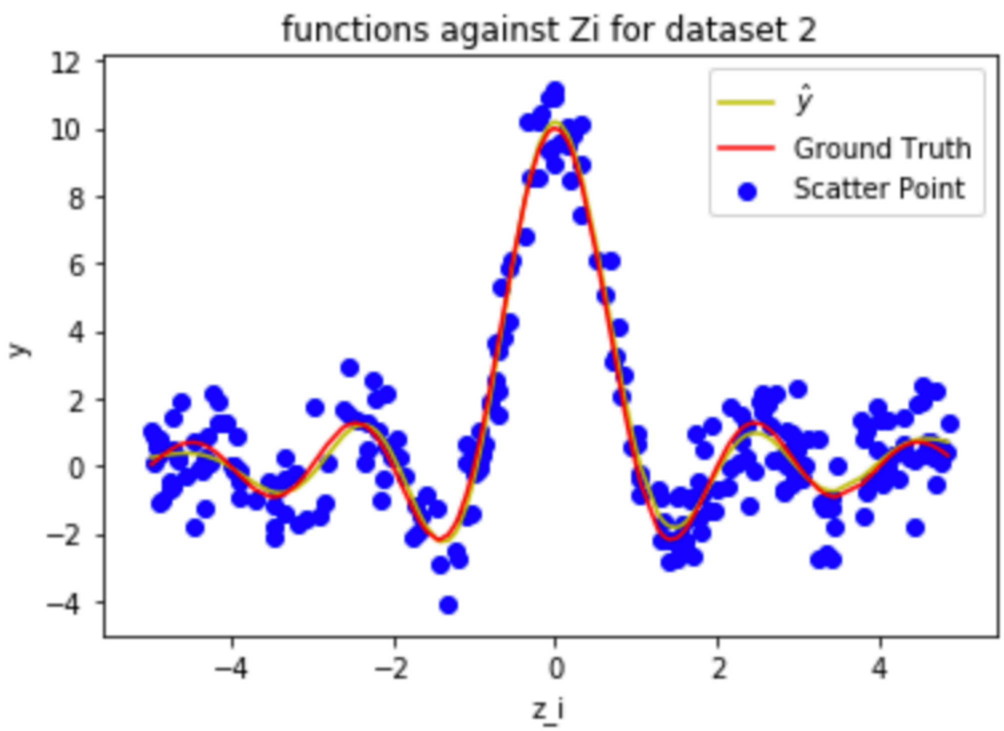
For dataset 2:


Variational Objective Function for dataset 2
Variational Objective Function
−9050
−9100
−9150
−9200
−9250
−9300
−9350
−9400
0
100
200
300
400
500
Number of Iterations

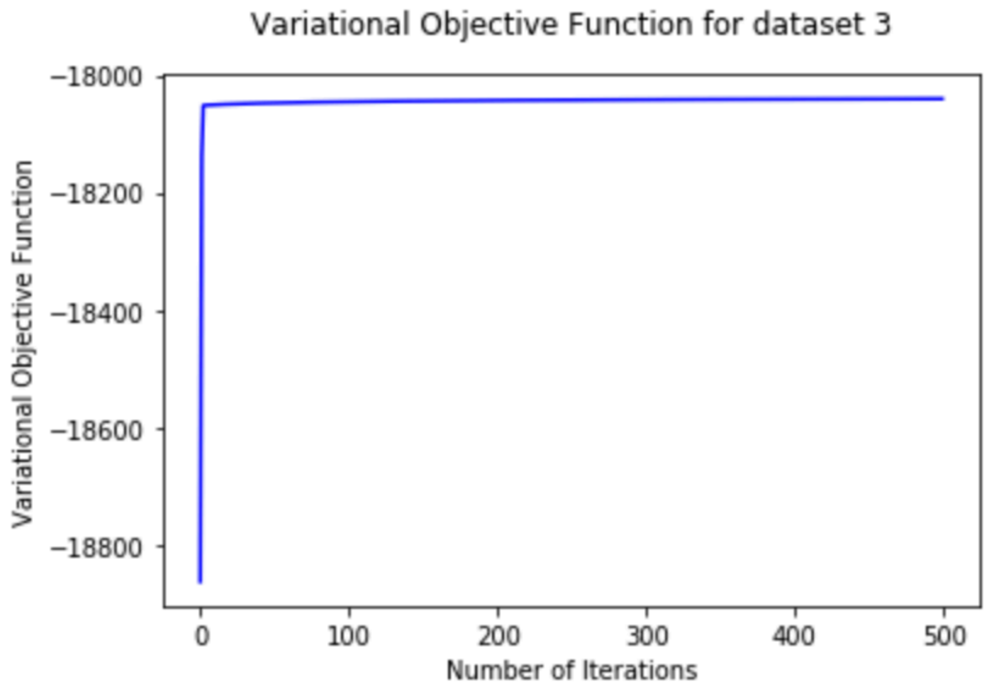

$1/\mathbb{E}_{q[\alpha_k]}$ for dataset 2
$1/\mathbb{E}_{q[\alpha_k]}$
160
140
120
100
80
60
40
20
0
0
50
100
150
200
250
k

```
f_2/e_2
```

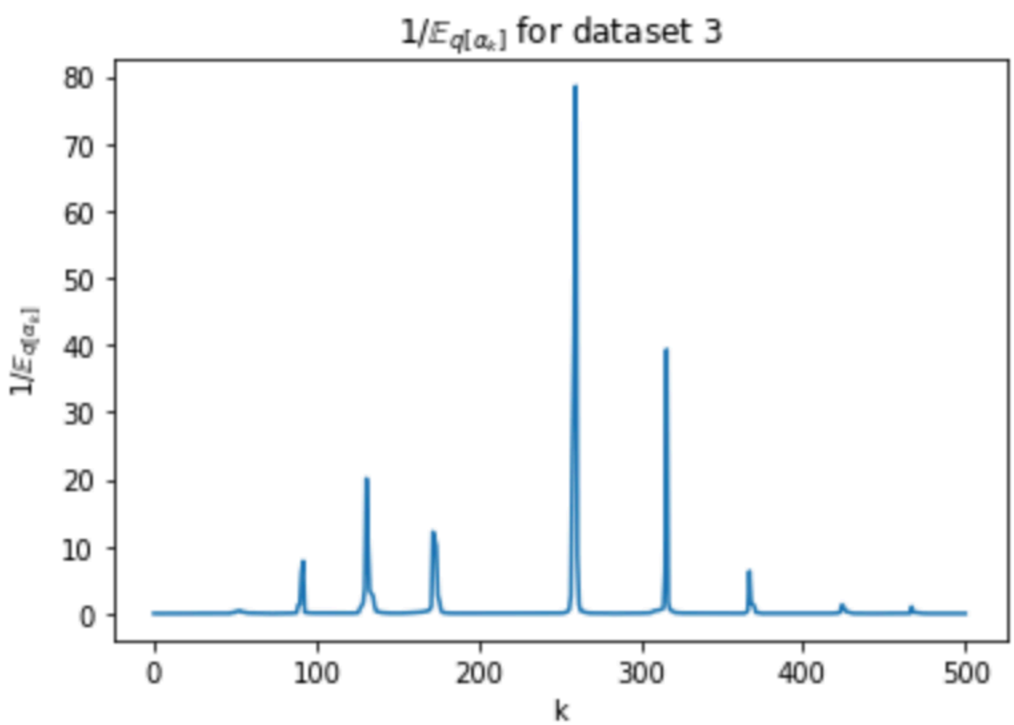
```
0.8994437867206255
```



For dataset 3:



Variational Objective Function for dataset 3
Variational Objective Function
−18000
−18200
−18400
−18600
−18800
0
100
200
300
400
500
Number of Iterations



$1/E_{q[\alpha_k]}$ for dataset 3
$1/E_{q[\alpha_k]}$
80
70
60
50
40
30
20
10
0
0
100
200
300
400
500
k

```
f_3/e_3
```

```
0.9781507541473099
```