Homework 3

Bohan Niu

November 4, 2018

Problem 1

a)By the definition of the model, we can factorize the joint likelihood:

$$p(w,\alpha_1,\dots,\alpha_d,\lambda|y,x)=p(w|\alpha_1,\dots,\alpha_d)p(\lambda)\prod_{k=1}^{d}p(\alpha_k)\prod_{i=1}^{N}p(y_i|x_i,w,\alpha_1,\dots,\alpha_d,\lambda) \quad (1)$$

From the question, we can also use the factorization:

$$p(w,\alpha_1,\dots,\alpha_d,\lambda|y,x)\approx q(w,\alpha_1,\dots,\alpha_d,\lambda)=q(w)q(\lambda)\prod_{k=1}^{d}q(\alpha_k) \quad (2)$$

By optimal method, we can find the distribution of $q(w), q(\lambda)$ and $q(\alpha_k)$ respectively:

$$\begin{aligned}
q(\lambda) &\propto \exp\{\mathbb{E}_{q(w)}[p(w,\alpha_1,\dots,\alpha_d,\lambda|y,x)]\}\\
&\propto \exp\{\mathbb{E}_{q(w)}[\ln p(y|x,w,\alpha_1,\dots,\alpha_d,\lambda)+\ln p(w|\alpha_1,\dots,\alpha_d)+\ln p(\lambda)+\sum_{k=1}^{d}\ln p(\alpha_k)]\}\\
&\propto \exp\{\mathbb{E}_{q(w)}[\ln p(y|x,w,\alpha_1,\dots,\alpha_d,\lambda)+\ln p(\lambda)]\}\\
&\propto \exp\{\mathbb{E}_{q(w)}[\ln p(y|x,w,\alpha_1,\dots,\alpha_d,\lambda)]\}p(\lambda)\\
&\propto \prod_{i=1}^{N}\lambda^{\frac{1}{2}}e^{-\frac{\lambda}{2}\mathbb{E}_{q(w)}[(y_i-x_i^Tw)^2]}\lambda^{e_0-1}e^{-f_0\lambda}\\
&\propto \lambda^{\frac{N}{2}+e_0-1}e^{(\frac{1}{2}\sum_{i=1}^{N}\mathbb{E}_{q(w)}[(y_i-x_i^Tw)^2]+f_0)\lambda}
\end{aligned} \quad (3)$$

So we can notice that:

$$q(\lambda)=\text{Gamma}(\lambda|e_0',f_0'),\ e_0'=\frac{N}{2}+e_0,\ f_0'=\frac{1}{2}\sum_{i=1}^{N}\mathbb{E}_{q(w)}[(y_i-x_i^Tw)^2]+f_0 \quad (4)$$

$$\begin{aligned}
q(\alpha_k) &\propto \exp\{\mathbb{E}_{q(w_k)}[p(w,\alpha_1,\dots,\alpha_d,\lambda|y,x)]\}\\
&\propto \exp\{\mathbb{E}_{q(w_k)}[\ln p(w_k|\alpha_k)+\ln p(\alpha_k)]\}\\
&\propto \exp\{\mathbb{E}_{q(w_k)}[\ln p(w_k|\alpha_k)]\}p(\alpha_k)\\
&\propto \alpha_k^{\frac{1}{2}}e^{-\frac{\alpha_k}{2}\mathbb{E}_{q(w_k)}[w_k^2]}\alpha_k^{a_0-1}e^{-b_0\alpha_k}\\
&\propto \alpha_k^{\frac{1}{2}+a_0-1}e^{[-\frac{1}{2}\mathbb{E}_{q(w_k)}[w_k^2]+b_0]\alpha_k}
\end{aligned} \quad (5)$$

That is:

$$q(\alpha_k) = \text{Gamma}(\alpha_k | a_0^{(k)'}, b_0^{(k)'}),\ a_0^{(k)'} = \frac{1}{2} + a_0,\ b_0^{(k)'} = \frac{1}{2}\mathbb{E}_{q(w_k)}[w_k^2] + b_0 \tag{6}$$

$$\begin{aligned}
q(w) &\propto \exp\{\mathbb{E}_{q(\lambda,\alpha_1,\ldots,\alpha_d)}[p(w,\alpha_1,\ldots,\alpha_d,\lambda|y,x)]\} \\
&\propto \exp\{\mathbb{E}_{q(\lambda)}[\ln p(y|x,w,\lambda)] + \mathbb{E}_{q(\alpha_1,\ldots,\alpha_d)}[\ln p(w|\alpha_1,\ldots,\alpha_d)]\} \\
&\propto \mathbb{E}_{q(\lambda)}[p(y|x,w,\lambda)]\mathbb{E}_{q(\alpha_1,\ldots,\alpha_d)}[p(w|\alpha_1,\ldots,\alpha_d)] \\
&\propto \prod_{i=1}^{N} \mathbb{E}_{q(\lambda)}[\lambda^{\frac{1}{2}} e^{-\frac{\lambda}{2}[(y_i - x_i^T w)^2]}]\mathbb{E}_{q(\alpha_1,\ldots,\alpha_d)}[A^{\frac{1}{2}} e^{\frac{1}{2}w^T A w}] \\
&\propto e^{-\frac{1}{2}\sum_{i=1}^{N}\mathbb{E}_{q(\lambda)}[\lambda](y_i - x_i^T w)^2 - \frac{1}{2}w^T\mathbb{E}_{q(\alpha_1,\ldots,\alpha_d)}[A]w} \qquad (7) \\
&\propto \exp\{-\frac{1}{2}\mathbb{E}_{q(\lambda)}[\lambda]\sum_{i=1}^{N}(y_i - x_i^T w)^2 - \frac{1}{2}w^T\mathbb{E}_{q(\alpha_1,\ldots,\alpha_d)}[A]w\} \\
&\propto \exp\{-\frac{1}{2}w^T(\mathbb{E}_{q(\lambda)}[\lambda]\sum_{i=1}^{N}x_i x_i^T + \mathbb{E}_{q(\alpha_1,\ldots,\alpha_d)}[A])w - 2w^T\mathbb{E}_{q(\lambda)}[\lambda]\sum_{i=1}^{N}y_i x_i\} \\
&\text{where } A = diag(\alpha_1,\ldots,\alpha_d)
\end{aligned}$$

From this, we can say that $q(w)$ is a Gaussian Distribution:

$$q(w) = \text{Normal}(w|\mu', \Sigma'),\ \Sigma' = (\mathbb{E}_{q(\lambda)}[\lambda]\sum_{i=1}^{N}x_i x_i^T + \mathbb{E}_{q(\alpha_1,\ldots,\alpha_d)}[A])^{-1},\ \mu' = \Sigma'(\mathbb{E}_{q(\lambda)}[\lambda]\sum_{i=1}^{N}y_i x_i) \tag{8}$$

We can also solve the expectation once we found all the distribution:

- $\mathbb{E}_{q(\lambda)}[\lambda] = \frac{e'_t}{f'_t}$
- $\mathbb{E}_{q(w)}[(y_i - x_i^T w)^2] = (y_i - x_i^T \mu')^2 + x_i^T \Sigma' x_i$
- $\mathbb{E}_{q(\alpha_1,\ldots,\alpha_d)}[A] = diag(a_t^{(1)'}/b_t^{(1)'}, \ldots, a_t^{(k)'}/b_t^{(k)'})$
- $\mathbb{E}_{q(w_k)}[w_k^2] = \Sigma_t^{(kk)'} + \mu_t^{(k)'}\mu_t^{(k)'T}$

b)

VI algorithm for Bayesian regression model:

1. Initialize $a'_0, b'_0, e'_0, f'_0, \mu'_0$ and Σ'_0 in some way

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

   - Update $q(\lambda)$ by setting

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

$$f'_t = f_0 + \frac{1}{2}\sum_{i=1}^{N}(y_i - x_i^T\mu')^2 + x_i^T\Sigma' x_i$$

   - Update $q(\alpha_k)$ by setting

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

$$b_t^{(k)'} = \frac{1}{2}(\Sigma_t^{(kk)'} + \mu_t^{(k)'}\mu_t^{(k)'T}) + b_0$$

   - Update $q(w)$ by setting

$$\Sigma'_t = (\frac{e'_t}{f'_t}\sum_{i=1}^{N} x_i x_i^T + diag(a_t^{(1)'}/b_t^{(1)'}, \ldots, a_t^{(k)'}/b_t^{(k)'}))^{-1}$$

$$\mu'_t = \Sigma'_t(\frac{e'_t}{f'_t}\sum_{i=1}^{N} y_i x_i)$$

   - Evaluate $\mathcal{L}(a'_t, b'_t, e'_t, f'_t, \mu'_t, \Sigma'_t)$ to access convergence c)

$$
\begin{aligned}
&\mathcal{L}(a_t', b_t', e_t', f_t', \mu_t', \Sigma_t') \\
&= \int\int\int q(w, \alpha_1, \dots, \alpha_d, \lambda) \ln \frac{p(w, \alpha_1, \dots, \alpha_d, \lambda | y, x)}{q(w, \alpha_1, \dots, \alpha_d, \lambda)} d\alpha_1, \dots, \alpha_d d\lambda dw \\
&= \int\int\int q(w)q(\lambda) \prod_{k=1}^{d} q(\alpha_k) \ln \frac{p(w, \alpha_1, \dots, \alpha_d, \lambda | y, x)}{q(w)q(\lambda)\prod_{k=1}^{d} q(\alpha_k)} d\alpha_1, \dots, \alpha_d d\lambda dw \\
&= \int q(\lambda) \ln p(\lambda) d\lambda + \int\int q(\lambda)q(w) \ln p(y|x, w, \lambda) d\lambda dw \\
&+ \int\int q(w) \prod_{i=1}^{d} q(\alpha_i) \ln p(w|\alpha_1, \dots, \alpha_d) d\alpha_1, \dots, \alpha_d dw + \int \prod_{i=1}^{d} q(\alpha_i) \ln p(\alpha_i) d\alpha_i \\
&- \int q(\lambda) \ln q(\lambda) d\lambda - \int q(w) \ln q(w) dw - \int \prod_{i=1}^{d} q(\alpha_i) \ln q(\alpha_i) d\alpha_i
\end{aligned}
\tag{9}
$$

For convince, we solve them separately.

$$
\begin{aligned}
\int q(\lambda) \ln p(\lambda) d\lambda &= e_0 \ln f_0 - \ln \Gamma(e_0) + (e_0 - 1)\mathbb{E}_{q(\lambda)}[\ln \lambda] - \mathbb{E}_{q(\lambda)}[\ln \lambda] f_0 \\
&= e_0 \ln f_0 - \ln \Gamma(e_0) + (e_0 - 1)(\psi(e_t') - \ln f_t') - f_0 \frac{e_t'}{f_t'}
\end{aligned}
\tag{10}
$$

$$
\begin{aligned}
\int \prod_{i=1}^{d} q(\alpha_i) \ln p(\alpha_i) d\alpha_i &= \sum_{i=1}^{d} \int q(\alpha_i) \ln p(\alpha_i) d\alpha_i \\
&= \sum_{i=1}^{d} (a_0 \ln b_0 - \ln \Gamma(a_0) + (a_0 - 1)(\psi(a_t^{(k)'}) - \ln b_t^{(k)'}) - b_0 \frac{a_t^{(k)'}}{b_t^{(k)'}})
\end{aligned}
\tag{11}
$$

$$
\begin{aligned}
&\int\int q(w) \prod_{i=1}^{d} q(\alpha_i) \ln p(w|\alpha_1, \dots, \alpha_d) d\alpha_1, \dots, \alpha_d dw \\
&= \int\int q(w) \prod_{i=1}^{d} q(\alpha_i) [-\frac{1}{2}(\ln 2\pi - \ln A - w^T A w)] d\alpha_1, \dots, \alpha_d dw \\
&= \int q(w) [-\frac{d}{2} \ln 2\pi + \frac{1}{2} \sum_{i=1}^{d} \mathbb{E}_{q(\alpha_i)} \ln[\alpha_i]] + \frac{1}{2} \mathbb{E}_{q(\alpha_1, \dots, \alpha_d)}[w^T A w] dw \\
&= -\frac{d}{2} \ln 2\pi + \frac{1}{2} \sum_{i=1}^{d} \mathbb{E}_{q(\alpha_i)}[\ln \alpha_i] + \frac{1}{2} \mathbb{E}_{q(\alpha_1, \dots, \alpha_d)q(w)}[w^T A w] \\
&= -\frac{d}{2} \ln 2\pi + \frac{1}{2} \sum_{i=1}^{d} (\psi(a_t^{(k)'}) - \ln b_t^{(k)'})) + \frac{1}{2} tr((\mu_t' \mu_t'^T + \Sigma_t') A_t')
\end{aligned}
\tag{12}
$$

where $A_t' = diag(a_t^{(0)'}/b_t^{(0)'}, \dots, a_t^{(d)'}/b_t^{(d)'})$

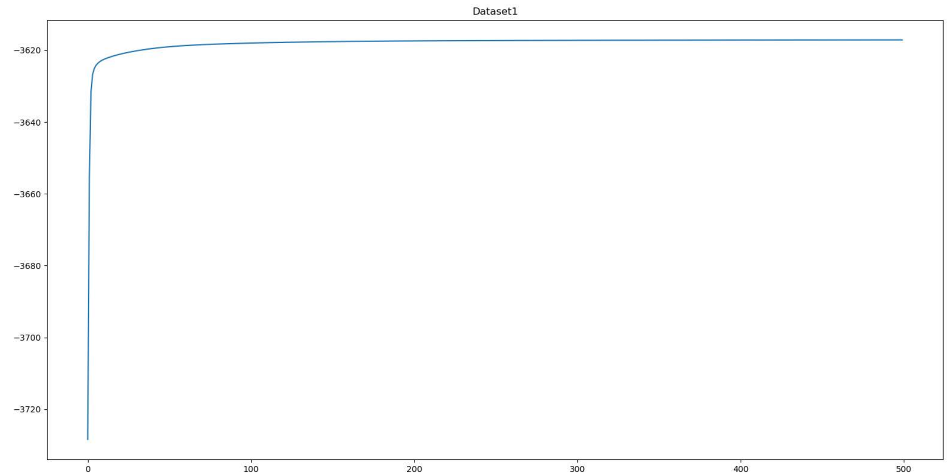
$$
\begin{aligned}
&\int\int q(\lambda)q(w)\ln p(y|x,w,\lambda)d\lambda dw \\
&= \int\int q(\lambda)q(w)[-\frac{N}{2}\ln 2\pi + \frac{N}{2}\ln\lambda - \sum_{i=1}^{N}\frac{\lambda}{2}(y_i - x_i^T w)^2]d\lambda dw \\
&= \int q(w)[-\frac{N}{2}\ln 2\pi + \frac{N}{2}\mathbb{E}_{q(\lambda)}[\ln\lambda] - \sum_{i=1}^{N}\frac{\mathbb{E}_{q(\lambda)}[\lambda]}{2}(y_i - x_i^T w)^2]dw \\
&= -\frac{N}{2}\ln 2\pi + \frac{N}{2}\mathbb{E}_{q(\lambda)}[\ln\lambda] - \sum_{i=1}^{N}\frac{\mathbb{E}_{q(\lambda)}[\lambda]}{2}\mathbb{E}_{q(w)}[(y_i - x_i^T w)^2] \\
&= -\frac{N}{2}\ln 2\pi + \frac{N}{2}(\psi(e_t') - \ln f_t') - \sum_{i=1}^{N}\frac{e_t'}{2f_t'}[(y_i - x_i^T\mu')^2 + x_i^T\Sigma_t' x_i]
\end{aligned}
\tag{13}
$$

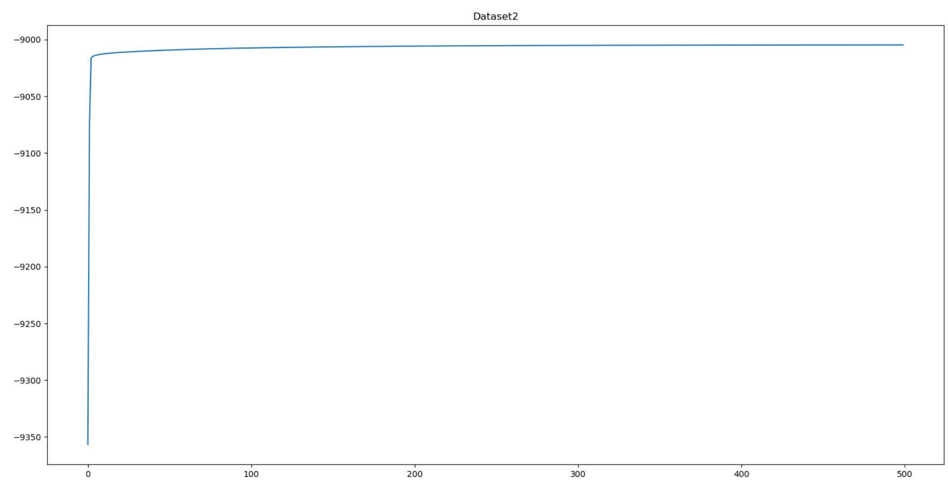
$$
\int q(\lambda)\ln q(\lambda)d\lambda = -e_t' + \ln f_t' - \ln\Gamma(e_t') - (1 - e_t')\psi(e_t') \tag{14}
$$

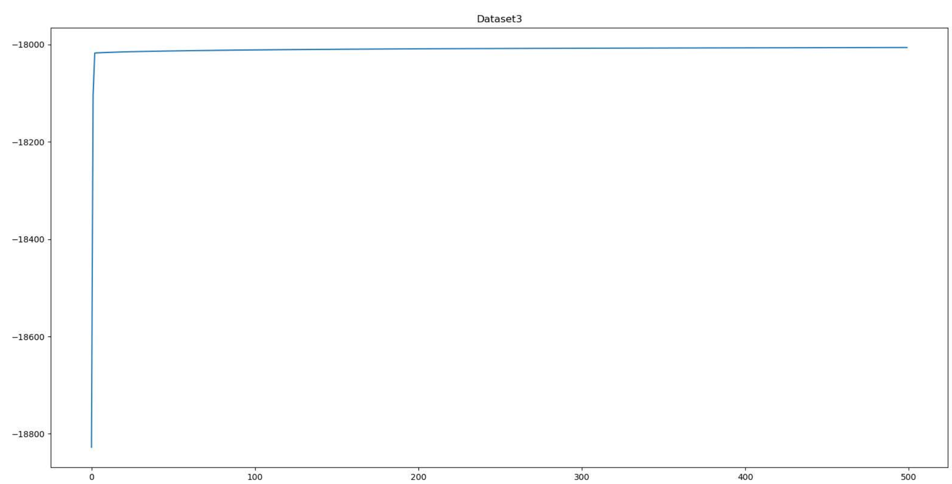
$$
\int q(w)\ln q(w)dw = -\frac{1}{2}\ln((2\pi e)^n|\Sigma_t'|) \tag{15}
$$

$$
\int\prod_{i=1}^{d} q(\alpha_i)\ln q(\alpha_i)dq(\alpha_i) = \sum_{i=1}^{d} -a_t^{(i)'} + \ln b_t^{(i)'} - \ln\Gamma(a_t^{(i)'}) - (1 - a_t^{(i)'})\psi(a_t^{(i)'}) \tag{16}
$$

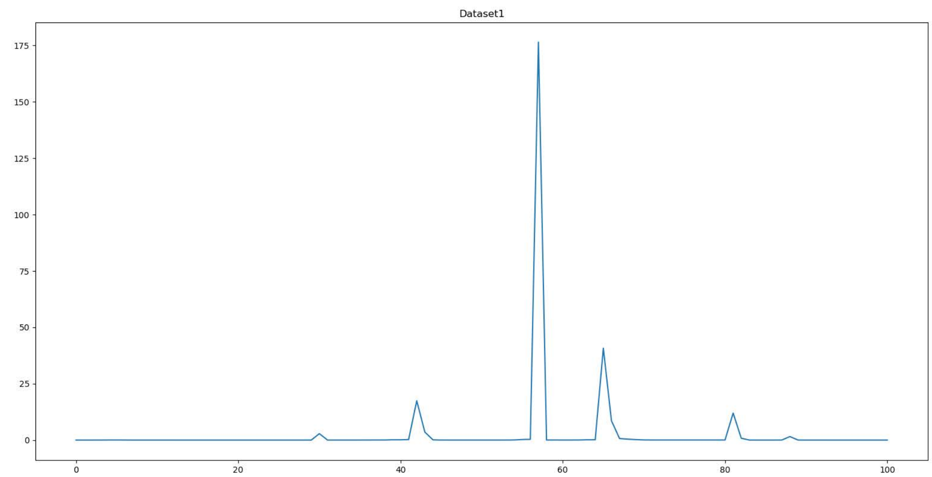
Problem 2
a)

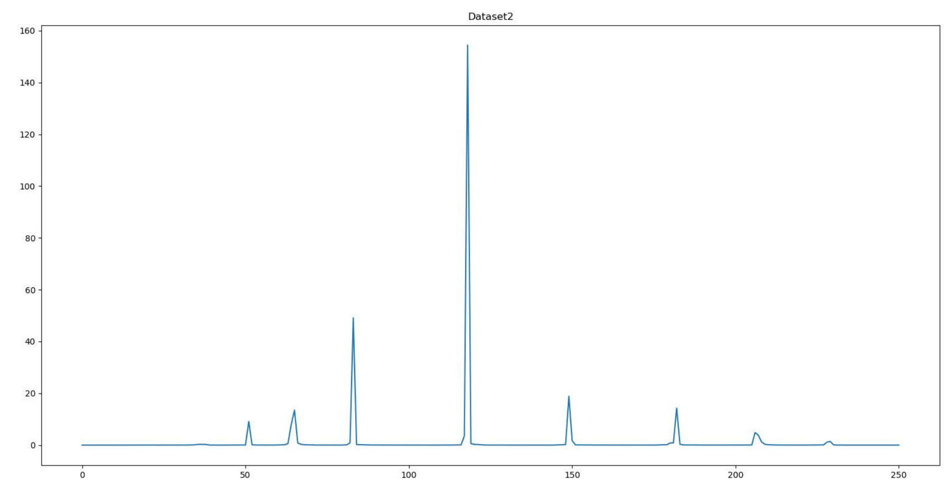

Dataset1
−3620
−3640
−3660
−3680
−3700
−3720
0
100
200
300
400
500

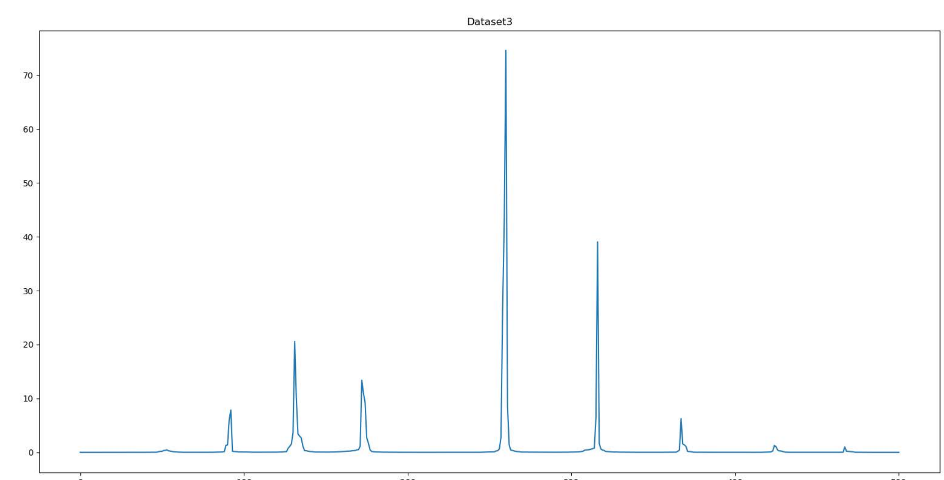

Dataset2
−9000
−9050
−9100
−9150
−9200
−9250
−9300
−9350
0
100
200
300
400
500


Dataset3
−18000
−18200
−18400
−18600
−18800
0
100
200
300
400
500

b)


Dataset1
175
150
125
100
75
50
25
0
0
20
40
60
80
100

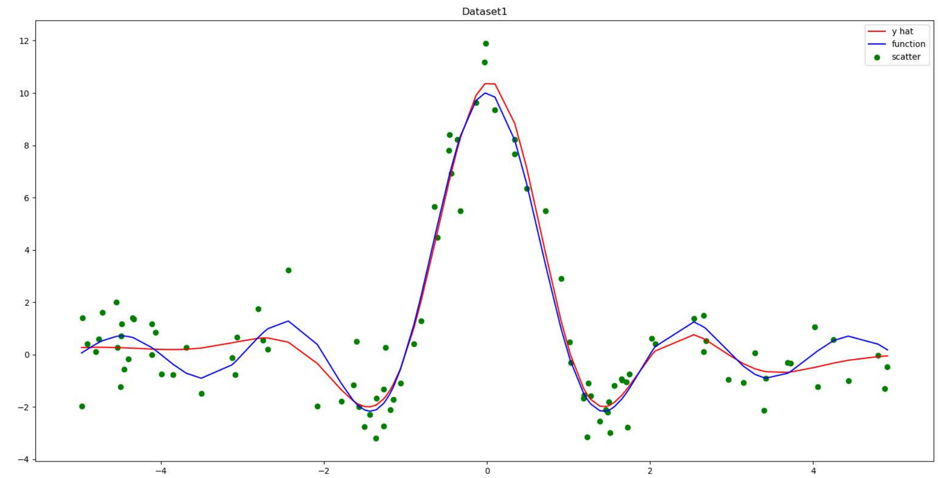

Dataset2
160
140
120
100
80
60
40
20
0
0
50
100
150
200
250

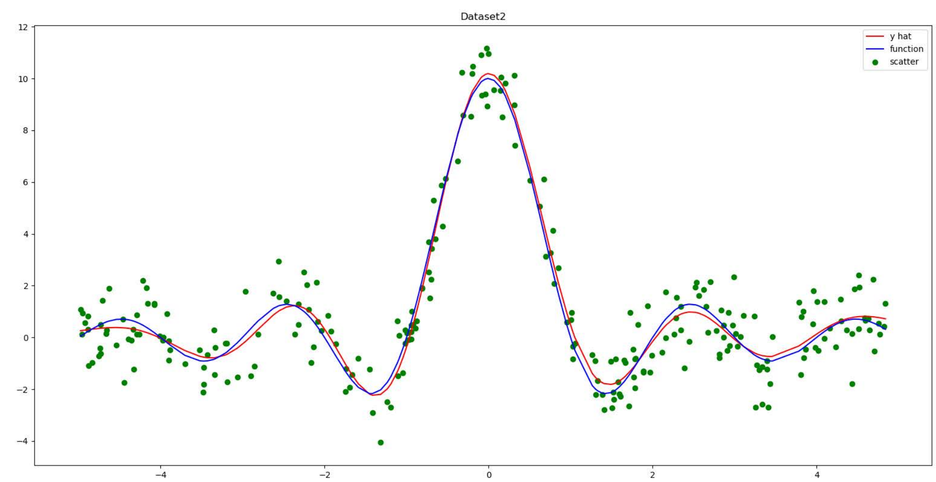

Dataset3
70
60
50
40
30
20
10
0
0
100
200
300
400
500

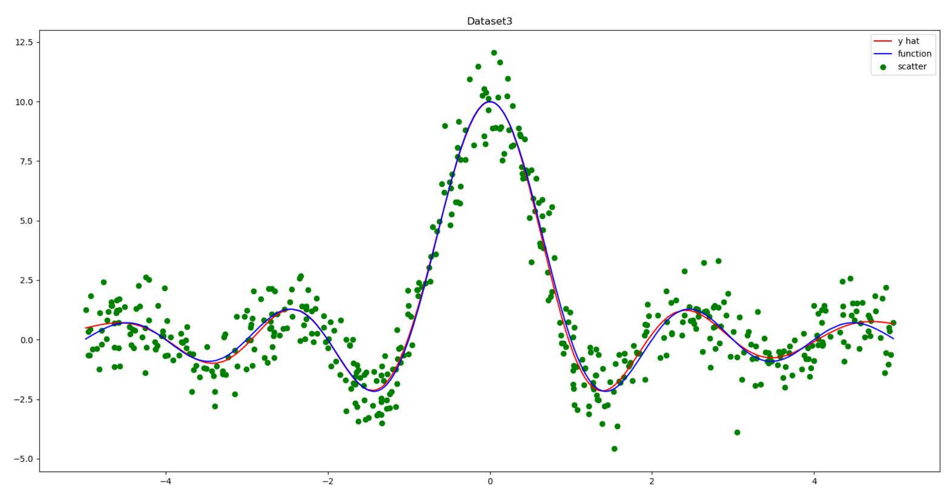
c) $1/\mathbb{E}_q[\lambda]$ in the final iteration:

- Dataset1: 1.0800299564504465
- Dataset2: 0.8994629800788205
- Dataset3: 0.9781435918476379

d)


Dataset1
y hat
function
scatter


Dataset2
y hat
function
scatter


Dataset3
y hat
function
scatter