EECS E6720: Bayesian Models for Machine Learning
Homework 3

Josh Rutta

November 4, 2018

Problem 1

Josh Rutta

Data: $\{(x_i, y_i)\}_{i=1}^{N}$, $y \in \mathbb{R}$, $x \in \mathbb{R}^d$

1. $y_i \overset{ind}{\sim} N(x_i^T w, \lambda^{-1})$, $w \sim N(0, \mathrm{diag}(\alpha_1, \dots, \alpha_d)^{-1})$, $\alpha_k \overset{iid}{\sim} \mathrm{Gamma}(a_0, b_0)$, $\lambda \sim \mathrm{Gamma}(e_0, f_0)$

$\mathrm{Gamma}(\eta \mid \tau_1, \tau_2) = \frac{\tau_2^{\tau_1}}{\Gamma(\tau_1)} \eta^{\tau_1 - 1} e^{-\tau_2 \eta}$ $\quad q(w, \alpha_1, \dots, \alpha_d, \lambda) \approx p(w, \alpha_1, \dots, \alpha_d, \lambda \mid y, X)$

a) $q(w, \lambda, \alpha_1, \dots, \alpha_d) = q(w) q(\lambda) \prod_{k=1}^{d} q(\alpha_k)$ $\qquad \alpha = \mathrm{diag}(\alpha_1, \dots, \alpha_k)$, $\vec{\alpha} = (\alpha_1, \alpha_2, \dots, \alpha_d)$

$$\ln p(y, w, \alpha, \lambda \mid X) = \ln[p(y \mid w, \vec{\alpha}, \lambda, x) p(w \mid \alpha) p(\vec{\alpha}) p(\lambda)]$$

$$p(y \mid w, \alpha, \lambda, x) = \prod_{i=1}^{N} \left(\frac{\lambda}{2\pi}\right)^{\frac{1}{2}} e^{-\frac{\lambda}{2}(y_i - x_i^T w)^2} = \left(\frac{\lambda}{2\pi}\right)^{\frac{N}{2}} e^{-\frac{\lambda}{2}\sum_{i=1}^{N}(y_i - x_i^T w)^2}$$

$$p(w \mid \alpha) = (2\pi)^{-\frac{d}{2}} \left(\prod_{i=1}^{d} \alpha_i\right)^{\frac{1}{2}} e^{-\frac{1}{2} w^T \alpha w} \qquad p(\alpha) = \prod_{i=1}^{d} \frac{b_0^{a_0}}{\Gamma(a_0)} \alpha_i^{a_0 - 1} e^{-b_0 \alpha_i}$$

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

$$q(w) \propto \exp\{\mathbb{E}_{-q}[\ln p(y \mid w, \lambda, x) + \ln p(w \mid \vec{\alpha})]\}$$

$$q(w) \propto \exp\left\{\mathbb{E}_{-q}\left[\frac{N}{2}\ln\left(\frac{\lambda}{2\pi}\right) - \frac{\lambda}{2}\sum_{i=1}^{N}(y_i - x_i^T w)^2 - \frac{d}{2}\ln(2\pi) + \frac{1}{2}\sum_{i=1}^{d}\ln\alpha_i - \frac{1}{2} w^T \alpha w\right]\right\}$$

$$q(w) \propto \exp\left\{\underbrace{\cancel{\frac{N}{2}\mathbb{E}_{q(\lambda)}[\ln(\lambda)]}}_{\text{no } w} - \frac{\mathbb{E}_{q(\lambda)}[\lambda]}{2} \underbrace{\sum_{i=1}^{N}(y_i - x_i^T w)^2}_{\sum_{i=1}^{N}(y_i^2 - 2w^T x_i y_i + w^T x_i x_i^T w)} + \underbrace{\cancel{\frac{1}{2}\sum_{i=1}^{d}\mathbb{E}_{q(\alpha_i)}[\ln\alpha_i]}}_{\text{no } w} - \frac{1}{2} w^T \mathbb{E}_{-q}[\alpha] w\right\}$$

$$\underbrace{\sum_{i=1}^{N} y_i^2}_{\text{no } w}$$

$$q(w) \propto \exp\left\{-\frac{1}{2}\left[w^T\left(\mathbb{E}_{q(\lambda)}(\lambda)\sum_{i=1}^{N} x_i x_i^T + \mathbb{E}_{-q}[\alpha]\right) w - 2w^T\left(\mathbb{E}_{q(\lambda)}[\lambda]\sum_{i=1}^{N} x_i y_i\right)\right]\right\}$$

let $A = \left(\mathbb{E}_{q(\lambda)}[\lambda]\sum_{i=1}^{N} x_i x_i^T + \mathbb{E}_{-q}[\alpha]\right)$ - constant w.r.t. w

$$N(\mu, \Sigma) = (2\pi)^{-\frac{d}{2}} |\Sigma|^{-\frac{1}{2}} \exp\{-\tfrac{1}{2}(w - \mu)^T \Sigma^{-1}(w - \mu)\} \propto \exp\{-\tfrac{1}{2}(w^T \Sigma^{-1} w - 2w^T \Sigma^{-1}\mu + \mu^T \Sigma^{-1}\mu)\}$$

$$q(w) \propto \exp\left\{-\frac{1}{2}\left[w^T \underset{\Sigma^{-1}}{A} w - 2w^T A \underset{\mu}{\left(\mathbb{E}_{q(\lambda)}[\lambda] A^{-1}\sum_{i=1}^{N} x_i y_i\right)} + \underset{\mu^T}{\left(\mathbb{E}_{q(\lambda)}[\lambda] A^{-1}\sum_{i=1}^{N} x_i y_i\right)^T} \underset{\Sigma^{-1}}{A} \underset{\mu}{\left(\mathbb{E}_{q(\lambda)}[\lambda] A^{-1}\sum_{i=1}^{N} x_i y_i\right)}\right]\right\}$$

$$q(w) = N\Big(\underbrace{\mathbb{E}_{q(\lambda)}[\lambda] A^{-1}\sum_{i=1}^{N} x_i y_i}_{\mu'}, \underbrace{A^{-1}}_{\Sigma'}\Big), \quad A = \left(\mathbb{E}_{q(\lambda)}(\lambda)\sum_{i=1}^{N} x_i x_i^T + \mathbb{E}_{-q}[\alpha]\right)$$

$$q(\lambda) \propto \exp\{\mathbb{E}_{-q}[\ln p(y \mid w, \lambda, x) + \ln p(\lambda)]\} \propto \exp\left\{\mathbb{E}_{-q}\left[\frac{N}{2}\ln\left(\frac{\lambda}{2\pi}\right) - \frac{\lambda}{2}\sum_{i=1}^{N}(y_i - x_i^T w)^2 + (e_0 - 1)\ln\lambda - f_0\lambda\right]\right\}$$

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

$w^T x_i x_i^T w = tr(w^T x_i x_i^T w) = tr(w w^T x_i x_i^T)$; $\mathbb{E}[tr(X)] = tr[\mathbb{E}(X)]$

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

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

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

$$p(w_i|\alpha_i) = \left(\frac{\alpha_i}{2\pi}\right)^{\frac{1}{2}} e^{-\frac{1}{2}w_i^2\alpha_i} \qquad p(\alpha_i) = \frac{b_0^{a_0}}{\Gamma(a_0)}\alpha_i^{a_0-1}e^{-b_0\alpha_i}$$

b/c $\Sigma = \text{diag}(\alpha_1, \ldots, \alpha_d)$

$$q(\alpha_i) \propto \exp\left\{\mathbb{E}_{-q}\left[\ln p(w_i|\alpha_i) + \ln p(\alpha_i)\right]\right\} \propto \exp\left\{\mathbb{E}_{-q}\left[\tfrac{1}{2}\ln(\alpha_i) - \tfrac{1}{2}w_i^2\alpha_i + (a_0-1)\ln\alpha_i - b_0\alpha_i\right]\right.$$

$$q(\alpha_i) \propto \exp\left\{(a_0 - \tfrac{1}{2})\ln(\alpha_i) - \alpha_i\left(\tfrac{1}{2}\mathbb{E}_{q(w_i)}[w_i^2] + b_0\right)\right\} = \alpha_i^{(a_0+\frac{1}{2})-1} e^{-\alpha_i\left(\frac{1}{2}\mathbb{E}_{q(w_i)}[w_i^2] + b_0\right)}$$

$$q(\alpha_i) = \text{Gamma}\Big(\underbrace{a_0 + \tfrac{1}{2}}_{a'_{(i)}}, \underbrace{\tfrac{1}{2}\mathbb{E}_{q(w_i)}(w_i^2) + b_0}_{b'_{(i)}}\Big)$$

$$A = \left(\mathbb{E}_{q(\lambda)}[\lambda]\sum_{i=1}^{N} x_i x_i^T + \mathbb{E}_{\prod_{i=1}^{d} q(\alpha)}[\alpha]\right)$$

$$\mathbb{E}_{q(\lambda)}[\lambda] = \frac{e'}{f'} \quad ; \quad \mathbb{E}_{\prod_{i=1}^{d} q(\alpha_i)}[\alpha] = \text{diag}\left(\frac{a'_{(1)}}{b'_{(1)}}, \frac{a'_{(2)}}{b'_{(2)}}, \ldots, \frac{a'_{(d)}}{b'_{(d)}}\right) ; \quad \mathbb{E}_{q(w)}[w] = \mathbb{E}_{q(\lambda)}[\lambda] A^{-1} \sum_{i=1}^{N} x_i y_i$$

$$A^{-1} = \mathbb{E}_{q(w)}[ww^T] - \mathbb{E}_{q(w)}[w]\left(\mathbb{E}_{q(w)}[w]\right)^T \Rightarrow \mathbb{E}_{q(w)}[ww^T] = A^{-1} + \mathbb{E}_{q(w)}[w]\left(\mathbb{E}_{q(w)}[w]\right)^T$$

$$\mathbb{E}_{q(w_i)}(w_i^2) = (A^{-1})_{ii} + \left(\mathbb{E}_{q(w)}[w]\right)_i^2$$

b.) Inputs: Data and definitions $q(w) = N(w|\mu', \Sigma')$, $q(\alpha_i) = \text{Gamma}(\alpha_i | a'_{(i)}, b'_{(i)})$, $q(\lambda) = \text{Gamma}(\lambda | e', f')$

Outputs: Values for μ', Σ'; $a'_{(i)}, b'_{(i)} \forall i \in \{1, \ldots, d\}$, e', f'

1. initialize $\mu'_0, \Sigma'_0, a'_{0(i)}, b'_{0(i)} \in \{1, \ldots, d\}, e'_0, f'_0$

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

   a) Update $q(w)$ by setting: $\Sigma'_t = \left(\frac{e'_{t-1}}{f'_{t-1}}\sum_{i=1}^{N} x_i x_i^T + \text{diag}\left(\frac{a'_{t-1(1)}}{b'_{t-1(1)}}, \frac{a'_{t-1(2)}}{b'_{t-1(2)}}, \ldots, \frac{a'_{t-1(d)}}{b'_{t-1(d)}}\right)\right)^{-1}$

   $$\mu'_t = \frac{e'_{t-1}}{f_{t-1}}\left(\Sigma'_t\right)\sum_{i=1}^{d} x_i y_i$$

   b) update $q(\lambda)$ by setting: $e'_t = \frac{N}{2} + e_0$, $f'_t = \frac{1}{2}\left(\sum_{i=1}^{N}(y_i^2 - 2y_i x_i^T \mu'_{t-1}) + tr\left((\Sigma'_{t-1} + \mu'_{t-1}\mu'^T_{t-1})\sum_{i=1}^{N} x_i x_i^T\right) + 2f_0\right)$

   c) update $q(\alpha_i)$ $\forall i \in \{1, \ldots, d\}$ by setting: $a'_{t(i)} = a_0 + \frac{1}{2}$

   $$b'_{t(i)} = b_0 + \tfrac{1}{2}\left[\left(\Sigma'_{t-1}\right)_{ii} + \left(\mu'_{t-1}\right)_i^2\right]$$

   d) Evaluate $\mathcal{L}(\mu'_t, \Sigma'_t; a'_{t(i)}, b'_{t(i)} \forall i \in \{1, \ldots, d\}; e'_t, f'_t)$

Josh Rutta

1.c.) $\mathcal{L}(\mu_t', \Sigma_t'; a_{t(i)}', b_{t(i)}' \ \forall i \in \{1, \dots, d\}, e_t', f_t') = \int\int\int q(w)q(\lambda)\prod_{i=1}^{d} q(\alpha_i) \ln \frac{p(y, w, \vec{\alpha}, \lambda \mid x)}{q(w)q(\lambda)\prod_{i=1}^{d} q(\alpha_i)} \, d\vec{\alpha}\, d\lambda\, dw$

$$\mathcal{L} = \int\int\int q(w)q(\lambda)\prod_{i=1}^{d} q(\alpha_i) \ln \frac{p(w|\vec{\alpha})p(\lambda)p(\alpha_1)p(\alpha_2)\dots p(\alpha_d)\, p(y|w,\lambda,X)}{q(w)q(\lambda)\prod_{i=1}^{d} q(\alpha_i)} \, d\vec{\alpha}\, d\lambda\, dw$$

$$\mathcal{L} = \underbrace{\int\int q(w)\prod_{i=1}^{d} q(\alpha_i) \ln p(w|\vec{\alpha})\, dw\, d\vec{\alpha}}_{①} + \underbrace{\int \left(q(\lambda)\ln p(\lambda) - q(\lambda)\ln q(\lambda)\right) d\lambda}_{②} + \underbrace{\int \prod_{i=1}^{d} q(\alpha_i)\ln(\alpha_i) - q(\alpha_i)\ln q(\alpha_i)\, d\alpha_i}_{③}$$
$$+ \underbrace{\int\int q(w)q(\lambda)\ln p(y|w,\lambda,X)\, d\lambda\, dw}_{④} - \underbrace{\int q(w)\ln q(w)\, dw}_{⑤}$$

① $\int\int \prod_{i=1}^{d} q(\alpha_i)q(w)\ln p(w|\vec{\alpha})\, dw\, d\vec{\alpha} = \mathbb{E}_{q(w,\vec{\alpha})}\left[-\frac{d}{2}\ln 2\pi + \frac{1}{2}\sum_{i=1}^{d}\ln\alpha_i - \frac{1}{2}w^T\alpha w\right] = -\frac{d}{2}\ln 2\pi + \frac{1}{2}\sum_{i=1}^{d}\mathbb{E}_{q(\alpha_i)}[\ln\alpha_i] - \frac{1}{2}\mathbb{E}_{q(w,\vec{\alpha})}[w^T\alpha w]$

$= -\frac{d}{2}\ln 2\pi + \frac{1}{2}\sum_{i=1}^{d}\left(\psi(a_{t(i)}') - \ln(b_{t(i)}')\right) - \frac{1}{2}tr\left((\Sigma_t' + \mu_t'\mu_t'^T)\left(\mathrm{diag}\left(\frac{a_{t(1)}'}{b_{t(1)}'}, \dots, \frac{a_{t(d)}'}{b_{t(d)}'}\right)\right)\right)$

② $\int q(\lambda)\ln p(\lambda) - q(\lambda)\ln q(\lambda)\, d\lambda = \mathbb{E}_{q(\lambda)}\left[e_0\ln f_0 - \ln\Gamma(e_0) + (e_0-1)\ln\lambda - f_0\lambda - e_t'\ln f_t' - \ln\Gamma(e_t') + (e_t'-1)\lambda - f_t'\lambda\right]$

$= e_0\ln f_0 - \ln\Gamma(e_0) + (e_0-1)(\psi(e_t') - \ln(f_t')) - f_0\frac{e_t'}{f_t'} - e_t'\ln f_t' - \ln\Gamma(e_t') + (e_t'-1)(\psi(e_t') - \ln f_t') - f_t'\frac{e_t'}{f_t'}$

$= e_0\ln f_0 - \ln\Gamma(e_0) + (e_0-1)(\psi(e_t') - \ln(f_t')) - f_0\frac{e_t'}{f_t'} + \ln f_t' - \ln\Gamma(e_t') + (e_t'-1)\psi(e_t') - e_t'$

③ $\int \prod_{i=1}^{d} q(\alpha_i)\ln(\alpha_i) - q(\alpha_i)\ln q(\alpha_i)\, d\alpha_i = \mathbb{E}_{\prod_{i=1}^{d} q(\alpha_i)}\left[\sum_{i=1}^{d} a_0\ln b_0 - \ln\Gamma(a_0) + (a_0-1)\ln\alpha_i - b_0\alpha_i - a_{t(i)}'\ln b_{t(i)}' - \ln\Gamma(a_{t(i)}') + (a_{t(i)}'-1)\ln\alpha_i - b_{t(i)}'\alpha_i\right]$

$= \sum_{i=1}^{d} a_0\ln b_0 - \ln\Gamma(a_0) + (a_0-1)(\psi(a_{t(i)}') - \ln b_{t(i)}') - b_0\frac{a_{t(i)}'}{b_{t(i)}'} - a_{t(i)}'\ln b_{t(i)}' - \ln\Gamma(a_{t(i)}')$
$+ (a_{t(i)}'-1)(\psi(a_{t(i)}') - \ln b_{t(i)}') - b_{t(i)}'\frac{a_{t(i)}'}{b_{t(i)}'}$

$= \sum_{i=1}^{d} a_0\ln b_0 - \ln\Gamma(a_0) + (a_0-1)(\psi(a_{t(i)}') - \ln b_{t(i)}') - b_0\frac{a_{t(i)}'}{b_{t(i)}'} + \ln b_{t(i)}' - \ln\Gamma(a_{t(i)}') + (a_{t(i)}'-1)\psi(a_{t(i)}') - a_{t(i)}'$

④ $\int\int q(w)q(\lambda)\ln p(y|w,\lambda,X)\, d\lambda\, dw = \mathbb{E}_{q(w)q(\lambda)}\left[\frac{N}{2}\ln(\lambda) - \frac{N}{2}\ln(2\pi) - \frac{\lambda}{2}\sum_{i=1}^{N}(y_i^2 - 2y_i x_i^T w + w^T x_i x_i^T w)\right]$

$= \frac{N}{2}(\psi(e_t') - \ln f_t') - \frac{N}{2}\ln(2\pi) - \frac{e_t'}{2f_t'}\left[\sum_{i=1}^{N} y_i^2 - 2y_i x_i^T\mu_t' + tr\left((\Sigma_t' + \mu_t'\mu_t'^T)\sum_{i=1}^{N} x_i x_i^T\right)\right]$

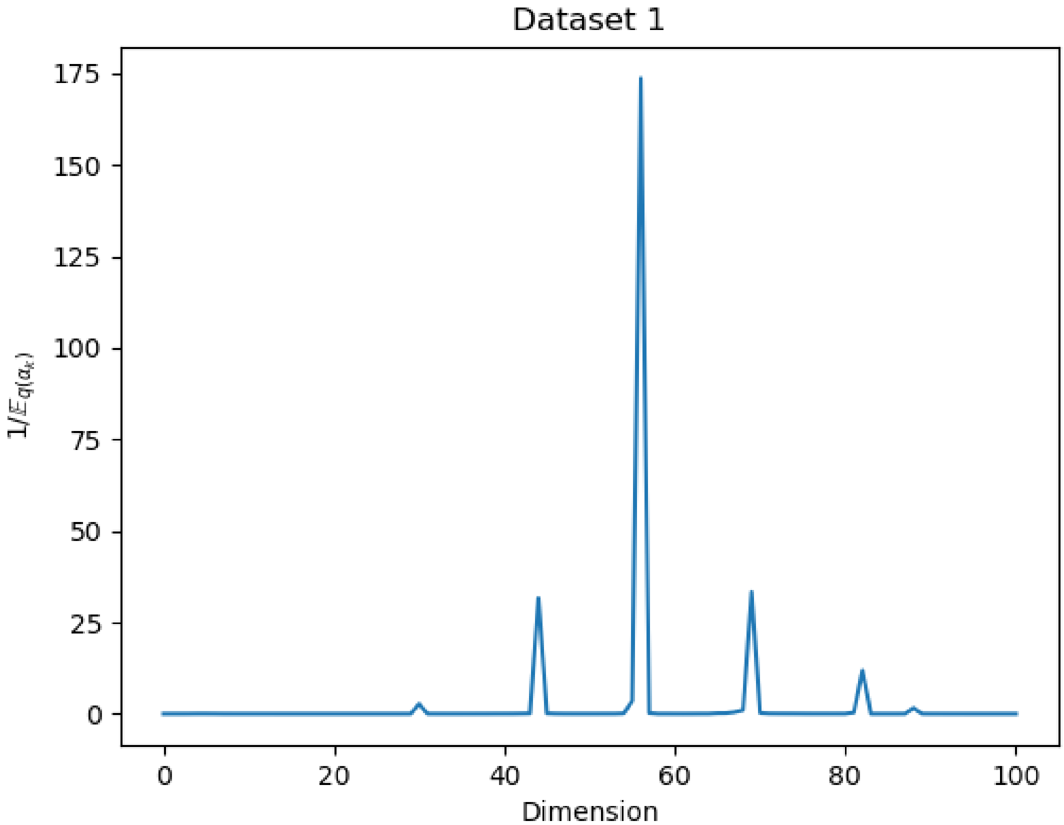
⑤ $\int q(w)\ln q(w)\, dw = \mathbb{E}_{q(w)}\left[-\frac{1}{2}\left[d\ln(2\pi) + \ln|\Sigma_t'|\right] - \frac{1}{2}\left((w-\mu_t')^T\Sigma_t'^{-1}(w-\mu_t')\right)\right]$

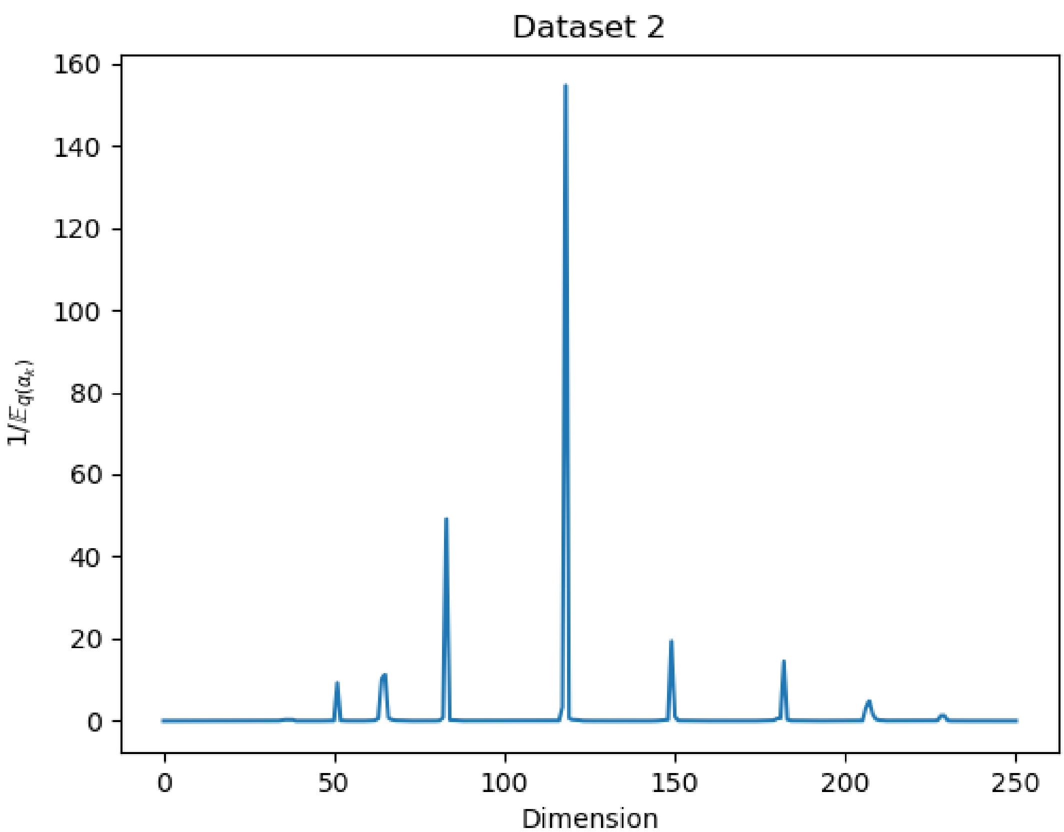
$= -\frac{1}{2}\left[d\ln(2\pi) + \ln|\Sigma_t'|\right] - \frac{1}{2}\left[tr\left(\Sigma_t'^{-1}\mathbb{E}_{q(w)}[(w-\mu_t')(w-\mu_t')^T]\right)\right] = -\frac{1}{2}\left[d\ln(2\pi) + \ln|\Sigma_t'|\right] - \frac{1}{2}\left(tr\left(\overbrace{\Sigma_t'^{-1}\Sigma_t'}^{I}\right)\right)$

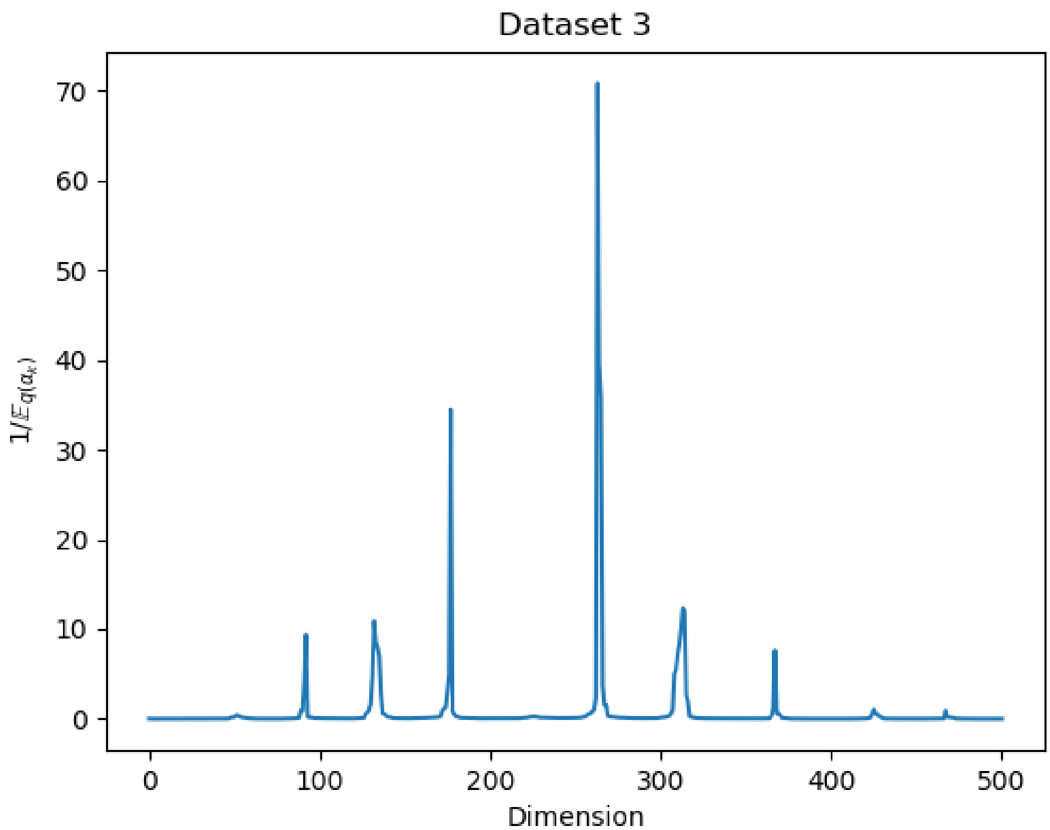
$= -\frac{1}{2}\left[d\ln(2\pi) + \ln|\Sigma_t'|\right] - \frac{1}{2}[d] = -\frac{1}{2}\left[d(\ln(2\pi)+1) + \ln|\Sigma_t'|\right]$

Problem 2

2b.


Dataset 1
175
150
125
100
75
50
25
0
0
20
40
60
80
100
Dimension
$1/\mathbb{E}_{q(a_k)}$


Dataset 2
160
140
120
100
80
60
40
20
0
0
50
100
150
200
250
Dimension
$1/\mathbb{E}_{q(a_k)}$



2c. Dataset 1:

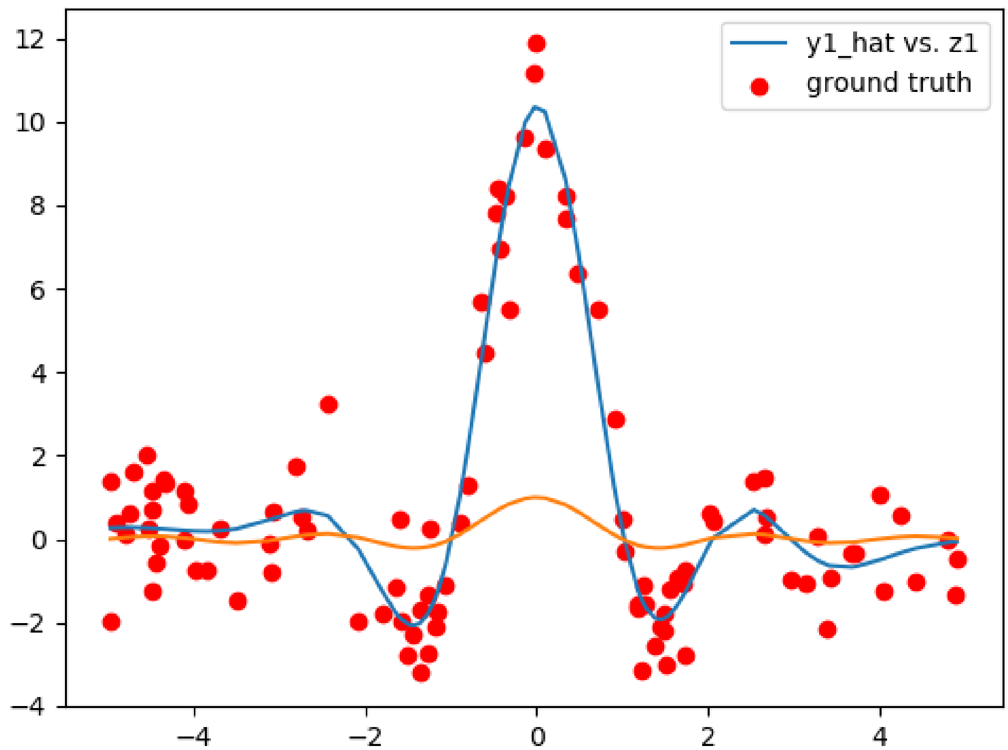
$$\frac{1}{E_{q(\lambda)}(\lambda)} = 1.0783383662230435$$

Dataset 2:

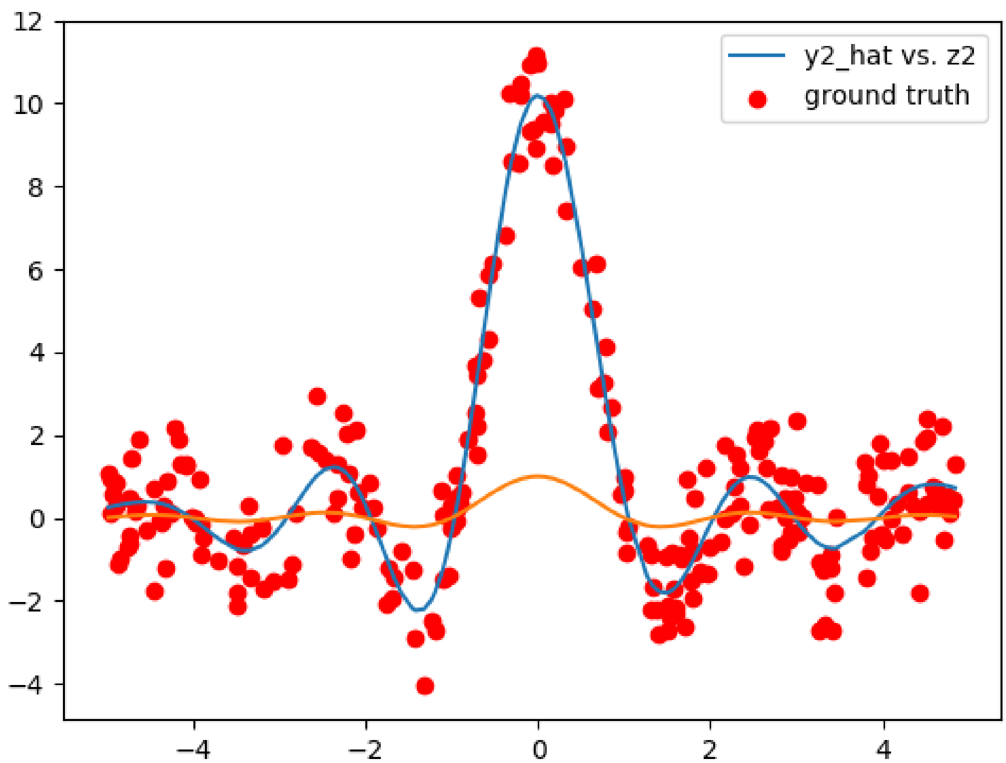
$$\frac{1}{E_{q(\lambda)}(\lambda)} = 0.8994164236209459$$

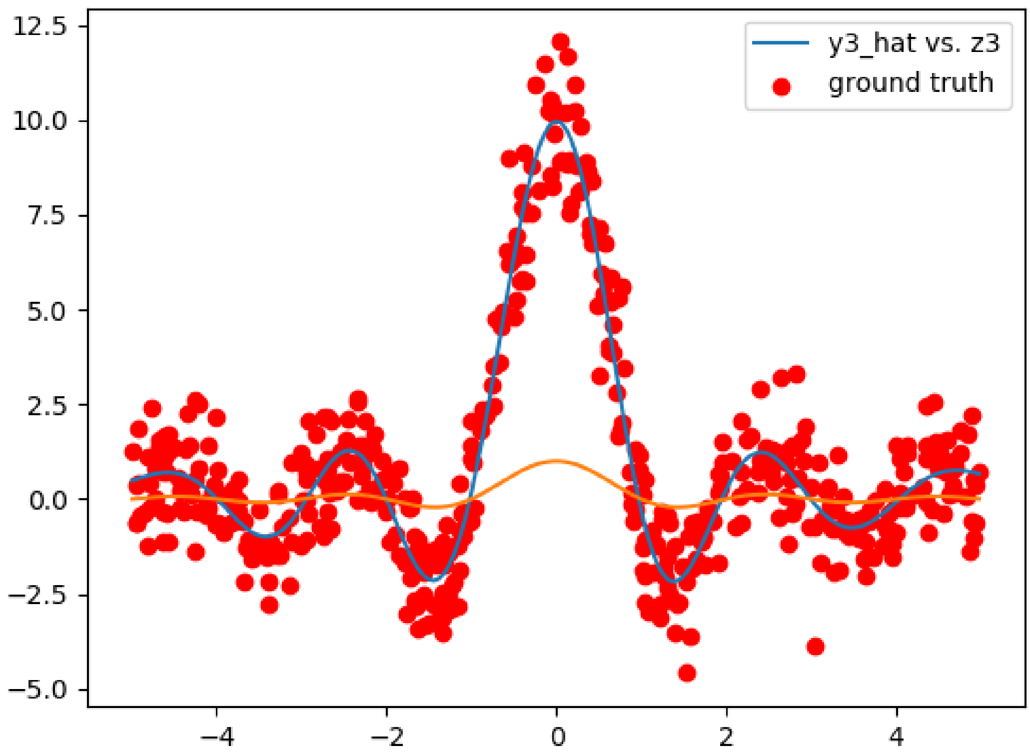
Dataset 3:

$$\frac{1}{E_{q(\lambda)}(\lambda)} = 0.9771899941954644$$



2d.




y3_hat vs. z3
ground truth
12.5
10.0
7.5
5.0
2.5
0.0
−2.5
−5.0
−4
−2
0
2
4