Kate Lassiter

Problem 1. (40 points)

We have a data set of the form $\{(x_i, y_i)\}_{i=1}^N$, where $y \in \mathbb{R}$ and $x \in \mathbb{R}^d$. We assume d is large and not all dimensions of x are informative in predicting y. Consider the following regression model for this problem:

$$y_i \overset{ind}{\sim} \text{Normal}(x_i^T w, \lambda^{-1}), \quad w \sim \text{Normal}(0, \text{diag}(\alpha_1, \dots, \alpha_d)^{-1}),$$

$$\alpha_k \overset{iid}{\sim} \text{Gamma}(a_0, b_0), \quad \lambda \sim \text{Gamma}(e_0, f_0).$$

Use the density function $\text{Gamma}(\eta|\tau_1, \tau_2) = \frac{\tau_2^{\tau_1}}{\Gamma(\tau_1)} \eta^{\tau_1 - 1} e^{-\tau_2 \eta}$. In this homework, you will derive a variational inference algorithm for approximating the posterior distribution with

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

a) Using the factorization $q(w, \alpha_1, \dots, \alpha_d, \lambda) = q(w)q(\lambda) \prod_{k=1}^d q(\alpha_k)$, derive the optimal form of each q distribution. Use these optimal q distributions to derive a variational inference algorithm for approximating the posterior.

$$q(a) = e^{E_{q-\alpha}[lnp(y_i|x_i^T w, \lambda^{-1}) + lnp(w|0, diag(\alpha_1, ..., \alpha_d)^{-1}) + lnp(a_k|a0, b0) + lnp(\lambda|e0, f0)]}$$

$$q(a) = e^{E_{q-\alpha}[lnp(w|0, diag(\alpha_1, ..., \alpha_d)^{-1}) + lnp(a_k|a0, b0)]}$$

$$q(a) = e^{E_{q-\alpha}[\sum_{i=1}^{d} ln \frac{1}{\sqrt{2\pi\alpha_i}} e^{-\frac{(w)^2}{2\alpha_i}} + \sum_{i=1}^{d} ln \frac{b0^{ao}}{\Gamma(ao)} a_i^{ao-1} e^{-b0a_i}]}$$

$$q(a) \propto e^{E_{q-\alpha}[\sum_{i=1}^{d} ln\alpha_i^{\frac{1}{2}} e^{-\frac{(w)^2}{2\alpha_i}} + \sum_{i=1}^{d} lna_i^{ao-1} e^{-b0a_i} + constant\ w.r.t\ \alpha]}$$

$$q(a) \propto e^{E_{q-\alpha}[\sum_{i=1}^{d} ln\alpha_i^{\frac{1}{2}} e^{-\frac{(w)^2}{2\alpha_i}} + lna_i^{ao-1} e^{-b0a_i}]}$$

$$q(a) \propto e^{E_{q-\alpha}[\prod_{i=1}^{d} ln\alpha_i^{\frac{1}{2}} e^{-\frac{(w)^2}{2\alpha_i}} lna_i^{ao-1} e^{-b0a_i}]}$$

$$q(a) \propto \prod_{i=1}^{d} \alpha_i^{\frac{1}{2}} e^{-\frac{a_i}{2}E_w[w^2]} a_i^{ao-1} e^{-b0a_i}$$

$$q(a) \propto \prod_{i=1}^{d} \alpha_i^{\frac{1}{2}+ao-1} e^{-\frac{a_i}{2}E_w[w^2] - b0a_i}$$

$$q(a) \propto \alpha_i^{\frac{1}{2}+ao-1} e^{\sum_{i=1}^{d} -\frac{a_i}{2}E_w[w^2] - b0a_i}$$

$\boxed{q(a) \sim \text{Gamma}(a', b') \quad a' = a0 + \frac{1}{2}}$

Isolating b0 for a single a_i:

$$0 = -\frac{a_i}{2}E_w[w^2] - b0a_i$$

$$b0a_i = -\frac{a_i}{2}E_w[w^2]$$

$$\frac{1}{a_i} b0a_i = -\frac{a_i}{2}E_w[w^2]\frac{1}{a_i}$$

$$b0 = -\frac{1}{2}E_w[w^2]$$

$$b' = b0 + \frac{1}{2}E_w[w^2]$$

Kate Lassiter

Recall:
var[y]= E[(y-u) (y-u) T]
var[y]= E[yyT] -uuT
E[yyT] = var[y] + uuT

Thus: $E[ww^T] = \Sigma' + u'u'^T$

$$\boxed{b' = b0 + \frac{1}{2}\sum_{i=1}^{d}(\Sigma_{ii}' + u_i'u_i'^T)}$$

$$q(w) = e^{E_{q-w}[lnp(y_i|x_i^T w,\lambda^{-1})+lnp(w|0,diag(\alpha_1,\ldots,\alpha_d)^{-1})+lnp(a_k|a0,b0)+lnp(\lambda|e0,f0)]}$$

$$q(w) = e^{E_{q-w}[lnp(y_i|x_i^T w,\lambda^{-1})+lnp(w|0,diag(\alpha_1,\ldots,\alpha_d)^{-1})]}$$

$$q(w) = e^{E_{q-w}[\sum_{i=1}^{N} ln\frac{1}{\sqrt{2\pi\lambda}}e^{-\frac{(y_i-x_i^T w)^2}{2\lambda}}+\sum_{i=1}^{d} ln\frac{1}{\sqrt{2\pi\alpha_i}}e^{-\frac{w^2}{2\alpha_i^2}}]}$$

$$q(w) \propto e^{E_{q-w}[e^{-\frac{\lambda}{2}\sum_{i=1}^{n}(y_i-x_i^T w)^2}+e^{-\frac{\alpha}{2}\sum_{i=1}^{d}(w)^2}+constant\ w.r.t\ w]}$$

$$q(w) \propto e^{E_{q-w}[e^{-\frac{\lambda}{2}\sum_{i=1}^{n}(y_i-x_i^T w)^2-\frac{\alpha}{2}\sum_{i=1}^{d}(w)^2}]}$$

$$q(w) \propto e^{-\frac{E_\lambda[\lambda]}{2}\sum_{i=1}^{n}(y_i-x_i^T w)^2-\frac{E_\alpha[\alpha]}{2}\sum_{i=1}^{d}(w)^2}$$

$$q(w) \propto e^{-\frac{E_\lambda[\lambda]}{2}\sum_{i=1}^{n}y_i^2-2x_i^T wy_i+x_i^T ww^T x_i)-\frac{E_\alpha[\alpha]}{2}w^T w}$$

$$q(w) \propto e^{-\frac{1}{2}[w^T(E_\alpha[\alpha]I+E_\lambda[\lambda]^{-1}\sum_{i=1}^{n}xx_i^T)w-2w^T(E_\lambda[\lambda]^{-1}\sum_{i=1}^{n}y_ix_i)]}e^{-\frac{1}{2}[E_\lambda[\lambda]^{-1}\sum_{i=1}^{n}y_i^2]}$$

$$q(w) \propto e^{-\frac{1}{2}[w^T(E_\alpha[\alpha]I+E_\lambda[\lambda]^{-1}\sum_{i=1}^{n}xx_i^T)w-2w^T(E_\lambda[\lambda]^{-1}\sum_{i=1}^{n}y_ix_i)]}$$

$$q(w) \propto e^{-\frac{1}{2}[w^T(E_\alpha[\alpha]I+E_\lambda[\lambda]^{-1}\sum_{i=1}^{n}xx_i^T)w-2w^T(E_\lambda[\lambda]^{-1}\sum_{i=1}^{n}y_ix_i)]}e^{-\frac{1}{2}[(E_\lambda[\lambda]^{-1}\sum_{i=1}^{n}y_ix_i)^T(E_\alpha[\alpha]I+E_\lambda[\lambda]^{-1}\sum_{i=1}^{n}xx_i^T)^{-1}(E_\lambda[\lambda]^{-1}\sum_{i=1}^{n}y_ix_i)]}$$

$$q(w) \propto e^{-\frac{1}{2}(w-u)^T\Sigma^{-1}(w-u)}$$

$q(w)$ ~ Normal(u',Σ')

$$\Sigma' = (E_\alpha[\alpha]I + E_\lambda[\lambda]^{-1}\sum_{i=1}^{n}xx_i^T)^{-1}$$

$$u' = \Sigma'(E_\lambda[\lambda]^{-1}\sum_{i=1}^{n}y_ix_i)$$

Kate Lassiter

$$q(\lambda) \; = \; e^{E_{q-\lambda}[lnp(y_i|x_i^T w,\lambda^{-1})+lnp(w|0,diag(\alpha_1,...,\alpha_d\;)^{-1})+lnp(a_k|a0,b0\;)+lnp(\lambda|e0,f0\;)]}$$

$$q(\lambda) \; = \; e^{E_{q-\lambda}[lnp(y_i|x_i^T w,\lambda^{-1})+lnp(\lambda|e0,f0\;)]}$$

$$q(\lambda) \; = \; e^{E_{q-\lambda}[\sum_{i=1}^{N} ln\frac{1}{\sqrt{2\pi\lambda}}e^{-\frac{(y_i-x_i^T w)^2}{2\lambda}}+\frac{f0^{e0}}{\Gamma(e0)}ln\lambda^{e0-1}e^{-f0\lambda}]}$$

$$q(\lambda) \; \propto \; e^{E_{q-\lambda}[\sum_{i=1}^{N} ln(\lambda)^{\frac{1}{2}}\; e^{-\frac{\lambda}{2}(y_i-x_i^T w)^2}+ln\lambda^{e0-1}e^{-f0\lambda}+constant\; w.r.t\; \lambda\,]}$$

$$q(\lambda) \propto e^{E_{q-\lambda}[ln\lambda^{\frac{n}{2}}\; e^{-\frac{\lambda}{2}\sum_{i=1}^{N}(y_i-x_i^T w)^2}+ln\lambda^{e0-1}e^{-f0\lambda}+constant\; w.r.t\; \lambda\,]}$$

$$q(\lambda) \propto e^{E_{q-\lambda}[ln\lambda^{\frac{n}{2}+e0-1}\; e^{-\frac{\lambda}{2}\sum_{i=1}^{N}(y_i-x_i^T w)^2-f0\lambda}\;]}$$

$$q(\lambda) \propto \lambda^{\frac{n}{2}+e0-1}\; e^{-\frac{\lambda}{2}\sum_{i=1}^{N}E_w[(y_i-x_i^T w)^2]-f0\lambda}$$

$$\boxed{\begin{array}{l} q(\lambda) \sim \text{Gamma } (e',f') \\ \\ e' = e0 + \frac{n}{2} \end{array}}$$

Isolating f0:

$$0 = -\;\frac{\lambda}{2}\sum_{i=1}^{N} E_w[(y_i - x_i^T w)^2] - f0\lambda$$

$$f0\lambda = -\;\frac{\lambda}{2}\sum_{i=1}^{N} E_w[(y_i - x_i^T w)^2]$$

$$\frac{1}{\lambda} f0\lambda = -\;\frac{\lambda}{2}\sum_{i=1}^{N} E_w[(y_i - x_i^T w)^2]\frac{1}{\lambda}$$

$$f0 = -\;\frac{1}{2}\sum_{i=1}^{N} E_w[(y_i - x_i^T w)^2]$$

$$f' = f0 + \frac{1}{2}\sum_{i=1}^{N} E_w[(y_i - x_i^T w)^2]$$

Recall:

$$\sum_{i=1}^{N} E_w[(y_i - x_i^T w)^2] = \sum_{i=1}^{N}\; (y_i - x_i^T u')^2 + x_i^T\Sigma' x_i$$

Thus:

$$\boxed{f' = f0 + \frac{1}{2}\,(\sum_{i=1}^{N}\; (y_i - x_i^T u')^2 + x_i^T\Sigma' x_i)}$$

Kate Lassiter

b) Summarize the algorithm derived in Part (a) using pseudo-code in a way similar to how algorithms are presented in the notes for the class.

VI algorithm for Bayesian Linear Regression:

Note: $y \sim N(x^T w, \lambda^{-1})$ with unknown precision λ^{-1} and $w \sim N(0,(a_1 ... a_d)^{-1})$ unknown precision $(a_1 ... a_d)^{-1}$

Inputs: Data and definitions $q(\alpha)$ = Gamma $(\alpha_k | a', b')$, $q(\lambda)$ = Gamma $(\lambda | e', f')$, $q(w)$ = Normal $(w | u', \Sigma')$

Output: Values for $a', b', e', f', u', \Sigma'$

1. Initialize $a_0', b_0', e_0', f_0', u_0', \Sigma'_0$ in some way
2. For iteration t=1,..., T
    a. Update $q(\alpha)$ by setting for each d
        i. $a_{dt}' = a + \frac{1}{2}$
        ii. $b_{dt}' = b + \frac{1}{2} \sum_{i=1}^{d} (\Sigma'_{i,i,t-1} + u'_{i,t-1} u'_{i,t-1})$
    b. Update $q(\lambda)$ by setting
        i. $e_t' = e + \frac{n}{2}$
        ii. $f_t' = f + \frac{1}{2} (\sum_{i=1}^{N} (y_i - x_i^T u_{t-1}')^2 + x_i^T \Sigma_{t-1}' x_i)$
    c. Update $q(w)$ by setting
        i. $\Sigma_t' = \left(\frac{a_t'}{b_t'} I + \frac{e_t'}{f_t'} \sum_{i=1}^{n} x x_i^T \right)^{-1}$
        ii. $u_t' = \Sigma_t' (\frac{e_t'}{f_t'} \sum_{i=1}^{n} y_i x_i)$
    d. Evaluate $L(a_t', b_t', e_t', f_t', u_t', \Sigma_t')$ to assess convergence
        i. $L_t = E_q[\ln p(y, w, \alpha, \lambda)] - E_q[\ln q(w, \alpha, \lambda)]$

Kate Lassiter

c) Using these q distributions, calculate the variational objective function. You will need to evaluate this function in the next problem to show the convergence of your algorithm.

VI Objective

$$L = E_q[lnp(y, w, \alpha, \lambda|x)] - E_q[lnq(w, \alpha, \lambda)]$$

Focus first on Decomposing: $E_q[lnp(y, w, \alpha, \lambda|x)]$

$$L = E_q[lnp(y_i|x_i^T w, \lambda^{-1}) + lnp(w|0, diag(\alpha_1, ..., \alpha_d)^{-1}) + lnp(a_k|a0, b0) + lnp(\lambda|e0, f0)] - E_q[lnq(w, \alpha, \lambda)]$$

$$L = E_q[ln\frac{1}{\sqrt{2\pi\lambda}}e^{-\frac{(y_i - x_i^T w)^2}{2\lambda^2}} + ln\frac{1}{\sqrt{2\pi\alpha}}e^{-\frac{(w)^2}{2\alpha^2}} + lna_k^{ao-1}e^{-b0a_k} + ln\lambda^{e0-1}e^{-f0\lambda}] - E_q[lnq(w, \alpha, \lambda)]$$

$$L = E_q[ln\left(\frac{\lambda}{2\pi}\right)^{\frac{n}{2}}e^{-\frac{\lambda}{2}\sum_{i=1}^{n}(y_i - x_i^T w)^2} + \sum_{i=1}^{d} ln\left(\frac{\alpha}{2\pi}\right)^{\frac{1}{2}}e^{-\frac{\alpha}{2}\sum_{i=1}^{d}(w)^2} + \sum_{i=1}^{d} lna_i^{ao-1}e^{-b0a_i} + ln\lambda^{e0-1}e^{-f0\lambda}] - E_q[lnq(w, \alpha, \lambda)]$$

$$L = E_q[\frac{n}{2}ln\left(\frac{\lambda}{2\pi}\right) - \frac{\lambda}{2}\sum_{i=1}^{n}(y_i - x_i^T w)^2 + \frac{1}{2}\sum_{i=1}^{d} ln\left(\frac{\alpha}{2\pi}\right) - \frac{\alpha}{2}\sum_{i=1}^{d}(w)^2 + \sum_{i=1}^{d}((ao - 1)lna_i - b0a_i) + (e0 - 1)ln\lambda - f0\lambda + constant] - E_q[lnq(w, \alpha, \lambda)]$$

$$L = [\frac{n}{2}E_\lambda[ln\left(\frac{\lambda}{2\pi}\right)] - \frac{E_\lambda[\lambda]}{2}E_w[\sum_{i=1}^{n}(y_i - x_i^T w)^2] + \frac{1}{2}\sum_{i=1}^{d} E_\alpha[ln\left(\frac{\alpha}{2\pi}\right)] - \frac{E_\alpha[\alpha]}{2}E_w[\sum_{i=1}^{d}(w)^2] + \sum_{i=1}^{d}((ao - 1)E_\alpha[lna_i] - b0E_\alpha[a_i]) + (e0 - 1)E_\lambda[ln\lambda] - f0E_\lambda[\lambda] - E_q[lnq(w, \alpha, \lambda)]$$

$$L = \frac{n}{2}E_\lambda[ln(\lambda)] - \frac{E_\lambda[\lambda]}{2}E_w[\sum_{i=1}^{n}(y_i - x_i^T w)^2] + \frac{1}{2}\sum_{i=1}^{d} E_\alpha[ln(\alpha)] - \frac{E_\alpha[\alpha]}{2}E_w[\sum_{i=1}^{d}(w)^2] + \sum_{i=1}^{d}((ao - 1)E_\alpha[lna_i] - b0E_\alpha[a_i]) + (e0 - 1)E_\lambda[ln\lambda] - f0E_\lambda[\lambda] - E_q[lnq(w, \alpha, \lambda)]$$

Recall:

$E_\lambda[\lambda_{e,f}] = \frac{e}{f}$

$E_\lambda[ln\lambda_{e,f}] = \varphi(e) - ln\, f$

$E_\alpha[\alpha_{a,b}] = \frac{a}{b}$

$E_\alpha[ln\alpha_{a,b}] = \varphi(a) - ln\, b$

$$L = [\frac{n}{2}(\varphi(e') - ln\, f') - \sum_{i=1}^{n}\frac{1}{2}\frac{e'}{f'}((y_i - x_i^T u')^2 + x_i^T\Sigma' x_i) + \frac{1}{2}(\sum_{i=1}^{d}(\varphi(a_i') - ln\, b_i') - \sum_{i=1}^{d}\frac{a_i'}{b_i'}\Sigma'_{i,i} + u'_i u'^T_i) + \sum_{i=1}^{d}((ao - 1)(\varphi(a_i') - ln\, b_i') - b0\frac{a_i'}{b_i'}) + (e0 - 1)(\varphi(e') - ln\, f') - f0\frac{e'}{f'}] - E_q[lnq(w, \alpha, \lambda)]$$

Now decomposing $E_q[lnq(w, \alpha, \lambda)]$

$$L = E_q[lnp(y, w, \alpha, \lambda|x)] - E_q[lnq(w, \alpha, \lambda)]$$

$$L = E_q[lnp(y, w, \alpha, \lambda|x)] - E_q[lnq(w|u', \Sigma') + lnq(\alpha_k|a', b') + lnq(\lambda|e', f')]$$

$$L = E_q[lnp(y, w, \alpha, \lambda|x)] - E_q[ln\frac{1}{\sqrt{2\pi\Sigma'}}e^{-\frac{(y_i - x_i^T u'')^2}{2\Sigma'}} + ln\frac{b_i'^{a_i'}}{\Gamma(a_i')}\alpha_k^{a_i'-1}e^{-b_i'\alpha_k} + ln\frac{f'^{e'}}{\Gamma(e')}\lambda^{e'-1}e^{-f'\lambda}]$$

$$L = E_q[lnp(y, w, \alpha, \lambda|x)] - E_q[\int ln\left(\frac{\Sigma'}{2\pi}\right)^{\frac{1}{2}}e^{-\frac{\Sigma'}{2}(y_i - x_i^T u')^2}dw + \sum_{i=1}^{d} ln\frac{b_i'^{a_i'}}{\Gamma(a_i')}\alpha_i^{a_i'-1}e^{-b_i'\alpha_i} + ln\lambda^{e'-1}e^{-f'\lambda}]$$

$$L = E_q[lnp(y, w, \alpha, \lambda|x)] - E_q[\int ln\left(\frac{\Sigma'}{2\pi}\right)^{\frac{1}{2}}e^{-\frac{1}{2}(w-u')^T\Sigma'^{-1}(w-u')}dw + \sum_{i=1}^{d} ln\frac{b_i'^{a_i'}}{\Gamma(a_i')}\alpha_i^{a_i'-1}e^{-b_i'\alpha_i} + ln\lambda^{e'-1}e^{-f'\lambda}]$$

Recall: $\int e^{-\frac{1}{2}(w-u)^T\Sigma^{-1}(w-u)}dw = (2\pi)^{\frac{d}{2}}|\Sigma|^{\frac{1}{2}}$

$$L = E_q[lnp(y, w, \alpha, \lambda|x)] - E_q[ln(2\pi)^{-\frac{d}{2}}|\Sigma'|^{-\frac{1}{2}}e^{-\frac{1}{2}(w-u')^T\Sigma'^{-1}(w-u')}) + (\sum_{i=1}^{d}(a_i'lnb_i' - ln\Gamma(a_i') + (a_i' - 1)ln\alpha_i - b_i'\alpha_i)) + ((e'lnf' - ln\Gamma(e') + (e' - 1)ln\lambda - f'\lambda))]$$

$$L = E_q[lnp(y, w, \alpha, \lambda|x)] - E_q[(-\frac{d}{2}ln(2\pi) - \frac{1}{2}ln|\Sigma'| - \frac{1}{2}(w - u')^T\Sigma'^{-1}(w - u')) + (\sum_{i=1}^{d}(a_i'lnb_i' - ln\Gamma(a_i') + (a_i' - 1)ln\alpha_i - b_i'\alpha_i)) + ((e'lnf' - ln\Gamma(e') + (e' - 1)ln\lambda - f'\lambda))]$$

$$L = E_q[lnp(y, w, \alpha, \lambda|x)] - E_q[(-\frac{d}{2}ln(2\pi) - \frac{1}{2}ln|\Sigma'| - \frac{1}{2}(w - u')^T\Sigma'^{-1}(w - u')) + (\sum_{i=1}^{d}(a_i'lnb_i' - ln\Gamma(a_i') + (a_i' - 1)ln\alpha_i - b_i'\alpha_i)) + ((e'lnf' - ln\Gamma(e') + (e' - 1)ln\lambda - f'\lambda))]$$

$$L = E_q[lnp(y, w, \alpha, \lambda|x)] - [(-\frac{d}{2}ln(2\pi) - \frac{1}{2}ln|\Sigma'| - E_w[\frac{1}{2}(w - u')^T\Sigma'^{-1}(w - u')]) + (\sum_{i=1}^{d}(a_i'lnb_i' - ln\Gamma(a_i') + (a_i' - 1)E_\alpha[ln\alpha_i] - b_i'E_\alpha[\alpha_i])) + ((e'lnf' - ln\Gamma(e') + (e' - 1)E_\lambda[ln\lambda] - f'E_\lambda[\lambda])]$$

Recall:

$E_w[(w - u')^T\Sigma'^{-1}(w - u')]$ **=** $E_w[trace((w - u')^T\Sigma'^{-1}(w - u'))]$

= $trace(\Sigma'^{-1}E_w[(w - u')(w - u')^T])$

= $trace(\Sigma'^{-1}\Sigma') = trace(I) = d$

$$L = E_q[lnp(y, w, \alpha, \lambda|x)] - [(-\frac{d}{2}ln(2\pi) - \frac{1}{2}ln|\Sigma'| - \frac{d}{2}) + (\sum_{i=1}^{d}(a_i'lnb_i' - ln\Gamma(a_i') + (a_i' - 1)(\varphi(a_i') - ln\, b_i') - b_i'\frac{a_i'}{b_i'})) + ((e'lnf' - ln\Gamma(e') + (e' - 1)(\varphi(e') - ln\, f') - f'\frac{e'}{f'}))]$$

$$L = E_q[lnp(y, w, \alpha, \lambda|x)] - [(-\frac{1}{2}ln|\Sigma'|) + (\sum_{i=1}^{d}(a_i'lnb_i' - ln\Gamma(a_i') + a_i'\varphi(a_i') - a_i'ln\, b_i' - \varphi(a_i') + ln\, b_i' - b_i'\frac{a_i'}{b_i'})) + ((e'lnf' - ln\Gamma(e') + e'\varphi(e') - e'ln\, f' - \varphi(e') + ln\, f' - e'))]$$

$$L = E_q[lnp(y, w, \alpha, \lambda|x)] - [(-\frac{1}{2}ln|\Sigma'|) + (\sum_{i=1}^{d}(a_i'lnb_i' - a_i'ln\, b_i' - ln\Gamma(a_i') + \varphi(a_i')(a_i' - 1) + ln\, b_i' - a_i'))) + ((e'lnf' - e'lnf' - ln\Gamma(e') + \varphi(e')(e' - 1) + ln\, f' - e'))]$$

$$L = E_q[lnp(y, w, \alpha, \lambda|x)] - [-\frac{1}{2}ln|\Sigma'|] - [\sum_{i=1}^{d}(-ln\Gamma(a_i') + \varphi(a_i')(a_i' - 1) + ln\, b_i' - a_i'))] - [(-ln\Gamma(e') + \varphi(e')(e' - 1) + ln\, f' - e')]$$

$$L = E_q[lnp(y, w, \alpha, \lambda|x)] + [\frac{1}{2}ln|\Sigma'|] + [\sum_{i=1}^{d}(ln\Gamma(a_i') + (1 - a_i')\varphi(a_i') - ln\, b_i' + a_i'))] + [(ln\Gamma(e') + (1 - e')\varphi(e') - ln\, f' + e')]$$

VI:

Kate Lassiter

$$L = [\frac{n}{2}(\varphi(e') - \ln f') - \sum_{i=1}^{n} \frac{1}{2}\frac{e'}{f'}((y_i - x_i^T u')^2 + x_i^T \Sigma' x_i) + \frac{1}{2}(\sum_{i=1}^{d}(\varphi(a_i') - \ln b_i') - \sum_{i=1}^{d} \frac{a_i'}{b_i'}\Sigma'_{i,i} + u'_i u'^T_i) + \sum_{i=1}^{d}((ao - 1)(\varphi(a_i') - \ln b_i') - b0\frac{a_i'}{b_i'}) + (e0 - 1)(\varphi(e') - \ln f') - f0\frac{e'}{f'}]$$
$$+ [\frac{1}{2}\ln|\Sigma'|] + [\sum_{i=1}^{d}(\ln\Gamma(a_i') + (1 - a_i')\varphi(a_i') - \ln b_i' + a_i'))] + [(\ln\Gamma(e') + (1 - e')\varphi(e') - \ln f' + e')] + \text{constant}$$

Problem 2. (35 points)

Implement the algorithm derived in Problem 1 and run it on the three data sets provided. Set the prior parameters $a_0 = b_0 = 10^{-16}$ and $e_0 = f_0 = 1$. We will not discuss sparsity-promoting "ARD" priors in detail in this course, but setting a_0 and b_0 in this way will encourage only a few dimensions of w to be significantly non-zero since many α_k should be extremely large according to $q(\alpha_k)$.

For each of the three data sets provided, show the following:

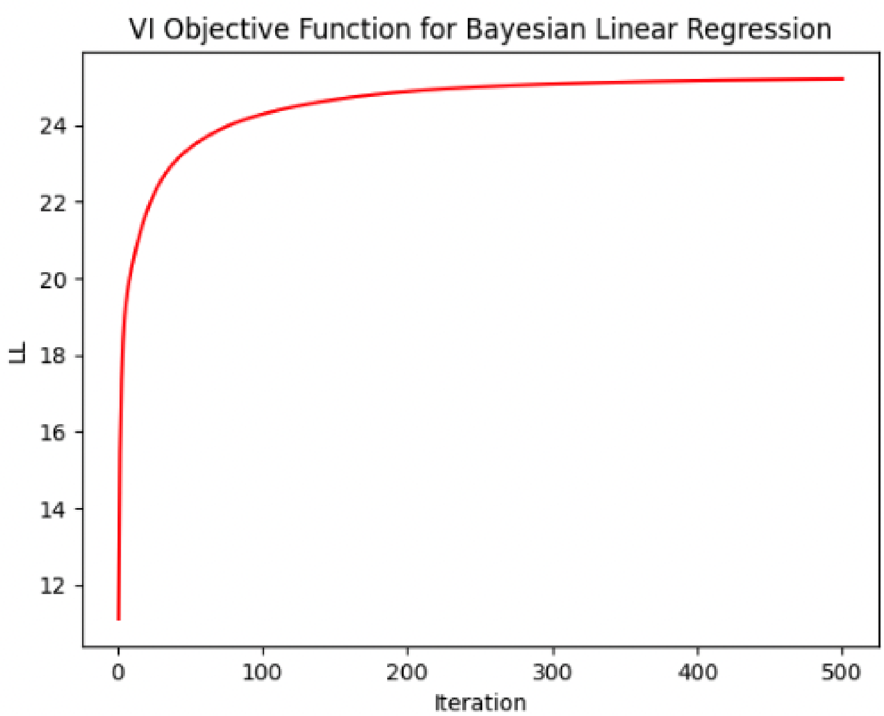
a) Run your algorithm for 500 iterations and plot the variational objective function.

b) Using the final iteration, plot $1/\mathbb{E}_q[\alpha_k]$ as a function of k.

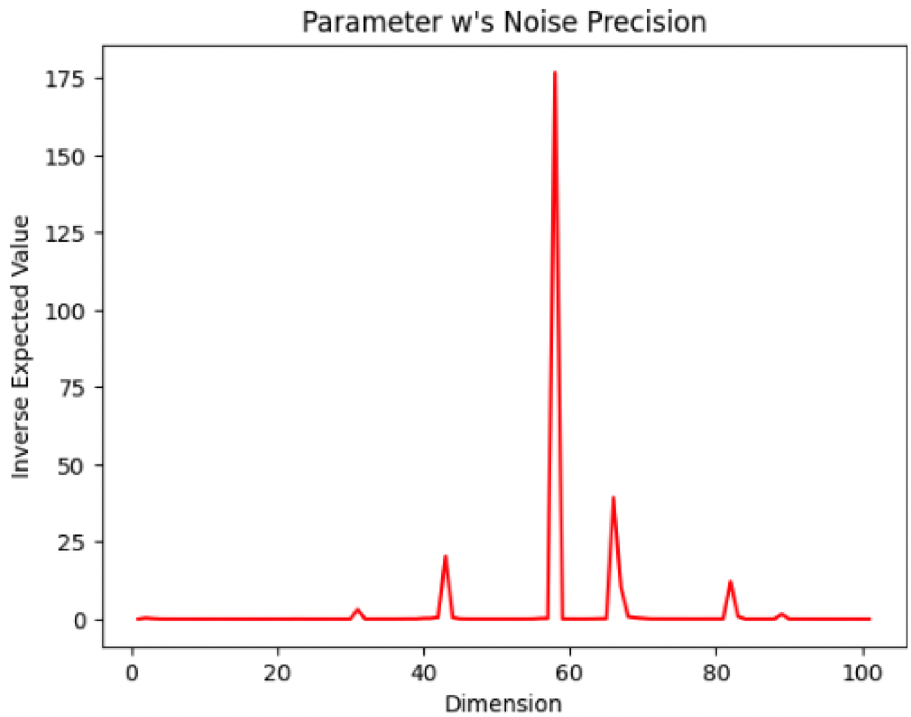
c) Give the value of $1/\mathbb{E}_q[\lambda]$ for the final iteration.

d) Using $\hat{w} = \mathbb{E}_{q(w)}[w]$, calculate $\hat{y}_i = x_i^T\hat{w}$ for each data point. Using the z_i associated with y_i (see below), plot $\hat{y}_i$ vs z_i as a solid line. On the same plot show (z_i, y_i) as a scatter plot. Also show the function $(z_i, 10 * \text{sinc}(z_i))$ as a solid line in a different color.

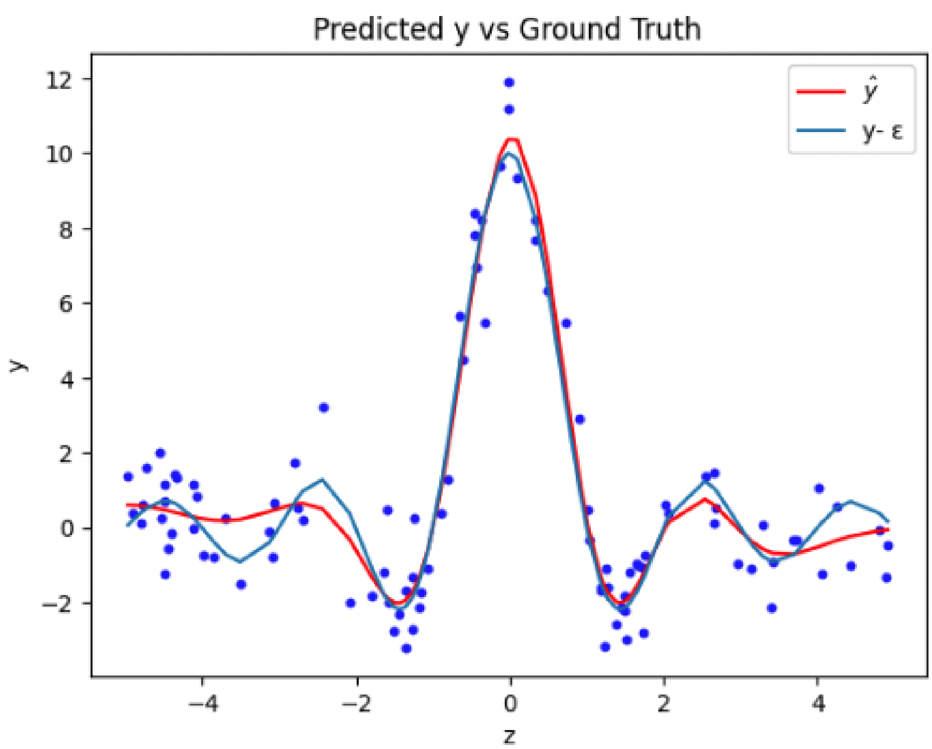
<u>Dataset 1</u>



Kate Lassiter

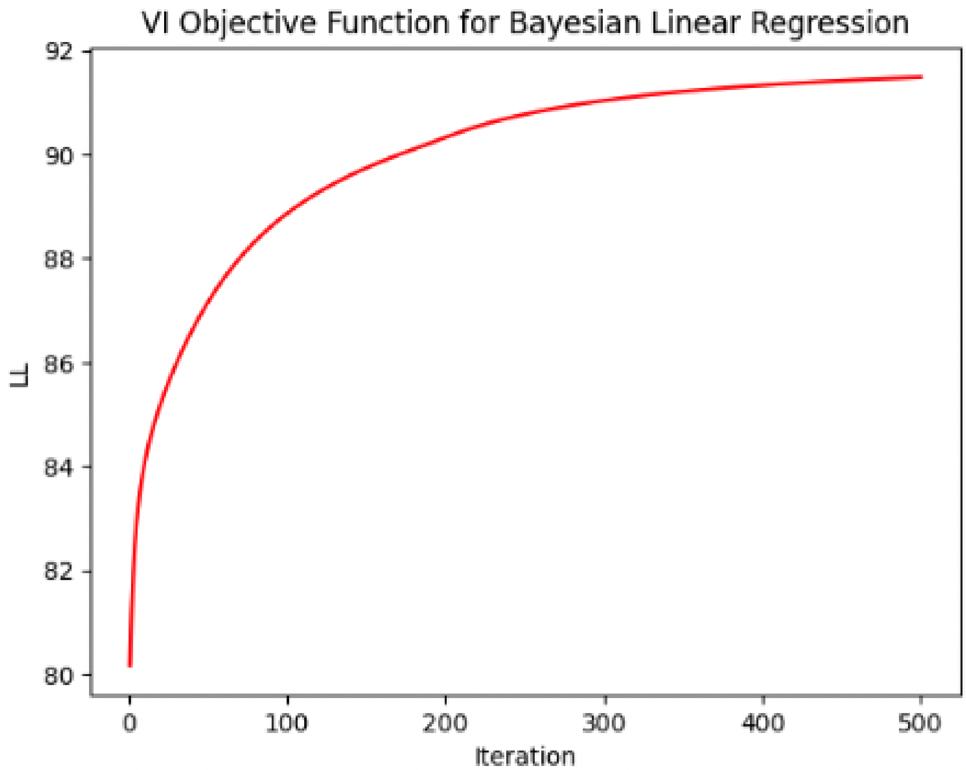


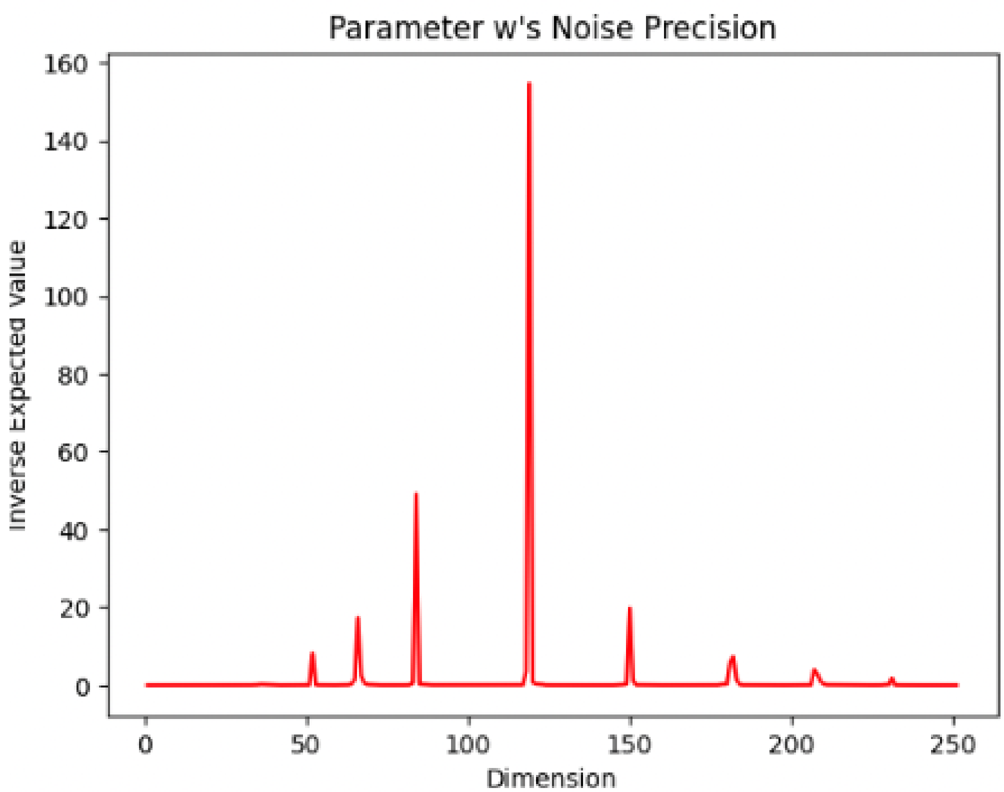
1/ E[λ]: 1.0155304519679365



Kate Lassiter

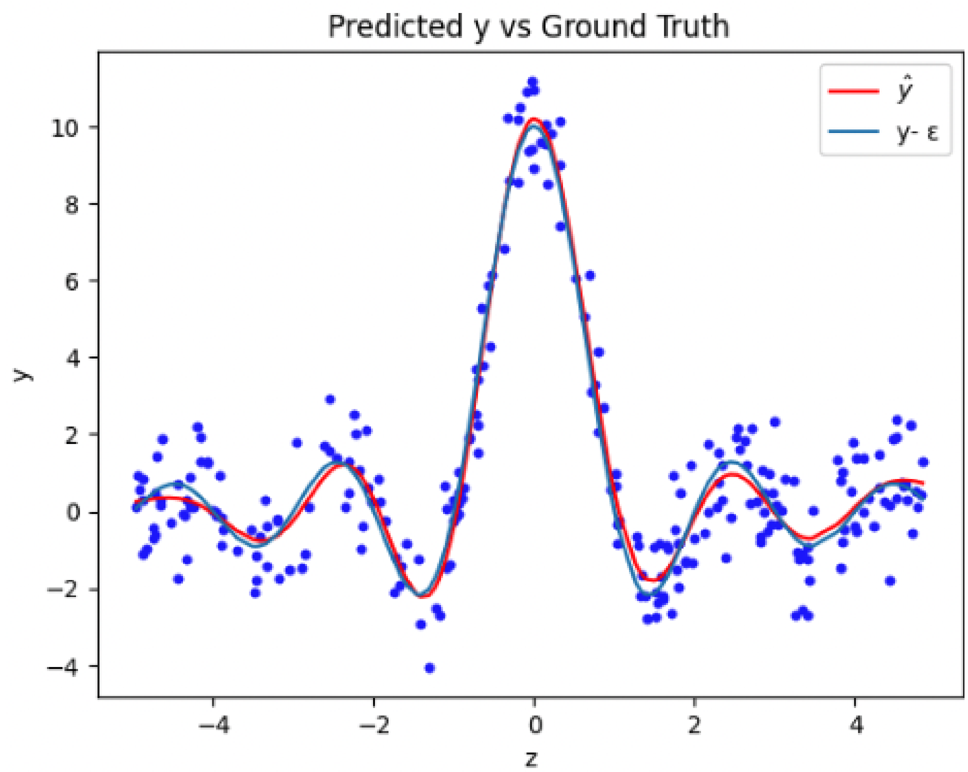
Dataset 2



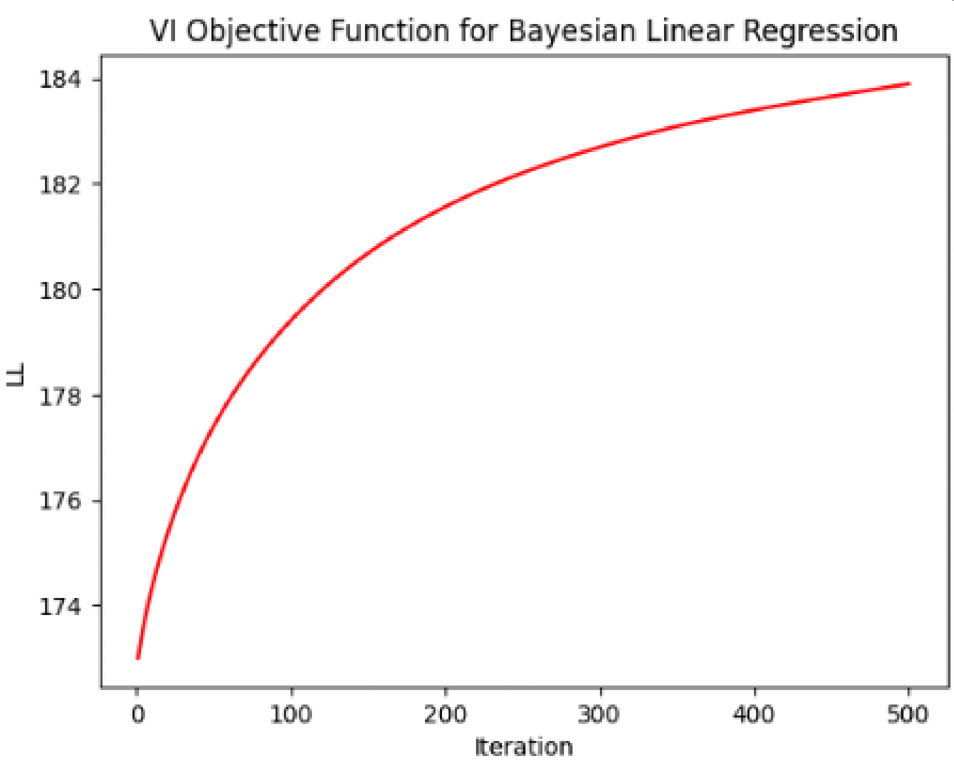


Kate Lassiter

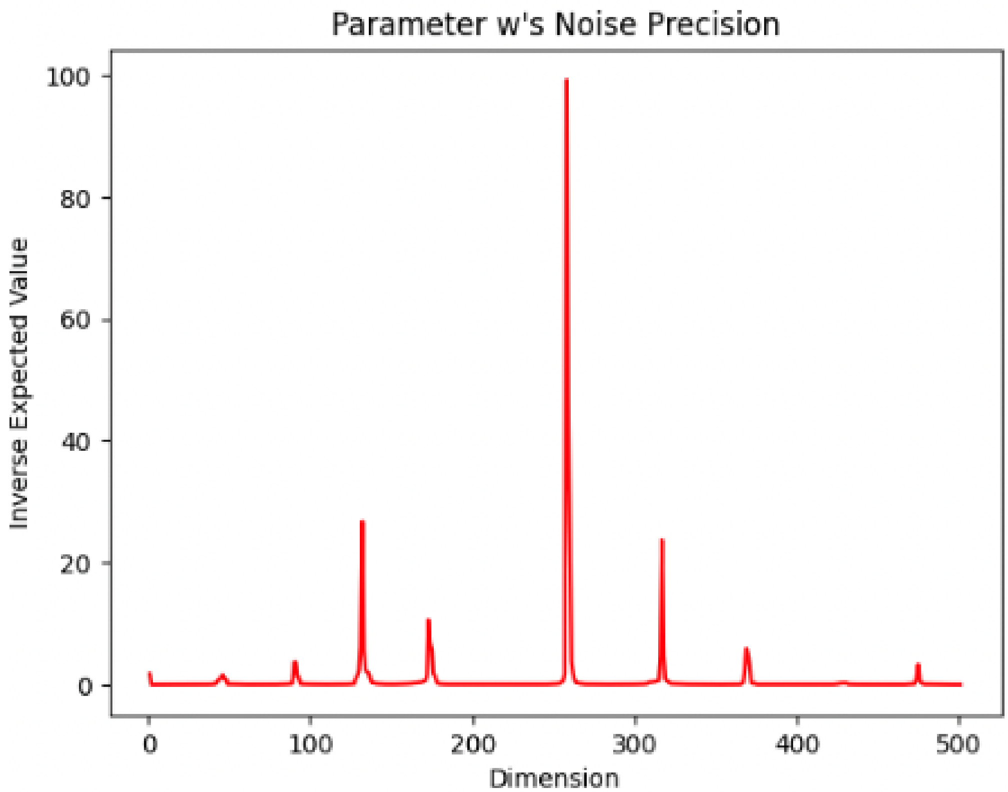
1/ E[λ]: 0.9100447162731942



Dataset 3



Kate Lassiter



1/ E[λ]: 0.9744161777751516