Homework 3 SOLUTIONS (WRITTEN)

Problem 1. Problem 1(a).

$q(\alpha_k)$ We know from the general approach that we can find $q(\alpha_k)$ as follows

$$q(\alpha_k) \propto p(\alpha_k) \cdot \exp\{\mathbb{E}[\ln p(w|\alpha)]\}$$

– Note that $p(w|\alpha) = \prod_{k=1}^{d} p(w_k|\alpha_k)$, where $p(w_k|\alpha_k) = \sqrt{\alpha_k/2\pi}\exp\{-\alpha_k w_k^2/2\}$. We have

$$\begin{aligned} q(\alpha_k) &\propto p(\alpha_k) \cdot \exp\{\mathbb{E}[\ln p(w_k|\alpha_k)]\} \\ &\propto \alpha_k^{a_0-1}\exp\{-b_0\alpha_k\} \cdot \exp\Big\{\frac{\ln \alpha_k}{2} - \frac{\alpha_k \mathbb{E}[w_k^2]}{2}\Big\} \\ &= \alpha_k^{a_0+1/2-1}\exp\Big\{-\Big(b_0 + \frac{\mathbb{E}[w_k^2]}{2}\Big)\alpha_k\Big\}. \end{aligned}$$

We don't know the exact form of $q(w)$, but we can specify the optimal form of $q(\alpha_k)$ by the above equation. That is, $q(\alpha_k) \sim \text{Gamma}(a_0 + \frac{1}{2},\ b_0 + \frac{\mathbb{E}[w_k^2]}{2}) = \text{Gamma}(a_k', b_k')$.

$q(\lambda)$ Next we know form general approach that we can find $q(\lambda)$

$$\begin{aligned} q(\lambda) &\propto p(\lambda) \cdot \exp\{\sum_{i=1}^{N} \mathbb{E}\ln p(y_i|x_i, w, \lambda)\}. \\ &\propto p(\lambda) \cdot \exp\Big\{\sum_{i=1}^{N}\Big(\frac{\ln\lambda}{2} - \frac{\lambda\cdot\mathbb{E}(x_i^\top w - y_i)^2}{2}\Big)\Big\} \\ &= \lambda^{e_0+N/2-1}\exp\Big\{-\Big(f_0 + \frac{\sum_{i=1}^{N}\mathbb{E}[(x_i^\top w - y_i)^2]}{2}\Big)\lambda\Big\} \end{aligned}$$

We don't know the exact form of $q(w)$, but we can specify the optimal form of $q(\lambda)$ by the above equation. That is, $q(\lambda) \sim \text{Gamma}(e_0 + \frac{N}{2},\ f_0 + \frac{\sum_{i=1}^{N}\mathbb{E}[(x_i^\top w - y_i)^2]}{2}) = \text{Gamma}(e', f')$.

$q(w)$ Finally we know the general approach that we can find $q(w)$

$$\begin{aligned} q(w) &\propto \exp\{\mathbb{E}\ln p(w|\alpha) + \sum_{i=1}^{N}\mathbb{E}\ln p(y_i|x_i, w, \lambda)\} \\ &\propto \exp\Big\{-\frac{\mathbb{E}[w^\top \text{diag}(\alpha_1, \dots, \alpha_d)w]}{2}\Big\} \cdot \exp\Big\{-\sum_{i=1}^{N}\frac{\mathbb{E}[\lambda(x_i^\top w - y_i)^2]}{2}\Big\} \\ &\propto \exp\Big\{-\frac{w^\top\Big(\text{diag}(\mathbb{E}[\alpha_1], \dots, \mathbb{E}[\alpha_d])\Big)w}{2} + \mathbb{E}[\lambda]\cdot\sum_{i=1}^{N} y_i x_i^\top w\Big\} \end{aligned}$$

We have derived the optimal form of $q(\alpha_k)$ and $q(\lambda)$. Thus, we can specify that $q(w) \sim \text{Normal}(\mu', \Sigma')$, where $\mu' = \Sigma' \cdot (\mathbb{E}[\lambda]\sum_{i=1}^{N} y_i x_i)$, $\Sigma' = \Big(\text{diag}(\mathbb{E}[\alpha_1], \dots, \mathbb{E}[\alpha_d]) + \mathbb{E}[\lambda]\cdot\sum_{i=1}^{N} x_i x_i^\top\Big)^{-1}$.

Problem 1(b).

VI algorithm for Bayesian linear regression

Inputs: Data and definitions $q(\alpha_k) = \text{Gamma}(\alpha_k | a'_k, b'_k)$, $q(\lambda) = \text{Gamma}(e', f')$, and $q(w) = \text{Normal}(\mu', \Sigma')$

Output: Values for a'_k, b'_k, e', f', μ' and Σ' in some way

1 Initialize $a'_{k,0}, b'_{k,0}, e'_0, f'_0, \mu'_0$ and Σ'_0 in some way

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

- Update $q_t(\alpha_k)$, for $1 \leq k \leq d$, by setting

$$a'_{k,t} = a_0 + \frac{1}{2}$$
$$b'_{k,t} = b_0 + \frac{(\mu'_{k,t-1})^2 + \Sigma'_{kk,t-1}}{2}$$

- Update $q_t(\lambda)$ by setting

$$e'_t = e_0 + \frac{N}{2}$$
$$f'_t = f_0 + \frac{\sum_{i=1}^N \Big((x_i^\top \mu'_{t-1} - y_i)^2 + x_i^\top \Sigma'_{t-1} x_i \Big)}{2}$$

- Update $q_t(w)$ by setting

$$\Sigma'_t = \Big(\text{diag}(\frac{a'_{1,t}}{b'_{1,t}}, \ldots, \frac{a'_{d,t}}{b'_{d,t}}) + \frac{e'_t}{f'_t} \cdot \sum_{i=1}^N x_i x_i^\top \Big)^{-1}$$
$$\mu'_t = \Sigma'_t \cdot \Big(\frac{e'_t}{f'_t} \sum_{i=1}^N y_i x_i \Big)$$

- Evaluate $\mathcal{L}(a'_{k,t}, b'_{k,t}, e'_t, f'_t, \mu'_t, \Sigma'_t)$ to assess convergence (i.e. decide T).

Problem 1(c).

$$
\begin{aligned}
\mathcal{L}_t = & \sum_{i=1}^{N} \mathbb{E} \ln p(y_i|x_i, w, \lambda) + \sum_{k=1}^{d} \mathbb{E} \ln p(w_k|\alpha_k) + \sum_{k=1}^{d} \mathbb{E} \ln p(\alpha_k) + \mathbb{E} \ln p(\lambda) \\
& - \mathbb{E} \ln q(w) - \sum_{k=1}^{d} \mathbb{E} \ln q(\alpha_k) - \mathbb{E} \ln q(\lambda),
\end{aligned}
$$

where

$$
\begin{aligned}
\mathbb{E} \ln p(y_i|x_i, w, \lambda) &= \frac{1}{2}\mathbb{E}[\ln \lambda] - \frac{1}{2} \ln(2\pi) - \frac{1}{2}\mathbb{E}[\lambda] \cdot \mathbb{E}[(x_i^\top w - y_i)^2] \\
&= \frac{1}{2}(\psi(e') - \ln(f')) - \frac{1}{2} \ln(2\pi) - \frac{1}{2} \cdot \frac{e'}{f'} \cdot \Big((x_i^\top \mu' - y_i)^2 + x_i^\top \Sigma' x_i\Big) \\
\mathbb{E} \ln p(w_k|\alpha_k) &= \frac{1}{2}\mathbb{E}[\ln \alpha_k] - \frac{1}{2} \ln(2\pi) - \frac{1}{2}\mathbb{E}[\alpha_k] \cdot \mathbb{E}[w_k^2] \\
&= \frac{1}{2}(\psi(a_k') - \ln(b_k')) - \frac{1}{2} \ln(2\pi) - \frac{1}{2} \cdot \frac{a_k'}{b_k'} \cdot ((\mu_k')^2 + \Sigma_{kk}') \\
\mathbb{E} \ln p(\alpha_k) &= a_0 \ln b_0 - \ln \Gamma(a_0) + (a_0 - 1)\mathbb{E}[\ln \alpha_k] - b_0\mathbb{E}[\alpha_k] \\
&= a_0 \ln b_0 - \ln \Gamma(a_0) + (a_0 - 1)(\psi(a_k') - \ln(b_k')) - b_0 \cdot \frac{a_k'}{b_k'} \\
\mathbb{E} \ln p(\lambda) &= e_0 \ln f_0 - \ln \Gamma(e_0) + (e_0 - 1)(\psi(e') - \ln(f')) - f_0 \cdot \frac{e'}{f'}
\end{aligned}
$$

$$
\begin{aligned}
-\mathbb{E} \ln q(w) &= H(q(w)) = \frac{d(1 + \ln(2\pi))}{2} + \frac{\ln |\Sigma'|}{2} \\
-\mathbb{E} \ln q(\alpha_k) &= H(q(\alpha_k)) = a_k' - \ln b_k' + \ln \Gamma(a_k') + (1 - a_k')\psi(a_k') \\
-\mathbb{E} \ln q(\lambda) &= H(q(\lambda)) = e' - \ln f' + \ln \Gamma(e') + (1 - e')\psi(e').
\end{aligned}
$$