Bayesian Models for Machine Learning

Haoyang Chen hc2812

a) Denote $X=\{x_1,\cdots,x_N\}$, $Y=\{y_1,\cdots,y_N\}$, $\alpha=\{\alpha_1,\cdots,\alpha_d\}$

$$P(Y,W,\alpha,\lambda|X)=P(Y|X,W,\lambda)P(W|\alpha)P(\lambda)P(\alpha)$$

$$\therefore \ln P(Y,W,\alpha,\lambda|X)=\ln P(Y|X,W,\lambda)+\ln P(W|\alpha)+\ln P(\lambda)+\ln P(\alpha)$$

$$=\sum_{i=1}^{N}\ln P(y_i|x_i,W,\lambda)+\ln P(W|\alpha)+\ln P(\lambda)+\sum_{i=1}^{d}\ln P(\alpha_i)$$

$$=\frac{N}{2}\ln\lambda-\frac{N}{2}\ln 2\pi-\frac{\lambda}{2}\sum_{i=1}^{N}(y_i^2-2W^Tx_iy_i+W^Tx_ix_i^TW)+\frac{1}{2}\sum_{i=1}^{d}\ln\alpha_i-\frac{1}{2}\ln 2\pi-\frac{1}{2}(W^T diag(\alpha_1,\cdots,\alpha_d)W)$$

$$+e_0\ln f_0-\ln\Gamma(e_0)+(e_0-1)\ln\lambda-f_0\lambda+\sum_{i=1}^{d}(a_0\ln b_0-\ln\Gamma(a_0)+(a_0-1)\ln\alpha_i-b_0\alpha_i)$$

- $q(W)=\exp\{E_{q(\alpha,\lambda)}[\ln P(Y,W,\alpha,\lambda|X)]\}\propto\exp\{E_{q(\alpha,\lambda)}[-\frac{\lambda}{2}\sum_{i=1}^{N}(y_i^2-2W^Tx_iy_i+W^Tx_ix_i^TW)-\frac{1}{2}(W^T diag(\alpha_1,\cdots,\alpha_d)W)]\}$

$$=\exp\{-\frac{1}{2}[E_{q(\lambda)}\lambda\sum_{i=1}^{N}(y_i^2-2W^Tx_iy_i+W^Tx_ix_i^TW)+(W^TE_{q(\alpha)}(diag(\alpha_1,\cdots,\alpha_d))W)]$$

$$\propto\exp\{-\frac{1}{2}[W^T(E_{q(\alpha)}(diag(\alpha_1,\cdots,\alpha_d)+E_{q(\lambda)}\lambda\sum_{i=1}^{N}x_ix_i^T)W-2W^T(E_{q(\lambda)}\lambda\sum_{i=1}^{N}x_iy_i)]$$

$\therefore q(W)=Normal(\mu',\Sigma')$ where $\Sigma'=(E_{q(\alpha)}[diag(\alpha_1,\cdots,\alpha_d)]+E_{q(\lambda)}\lambda\sum_{i=1}^{N}x_ix_i^T)^{-1}$

$$\mu'=\Sigma'(E_{q(\lambda)}\lambda\sum_{i=1}^{N}y_ix_i)$$

- $q(\lambda)=\exp\{E_{q(W,\alpha)}[\ln P(Y,W,\alpha,\lambda|X)]\}\propto\exp\{E_{q(W,\alpha)}[\frac{N}{2}\ln\lambda-\frac{\lambda}{2}\sum_{i=1}^{N}(y_i^2-2W^Tx_iy_i+W^Tx_ix_i^TW)+(e_0-1)\ln\lambda-f_0\lambda]\}$

$$=\exp\{\frac{N}{2}\ln\lambda-\frac{\lambda}{2}\sum_{i=1}^{N}(y_i^2-2E_{q(W)}(W^T)x_iy_i+E_{q(W)}(W^T)x_ix_i^TE_{q(W)}(W))+(e_0-1)\ln\lambda-f_0\lambda\}$$

$$=\exp\{\ln\lambda(\frac{N}{2}+e_0-1)-\lambda(\frac{1}{2}\sum_{i=1}^{N}(y_i^2-2E_{q(W)}(W^T)x_iy_i+E_{q(W)}(W^T)x_ix_i^TE_{q(W)}(W))+f_0)\}$$

$\therefore q(\lambda)=Gamma(e',f')$ where $e'=\frac{N}{2}+e_0$ $\quad f'=\frac{1}{2}\sum_{i=1}^{N}(y_i^2-2E_{q(W)}(W^T)x_iy_i+E_{q(W)}(W^T)x_ix_i^TE_{q(W)}(W))+f_0$

- $q(\alpha)=\exp\{E_{q(W,\lambda)}[\ln P(Y,W,\alpha,\lambda|X)]\}\propto\exp\{E_{q(W,\lambda)}[\frac{1}{2}\sum_{i=1}^{d}\ln\alpha_i-\frac{1}{2}(W^T diag(\alpha_1,\cdots,\alpha_d)W)+(a_0-1)\sum_{i=1}^{d}\ln\alpha_i-b_0\sum_{i=1}^{d}\alpha_i]\}$

Since $q(\alpha)=\sum_{i=1}^{d}q(\alpha_i)$

$$q(\alpha_i)\propto\exp\{E_{q(W,\lambda)}[\frac{1}{2}\ln\alpha_i-\frac{1}{2}(W^T diag(\alpha_1,\cdots,\alpha_d)W)+(a_0-1)\ln\alpha_i-b_0\alpha_i]\}$$

$$=\exp\{\frac{1}{2}\ln\alpha_i-\frac{1}{2}Tr(diag(\alpha_1,\cdots,\alpha_d))E_{q(W)}(W^TW)+(a_0-1)\ln\alpha_i-b_0\alpha_i]\}\propto\exp\{\frac{1}{2}\ln\alpha_i-\frac{1}{2}E_{q(W)}(W_i^2)\alpha_i+(a_0-1)\ln\alpha_i-b_0\alpha_i\}$$

$\therefore q(\alpha_i) \propto \exp\{\ln\alpha_i(\frac{1}{2}+a_0-1) - \alpha_i(\frac{1}{2}E_{q(w)}[w_i^2]+b_0)\}$

$\therefore q(\alpha_i) = \text{Gamma}(a_i', b_i')$ where $a_i' = \frac{1}{2}+a_0$. $b_i' = \frac{1}{2}E_{q(w)}[w_i^2]+b_0$

For this problem, $E_{q(\lambda)}[\lambda] = e'/f'$, $E_{q(\alpha)}[\alpha_i] = \frac{a'}{b'}$, $E_{q(w)}[w] = \mu'$, $E_{q(w)}[ww^T] = \Sigma' + \mu'\mu'^T$

Thus $q(w) = \text{Normal}(\mu', \Sigma')$ where $\Sigma' = (\text{diag}(\frac{a_1'}{b_1'}, \dots, \frac{a_d'}{b_d'}) + \frac{e'}{f'}\sum_{i=1}^N x_i x_i^T)^{-1}$, $\mu' = \Sigma'(\frac{e'}{f'}\sum_{i=1}^N y_i x_i)$

$q(\lambda) = \text{Gamma}(e', f')$ where $e' = \frac{N}{2}+e_0$, $f' = \frac{1}{2}\sum_{i=1}^N[(y_i - \mu'^T x_i)^2 + x_i^T\Sigma' x_i] + f_0$

$q(\alpha_i) = \text{Gamma}(a_i', b_i')$ where $a_i' = \frac{1}{2}+a_0$, $b_i' = \frac{1}{2}(\Sigma' + \mu'\mu'^T)_{[i,i]} + b_0$

(b). Pseudo-code:

1. Initialize $a_0', b_0', e_0', f_0', \mu_0'$ and Σ_0' in some way.

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

   - for $k = 1, \dots, d$, update $q(\alpha_k)$ by setting:

     $a_{kt}' = a_0 + \frac{1}{2}$, $b_{kt}' = \frac{1}{2}(\Sigma_{t-1}' + \mu_{t-1}'\mu_{t-1}'^T)_{[i,i]} + b_0$

   - update $q(\lambda)$ by setting:

     $e_t' = e_0 + \frac{N}{2}$, $f_t' = \frac{1}{2}\sum_{i=1}^N[(y_i - \mu_{t-1}'^T x_i)^2 + x_i^T\Sigma_{t-1}' x_i] + f_0$

   - update $q(w)$ by setting

     $\Sigma_t' = (\text{diag}(\frac{a_{1t}'}{b_{1t}'}, \dots, \frac{a_{dt}'}{b_{dt}'}) + \frac{e_t'}{f_t'}\sum_{i=1}^N x_i x_i^T)^{-1}$. $\mu_t' = \Sigma_t'(\frac{e_t'}{f_t'}\sum_{i=1}^N y_i x_i)$

   - Evaluate $\mathcal{L}(a_t', b_t', e_t', f_t', \Sigma_t', \mu_t')$ to assess convergence.

(c). $\mathcal{L} = \int q(w,\lambda,\alpha) \ln \frac{p(Y,w,\alpha,\lambda|X)}{q(w,\lambda,\alpha)} dw d\lambda d\alpha = \int q(w) q(\lambda) q(\alpha) \ln \frac{p(Y,w,\alpha,\lambda|X)}{q(w)q(\lambda)q(\alpha)} dw d\lambda d\alpha$

$= \int q(w,\lambda,\alpha) \ln p(Y,w,\alpha,\lambda|X) dw d\lambda d\alpha - \int q(w) \ln q(w) dw - \int q(\lambda) \ln q(\lambda) d\lambda - \int q(\alpha) \ln q(\alpha) d\alpha.$

$= E_q[\ln P(Y,w,\alpha,\lambda|X)] - E_q[\ln q(w)] - E_q[\ln q(\lambda)] - E_q[\ln q(\alpha)]$

$= E_q[\sum_{i=1}^{N} \ln p(y_i|w,\lambda,x_i)] + E_q[\ln p(\lambda|e_0,f_0)] + E_q[\ln p(w|\alpha)] + E_q[\sum_{k=1}^{d} \ln p(\alpha_k|a_0,b_0)]$
$- E_q[\ln q(w|\mu',\Sigma')] - E_q[\ln q(\lambda|e',f')] - E_q[\sum_{k=1}^{d} \ln q(\alpha_k|a_k',b_k')]$

$= (E_q[\ln p(\lambda|e_0,f_0)] - E_q[\ln q(\lambda|e',f')]) + (E_q[\ln p(w|\alpha)] - E_q[\ln q(w|\mu',\Sigma')])$
$+ (E_q[\sum_{i=1}^{d} p(\alpha_k|a_0,b_0)] - E_q[\sum_{k=1}^{d} \ln q(\alpha_k|a_k',b_k')]) + E_q[\sum_{i=1}^{N} \ln p(y_i|w,\lambda,x_i)]$

① $E_q[\ln p(\lambda|e_0,f_0)] - E_q[\ln p(\lambda|e',f')] = E_q[(e_0 \ln f_0 - \ln\Gamma(e_0) + (e_0-1)\ln\lambda - f_0\lambda) - (e'\ln f' - \ln\Gamma(e') + (e'-1)\ln\lambda - f'\lambda)]$

$= e_0 \ln f_0 - \ln\Gamma(e_0) + (e_0-1) E_q[\ln\lambda] - f_0 E_q[\lambda] - e'\ln f' + \ln\Gamma(e') - (e'-1) E_q[\ln\lambda] + f' E_q[\lambda]$

$= e_0 \ln f_0 - \ln\Gamma(e_0) + (e_0-1)(\psi(e') - \ln f') - f_0 \frac{e'}{f'} - e'\ln f' + \ln\Gamma(e') - (e'-1)(\psi(e') - \ln f') + f' \frac{e'}{f'}$ where ψ is the Digamma function

$= e_0 \ln f_0 - \ln\Gamma(e_0) + (e_0 - e')(\psi(e') - \ln f') - (f_0 - f')\frac{e'}{f'} - e'\ln f' + \ln\Gamma(e')$

② $E_q[\ln p(w|\alpha)] - E_q[\ln q(w|\mu',\Sigma')] = \frac{1}{2}\sum_{k=1}^{d} E_q[\ln\alpha_k] - \frac{d}{2}\ln 2\pi - \frac{1}{2}(E_q[w^T \mathrm{diag}(\alpha_1,\dots,\alpha_d) w]) + \frac{1}{2}\ln|\Sigma'| + \frac{d}{2}\ln 2\pi + \frac{1}{2}E_q[(w-\mu')^T \Sigma'^{-1}(w-\mu')]$

$= \frac{1}{2}\sum_{k=1}^{d} E_q[\ln\alpha_k] - \frac{1}{2}\sum_{k=1}^{d} E_q(w_k^2) E_q(\alpha_k) + \frac{1}{2}\ln|\Sigma'| + \frac{1}{2}E_q[w^T\Sigma'^{-1}w - 2w\Sigma'^{-1}\mu' + \mu'^T\Sigma'^{-1}\mu']$

$= \frac{1}{2}\sum_{k=1}^{d}(\psi(a_k') - \ln b_k') - \frac{1}{2}\sum_{k=1}^{d}[\Sigma' + \mu'\mu'^T]_{(k,k)} \frac{a_k'}{b_k'} + \frac{1}{2}\ln|\Sigma'| + \frac{1}{2}(E_q(w^T\Sigma'^{-1}w) - \mu'^T\Sigma'^{-1}\mu')$

Since $E_q(w^T\Sigma'^{-1}w) - \mu'^T\Sigma'^{-1}\mu' = E_q[\mathrm{Tr}(\Sigma'^{-1}ww^T)] - \mathrm{Tr}(\Sigma'^{-1}\mu'\mu'^T) = \sum_{k=1}^{d}[\Sigma'^{-1}_{(k,k)}(E_q(w_k^2) - (\mu'\mu'^T)_{(k,k)})]$

$= \sum_{k=1}^{d}[\Sigma'^{-1}_{(k,k)}(\Sigma'_{(k,k)} + (\mu'\mu'^T)_{(k,k)} - (\mu'\mu'^T)_{(k,k)})] = \sum_{k=1}^{d}[\Sigma'^{-1}_{(k,k)}\Sigma'_{(k,k)}] = d$

$\therefore E_q[\ln p(w|\alpha)] - E_q[\ln q(w|\mu',\Sigma')] = \frac{1}{2}\sum_{k=1}^{d}(\psi(a_k') - \ln b_k') - \frac{1}{2}\sum_{k=1}^{d}[\Sigma' + \mu'\mu'^T]_{(k,k)}\frac{a_k'}{b_k'} + \frac{1}{2}\ln|\Sigma'| + \frac{d}{2}$

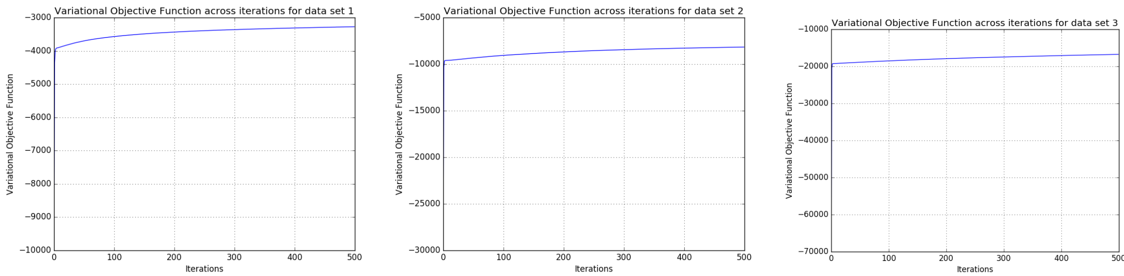
③ $$E_q\left[\sum_{k=1}^{d}\ln p(\alpha_k|a_0,b_0)\right]-E_q\left[\sum_{k=1}^{d}\ln q(\alpha_k|a_k',b_k')\right]=da_0\ln b_0-d\ln\Gamma(a_0)+(a_0-1)\sum_{k=1}^{d}E_q(\ln\alpha_k)-b_0\sum_{k=1}^{d}E_q(\alpha_k)$$
$$-\sum_{k=1}^{d}\left(a_k'\ln b_k'-\ln\Gamma(a_k')+(a_k'-1)E_q[\ln\alpha_k]-b_k'E_q[\alpha_k]\right)=d(a_0\ln b_0-\ln\Gamma(a_0))-\left(\sum_{k=1}^{d}a_k'\ln b_k'-\sum_{k=1}^{d}\ln\Gamma(a_k')\right)$$
$$+\sum_{k=1}^{d}(a_0-a_k')(\psi(a_k')-\ln b_k')-b_0\sum_{k=1}^{d}\frac{a_k'}{b_k'}+\sum_{k=1}^{d}a_k'$$

④ $$E_q\left[\sum_{k=1}^{N}\ln p(y_i|x_i,w,\lambda)\right]=\frac{N}{2}E_q[\ln\lambda]-\frac{n}{2}\ln 2\pi-\frac{E_q[\lambda]}{2}\sum_{i=1}^{N}(y_i^2-2E_q(w^T)x_iy_i+x_i^TE[ww^T]x_i)$$
$$=\frac{N}{2}(\psi(e')-\ln f')-\frac{N}{2}\ln 2\pi-\frac{f'}{2e'}\sum_{i=1}^{N}(y_i^2-2\mu'^Tx_iy_i+x_i^T[\Sigma'+\mu'\mu'^T]x_i)$$

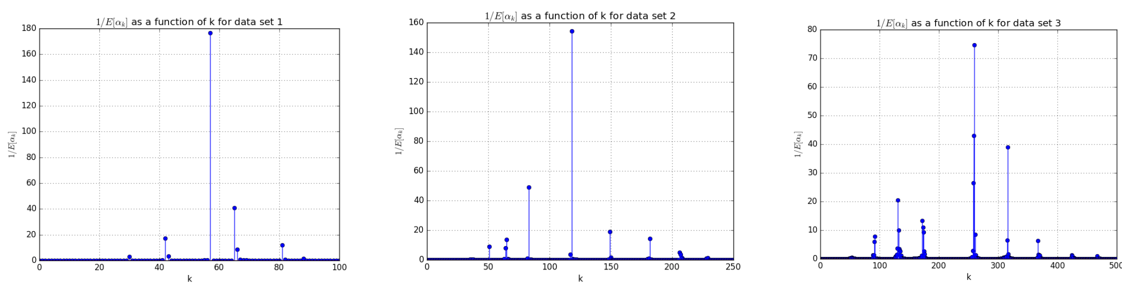
Thus we evaluate $\mathcal{L}$ by ①+②+③+④

Problem 2

a)



b)



c)

$$\frac{1}{E_q[\lambda]} = 1.08002996 \qquad \frac{1}{E_q[\lambda]} = 0.89946298 \qquad \frac{1}{E_q[\lambda]} = 0.97814359$$

d)