Bayesian Models for Machine Learning

Haoyang Chen hc2812

Question 1.

a). $p(\pi|x_1,\cdots,x_n) = p(\pi|x_1,\cdots,x_n,r) \propto \prod_{i=1}^{n} p(x_i|\pi,r)\cdot p(\pi) = \prod_{i=1}^{n}\binom{x_i+r-1}{x_i}\pi^{x_i}(1-\pi)^r \cdot \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)}\pi^{a-1}(1-\pi)^{b-1}$

$\propto \pi^{\sum_{i=1}^{n}x_i+a-1}(1-\pi)^{nr+b-1}$

$\therefore \pi|x_1,\cdots,x_n \sim Beta(\sum_{i=1}^{n}x_i+a,\ nr+b)$

$p(\pi|x_1,\cdots,x_n) = \frac{\Gamma(\sum_{i=1}^{n}x_i+a+nr+b)}{\Gamma(\sum_{i=1}^{n}x_i+a)\Gamma(nr+b)}\pi^{\sum_{i=1}^{n}x_i+a-1}(1-\pi)^{nr+b-1}$ #

b). $p(x_{n+1}|x_1,\cdots,x_n) = \int p(x_{n+1}|\pi)\,p(\pi|x_1,\cdots,x_n)\,d\pi = \int p(x_{n+1}|\pi,r)\,p(\pi|x_1,\cdots,x_n,r)\,d\pi$

$= \int \binom{x_{n+1}+r-1}{x_{n+1}}\pi^{x_{n+1}}(1-\pi)^r \frac{\Gamma(\sum_{i=1}^{n}x_i+a+nr+b)}{\Gamma(\sum_{i=1}^{n}x_i+a)\Gamma(nr+b)}\pi^{\sum_{i=1}^{n}x_i+a-1}(1-\pi)^{nr+b-1}\,d\pi.$

$= \binom{x_{n+1}+r-1}{x_{n+1}}\frac{\Gamma(\sum_{i=1}^{n}x_i+a+nr+b)}{\Gamma(\sum_{i=1}^{n}x_i+a)\Gamma(nr+b)}\int \pi^{\sum_{i=1}^{n+1}x_i+a-1}(1-\pi)^{(n+1)r+b-1}\,d\pi$

$= \binom{x_{n+1}+r-1}{x_{n+1}}\frac{\Gamma(\sum_{i=1}^{n}x_i+a+nr+b)}{\Gamma(\sum_{i=1}^{n}x_i+a)\Gamma(nr+b)}B(\sum_{i=1}^{n+1}x_i+a,\ (n+1)r+b)$

$= \binom{x_{n+1}+r-1}{x_{n+1}}\frac{\Gamma(\sum_{i=1}^{n}x_i+a+nr+b)}{\Gamma(\sum_{i=1}^{n}x_i+a)\Gamma(nr+b)}\cdot\frac{\Gamma(\sum_{i=1}^{n+1}x_i+a)\Gamma((n+1)r+b)}{\Gamma(\sum_{i=1}^{n+1}x_i+a+(n+1)r+b)}$ #

Question 2.

Denote $X=\{x_1,\cdots,x_N\}$, $Y=\{y_1,\cdots,y_N\}$

$$\ln p(Y,\lambda|X)=\underbrace{\int q(w)\ln\frac{p(Y,\lambda,w|X)}{q(w)}dw}_{L(\lambda)}+\underbrace{\int q(w)\ln\frac{q(w)}{p(w|Y,\lambda,X)}dw}_{KL(q\|p)}$$

(a) $\boxed{\text{E - Step}}$

$$p(w|Y,\lambda,X)\propto p(Y|w,\lambda,X)\cdot p(w)=\prod_{i=1}^{N}p(y_i|w,\lambda,x_i)\cdot p(w)=\prod_{i=1}^{N}N(x_i^Tw,\alpha^{-1})\cdot N(0,\lambda^{-1}I)$$

$$\propto\prod_{i=1}^{N}\exp\left(-\frac{\alpha}{2}(y_i-x_i^Tw)^2\right)\cdot\exp\left(-\frac{\lambda}{2}w^Tw\right)$$

$$=\exp\left\{-\frac{\alpha}{2}\sum_{i=1}^{N}(y_i-x_i^Tw)-\frac{\lambda}{2}w^Tw\right\}$$

Thus $p(w|Y,\lambda,X)=N(w|\mu,\Sigma)$, where $\Sigma=(\lambda I+\alpha\sum_{i=1}^{N}x_ix_i^T)^{-1}$, $\mu=\Sigma\cdot(\alpha\sum_{i=1}^{N}y_ix_i)$

$E_{q_t(w)}[w]=\mu$, $E_{q_t(w)}[ww^T]=\mu\mu^T+\Sigma$

$$L_t(\lambda)=\int q_t(w)\ln p(Y,\lambda,w|X)dw-\int q_t(w)\ln q_t(w)dw$$

$=E_{q_t(w)}[\ln p(Y,\lambda,w|X)]+\text{constant}$ Since $\int q_t(w)\ln q_t(w)dw$ does not contain λ

$=E_{q_t(w)}[\ln p(Y|w,\lambda,X)+\ln p(w|\lambda)+\ln p(\lambda)]+\text{constant}$

$=\ln p(\lambda)+E_{q_t(w)}[\ln p(w|\lambda)]+\text{constant}$ Since $\ln p(Y|w,\lambda,X)$ does not contain λ

$=(a-1)\ln\lambda-b\lambda+E_{q_t(w)}\left[-\frac{1}{2}\ln|\lambda^{-1}I|-\frac{\lambda}{2}w^Tw\right]+\text{constant}$

$=(a-1)\ln\lambda-b\lambda+\frac{d}{2}\ln\lambda-\frac{\lambda}{2}E_{q_t(w)}[tr(ww^T)]+\text{constant}$

$=(a+\frac{d}{2}-1)\ln\lambda-b\lambda-\frac{\lambda}{2}\cdot tr\left(E_{q_t(w)}(ww^T)\right)+\text{constant}$

$=(a+\frac{d}{2}-1)\ln\lambda-b\lambda-\frac{\lambda}{2}tr(\mu\mu^T+\Sigma)+\text{constant}$

(b) $\boxed{\text{M-Step}}$

$$\lambda_t = \arg\max_{\lambda} L_t(\lambda)$$

$$\frac{\partial L_t(\lambda)}{\partial \lambda} = (a+\frac{d}{2}-1)\frac{1}{\lambda} - b + \frac{1}{2}tr(\mu\mu^T+\Sigma) = 0$$

$$\lambda = \frac{a+\frac{d}{2}-1}{b+\frac{1}{2}tr(\mu\mu^T+\Sigma)}$$

(c) $\boxed{\text{The marginal objective}}$

$$\ln p(Y, \lambda_{t-1}|X) = L_t(\lambda_{t-1}) = E_{q_t(W)}[\ln p(Y, \lambda_{t-1}, W|X) - \ln q_t(W)]$$

$$= E_{q_t(W)}[\ln p(Y|W, \lambda_{t-1}, X) + \ln p(W|\lambda_{t-1}) + \ln p(\lambda_{t-1}) - \ln q_t(W)]$$

$$E_{q_t(W)}[\ln p(Y|W, \lambda_{t-1}, X)] = \frac{dN}{2}\ln(\frac{\alpha}{2\pi}) - \frac{\alpha}{2}\sum_{i=1}^{N}(y_i)^2 - \frac{\alpha}{2}\sum_{i=1}^{N}x_i^T(\mu\mu^T+\Sigma)x_i + \alpha\sum_{i=1}^{N}y_i x_i^T\mu$$

$$E_{q_t(W)}[\ln p(W|\lambda_{t-1})] = \frac{d}{2}\ln\lambda_{t-1} - \frac{\lambda_{t-1}}{2}tr(\mu\mu^T+\Sigma)$$

$$E_{q_t(W)}[\ln p(\lambda_{t-1})] = (a-1)\ln\lambda_{t-1} - b\lambda_{t-1}$$

$$E_{q_t(W)}[-\ln q_t(W)] = \frac{1}{2}\ln\det(2\pi e\Sigma)$$

$\boxed{\text{Pseudo-code}}$

1. Initialized λ_0 to zero
2. For $t=1,\cdots,T$ do:
   (a) E-Step. Calculate $\Sigma = (\lambda_{t-1}I + \alpha\sum_{i=1}^{N}x_i x_i^T)^{-1}$, $\mu = \Sigma(\alpha\sum_{i=1}^{N}y_i x_i)$
   (b) M-Step. Calculate $\lambda_t = \dfrac{a+\frac{d}{2}-1}{b+\frac{1}{2}tr(\mu\mu^T+\Sigma)}$
   (c) Calculate marginal objective $\ln p(Y, \lambda_{t-1}|X) = L_t(\lambda_{t-1}) = E_{q_t(W)}[\ln p(Y|W, \lambda_{t-1}, X) + \ln p(W|\lambda_{t-1}) + \ln p(\lambda_{t-1}) - \ln q_t(W)]$
   if $\ln p(Y, \lambda_{t-1}|X) < \ln p(Y, \lambda_{t-2}|X)$, terminate

Question 3.

Denote $X=\{x_1,\cdots,x_N\}$, $Y=\{y_1,\cdots,y_N\}$

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

$$= \prod_{i=1}^{N}(2\pi\alpha^{-1})^{-\frac{1}{2}}\exp\{-\tfrac{\alpha}{2}(y_i-x_i^Tw)^2\}\cdot\lambda^{\frac{d}{2}}\exp\{-\tfrac{\lambda}{2}w^Tw\}\cdot\lambda^{e-1}e^{-\lambda f}\;\alpha^{a-1}e^{-\alpha b}$$

$$\propto \alpha^{\frac{N}{2}+a-1}\cdot\exp\{-\alpha[\tfrac{1}{2}\sum_{i=1}^{N}(y_i-x_i^Tw)^2+b]\}\;\lambda^{\frac{d}{2}+e-1}\cdot\exp\{-\lambda(\tfrac{w^Tw}{2}+f)\}$$

• $q(\lambda)\propto\exp\{E_{q(\alpha,w)}[\ln p(Y,W,\alpha,\lambda|X)]\}\propto\exp\{E_{q(\alpha,w)}[(\tfrac{N}{2}+a-1)\ln\alpha-\alpha(\tfrac{1}{2}\sum_{i=1}^{N}(y_i-x_i^Tw)^2+b)+(\tfrac{d}{2}+e-1)\ln\lambda-\lambda(\tfrac{w^Tw}{2}+f)]\}$

$\propto \lambda^{(\frac{d}{2}+e-1)}\exp\{-\lambda(E_{q(w)}[\tfrac{w^Tw}{2}]+f)\}$ Since other terms do not contain λ.

$q(\lambda)=\mathrm{Gamma}(e',f')$, $e'=\frac{d}{2}+e$, $f'=E_{q(w)}[\frac{w^Tw}{2}]+f$.

• $q(\alpha)\propto\exp\{E_{q(\lambda,w)}[\ln p(Y,W,\alpha,\lambda|X)]\}\propto\exp\{E_{q(\lambda,w)}[(\tfrac{N}{2}+a-1)\ln\alpha-\alpha(\tfrac{1}{2}\sum_{i=1}^{N}(y_i-x_i^Tw)^2+b)+(\tfrac{d}{2}+e-1)\ln\lambda-\lambda(\tfrac{w^Tw}{2}+f)]\}$

$\propto\alpha^{\frac{N}{2}+a-1}\exp\{-\alpha(\tfrac{1}{2}E_{q(w)}[\sum_{i=1}^{N}(y_i-x_i^Tw)^2]+b)\}$ Since other terms do not contain α

$q(\alpha)=\mathrm{Gamma}(a',b')$, $a'=\frac{N}{2}+a$, $b'=\frac{1}{2}E_{q(w)}[\sum_{i=1}^{N}(y_i-x_i^Tw)^2]+b=\frac{1}{2}\sum_{i=1}^{N}E_{q(w)}[y_i^2+x_i^Tww^Tx_i-2y_ix_i^Tw]+b$

• $q(w)\propto\exp\{E_{q(\alpha,\lambda)}[\ln p(Y,W,\alpha,\lambda|X)]\}\propto\exp\{E_{q(\alpha,\lambda)}[(\tfrac{N}{2}+a-1)\ln\alpha-\alpha(\tfrac{1}{2}\sum_{i=1}^{N}(y_i-x_i^Tw)^2+b)+(\tfrac{d}{2}+e-1)\ln\lambda-\lambda(\tfrac{w^Tw}{2}+f)]\}$

$\propto\exp\{-\frac{E_{q(\alpha)}[\alpha]}{2}\sum_{i=1}^{N}(y_i-x_i^Tw)^2\}\exp\{-\frac{E_{q(\lambda)}[\lambda]}{2}w^Tw\}$ Since other terms do not contain w

$q(w)=\mathrm{Normal}(\mu',\Sigma')$, $\Sigma'=(E_{q(\lambda)}[\lambda]\cdot I+E_{q(\alpha)}[\alpha]\cdot\sum_{i=1}^{N}x_ix_i^T)^{-1}$, $\mu'=\Sigma'(E_{q(\alpha)}[\alpha]\sum_{i=1}^{N}y_ix_i)$

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

Thus, $e'=\frac{d}{2}+e$, $f'=\frac{1}{2}\mathrm{tr}(\mu'\mu'^T+\Sigma')+f$, $a'=\frac{N}{2}+a$, $b'=\frac{1}{2}\sum_{i=1}^{N}(y_i^2+x_i^T(\mu'\mu'^T+\Sigma')x_i-\sum_{i=1}^{N}y_ix_i^T\mu')$

$\Sigma'=(\frac{e'}{f'}\cdot I+\frac{a'}{b'}\sum_{i=1}^{N}x_ix_i^T)^{-1}$, $\mu'=\Sigma'(\frac{a'}{b'}\sum_{i=1}^{N}y_ix_i)$

(2) Evaluate $L(a'_t, b'_t, e'_t, f'_t, \mu'_t, \Sigma'_t)$

$$L(a'_t, b'_t, e'_t, f'_t, \mu'_t, \Sigma'_t) = E_q[\ln p(Y, a'_t, b'_t, e'_t, f'_t, \mu'_t, \Sigma'_t \mid X)] - E_{q(a'_t)}[\ln(a'_t)] - E_{q(b'_t)}[\ln(b'_t)]$$
$$- E_{q(e'_t)}[\ln(e'_t)] - E_{q(f'_t)}[\ln(f'_t)] - E_{q(\mu'_t)}[\ln(\mu'_t)] - E_{q(\Sigma'_t)}[\ln(\Sigma'_t)]$$

$\boxed{\text{Pseudo-code:}}$

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

2. For iteration $t = 1, \cdots, T$:
   - Update $q(\alpha)$ by setting:

$$a'_t = \frac{N}{2} + a, \quad b'_t = \frac{1}{2}\sum_{i=1}^{N}\left(y_i^2 + x_i^T(\mu'_{t-1}\mu'^T_{t-1} + \Sigma'_{t-1})x_i - \sum_{i=1}^{N} y_i x_i^T \mu'_{t-1}\right)$$

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

$$e'_t = \frac{d}{2} + e, \quad f'_t = \frac{1}{2}\mathrm{tr}(\mu'_{t-1}\mu'^T_{t-1} + \Sigma'_{t-1}) + f$$

   - update $q(w)$ by setting:

$$\Sigma'_t = \left(\frac{e'_t}{f'_t}\cdot I + \frac{a'_t}{b'_t}\sum_{i=1}^{N} x_i x_i^T\right)^{-1}, \quad \mu'_t = \Sigma'_t\left(\frac{a'_t}{b'_t}\sum_{i=1}^{N} y_i x_i\right)$$

   - Evaluate $L(a'_t, b'_t, e'_t, f'_t, \mu'_t, \Sigma'_t)$ to assess convergence.