EE6720 Midterm.

Jingying Zhou JZ2673.

Question 1.

a) Posterior of $\pi$:

$$P(\pi | x_1 \cdots x_n) \propto P(x_1 \cdots x_n | \pi) P(\pi) = \prod_{i=1}^{N} \binom{x_i+r-1}{x_i} \pi^{x_i} (1-\pi)^r \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} \pi^{a-1} (1-\pi)^{b-1}$$

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

which is in the form of a Beta $(a+\sum_{i=1}^{N} x_i, b+nr)$

b) Posterior predictive for $x_{n+1}$

$$P(x_{n+1} | x_1 \cdots x_n) = \int P(x_{n+1}|\pi) P(\pi | x_1 \cdots x_n) d\pi$$

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

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

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

which follows a beta negative binomial distribution.

EE6720
Midterm.
Jingyiy Zhou
JZ2673.

Question 2

Pseudo Code:

Algorithm: 1: Initialize. $\lambda_0$.

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

1) E-step: ~~Update~~ Calculate. vector $q_t(w)=P(w|y,\lambda_{t-1},x)=\prod_{i=1}^{N}P(y_i|w,x_i,\lambda_{t-1})P(w)$

2) M-step: Update $\lambda_t=\arg\max_{\lambda} E_t[\ln(P(y,\lambda,w|x)-\ln q_t(w)]$

3). Calculate $\ln(y,\lambda|x)=\mathcal{L}_t(\lambda_{t-1})$

1) Estep $P(w|y,x,\lambda)\propto P(w)P(y|w,x,\lambda)$

$$\propto \text{Normal}(0,\lambda^{-1}I) * \prod_{i=1}^{N}\text{Normal}(x_i^Tw,\alpha^{-1})$$
$$\propto \exp\{-\tfrac{\lambda}{2}(w^Tw)\}\cdot\exp\{\sum_{i=1}^{N}(y_i-x_i^Tw)^2(\tfrac{-\alpha}{2})\}.$$

Which is in the form of Normal $(\mu,\Sigma)$ with.

$$\mu=\Sigma(\alpha\sum_{i=1}^{N}y_ix_i)$$
$$\Sigma=(\lambda I+\alpha\sum_{i=1}^{N}x_ix_i^T)^{-1}$$

$$\mathcal{L}_t(\lambda)=E_{q_t}[\ln P(y,w,\lambda|x)-\ln q_t(w)]$$
$$=\sum_{i=1}^{N}E_q[\ln(P(y_i|w,\lambda,x)]+E_q[\ln P(w|\lambda)]+E_q[\ln P(\lambda)]+\text{constant}$$
$$=\sum_{i=1}^{N}E_q[(y_i-x_i^Tw)^2]+\text{constant}+E_q[-\tfrac{1}{2}\ln(\tfrac{I}{\lambda})-\tfrac{\lambda}{2}w^Tw]+E_q[(a-1)\ln\lambda-b\lambda]+\text{constant}$$
$$=\frac{d}{2}\ln(\lambda)-\frac{\lambda}{2}E_q[w^Tw]+(a-1)\ln(\lambda)-b(\lambda)+\text{constant}$$

$E_q(w^Tw)=tr(\Sigma)+\mu^T\mu$; then.

$$\mathcal{L}_t(\lambda)=\ln(\lambda)(a+\tfrac{d}{2}-1)-b\cdot\lambda-\tfrac{\lambda}{2}(tr(\Sigma)+\mu^T\mu)+\text{constant}$$

2) M-step: take partial derivative w.r.t. $\lambda$ and set to 0:

$$\frac{\partial\mathcal{L}_t(\lambda)}{\partial\lambda}=0 \Rightarrow (a+\tfrac{d}{2}-1)\frac{1}{\lambda}-b-\tfrac{1}{2}(tr(\Sigma)+\mu^T\mu)+0=0$$
$$\Rightarrow \lambda_t=\left[\frac{a+\frac{d}{2}-1}{b+\frac{1}{2}(tr(\Sigma)+\mu^T\mu)}\right]$$

EE6720 Midterm.

Jingying Zhou . UNI: JZ2673

Question 3 | Pseudo-code:

Inputs: Data $D=\{(x_i, y_i)\}_{i=1}^N$ and, all question definitions.

Outputs: Values for $(\alpha'. b')$ $(e', f')$ $(\mu', \Sigma')$

Algorithm: 1. Initialize $a_0', b_0'$ $e_0'$ $f_0'$ $\mu_0', \Sigma_0'$

2. for iteration $t=1 \cdots T$

1). Update $q(\alpha)$ by setting $a_t' = a + \frac{N}{2}$

$$b_t' = b + \frac{1}{2}\sum_{i=1}^{N}(y_i - x_i^T \mu_{t-1}')^2 + x_i^T \Sigma_{t-1}' x_i$$

2) Update $q(\lambda)$ by setting $e_t' = e + \frac{d}{2}$

$$f_t' = f + \frac{1}{2}[tr(\Sigma') + \mu'^T \mu]$$

3) Update $q(w)$ by setting $\mu_t' = \Sigma'\left(\frac{a_t'}{b_t'}\sum_{i=1}^{N} x_i^T x_i\right)$

$$\Sigma_t' = \left(\frac{e_t'}{f_t'} I + \frac{a_t}{b_t} x_i^T x_i\right)^{-1}$$

4) Evaluate $\mathcal{L}(a_t', b_t', e_t', f_t', \mu_t', \Sigma_t')$ to assess convergence. If marginal increase is small, we terminate the process

1) Update $\alpha$:

$$q(\alpha) \propto \exp\left\{E_{q(w,\lambda)}[\ln P(y|x,\alpha,w,\lambda) + \ln P(\alpha) + \ln P(w|\lambda) + \ln(\lambda)]\right\}$$

$$\propto \exp\left\{E_{q(w,\lambda)}[\ln P(y|x,\alpha,w,\lambda) + \ln(\alpha)]\right\}$$ , for the rest terms do not involve $\alpha$.

$$\propto P(\alpha)\exp\left\{\sum_{i=1}^{N} E_{q(w,\lambda)}[\ln P(y_i|x_i,\alpha,w,\lambda)]\right\}.$$

$$\propto \alpha^{a-1} e^{-b\alpha} \cdot \left[\prod_{i=1}^{N} a^{\frac{1}{2}} e^{-\frac{\alpha}{2} E_{q(w,\lambda)}[(y_i - x_i^T w)^2]}\right]$$

$$= \alpha^{a+\frac{N}{2}-1} \cdot e^{-\alpha(b+\frac{1}{2}\sum_{i=1}^{N} E_{q(w,\lambda)}[(y_i - x_i^T w)^2])}$$ follows the form of a Gamma distributi

should follow Gamma $(a', b')$, according to the question setting;

We then have:

$$\begin{cases} a' = a + \frac{N}{2} \\ b' = b + \frac{1}{2}\sum_{i=1}^{N} E_{q(w,\lambda)}[(y_i - x_i^T w)^2]. \end{cases}$$

We cannot take expectation yet.

EE6720 Midterm:

Jingying Zhou. UNI: JZ2673

2) Update $\lambda$

$q(\lambda) \propto \exp\{E_{q(\alpha,w)}[\ln P(y|x.\alpha.w.\lambda) + \ln P(\alpha) + \ln P(w|\lambda) + \ln P(\lambda)]\}$

$\propto \exp\{E_{q(\alpha,w)}[\ln P(w|\lambda) + \ln P(\lambda)]\}$ ; since $\ln P(\alpha)$ does not involve $\lambda$; and that $E_{q(\alpha,w)}[\ln P(y|x\alpha w\lambda)]$ does not either.

$\propto P(\lambda) \exp\{E_{q(\alpha,w)}[\ln P(w|\lambda)]\}$

$\propto \lambda^{e-1} e^{-\lambda f} \cdot \lambda^{\frac{d}{2}} \cdot \exp\{-\frac{\lambda}{2} E_{q(\alpha,w)}[w^T w]\}$

$= \lambda^{e+\frac{d}{2}-1} \cdot e^{-\lambda\{f + \frac{1}{2} E_{q(\alpha,w)}[w^T w]\}}$. which follows a Gamma distribution form.

As suggested by the question $q(\lambda) \sim Gamma(e', f')$; then we have.

$e' = e + \frac{d}{2}$

$f' = f + \frac{1}{2} E_{q(\alpha,w)}[w^T w]$

3) Update $w$

$q(w) \propto \exp\{E_{q(\alpha,\lambda)}[\ln P(y|x,\alpha,w,\lambda) + \ln P(w|\lambda) + \ln P(\alpha) + \ln(\alpha)]\}$

$\propto \exp\{E_{q(\alpha,\lambda)}[\ln P(y|x.\alpha.w.\lambda) + \ln P(w|\lambda)]\}$. for the rest two terms do not contain $w$.

$\propto \exp\{\sum_{i=1}^{N} E_{q(\alpha,\lambda)}[\ln P(y_i|x_i,\alpha,w,\lambda)]\} \cdot \exp\{E_{q(\alpha,\lambda)}[\ln(P(w|\lambda))]\}$.

$\propto \exp\{-\frac{E_{q(\alpha,\lambda)}[\alpha]}{2} \sum_{i=1}^{N}(y_i - x_i^T w)^2\} \cdot \exp\{-\frac{1}{2} E_{q(\alpha,\lambda)}[\lambda](w^T w)\}$.

Complete the squares we have. $q(w)$ follows a normal distribution with:

~~$q(w) \propto \exp\{-\frac{1}{2}[E_{q(\alpha,\lambda)}[\alpha]$~~

$\Sigma' = (E_{q(\alpha,\lambda)}[\lambda] I + E_{q(\alpha)}[\alpha] \sum_{i=1}^{N} x_i x_i^T)^{-1}$.

$\mu' = \Sigma' (E_{q(\alpha,\lambda)}[\alpha] \sum_{i=1}^{N} y_i x_i)$

Where the expectations are:

$$E_{q(\alpha,\lambda)}[\lambda] = \frac{e'}{f'}$$

$$E_{q(\alpha,\lambda)}[\alpha] = \frac{\alpha'}{b'}$$

$$E_{q(\alpha,w)}[w^T w] = tr(\Sigma') + \mu'^T \mu'$$

$$E_{q(w,\lambda)}[(y_i - x_i^T w)^2] = \sum_{i=1}^{N} (y_i - x_i^T \mu')^2 + x_i^T \Sigma' x_i$$

4). $$\mathcal{L}(a_t', b_t', e_t', f_t', \mu_t', \Sigma_t') = E_q[\ln p(y, a_t', b_t', e_t', f_t', \mu_t', \Sigma_t' \mid X)] - E_{q(a_t')}[\ln(a_t')]$$
$$- E_{q(b_t')}[\ln(b_t')] - E_{q(e_t')}[\ln(e_t')] - E_{q(f_t')}[\ln(f_t')]$$
$$- E_{q(\mu_t')}[\ln(\mu_t')] - E_{q(\Sigma_t')}[\ln(\Sigma_t')]$$