Midterm

UNI: yb2356 Name: Yang Bai

Problem 1

(a)

$$p(\theta|x_1, x_2, \dots, x_n) \propto p(x_1, \dots, x_n|\theta)p(\theta)$$

$$x\ iid \sim Gamma(a, \theta), \qquad so\ \ p(x_1, \dots, x_n|\theta) = \prod_{i=1}^{N} p(x_i|\theta)$$

$$p(x_1, \dots, x_n|\theta) = \prod_{i=1}^{N} Gamma(a, \theta) = \prod_{i=1}^{N} \frac{\theta^a}{\Gamma(a)} {x_i}^{a-1} e^{-\theta x_i}$$

$$p(x_1, \dots, x_n|\theta) \propto \theta^{aN} e^{-\theta \sum_{i=1}^{N} x_i} \prod_{i=1}^{N} {x_i}^{a-1}$$

$$We\ know\ that\ \ p(\theta) \propto \frac{c^b}{\Gamma(b)} \theta^{b-1} e^{-c\theta}$$

$$so\ \ p(\theta|x_1, x_2, \dots, x_n) \propto \theta^{b-1} e^{-c\theta} \theta^{aN} e^{-\theta \sum_{i=1}^{N} x_i} \prod_{i=1}^{N} {x_i}^{a-1}$$

$$p(\theta|x_1, x_2, \dots, x_n) \propto \theta^{aN+b-1} e^{-\theta(\sum_{i=1}^{N} x_i + c)} \prod_{i=1}^{N} {x_i}^{a-1}$$

We know the values of all x_i , so treat $\prod_{i=1}^{N} {x_i}^{a-1}$ as a constant. Then we find

$$p(\theta|x_1, x_2, \dots, x_n) \propto \theta^{aN+b-1} e^{-\theta(\sum_{i=1}^{N} x_i + c)}$$

which is a gamma distribution. $p(\theta|x_1, x_2, \dots, x_n) \propto Gamma(aN + b, \sum_{i=1}^{N} x_i + c)$

(b)

$$p(x_{n+1}|x_1, x_2, \dots, x_n) = \int p(x_{n+1}|\theta)\, p(\theta|x_1, x_2, \dots, x_n) d\theta$$

$$p(x_{n+1}|x_1, x_2, \dots, x_n) = \int Gamma(x_{n+1}|a, \theta) Gamma(\theta|aN + b, \sum_{i=1}^{N} x_i + c)\, d\theta$$

$$= \int \frac{\theta^a}{\Gamma(a)} {x_{n+1}}^{a-1} e^{-\theta x_{n+1}} \frac{(\sum_{i=1}^{N} x_i + c)^{aN+b}}{\Gamma(aN + b)} \theta^{aN+b-1} e^{-\theta(\sum_{i=1}^{N} x_i + c)} d\theta$$

$$= \frac{(\sum_{i=1}^{N} x_i + c)^{aN+b}}{\Gamma(a)\Gamma(aN + b)} {x_{n+1}}^{a-1} \int e^{-\theta x_{n+1}} \theta^a \theta^{aN+b-1} e^{-\theta(\sum_{i=1}^{N} x_i + c)} d\theta$$

$$p(x_{n+1}|x_1, x_2, \dots, x_n) = \gamma \int \theta^{\alpha-1} e^{-\theta\beta} d\theta$$

$$p(x_{n+1}|x_1, x_2, \dots, x_n) = \gamma \frac{\Gamma(\alpha)}{\beta^\alpha} {x_{n+1}}^{a-1} \int_0^\infty \frac{\beta^\alpha}{\Gamma(\alpha)} \theta^{\alpha-1} e^{-\theta\beta} d\theta$$

It obvious that $\int_0^\infty \frac{\beta^\alpha}{\Gamma(\alpha)} \theta^{\alpha-1} e^{-\theta\beta} d\theta = 1$ *because it's a* integration *of gamma distribution function.*

In the equation, $\alpha = aN + b + a$, $\beta = \sum_{i=1}^{N} x_i + c + x_{n+1}$, $\gamma = \frac{(\sum_{i=1}^{N} x_i + c)^{aN+b}}{\Gamma(a)\Gamma(aN+b)}$

$$p(x_{n+1}|x_1, x_2, \dots, x_n) = \gamma \frac{\Gamma(\alpha)}{\beta^\alpha} x_{n+1} = \gamma \frac{\Gamma(\alpha)}{(\sum_{i=1}^{N} x_i + c + x_{n+1})^\alpha} {x_{n+1}}^{a-1}$$

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

Problem 2

We want to maximize $\ln\ (y, \alpha|x) = \int \ln\ (y, \alpha, \omega|x)\, d\omega$ over α. EM equation in this case is:

$$\ln\ (y, \alpha|x) = \int q(\omega) \ln \frac{p(y, \alpha, \omega|x)}{q(\omega)} d\omega + \int q(\omega) \ln \frac{q(\omega)}{p(\omega|y, \alpha, x)} d\omega$$

$$\mathcal{L}(\alpha) = \int q(\omega) \ln \frac{p(y, \alpha, \omega|x)}{q(\omega)} d\omega$$

$$KL(q||p) = \int q(\omega) \ln \frac{q(\omega)}{p(\omega|y, \alpha, x)} d\omega$$

Algorithm outline:

1. Initialize α_0 to a vector of all zero
2. For iteration t:
   (a) E-step: Calculate the vector $q_t(\omega) = p(\omega|y, \alpha_{t-1}, x) = \prod_{i=1}^{N} p(y_i|\omega, \alpha_{t-1}, x_i)p(\omega)$
   (b) M-step: Update $\alpha_t = \arg\max_{\alpha} \mathcal{L}_t(\alpha) = \arg\max_{\alpha} E_t[\ln\ p(y, \alpha, \omega|x) - \ln\ q_t(\omega)]$
   (c) Calculate $\ln\ (y, \alpha|x) = \mathcal{L}_t(\alpha_{t-1})$

(a) E-step:

Set $q(\omega)$.
First calculate $p(\omega|y, \alpha, x)$

$$p(\omega|y, \alpha, x) \propto p(y|\omega, \alpha, x)p(\omega)$$

$$\propto \prod_{i=1}^{N} p(y_i|\omega, \alpha, x_i)p(\omega)$$

$$\propto \exp\left[-\frac{\alpha}{2}\sum_{i=1}^{N}(y_i - x_i^T\omega)^2\right]\exp\left(-\frac{\lambda}{2}\omega^T\omega\right)$$

$$\propto \exp\left[-\frac{1}{2}(\omega - \mu)^T\Sigma\quad(\omega - \mu)\right]$$

where

$$\Sigma = (\lambda\mathrm{I} + \alpha\sum_{i=1}^{N} x_i {x_i}^T)^{-1} \quad and \quad \mu = \Sigma\cdot(\alpha\sum_{i=1}^{N} y_i x_i)$$

$$p(\omega|y, \alpha, x) = Norm(\omega|\mu, \Sigma)$$

$$\mathrm{E}_{qt}[\omega] = \mu$$

$$\mathrm{E}_{qt}[\omega\omega^T] = \mu\mu^T + \Sigma$$

Set $q_t(\omega) = p(\omega|y, \alpha_{t-1}, x)$ at iteration t. Then calculate the expectation:

$$\mathcal{L}(\alpha) = E_t[\ln\ p(y, \alpha, \omega|x) - \ln\ q_t(\omega)]$$

$$= \mathrm{E}_q[\ln \mathrm{p}(y|\alpha, x, \omega)] + \mathrm{E}_q[\ln\ p(\alpha)] + \mathrm{E}_q[\ln \mathrm{p}(\omega)] - \mathrm{E}_q[\ln \mathrm{q}(\omega)]$$

$$= \mathrm{E}_{qt(\omega)}\left[\ln\left(\alpha^{\frac{N}{2}}\right)\right] + \mathrm{E}_{qt(\omega)}\left[-\frac{\alpha}{2}\sum_{i=1}^{N}(y_i - x_i^T\omega)^2\right] + (a-1)\ln(\alpha) - b\alpha + const\ w.r.t\ \alpha$$

$$= \left(a - 1 + \frac{N}{2}\right)\ln\alpha - b\alpha + \sum_{i=1}^{N} -\frac{\alpha}{2}\{tr[x_i x_i^T(\mu\mu^T + \Sigma)] - 2\mathrm{tr}(\mu y_i x_i)\} + const$$

(b) M-step:

Maximize $\mathcal{L}(\alpha)$. Differentiating $\nabla_\alpha \mathcal{L}(\alpha)$ and setting to zero:

$$\nabla_\alpha \mathcal{L}(\alpha) = 0$$

$$\frac{\left(\frac{N}{2} + a - 1\right)}{\alpha} - b + \sum_{i=1}^{N} -\frac{1}{2}\{tr[x_i x_i^T(\mu\mu^T + \Sigma)] - 2\mathrm{tr}(\mu y_i x_i)\} = 0$$

$$\Rightarrow \alpha = \frac{\frac{N}{2} + a - 1}{b + \sum_{i=1}^{N} \frac{1}{2}\{tr[x_i x_i^T(\mu\mu^T + \Sigma)] - 2\mathrm{tr}(\mu y_i x_i)\}}$$

(c)

The marginal objective is:

$$\ln\ (y, \alpha | x) = \mathcal{L}(\alpha_{t-1}) = E_t[\ln\ \ p(y, \alpha_{t-1}, \omega | x) - \ln\ \ q_t(\omega)]$$

$$= \mathrm{E}_q[\ln \mathrm{p}(y|\alpha_{t-1}, x, \omega)] + \mathrm{E}_q[\ln\ \ p(\alpha_{t-1})] + \mathrm{E}_q[\ln \mathrm{p}(\omega)] - \mathrm{E}_q[\ln \mathrm{q}(\omega)]$$

$$\mathrm{E}_q[\ln \mathrm{p}(y|\alpha_{t-1}, x, \omega)]$$

$$= \frac{dN}{2}\ln\left(\frac{\alpha}{2\pi}\right) - \frac{\alpha_{t-1}}{2}\sum_{i=1}^{N}(y_i)^2 - \frac{\alpha_{t-1}}{2}\sum_{i=1}^{N}\{tr[x_i x_i^T(\mu\mu^T + \Sigma)] - 2\mathrm{tr}(\mu y_i x_i)\}$$

$$\mathrm{E}_q[\ln\ \ p(\alpha_{t-1})] = \ln\left[\frac{b^a}{\Gamma(a)}\alpha_{t-1}{}^{a-1}e^{-\alpha_{t-1}b}\right] = a\ln b - \ln\Gamma(a) + (a-1)\ln(\alpha_{t-1}) - b\alpha_{t-1}$$

$$\mathrm{E}_q[\ln \mathrm{p}(\omega)] = \frac{d}{2}\ln\left(\frac{\lambda}{2\pi}\right) - \frac{\lambda}{2}tr(\mu\mu^T + \Sigma)$$

$$\mathrm{E}_q[-\ln \mathrm{q}(\omega)] = \frac{1}{2}\ln\det(2\pi e\Sigma)$$

$\det(\cdot)$ represents the matrix determinant.

Problem 3

Algorithm outline

Inputs: Data X, Y and $q(\lambda) = Gamma(\lambda|a', b')\ and\ q(\omega) = Normal(\omega|\mu', \Sigma')$
Outputs: Values for a', b', μ', Σ'

1. Initialize $a'_0, b'_0, \mu'_0, \Sigma'_0$
2. For iteration t=1, …, T
   (a) Update $q(\lambda)$ by setting
   $$a'_t = a + \frac{d}{2}, \qquad b'_t = \frac{1}{2}(\mu'_{t-1}\mu'^T_{t-1} + \Sigma'_{t-1}) + b$$
   (b) Update $q(\omega)$ by setting
   $$\Sigma'_t = \left(\frac{a'_t}{b'_t}\mathrm{I} + \alpha\sum_{i=1}^{N} x_i x_i^T\right)^{-1}, \quad \mu'_t = \Sigma'_t \cdot (\alpha \sum_{i=1}^{N} y_i x_i)$$
   (c) Evaluate $\mathcal{L}(a'_t, b'_t, \mu'_t, \Sigma'_t)$ to assess convergence. If the marginal increase in $\mathcal{L}_t$ compared with $\mathcal{L}_{t-1}$ is "small", terminate. Continue to the next iteration otherwise.

(a)

We want to approximate the posterior $p(\omega, \lambda|y, x)$ with $q(\omega, \lambda)$ using variational inference.
In general setup, we learned that
$$q_i(\theta_i|\phi_i) = \frac{1}{Z}\exp\left[\ln p(y, \theta_1, \theta_2 \ldots |x)\right]$$

In this case,

- $q(\lambda)$

$$q(\lambda) \propto \exp\{E_{q(\omega)}[\ln p(y, \omega, \lambda|x)]\}$$

$$p(y, \omega, \lambda|x) \propto \prod_{i=1}^{N} p(y_i|\omega, \lambda, x_i)\, p(\omega|\lambda) p(\lambda)$$
$$\propto \exp\left[-\frac{\alpha}{2}\sum_{i=1}^{N}(y_i - x_i^T\omega)^2\right]\lambda^{\frac{d}{2}}\exp\left(-\frac{\lambda}{2}\omega^T\omega\right)\lambda^{a-1}e^{-\lambda b}$$

$$\propto \exp\left[-\frac{\alpha}{2}\sum_{i=1}^{N}(y_i - x_i^T\omega)^2\right]\lambda^{a+\frac{d}{2}-1}\exp\left[-\lambda(\frac{\omega^T\omega}{2}+b)\right]$$

Notice that $\exp\left[-\frac{\alpha}{2}\sum_{i=1}^{N}(y_i - x_i^T\omega)^2\right]$ does not contain λ.

$$q(\lambda) \propto \lambda^{a+\frac{d}{2}-1}\exp\left\{-\lambda\left(E_{q(\omega)}\left[\frac{\omega^T\omega}{2}\right]+b\right)\right\}$$

$$q(\lambda) = Gamma(\lambda|a',b') \quad a' = a+\frac{d}{2}, \quad b' = \frac{1}{2}E_{q(\omega)}[\omega^T\omega]+b$$

(b)

- $q(\omega)$

$$q(\omega) \propto \exp\{E_{q(\lambda)}[\ln p(y,\omega,\lambda|x)]\}$$

$$p(y,\omega,\lambda|x) \propto \prod_{i=1}^{N} p(y_i|\omega,\lambda,x_i)\, p(\omega|\lambda)p(\lambda)$$

$$\propto \exp\left[-\frac{\alpha}{2}\sum_{i=1}^{N}(y_i - x_i^T\omega)^2\right]\lambda^{\frac{d}{2}}\exp\left(-\frac{\lambda}{2}\omega^T\omega\right)p(\lambda)$$

So

$$q(\omega) \propto \exp\left\{E_{q(\lambda)}\left[ln\ \prod p(y_i|\omega,\lambda,x_i)\right] - E_{q(\lambda)}[ln\, p(\omega|\lambda)\ ]\right\}$$

$$\propto \exp\left[-\frac{\alpha}{2}\sum_{i=1}^{N}(y_i - x_i^T\omega)^2\right]\exp\left\{-\frac{E_{q(\lambda)}[\lambda]}{2}\omega^T\omega\right\}$$

$$q(\omega) = Normal(\omega|\mu',\Sigma')$$

$$\Sigma' = \left(E_{q(\lambda)}[\lambda]\mathrm{I} + \alpha\sum_{i=1}^{N}x_i{x_i}^T\right)^{-1} \quad and \quad \mu' = \Sigma'\cdot(\alpha\sum_{i=1}^{N}y_i x_i)$$

For this problem,

$$E_{q(\lambda)}[\lambda] = a'/b'$$

$$E_{q(\omega)}[\omega^T\omega] = \mu'\mu'^T + \Sigma'$$

(c)

Evaluating $\mathcal{L}(a'_t, b'_t, \mu'_t, \Sigma'_t)$

$$\mathcal{L}_t = E_q\left[\ln p\left(y, a'_t, b'_t, \mu'_t, \Sigma'_t \middle| x\right)\right] - E_{q(a'_t)}[\ln(a'_t)] - E_{q(b'_t)}[\ln(b_t)] - E_{q(\mu'_t)}\left[\ln\left(\mu'_t\right)\right]$$

$$- E_{q(\Sigma'_t)}[\ln(\Sigma'_t)]$$