CMU CS 10-701 HW 5.

Q1.

$P(\theta; m_0, \phi_0)$

$$= \exp\left[\left\langle \underbrace{\begin{pmatrix} \theta \\ -g(\theta) \end{pmatrix}}_{\phi(x)}, \underbrace{\begin{pmatrix} \phi_0 \\ m_0 \end{pmatrix}}_{\theta} \right\rangle - \underbrace{h(m_0, \phi_0)}_{1^T g(\theta)} \right]$$

following the exponential format.

Q2 Assume $x_i \overset{iid}{\sim} p(\cdot|\theta)$

Likelihood:

$$P(X|\theta) = \prod_{i=1}^{n} p(x_i|\theta)$$

$$= \prod_{i=1}^{n} \exp(\langle \phi(x_i), \theta \rangle - 1^T g(\theta))$$

$$= \exp\left(\sum_{i=1}^{n} \langle \phi(x_i), \theta \rangle - n 1^T g(\theta) \right)$$

$$= \exp\left(\left\langle \sum_{i=1}^{n} \phi(x_i), \theta \right\rangle - n 1^T g(\theta) \right)$$

$$= \exp\left[\left\langle \begin{pmatrix} \theta \\ -g(\theta) \end{pmatrix}, \begin{pmatrix} \sum_{i=1}^{n} \phi(x_i) \\ n 1 \end{pmatrix} \right\rangle \right]$$

Since $\int P(\theta; m_0, \phi_0) = 1$

$$\Rightarrow h(m_0, \phi_0) = \log \int \exp\left[\left\langle \begin{pmatrix} \theta \\ -g(\theta) \end{pmatrix}, \begin{pmatrix} \phi_0 \\ m_0 \end{pmatrix} \right\rangle \right] d$$

The posterior:

$$P(\theta|X) \propto P(X|\theta)P(\theta; m_0, \phi_0)$$

$$= \exp\left[\left\langle \begin{pmatrix} \theta \\ -g(\theta) \end{pmatrix}, \begin{pmatrix} \phi_0 + \sum_{i=1}^{n} \phi(x_i) \\ m_0 + n1 \end{pmatrix} \right\rangle - h(m_0, \phi_0) \right]$$

To normalize it

$$P(\theta|X)$$

$$= \exp\left[\left\langle \begin{pmatrix} \theta \\ -g(\theta) \end{pmatrix}, \begin{pmatrix} \phi_0 + \sum_{i=1}^{n} \phi(x_i) \\ m_0 + n1 \end{pmatrix} \right\rangle - h\left(m_0 + n1, \phi_0 + \sum_{i=1}^{n} \phi(x_i)\right) \right]$$

Q3: Let $m_n = m_0 + n1$

$$\phi_n = \phi_0 + \sum_{i=1}^{n} \phi(x_i)$$

$$P(\theta|X) = \exp\left[\left\langle \begin{pmatrix} \theta \\ -g(\theta) \end{pmatrix}, \begin{pmatrix} \phi_n \\ m_n \end{pmatrix} \right\rangle - h(m_n, \phi_n) \right]$$

Q4. $p(x|\mu,\Sigma) = \frac{1}{\sqrt{(2\pi)^d|\Sigma|}} \exp\left[-\frac{1}{2}(x-\mu)^T\Sigma^{-1}(x-\mu)\right]$

$= \exp\left[-\frac{1}{2}x^T\Sigma^{-1}x + x^T\Sigma^{-1}\mu - \frac{1}{2}\log\left((2\pi)^d|\Sigma|\right) - \frac{1}{2}\mu^T\Sigma^{-1}\mu\right]$

where $x^T\Sigma^{-1}x = tr(x^T\Sigma^{-1}x)$

$= tr(xx^T\Sigma^{-1})$

$= \langle xx^T, \Sigma^{-1}\rangle$ Since $tr(A^TB) = \langle A, B\rangle$

$\therefore p(x|\mu,\Sigma)$

$= \exp\left[\left\langle \begin{pmatrix} x \\ xx^T \end{pmatrix}, \begin{pmatrix} \Sigma^{-1}\mu \\ -\frac{1}{2}\Sigma^{-1} \end{pmatrix} \right\rangle - \mathbf{1}^T \begin{pmatrix} \frac{1}{2}\mu^T\Sigma^{-1}\mu \\ \frac{1}{2}\log\left((2\pi)^d|\Sigma|\right) \end{pmatrix}\right]$

So $\phi(x) = \begin{pmatrix} x \\ xx^T \end{pmatrix}$ $\theta = \begin{pmatrix} \Sigma^{-1}\mu \\ -\frac{1}{2}\Sigma^{-1} \end{pmatrix}$

$g(\theta) = \begin{pmatrix} \frac{1}{2}\mu^T\Sigma^{-1}\mu \\ \frac{1}{2}\log\left((2\pi)^d|\Sigma|\right) \end{pmatrix}$

Q5 $P(\mu, \Sigma ; \mu_0, k_0, \Sigma_0, \nu_0)$

$$= \exp\Big[-\frac{k_0}{2}\mu^T\Sigma^{-1}\mu + k_0\,\mu_0^T\Sigma^{-1}\mu - \frac{k_0}{2}\mu_0^T\Sigma^{-1}\mu_0 - \frac{1}{2}\operatorname{tr}(\Sigma_0 \Sigma^{-1})$$

$$+ \frac{d}{2}\log k_0 + \frac{\nu_0}{2}\log|\Sigma_0| - \frac{\nu_0+d+2}{2}\log|\Sigma| - \frac{(\nu_0+1)d}{2}\log 2$$

$$- \frac{d}{2}\log\pi - \log\Gamma_d\left(\frac{\nu_0}{2}\right)\Big]$$

$$= \exp\Big[\Big\langle \overset{\to \phi_0}{\begin{pmatrix} k_0\mu_0 \\ k_0\mu_0\mu_0^T + \Sigma_0 \end{pmatrix}}, \overset{\to \theta}{\begin{pmatrix} \Sigma^{-1}\mu \\ -\frac{1}{2}\Sigma^{-1} \end{pmatrix}} \Big\rangle$$

$$- \Big\langle \overset{\to m_0}{\begin{pmatrix} k_0 \\ \nu_0+d+2 \end{pmatrix}}, \overset{\to g(\theta)}{\begin{pmatrix} \frac{1}{2}\mu^T\Sigma^{-1}\mu \\ \frac{1}{2}\log\left((2\pi)^d|\Sigma|\right) \end{pmatrix}} \Big\rangle$$

$$+ \frac{d}{2}\log k_0 + \frac{\nu_0}{2}\log|\Sigma_0| - \frac{(\nu_0+1)d}{2}\log 2 - \frac{d}{2}\log\pi - \log\Gamma_d\left(\frac{\nu_0}{2}\right)$$

$$-h(m_0, \phi_0) \qquad + \frac{\nu_0+d+2}{2}\log\left((2\pi)^d\right)\Big]$$

Simplify $h(m_0, \phi_0)$

$$h(m_0, \phi_0) = \log\Gamma_d\left(\frac{\nu_0}{2}\right) - \frac{d(d-\nu_0+1)}{2}\log 2\pi + \frac{\nu_0 d}{2}\log 2$$

$$- \frac{d}{2}\log k_0 - \frac{\nu_0}{2}\log|\Sigma_0|$$

Q6 $P(\theta|X) = \exp\left[\left\langle \begin{pmatrix} \theta \\ -g(\theta) \end{pmatrix}, \begin{pmatrix} \phi_n \\ m_n \end{pmatrix} \right\rangle - h(m_n, \phi_n) \right]$

where $\theta = \begin{pmatrix} \Sigma^{-1}\mu \\ -\frac{1}{2}\Sigma^{-1} \end{pmatrix}$ $\quad \phi(x) = \begin{pmatrix} x \\ xx^T \end{pmatrix}$ in Q4

$g(\theta) = \begin{pmatrix} \frac{1}{2}\mu^T\Sigma^{-1}\mu \\ \frac{1}{2}\log\left((2\pi)^d|\Sigma|\right) \end{pmatrix}$ in Q5

$m_0 = \begin{pmatrix} k_0 \\ v_0+d+2 \end{pmatrix} \quad \phi_0 = \begin{pmatrix} k_0\mu_0 \\ k_0\mu_0\mu_0^T + \Sigma_0 \end{pmatrix}$

use the update rule in Q3

$$m_n = m_0 + n\mathbb{1} = \begin{pmatrix} k_0+n \\ v_0+d+2+n \end{pmatrix} = \begin{pmatrix} k_n \\ v_n+d+2 \end{pmatrix}$$

$$\phi_n = \phi_0 + \sum_{i=1}^{n}\phi(x_i)$$
$$= \begin{pmatrix} k_0\mu_0 + \sum_{i=1}^{n} x_i \\ k_0\mu_0\mu_0^T + \Sigma_0 + \sum_{i=1}^{n} x_ix_i^T \end{pmatrix} = \begin{pmatrix} k_n\mu_n \\ k_n\mu_n\mu_n^T + \Sigma_n \end{pmatrix}$$

$\Rightarrow k_n = k_0 + n \quad v_n = v_0 + n$

$$k_n\mu_n = k_0\mu_0 + \sum_{i=1}^{n} x_i \Rightarrow \mu_n = \frac{k_0\mu_0 + \sum_{i=1}^{n} x_i}{k_0+n}$$

$$k_n\mu_n\mu_n^T + \Sigma_n = k_0\mu_0\mu_0^T + \Sigma_0 + \sum_{i=1}^{n} x_ix_i^T$$

$$\Rightarrow \Sigma_n = k_0\mu_0\mu_0^T - k_n\mu_n\mu_n^T + \Sigma_0 + \sum_{i=1}^{n} x_ix_i^T$$

Q7 $P(\tilde{x}|X) = \int P(\tilde{x}|\theta)\, P(\theta|X)\, d\theta$

where $P(\tilde{x}|\theta) = \exp\left(\langle \phi(\tilde{x}), \theta \rangle - 1^T g(\theta)\right)$

$P(\theta|X) = \exp\left(\langle \begin{pmatrix} \theta \\ -g(\theta) \end{pmatrix}, \begin{pmatrix} \phi_n \\ m_n \end{pmatrix} \rangle - h(m_n, \phi_n)\right)$

$\Rightarrow P(\tilde{x}|X)$

$= \int \exp\left(\langle \begin{pmatrix} \theta \\ -g(\theta) \end{pmatrix}, \begin{pmatrix} \phi(\tilde{x}) + \phi_n \\ m_n + 1 \end{pmatrix} \rangle - h(m_n, \phi_n) \right) d$

$= \exp(-h(m_n, \phi_n)) \int \exp\left(\langle \begin{pmatrix} \theta \\ -g(\theta) \end{pmatrix}, \begin{pmatrix} \phi(\tilde{x}) + \phi_n \\ m_n + 1 \end{pmatrix} \rangle \right) d\theta$

Recall that due to $\int p(x)\, dx = 1$

$h(m, \phi) = \log \int \exp\left[\langle \begin{pmatrix} \theta \\ -g(\theta) \end{pmatrix}, \begin{pmatrix} \phi \\ m \end{pmatrix} \rangle \right] d\theta$

$\Rightarrow P(\tilde{x}|X)$

$= \exp(-h(m_n, \phi_n))\, \exp(h(m_n + 1, \phi(\tilde{x}) + \phi_n))$

$= \exp\left[h(m_n + 1, \phi(\tilde{x}) + \phi_n) - h(m_n, \phi_n) \right]$

Q8 $\quad m_n = \begin{pmatrix} k_n \\ v_n + d + 2 \end{pmatrix} \qquad \phi_n = \begin{pmatrix} k_n \mu_n \\ k_n \mu_n \mu_n^T + \Sigma_n \end{pmatrix}$

$$\Rightarrow \exp\left[h(m_n, \phi_n) \right]$$

$$= \frac{2^{\frac{v_n d}{2}} \Gamma_d\left(\frac{v_n}{2}\right)}{(2\pi)^{\frac{d(d+v_n+1)}{2}} k_n^{\frac{d}{2}} |\Sigma_n|^{\frac{v_n}{2}}}$$

$$\Rightarrow \exp\left[h(m_n + 1), \phi(\tilde{x}) + \phi_n \right]$$

$$= \frac{2^{\frac{(v_n+1)d}{2}} \Gamma_d\left(\frac{v_n+1}{2}\right)}{(2\pi)^{\frac{d(d+v_n+2)}{2}} (k_n+1)^{\frac{d}{2}} |\tilde{\Sigma}|^{\frac{(v_n+1)}{2}}}$$

$$(k_n+1)\tilde{\mu} = k_n \mu_n + \tilde{x}$$

$$\Rightarrow \tilde{\mu} = \frac{k_n \mu_n + \tilde{x}}{k_n + 1}$$

$$(k_n+1)\tilde{\mu}\tilde{\mu}^T + \tilde{\Sigma} = k_n \mu_n \mu_n^T + \Sigma_n + \tilde{x}\tilde{x}^T$$

$$\Rightarrow \tilde{\Sigma} = \Sigma_n + \tilde{x}\tilde{x}^T + k_n \mu_n \mu_n^T - \frac{1}{(k_n+1)}(k_n \mu_n + \tilde{x})(k_n \mu_n + \tilde{x})^T$$

$$= \Sigma_n + \frac{k_n}{k_n+1}(\tilde{x} - \mu_n)(\tilde{x} - \mu_n)^T$$

$$\Rightarrow |\tilde{\Sigma}| = \left(1 + \frac{k_n}{k_n+1}(\tilde{x} - \mu_n)\Sigma_n^{-1}(\tilde{x} - \mu_n)^T\right)|\Sigma_n|$$

$$\Rightarrow p(\tilde{x} \mid \bar{x}) = \frac{\Gamma_d\left(\frac{v_n+1}{2}\right)\left(\frac{k_n}{k+1}\right)^{\frac{d}{2}}}{\Gamma_d\left(\frac{v_n}{2}\right)\pi^{\frac{d}{2}}|\Sigma_n|^{\frac{1}{2}}\left(1 + \frac{k_n}{k_n+1}(\tilde{x} - \mu_n)\Sigma_n^{-1}(\tilde{x} - \mu_n)^T\right)^{\frac{v_n+1}{2}}}$$