Hw 2.

Q1.

(1) True.

(2)

$$\mu = \begin{pmatrix} \mu_a \\ \mu_b \\ \mu_c \end{pmatrix}, \quad \Sigma = \begin{pmatrix} \Sigma_{aa} & \Sigma_{ab} & \Sigma_{ac} \\ \Sigma_{ba} & \Sigma_{bb} & \Sigma_{bc} \\ \Sigma_{ca} & \Sigma_{cb} & \Sigma_{cc} \end{pmatrix}.$$

We first of all take the joint distribution $p(x_a, x_b, x_c)$ and marginalize to obtain the distribution $p(x_a, x_b)$. Using the result of section 2.3.2 this is again a Gaussian distribution with mean and covariance given by

$$\mu = \begin{pmatrix} \mu_a \\ \mu_b \end{pmatrix} \qquad \Sigma = \begin{pmatrix} \Sigma_{aa} & \Sigma_{ab} \\ \Sigma_{ba} & \Sigma_{bb} \end{pmatrix}.$$

From section 2.3.1, the distribution $p(x_a, x_b)$ is then Gaussian with mean and covariance given by

$$\mu_{a|b} = \mu_a + \Sigma_{ab}\Sigma_{bb}^{-1}(x_b - \mu_b)$$

$$\Sigma_{a|b} = \Sigma_{aa} - \Sigma_{ab}\Sigma_{bb}^{-1}\Sigma_{ba}.$$

Q2.

(a) ~~Ema~~ For the marginal distribution $p(x)$ we see from $E[x] = \mu$ that the mean is given by the top $E[z] = \begin{pmatrix} \mu \\ A\mu + b \end{pmatrix}$

similarly the covariance is $\mathrm{cov}[x] = \cancel{\begin{pmatrix} \Lambda^{-1} \\ A\Lambda^{-1} \end{pmatrix}}\ \Lambda^{-1}$

a) Apply the result $\mu_{a|b} = \mu_a + \Sigma_{ab}\Sigma_{bb}^{-1}(x_b - \mu_b)$

$\mu_{y|x} = A\mu + b + A\Lambda^{-1}\Lambda(x - \mu) = Ax + b$ so $p(x) = \mathcal{N}(x|\mu, \Lambda^{-1})$ defined the pdf

b) Apply the result $\Sigma_{a|b} = \Sigma_{aa} - \Sigma_{ab}\Sigma_{bb}^{-1}\Sigma_{ba}$.

$\mathrm{cov}[y|x] = L^{-1} + A\Lambda^{-1}A^T - A\Lambda^{-1}\Lambda\Lambda^{-1}A^T = L^{-1}$

Q3. Differentiating a) with respect to Σ, we obtain two terms

$$-\frac{N}{2}\frac{\partial}{\partial\Sigma}\ln|\Sigma| - \frac{1}{2}\frac{\partial}{\partial\Sigma}\sum_{n=1}^{N}(x_n - \mu)^T\Sigma^{-1}(x_n - \mu)$$

we apply $\frac{\partial}{\partial A}\ln|A| = (A^{-1})^T$ to get $-\frac{N}{2}\frac{\partial}{\partial\Sigma}\ln|\Sigma| = -\frac{N}{2}(\Sigma^{-1})^T = -\frac{N}{2}\Sigma^{-1}$

$\sum_{n=1}^{N}(x_n - \mu)^T\Sigma^{-1}(x_n - \mu) = N\,\mathrm{Tr}[\Sigma^{-1}S]$. $\qquad S = \frac{1}{N}\sum_{n=1}^{N}(x_n - \mu)(x_n - \mu)^T$.

$$\frac{\partial}{\partial\Sigma_{ij}}\sum_{n=1}^{N}(x_n - \mu)^T\Sigma^{-1}(x_n - \mu) = N\frac{\partial}{\partial\Sigma_{ij}}\mathrm{Tr}[\Sigma^{-1}S] = N\,\mathrm{Tr}\left[\frac{\partial}{\partial\Sigma_{ij}}\Sigma^{-1}S\right] = -N\,\mathrm{Tr}\left[\Sigma^{-1}\frac{\partial\Sigma}{\partial\Sigma_{ij}}\Sigma^{-1}S\right]$$

$$= -N\,\mathrm{Tr}\left[\frac{\partial\Sigma}{\partial\Sigma_{ij}}\Sigma^{-1}S\Sigma^{-1}\right] = -N(\Sigma^{-1}S\Sigma^{-1})_{ij}$$

$$-\frac{1}{2}\frac{\partial}{\partial\Sigma}\sum_{n=1}^{N}(x_n - \mu)^T\Sigma^{-1}(x_n - \mu) = \frac{N}{2}\Sigma^{-1}S\Sigma^{-1}$$

so $\frac{N}{2}\Sigma^{-1} = \frac{N}{2}\Sigma^{-1}S\Sigma^{-1}$

$\frac{N}{2} = \frac{N}{2}\Sigma^{-1}S. \Rightarrow \Sigma = S.$

the result symmetric and positive defined

Q4. a). $\sigma^2(N) = \frac{1}{N}\sum_{n=1}^{N}(x_n-\mu)^2$

$$= \frac{1}{N}\sum_{n=1}^{N-1}(x_n-\mu)^2 + \frac{(x_N-\mu)^2}{N}$$

$$= \frac{N-1}{N}\sigma^2(N-1) + \frac{(x_N-\mu)^2}{N}$$

$$= \frac{N-1}{N}\sigma^2(N-1) + \frac{(x_N-\mu)^2}{N}$$

$$= \sigma^2(N-1) - \frac{1}{N}\sigma^2(N-1) + \frac{(x_N-\mu)^2}{N}$$

$$= \sigma^2(N-1) + \frac{1}{N}\{(x_N-\mu)^2 - \sigma^2(N-1)\}$$

substitute the expression for a Gaussian distribution into the result for the Robbins - Monro procedure applied to maximizing likelihood.

$$\sigma^2(N) = \sigma^2(N-1) + a_{N-1}\frac{\partial}{\partial\sigma^2(N-1)}\left\{-\frac{1}{2}\ln\sigma^2_{(N-1)} - \frac{(x_N-\mu)^2}{2\sigma^2(N-1)}\right\}$$

$$= \sigma^2(N-1) + a_{N-1}\left\{\frac{1}{2\sigma^2(N-1)} + \frac{(x_N-\mu)^2}{2\sigma^4(N-1)}\right\}$$

$$= \sigma^2(N-1) + \frac{a_{N-1}}{2\sigma^4(N-1)}\{(x_N-\mu)^2 - \sigma^2(N-1)\}. \qquad a_{N-1} = \frac{2\sigma^4(N-1)}{N}$$

b). $\Sigma_{ML}^{(N)} = \frac{1}{N}\sum_{n=1}^{N}(x_n-\mu)(x_n-\mu)^T$

$$= \frac{1}{N}\sum_{n=1}^{N-1}(x_n-\mu)(x_n-\mu)^T + \frac{1}{N}(x_N-\mu)(x_N-\mu)^T$$

$$= \frac{N-1}{N}\Sigma_{ML}^{(N-1)} + \frac{1}{N}(x_N-\mu)(x_N-\mu)^T$$

$$= \Sigma_{ML}^{(N-1)} + \frac{1}{N}\left((x_N-\mu)(x_N-\mu)^T - \Sigma_{ML}^{(N-1)}\right)$$

$$\frac{\partial}{\partial\Sigma_{ML}^{(N-1)}}\ln p(x_N|\mu,\Sigma_{ML}^{(N-1)})$$

$$= \frac{1}{2}\left(\Sigma_{ML}^{(N-1)}\right)^{-1}\left((x_N-\mu)(x_N-\mu)^T - \Sigma_{ML}^{(N-1)}\right)\left(\Sigma_{ML}^{(N-1)}\right)^{-1}$$

$$= \frac{1}{2}\left(\Sigma_{ML}^{(N-1)}\right)^{-2}\left((x_N-\mu)(x_N-\mu)^T - \Sigma_{ML}^{(N-1)}\right)$$

$\left(\Sigma_{ML}^{(N-1)}\right)^{-1}$ is diagonal.

$$\Sigma_{ML}^{(N)} = \Sigma_{ML}^{(N-1)} + A_{N-1}\frac{1}{2}\left(\Sigma_{ML}^{(N-1)}\right)^{-2}\left((x_N-\mu)(x_N-\mu)^T - \Sigma_{ML}^{(N-1)}\right)$$

if we chose $A_{N-1} = \frac{2}{N}\left(\Sigma_{ML}^{(N-1)}\right)^2$ it satisfy.

Q5. $p(\mu|X) \propto p(\mu) \prod_{n=1}^{N} p(x_n|\mu, \Sigma)$.

$$-\frac{1}{2}(\mu-\mu_0)^T \Sigma_0^{-1}(\mu-\mu_0) - \frac{1}{2}\sum_{n=1}^{N}(x_n-\mu)^T \Sigma^{-1}(x_n-\mu)$$

$$= -\frac{1}{2}\mu^T(\Sigma_0^{-1} + N\Sigma^{-1})\mu + \mu^T\left(\Sigma_0^{-1}\mu_0 + \Sigma^{-1}\sum_{n=1}^{N} x_n\right) + \text{const}.$$

$$\mu_N = (\Sigma_0^{-1} + N\Sigma^{-1})^{-1}(\Sigma_0^{-1}\mu_0 + \Sigma^{-1} N \mu_{ML}).$$

$$\Sigma_N^{-1} = \Sigma_0^{-1} + N\Sigma^{-1}.$$

$$\mu_{ML} = \frac{1}{N}\sum_{n=1}^{N} x_n.$$