Homework 2

Questions: https://github.com/SUSTech-ML-Course/Machine-Learning-Course/blob/05b4e0bd4063b9d486f34ce557720b519947cbe3/lesson-homework/hw2/HW2.md

Question 1

(a) True

(b) According to the marginalization of Gaussian distribution, if x_c has been marginalized out, we have

$$p(x_a, x_b) = N(x|\mu, \Sigma)$$
$$where \;\; \mu = \begin{bmatrix} \mu_a \\ \mu_b \end{bmatrix} \quad \Sigma = \begin{bmatrix} \Sigma_{aa} & \Sigma_{ab} \\ \Sigma_{ba} & \Sigma_{bb} \end{bmatrix}$$

And we define

$$\Lambda = \Sigma^{-1} = \begin{bmatrix} \Lambda_{aa} & \Lambda_{ab} \\ \Lambda_{ba} & \Lambda_{bb} \end{bmatrix}$$

By solving the equation, we have

$$\begin{cases} \Lambda_{aa} = (\Sigma_{aa} - \Sigma_{ab}\Sigma_{bb}^{-1}\Sigma_{ba})^{-1} \\ \Lambda_{ab} = -(\Sigma_{aa} - \Sigma_{ab}\Sigma_{bb}^{-1}\Sigma_{ba})^{-1}\Sigma_{ab}\Sigma_{bb}^{-1} \end{cases}$$

So we have

$$\begin{aligned} \mu_{a|b} &= \mu_a - \Lambda_{aa}^{-1}\Lambda_{ab}(x_b - \mu_b) \\ &= \mu_a + (\Sigma_{aa} - \Sigma_{ab}\Sigma_{bb}^{-1}\Sigma_{ba})(\Sigma_{aa} - \Sigma_{ab}\Sigma_{bb}^{-1}\Sigma_{ba})^{-1}\Sigma_{ab}\Sigma_{bb}^{-1} \\ &= \mu_a + \Sigma_{ab}\Sigma_{bb}^{-1} \end{aligned}$$

Finally we get

$$\begin{aligned} p(x_a|x_b) &= N(x|\mu_{a|b}, \Lambda_{aa}^{-1}) \\ &= N(x|\mu_a + \Sigma_{ab}\Sigma_{bb}^{-1},\; \Sigma_{aa} - \Sigma_{ab}\Sigma_{bb}^{-1}\Sigma_{ba}) \end{aligned}$$

Question 2

(a) We have the property of Gaussian distribution

$$p(z) = N(z|\mu, \Sigma)$$
$$\mu = \begin{bmatrix} \mu_x \\ \mu_y \end{bmatrix} = \begin{bmatrix} \mu \\ A\mu + b \end{bmatrix} = E(z)$$
$$\Sigma = \begin{bmatrix} \Sigma_{xx} & \Sigma_{xy} \\ \Sigma_{yx} & \Sigma_{yy} \end{bmatrix} = \begin{bmatrix} \Lambda^{-1} & \Lambda^{-1}A^T \\ A\Lambda^{-1} & L^{-1} + A\Lambda^{-1}A^T \end{bmatrix}$$

Then we can get the marginal distribution

$$
\begin{aligned}
p(x) &= N(x|\mu_x, \Sigma_{xx}) \\
&= N(x|\mu, \Lambda^{-1})
\end{aligned}
$$

(b) We can get $\mu_{y|x}$ and Λ_{yy}^{-1} by using the equations above

$$
\begin{aligned}
\mu_{y|x} &= \mu_y + \Sigma_{yx}\Sigma_{xx}^{-1}(x - \mu_x) \\
&= (A\mu + b) - A\Lambda^{-1}\Lambda(x - \mu) \\
&= Ax + b
\end{aligned}
$$

$$
\begin{aligned}
\Lambda_{yy}^{-1} &= \Sigma_{yy} - \Sigma_{yx}\Sigma_{xx}^{-1}\Sigma_{xy} \\
&= L^{-1} + A\Lambda^{-1}A^T - A\Lambda^{-1}\Lambda\Lambda^{-1}A^T \\
&= L^{-1}
\end{aligned}
$$

Finally, by the property of Gaussian distribution, we have

$$
\begin{aligned}
p(y|x) &= N(y|y_{y|x}, \Lambda_{yy}^{-1}) \\
&= N(y|Ax + b, L^{-1})
\end{aligned}
$$

Question 3

$$
\begin{aligned}
lnp(X|\mu, \Sigma) &= -\frac{ND}{2}ln(2\pi) - \frac{N}{2}ln|\Sigma| - \frac{1}{2}\sum_{n=1}^{N}(x_n - \mu)^T\Sigma^{-1}(x_n - \mu) \\
\frac{\partial}{\partial \Sigma^{-1}}lnp(X|\mu, \Sigma) &= -\frac{N}{2}\frac{\partial ln|\Sigma|}{\partial \Sigma^{-1}} - \frac{1}{2}\frac{\partial}{\partial \Sigma^{-1}}\sum_{n=1}^{N}(x_n - \mu)^T\Sigma^{-1}(x_n - \mu) \\
&= \frac{N\Sigma^T}{2} - \frac{1}{2}\frac{\partial}{\partial \Sigma^{-1}}\sum_{n=1}^{N}tr((x_n - \mu)(x_n - \mu)^T\Sigma^{-1}) \\
&= \frac{N\Sigma^T}{2} - \frac{1}{2}\frac{\partial}{\partial \Sigma^{-1}}tr(\sum_{n=1}^{N}(x_n - \mu)(x_n - \mu)^T\Sigma^{-1}) \\
&= \frac{N\Sigma}{2} - \frac{1}{2}\sum_{n=1}^{N}(x_n - \mu)(x_n - \mu)^T \\
&= 0
\end{aligned}
$$

Finally we get

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

Question 4

(a) The Robbins-Monro algorithm for a sequence of random variable x_n is given by

$$
a_{n+1} = a_n + \alpha_n(x_n - a_n)
$$

where a_n is the estimate at the nth step, x_n is the nth data point, α_n is a sequence of positive step sizes that satisfy certain conditions

Question 5

From Bayes theorem, we have

$$p(\mu|X) = \frac{p(X|\mu)p(\mu)}{p(X)}$$

For the production of normal distribution, we consider the term in exp

$$\begin{aligned} R.H.S &= -\frac{1}{2}(\mu-\mu_0)^T\Sigma_0^{-1}(\mu-\mu_0)\sum_{n=1}^{N}-\frac{1}{2}(x_n-\mu)^T\Sigma_0^{-1}(x_n-\mu) \\ &= -\frac{1}{2}\mu^T(\Sigma_0^{-1}+N\Sigma^{-1})\mu+\mu^T(\Sigma_0^{-1}\mu_0+\Sigma^{-1}\sum_{n=1}^{N}x_n)+\ldots \end{aligned}$$

Because

$$\frac{1}{2}(x-\mu)^T\Sigma^{-1}(x-\mu) = \frac{1}{2}x^T\Sigma^{-1}x + x^T\Sigma^{-1}\mu+\ldots$$

We can finally get

$$\begin{cases} \mu_{post} = (\Sigma_0^{-1}+N\Sigma^{-1})^{-1}(\Sigma_0^{-1}\mu_0+\Sigma^{-1}\sum_{n=1}^{N}x_n) \\ \Sigma_{post} = (\Sigma_0^{-1}+N\Sigma^{-1})^{-1} \end{cases}$$