CS405 Homework #2

SID: 12110524

Name: 陈康睿

Question 1

1. True (Ref: PRML 2.3.1, 2.3.2)

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

2. Merge x_a, x_b, then

$$x = \begin{bmatrix} x_{a,b} \\ x_c \end{bmatrix} \quad \mu = \begin{bmatrix} \mu_{a,b} \\ \mu_c \end{bmatrix} \quad \Sigma = \begin{bmatrix} \Sigma_{(a,b)(a,b)} & \Sigma_{(a,b)c} \\ \Sigma_{c(a,b)} & \Sigma_{cc} \end{bmatrix}$$

Then we can take advantage of the marginal result in (1):

$$p(x_{a,b}) = \mathcal{N}(x_{a,b}|\mu_{a,b}, \varSigma_{(a,b)(a,b)})$$

Now we do the partition again:

$$x_{a,b} = \begin{bmatrix} x_a \\ x_b \end{bmatrix} \quad \mu_{a,b} = \begin{bmatrix} \mu_a \\ \mu_b \end{bmatrix} \quad \Sigma_{a,b} = \begin{bmatrix} \Sigma_{aa} & \Sigma_{ab} \\ \Sigma_{ba} & \Sigma_{bb} \end{bmatrix}$$

Finally we use the conditional result in (1):

$$p(\mathbf{x_a}|\mathbf{x_b}) = \mathcal{N}(\mathbf{x_a}|\mu_\mathbf{a} - \mathbf{\Lambda_{ab}^{-1}}\mathbf{\Lambda_{bb}}(\mathbf{x_b} - \mu_\mathbf{b}), \mathbf{\Lambda_{aa}^{-1}})$$

Question 2

1. Using the marginal result in Q1.1:

$$p(x) = \mathcal{N}(x|\mu_x, \varSigma_{xx}) = \mathcal{N}(x|\mu, \varLambda^{-1})$$

2. Using the conditional result in Q1.1:

$$\begin{aligned} p(\mathbf{y}|\mathbf{x}) &= \mathcal{N}(\mathbf{y}|\mu_\mathbf{y} - \mathbf{\Sigma_{yx}}\mathbf{\Sigma_{xx}^{-1}}(\mathbf{x} - \mu_\mathbf{x}), \mathbf{\Sigma_{yy}} - \mathbf{\Sigma_{yx}}\mathbf{\Sigma_{xx}^{-1}}\mathbf{\Sigma_{xy}}) \\ &= \mathcal{N}(\mathbf{y}|\mathbf{A}\mu + \mathbf{b} - \mathbf{A}\mathbf{\Lambda}^{-1}\mathbf{\Lambda}(\mathbf{x} - \mu), \mathbf{L}^{-1} + \mathbf{A}\mathbf{\Lambda}^{-1}\mathbf{A}^\mathbf{T} - \mathbf{A}\mathbf{\Lambda}^{-1}\mathbf{\Lambda}\mathbf{\Lambda}^{-1}\mathbf{A}^\mathbf{T}) \\ &= \mathcal{N}(\mathbf{y}|\mathbf{A}\mathbf{x} + \mathbf{b}, \mathbf{L}^{-1}) \end{aligned}$$

Question 3

Log likelihood function respect to Σ:

$$
\begin{aligned}
\frac{\partial \ln p(\mathbf{X}|\mu,\Sigma)}{\partial \Sigma} &= -\frac{\partial}{\partial \Sigma}\left(\frac{N}{2}\ln|\Sigma| + \frac{1}{2}\sum_{n=1}^{N}(\mathbf{x}_n-\mu)^T\Sigma^{-1}(\mathbf{x}_n-\mu)\right) \\
&= \frac{1}{2}\Sigma^{-1}\sum_{n=1}^{N}(\mathbf{x}_n-\mu)(\mathbf{x}_n-\mu)^T\Sigma^{-1} - \frac{N}{2}\Sigma^{-1} \qquad (with\ \Sigma^{-T}=\Sigma^{-1}) \\
&= 0
\end{aligned}
$$

Then we can get:

$$
\Sigma_{ML} = \sum_{n=1}^{N}(\mathbf{x}_n-\mu)(\mathbf{x}_n-\mu)^T
$$

Obviously, Σ_{ML} is symmetric, and $x^T\Sigma_{ML}x = \frac{1}{N}\sum_{n=1}^{N}(x^T(\mathbf{x}_n-\mu))^2 > 0$ for any $x \neq 0, \exists \mathbf{x}_n \neq \mu$.

So the final result is symmetric and positive definite.

Question 4

Robbins-Monro sequential estimation formula:

$$
\theta^{(N)} = \theta^{(N-1)} + a_{N-1}\frac{\partial}{\partial \theta^{N-1}}\ln p(x_N|\theta^{N-1})
$$

1. Dissecting out the contribution from the final data point, we obtain:

$$
(\sigma^2_{ML})^{(N)} = (\sigma^2_{ML})^{(N-1)} + \frac{1}{N}\left((x_N-\mu_{ML})^2 - (\sigma^2_{ML})^{(N-1)}\right)
$$

By likelihood:

$$
\lim_{N\to\infty}\frac{1}{N}\sum_{n=1}^{N}\ln p(x_n|\mu,\sigma^2) = \mathbb{E}_x\left[\frac{\partial}{\partial\sigma^2}\ln p(x_n|\mu,\sigma^2)\right]
$$

Substituting into thesequential formula:

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

So $a_N = \frac{2}{N+1}(\sigma^4_{ML})^{(N)}$

2. Similarly, dissecting out the last data point:

$$
\Sigma^{(N)}_{ML} = \Sigma^{(N-1)}_{ML} + \frac{1}{N}\left((\mathbf{x}_N-\mu_{ML})(\mathbf{x}_N-\mu_{ML})^T - \Sigma^{(N-1)}_{ML}\right)
$$

By substitution:

$$
\Sigma^{(N)}_{ML} = \Sigma^{(N-1)}_{ML} + \frac{a_{N-1}}{2}\left((\Sigma^{-1}_{ML})^{(N-1)}(\mathbf{x}_N-\mu_{ML})(\mathbf{x}_N-\mu_{ML})^T(\Sigma^{-1}_{ML})^{(N-1)} - (\Sigma^{-1}_{ML})^{(N-1)}\right)
$$

Hence $a_N = \frac{2}{N+1}(\Sigma^2_{ML})^{(N)}$

Question 5

$\textbf{Posterior} \propto \textbf{Prior} \times \textbf{Likelihood}$:

$$
\begin{aligned}
p(\mu|\mathbf{X}) &\propto p(\mu)\prod_{n=1}^{N} p(\mathbf{X}_n|\mu,\Sigma) \\
&\propto \exp\left\{-\frac{1}{2}(\mu-\mu_0)^{\mathbf{T}}\Sigma_0^{-1}(\mu-\mu_0)-\frac{1}{2}\sum_{n=1}^{N}(\mathbf{x}_n-\mu)^{\mathbf{T}}(\mathbf{x}_n-\mu)\right\}
\end{aligned}
$$

And the exponential terms can be rearranged by the power of μ (as we are finding the distribution of μ, and the constant term can be seen as coefficient of $\exp$):

$$
-\frac{1}{2}\mu^{\mathbf{T}}(\Sigma_0^{-1}+N\Sigma^{-1})\mu+\mu^{\mathbf{T}}\left(\Sigma_0^{-1}\mu_0+\Sigma^{-1}\sum_{n=1}^{N}\mathbf{x}_n\right)+C
$$

where C stands for constant.

As the exponent of $\mathcal{N}(\mathbf{x}|\mu,\Sigma)$ can be expressed by:

$$
-\frac{1}{2}\mathbf{x}^{\mathbf{T}}\Sigma^{-1}\mathbf{x}+\mathbf{x}^{\mathbf{T}}\Sigma^{-1}\mu+C
$$

So we can get:

$$
\begin{aligned}
\mu_{post} &= (\Sigma_0^{-1}+N\Sigma^{-1})^{-1}\left(\Sigma_0^{-1}\mu_0+\Sigma^{-1}\sum_{n=1}^{N}\mathbf{x}_n\right) \\
\Sigma_{post} &= (\Sigma_0^{-1}+N\Sigma^{-1})^{-1}
\end{aligned}
$$

Thus, $p(\mu|\mathbf{X})=\mathcal{N}(\mathbf{x}|\mu_{post},\Sigma_{post})$