CS405 Homework #2

12110644 周思呈

Question 1

(a) [True or False] If two sets of variables are jointly Gaussian, then the conditional distribution of one set conditioned on the other is again Gaussian. Similarly, the marginal distribution of either set is also Gaussian

True.

(b) (textbook exercise 2.25) Consider a partitioning of the components of x into three groups x_a, x_b, and x_c, with a corresponding partitioning of the mean vector μ and of the covariance matrix Σ in the form

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

Find an expression for the conditional distribution $p(x_a|x_b)$ in which x_c has been marginalized out.

First group $\boldsymbol{x}_a$ and $\boldsymbol{x}_b$ together:

$$\boldsymbol{x} = \begin{pmatrix} \boldsymbol{x}_{a,b} \\ \boldsymbol{x}_c \end{pmatrix}, \quad \mu = \begin{pmatrix} \mu_{a,b} \\ \mu_c \end{pmatrix}, \quad \Sigma = \begin{pmatrix} \Sigma_{(a,b)(a,b)} & \Sigma_{(a,b)c} \\ \Sigma_{(a,b)c} & \Sigma_{cc} \end{pmatrix}. \tag{1}$$

According to the formula for the marginal probability distribution, we have the joint distribution of $\boldsymbol{x}_a$ and $\boldsymbol{x}_b$:

$$p\left(\boldsymbol{x}_{a,b}\right) = \mathcal{N}\left(\boldsymbol{x}_{a,b} \mid \boldsymbol{\mu}_{a,b}, \boldsymbol{\Sigma}_{(a,b)(a,b)}\right). \tag{2}$$

So we have

$$p\left(\boldsymbol{x}_a \mid \boldsymbol{x}_b\right) = \mathscr{N}\left(\boldsymbol{x} \mid \boldsymbol{\mu}_{a|b}, \boldsymbol{\Lambda}_{aa}^{-1}\right) \tag{3}$$

where

$$\boldsymbol{\mu}_{a|b} = \boldsymbol{\mu}_a - \boldsymbol{\Lambda}_{aa}^{-1}\boldsymbol{\Lambda}_{ab}\left(\boldsymbol{x}_b - \boldsymbol{\mu}_b\right) \tag{4}$$

Question 2

(textbook exercise 2.28) Consider a joint distribution over the variable

$$\mathbf{z} = \begin{pmatrix} \mathbf{x} \\ \mathbf{y} \end{pmatrix} \tag{5}$$

whose mean and covariance are given by

$$\mathbb{E}[\mathbf{z}] = \begin{pmatrix} \mu \\ \mathbf{A}\mu + \mathbf{b} \end{pmatrix}, \quad \operatorname{cov}[\mathbf{z}] = \begin{pmatrix} \mathbf{\Lambda}^{-1} & \mathbf{\Lambda}^{-1}\mathbf{A}^{\mathrm{T}} \\ \mathbf{A}\mathbf{\Lambda}^{-1} & \mathbf{L}^{-1} + \mathbf{A}\mathbf{\Lambda}^{-1}\mathbf{A}^{\mathrm{T}} \end{pmatrix}. \tag{6}$$

(a) Show that the marginal distribution $p(\mathbf{x})$ is given by $p(\mathbf{x}) = \mathcal{N}(\mathbf{x}|\mu, \mathbf{\Lambda}^{-1})$.

$$p\left(\mathbf{x}\right) = \int p\left(\mathbf{x}, \mathbf{y}\right)\mathrm{d}\mathbf{y} \tag{7}$$

Considering the terms involving $\mathbf{y}$, we have

$$-\frac{1}{2}\mathbf{y}^{\mathrm{T}}\boldsymbol{\Lambda}_{bb}\mathbf{y} + \mathbf{y}^{T}\mathbf{m} = -\frac{1}{2}\left(\mathbf{y} - \boldsymbol{\Lambda}_{bb}^{-1}\mathbf{m}\right)^{\mathrm{T}}\boldsymbol{\Lambda}_{bb}\left(\mathbf{y} - \boldsymbol{\Lambda}_{bb}^{-1}\mathbf{m}\right) + \frac{1}{2}\mathbf{m}^{\mathrm{T}}\boldsymbol{\Lambda}_{bb}^{-1}\mathbf{m} \tag{8}$$

where we have defined

$$
\begin{aligned}
\mathbf{m} &= \mathbf{\Lambda}_{bb}\boldsymbol{\mu}_b - \mathbf{\Lambda}_{ba}\left(\mathbf{x}_a - \boldsymbol{\mu}_a\right), \\
\mathbf{\Lambda}_{bb} &= \mathbf{L}^{-1} + \mathbf{A}\mathbf{\Lambda}^{-1}\mathbf{A}^{\mathrm{T}}, \\
\mathbf{\Lambda}_{ba} &= \mathbf{A}\mathbf{\Lambda}^{-1}.
\end{aligned} \tag{9}
$$

We see that the dependence on $\mathbf{x}_b$ has been cast into the standard quadratic form of a Gaussian distribution, so we have

$$
p(\mathbf{x}) = \int \exp\left\{-\frac{1}{2}\left(\mathbf{y} - \mathbf{\Lambda}_{bb}^{-1}\mathbf{m}\right)^{\mathrm{T}}\mathbf{\Lambda}_{bb}\left(\mathbf{y} - \mathbf{\Lambda}_{bb}^{-1}\mathbf{m}\right)\right\}\mathrm{d}\mathbf{y}. \tag{10}
$$

The coefficient of the normalized Gaussian is independent of the mean and depends only on the determinant of the covariance matrix. So we have

$$
\begin{aligned}
\mathbf{\Sigma}_a &= \left(\mathbf{\Lambda}_{aa} - \mathbf{\Lambda}_{ab}\mathbf{\Lambda}_{bb}^{-1}\mathbf{\Lambda}_{ba}\right)^{-1}, \\
\boldsymbol{\mu}_a &= \mathbf{\Sigma}_a\left(\mathbf{\Lambda}_{aa} - \mathbf{\Lambda}_{ab}\mathbf{\Lambda}_{bb}^{-1}\mathbf{\Lambda}_{ba}\right)\boldsymbol{\mu}_a.
\end{aligned} \tag{11}
$$

Thus the marginal distribution $p\left(\mathbf{x}_a\right)$ has mean and covariance given by

$$
\begin{aligned}
\mathbb{E}\left[\mathbf{y}\right] &= \boldsymbol{\mu}_a \\
\operatorname{cov}\left[\mathbf{y}\right] &= \mathbf{\Sigma}_{aa}.
\end{aligned} \tag{12}
$$

Thus

$$
p(\mathbf{x}) = \mathcal{N}(\mathbf{x}|\mu, \mathbf{\Lambda}^{-1}). \tag{13}
$$

(b) Show that the conditional distribution $p(\mathbf{y}|\mathbf{x})$ is given by $p(\mathbf{y}|\mathbf{x}) = \mathcal{N}(\mathbf{y}|\mathbf{A}\mathbf{x} + \mathbf{b}, \mathbf{L}^{-1})$.

By using (2.98) in textbook, we can obtain:

$$
p(\boldsymbol{x}) = \mathscr{N}\left(\boldsymbol{x} \mid \boldsymbol{\mu}, \mathbf{\Lambda}^{-1}\right) \tag{14}
$$

And by using (2.96) and (2.97), we can obtain:

$$
p(\boldsymbol{y} \mid \boldsymbol{x}) = \mathscr{N}\left(\boldsymbol{y} \mid \boldsymbol{\mu}_{\boldsymbol{y}|\boldsymbol{x}}, \mathbf{\Lambda}_{\boldsymbol{yy}}^{-1}\right) \tag{15}
$$

where $\mathbf{\Lambda}_{yy} = \boldsymbol{L}^{-1}$. Finally the conditional mean is given by (2.97) :

$$
\boldsymbol{\mu}_{\boldsymbol{y}|\boldsymbol{x}} = \boldsymbol{A}\boldsymbol{\mu} + \boldsymbol{L} - \boldsymbol{L}^{-1}(-\boldsymbol{L}\boldsymbol{A})(\boldsymbol{x} - \boldsymbol{\mu}) = \boldsymbol{A}\boldsymbol{x} + \boldsymbol{L} \tag{16}
$$

Question 3

Show that the covariance matrix Σ that maximizes the log likelihood function is given by the sample covariance

$\ln p(\mathbf{X}|\mu, \Sigma) = -\frac{ND}{2}\ln(2\pi) - \frac{N}{2}\ln|\Sigma| - \frac{1}{2}\sum_{n=1}^{N}(\mathbf{x}_n - \mu)^{\mathrm{T}}\Sigma^{-1}(\mathbf{x}_n - \mu)$.

Is the final result symmetric and positive definite (provided the sample covariance is nonsingular)?

> ### Hints
>
> *(a)* To find the maximum likelihood solution for the covariance matrix of a multivariate Gaussian, we need to maximize the log likelihood function with respect to Σ. The log likelihood function is given by
>
> $\ln p(\mathbf{X}|\mu, \Sigma) = -\frac{ND}{2}\ln(2\pi) - \frac{N}{2}\ln|\Sigma| - \frac{1}{2}\sum_{n=1}^{N}(\mathbf{x}_n - \mu)^{\mathrm{T}}\Sigma^{-1}(\mathbf{x}_n - \mu)$.
>
> *(b)* The derivative of the inverse of a matrix can be expressed as
>
> $\frac{\partial}{\partial x}(\mathbf{A}^{-1}) = -\mathbf{A}^{-1}\frac{\partial \mathbf{A}}{\partial x}\mathbf{A}^{-1}$
>
> We have the following properties
>
> $\frac{\partial}{\partial \mathbf{A}}\mathrm{Tr}(\mathbf{A}) = \mathbf{I}, \quad \frac{\partial}{\partial \mathbf{A}}\ln|\mathbf{A}| = (\mathbf{A}^{-1})^{\mathrm{T}}$.

Let $C = -\frac{ND}{2}\ln(2\pi)$, which is a constant, we can rewrite the log likelihood function as

$$\begin{aligned} l &= \mathrm{C} - \frac{N}{2}\ln|\Sigma| - \frac{1}{2}\sum_{i=1}^{N}\left(\mathbf{x}^{(\mathbf{i})} - \mu\right)^{\mathbf{T}}\Sigma^{-\mathbf{1}}\left(\mathbf{x}^{(\mathbf{i})} - \mu\right) \\ &= \mathrm{C} + \frac{N}{2}\ln\left|\Sigma^{-1}\right| - \frac{1}{2}\sum_{i=1}^{N}\operatorname{tr}\left[\left(\mathbf{x}^{(\mathbf{i})} - \mu\right)\left(\mathbf{x}^{(\mathbf{i})} - \mu\right)^{\mathbf{T}}\Sigma^{-\mathbf{1}}\right]. \end{aligned} \tag{17}$$

Take derivate w.r.t Σ^{-1}

$$\begin{aligned} \frac{\partial}{\partial\Sigma^{-1}}l &= \frac{N}{2}\Sigma - \frac{1}{2}\sum_{i=1}^{N}\left(\mathbf{x}^{(\mathbf{i})} - \mu\right)\left(\mathbf{x}^{(\mathbf{i})} - \mu\right)^{T} \\ &= 0 \end{aligned} \tag{18}$$

$$\begin{aligned} 0 &= N\Sigma - \textstyle\sum_{i=1}^{N}\left(\mathbf{x}^{(\mathbf{i})} - \mu\right)\left(\mathbf{x}^{(\mathbf{i})} - \mu\right)^{T} \\ \hat{\Sigma} &= \textstyle\frac{1}{N}\sum_{i=1}^{N}\left(\mathbf{x}^{(\mathbf{i})} - \hat{\mu}\right)\left(\mathbf{x}^{(\mathbf{i})} - \hat{\mu}\right)^{T} \end{aligned} \tag{19}$$

This is symmetric and positive definite.

Reference for this answer

Question 4

(a) (textbook exercise 2.36) Derive an expression for the sequential estimation of the variance of a univariate Gaussian distribution, by starting with the maximum likelihood expression

$\sigma^2_{\mathrm{ML}} = \frac{1}{N}\sum_{n=1}^{N}(x_n - \mu)^2.$

Verify that substituting the expression for a Gaussian distribution into the Robbins-Monro sequential estimation formula gives a result of the same form, and hence obtain an expression for the corresponding coefficients a_N.

Derive an expression for the sequential estimation of the variance of a univariate Gaussian distribution:

$$\begin{aligned} \sigma_{ML}^{2(N)} &= \frac{1}{N}\sum_{n=1}^{N}\left(x_n - \mu_{ML}^{(N)}\right)^2 \\ &= \frac{1}{N}\left[\sum_{n=1}^{N-1}\left(x_n - \mu_{ML}^{(N)}\right)^2 + \left(x_N - \mu_{ML}^{(N)}\right)^2\right] \\ &= \frac{N-1}{N}\frac{1}{N-1}\sum_{n=1}^{N-1}\left(x_n - \mu_{ML}^{(N)}\right)^2 + \frac{1}{N}\left(x_N - \mu_{ML}^{(N)}\right)^2 \\ &= \frac{N-1}{N}\sigma_{ML}^{2(N-1)} + \frac{1}{N}\left(x_N - \mu_{ML}^{(N)}\right)^2 \\ &= \sigma_{ML}^{2(N-1)} + \frac{1}{N}\left[\left(x_N - \mu_{ML}^{(N)}\right)^2 - \sigma_{ML}^{2(N-1)}\right] \end{aligned} \tag{20}$$

which means we obtain our revised estimate by moving the old estimate a small amount, in the direction of the 'error signal'.

According to the Robbins-Monro sequential estimation formula

$$\begin{aligned} \sigma_{ML}^{2(N)} &= \sigma_{ML}^{2(N-1)} + a_{N-1}\frac{\partial}{\partial\sigma_{ML}^{2(N-1)}}\ln p\left(x_N \mid \mu_{ML}^{(N)}, \sigma_{ML}^{(N-1)}\right) \\ &= \sigma_{ML}^{2(N-1)} + a_{N-1}\left[-\frac{1}{2\sigma_{ML}^{2(N-1)}} + \frac{\left(x_N - \mu_{ML}^{(N)}\right)^2}{2\sigma_{ML}^{4(N-1)}}\right] \end{aligned} \tag{21}$$

If we choose :

$$a_{N-1} = \frac{2\sigma_{ML}^{4(N-1)}}{N} \tag{22}$$

then we will obtain :

$$\sigma_{ML}^{2(N)} = \sigma_{ML}^{2(N-1)} + \frac{1}{N}\left[-\sigma_{ML}^{2(N-1)} + \left(x_N - \mu_{ML}^{(N)}\right)^2\right] \tag{23}$$

So

$$a_N = \frac{2\sigma_{ML}^{4(N)}}{N+1} \tag{24}$$

(b) (textbook exercise 2.37) Derive an expression for the sequential estimation of the covariance of a multivariate Gaussian distribution, by starting with the maximum likelihood expression

$\Sigma_{\mathrm{ML}} = \frac{1}{N}\sum_{n=1}^{N}(\mathbf{x}_n - \mu_{\mathrm{ML}})(\mathbf{x}_n - \mu_{\mathrm{ML}})^{\mathrm{T}}$.

Verify that substituting the expression for a Gaussian distribution into the Robbins-Monro sequential estimation formula gives a result of the same form, and hence obtain an expression for the corresponding coefficients a_N.

> ### Hints
>
> *(a)* Consider the result $\mu_{\mathrm{ML}} = \frac{1}{N}\sum_{n=1}^{N}\mathbf{x}_n$ for the maximum likelihood estimator of the mean μ_{ML}, which we will denote by $\mu_{\mathrm{ML}}^{(N)}$ when it is based on N observations. If we dissect out the contribution from the final data point $\mathbf{x}_N$, we obtain
>
> $\mu_{\mathrm{ML}}^{(N)} = \frac{1}{N}\sum_{n=1}^{N}\mathbf{x}_n = \frac{1}{N}\mathbf{x}_N + \frac{1}{N}\sum_{n=1}^{N-1}\mathbf{x}_n = \frac{1}{N}\mathbf{x}_N + \frac{N-1}{N}\mu_{\mathrm{ML}}^{(N-1)}$
>
> *(b)* Robbins-Monro for maximum likelihood
>
> $\theta^{(N)} = \theta^{(N-1)} + a_{(N-1)}\frac{\partial}{\partial\theta^{(N-1)}}\ln p(x_N|\theta^{(N-1)})$.

Derive an expression for the sequential estimation of the covariance of a multivariate Gaussian distribution:

$$\begin{aligned}
\Sigma_{ML}^{(N)} &= \frac{1}{N}\sum_{n=1}^{N}\left(x_n - \mu_{ML}^{(N)}\right)\left(x_n - \mu_{ML}^{(N)}\right)^T \\
&= \frac{1}{N}\left[\sum_{n=1}^{N-1}\left(x_n - \mu_{ML}^{(N)}\right)\left(x_n - \mu_{ML}^{(N)}\right)^T + \left(x_N - \mu_{ML}^{(N)}\right)\left(x_N - \mu_{ML}^{(N)}\right)^T\right] \\
&= \frac{N-1}{N}\Sigma_{ML}^{(N-1)} + \frac{1}{N}\left(x_N - \mu_{ML}^{(N)}\right)\left(x_N - \mu_{ML}^{(N)}\right)^T \\
&= \Sigma_{ML}^{(N-1)} + \frac{1}{N}\left[\left(x_N - \mu_{ML}^{(N)}\right)\left(x_N - \mu_{ML}^{(N)}\right)^T - \Sigma_{ML}^{(N-1)}\right]
\end{aligned} \tag{25}$$

If we use Robbins-Monro sequential estimation formula, we can obtain:

$$\begin{aligned}
\Sigma_{ML}^{(N)} &= \Sigma_{ML}^{(N-1)} + a_{N-1}\frac{\partial}{\partial\Sigma_{ML}^{(N-1)}}\ln p\left(x_N \mid \mu_{ML}^{(N)}, \Sigma_{ML}^{(N-1)}\right) \\
&= \Sigma_{ML}^{(N-1)} + a_{N-1}\frac{\partial}{\partial\Sigma_{ML}^{(N-1)}}\ln p\left(x_N \mid \mu_{ML}^{(N)}, \Sigma_{ML}^{(N-1)}\right) \\
&= \Sigma_{ML}^{(N-1)} + a_{N-1}\left[-\frac{1}{2}\left[\Sigma_{ML}^{(N-1)}\right]^{-1} + \frac{1}{2}\left[\Sigma_{ML}^{(N-1)}\right]^{-1}\left(x_n - \mu_{ML}^{(N-1)}\right)\left(x_n - \mu_{ML}^{(N-1)}\right)^T\left[\Sigma_{ML}^{(N-1)}\right]^{-1}\right]
\end{aligned} \tag{26}$$

so

$$\begin{aligned}
a_{N-1} &= \frac{2}{N}\Sigma_{ML}^{2(N-1)} \\
a_N &= \frac{2}{N+1}\Sigma_{ML}^{2(N)}
\end{aligned} \tag{27}$$

Question 5

(textbook exercise 2.40) Consider a D-dimensional Gaussian random variable x with distribution $N(x|\mu, \Sigma)$ in which the covariance Σ is known and for which we wish to infer the mean μ from a set of observations $\mathbf{X} = \{x_1, x_2, \ldots\ldots, x_N\}$. Given a prior distribution $p(\mu) = N(\mu|\mu_0, \Sigma_0)$, find the corresponding posterior distribution $p(\mu|\mathbf{X})$.

According to Bayes Theorem

$$p(\boldsymbol{\mu} \mid \boldsymbol{X}) \propto p(\boldsymbol{X} \mid \boldsymbol{\mu}) p(\boldsymbol{\mu}) \tag{28}$$

Focus only on the exponential term on the right side and rearrange it regarding to $\boldsymbol{\mu}$.

$$\begin{aligned}
&\left[\sum_{n=1}^{N} -\frac{1}{2}(\boldsymbol{x}_n - \boldsymbol{\mu})^T \boldsymbol{\Sigma}^{-1} (\boldsymbol{x}_n - \boldsymbol{\mu})\right] - \frac{1}{2}(\boldsymbol{\mu} - \boldsymbol{\mu}_0)^T \boldsymbol{\Sigma}_0^{-1} (\boldsymbol{\mu} - \boldsymbol{\mu}_0) \\
&= -\frac{1}{2}\boldsymbol{\mu} \left(\boldsymbol{\Sigma}_0^{-1} + N\boldsymbol{\Sigma}^{-1}\right)\boldsymbol{\mu} + \boldsymbol{\mu}^T \left(\boldsymbol{\Sigma}_0^{-1}\boldsymbol{\mu}_0 + \boldsymbol{\Sigma}^{-1}\sum_{n=1}^{N}\boldsymbol{x}_n\right) + C
\end{aligned} \tag{29}$$

According to the quadratic term

$$\boldsymbol{\Sigma}_N^{-1} = \boldsymbol{\Sigma}_0^{-1} + N\boldsymbol{\Sigma}^{-1} \tag{30}$$

According to the linear term

$$\boldsymbol{\Sigma}_N^{-1}\boldsymbol{\mu}_N = \left(\boldsymbol{\Sigma}_0^{-1}\boldsymbol{\mu}_0 + \boldsymbol{\Sigma}^{-1}\sum_{n=1}^{N}\boldsymbol{x}_n\right) \tag{31}$$

So we have

$$\boldsymbol{\mu}_N = \left(\boldsymbol{\Sigma}_0^{-1} + N\boldsymbol{\Sigma}^{-1}\right)^{-1}\left(\boldsymbol{\Sigma}_0^{-1}\boldsymbol{\mu}_0 + \boldsymbol{\Sigma}^{-1}\sum_{n=1}^{N}\boldsymbol{x}_n\right) \tag{32}$$