Machine Learning (H)
Southern University of Science and Technology
Mengxuan Wu
12212006

Assignment 2

Mengxuan Wu

Question 1

(a)

True.
If two sets of random variables are jointly Gaussian, then the conditional distribution of one set given the other set is still Gaussian:

$$p(x_a|x_b) = \mathcal{N}(\mu_a + \Sigma_{ab}\Sigma_{bb}^{-1}(x_b - \mu_b), \Sigma_{aa} - \Sigma_{ab}\Sigma_{bb}^{-1}\Sigma_{ba})$$

Also, the marginal distribution of the two sets is still Gaussian:

$$p(x_a) = \mathcal{N}(\mu_a, \Sigma_{aa})$$

(b)

We can rewrite the mean and variance as:

$$\mu = \begin{pmatrix} \mu_{ab} \mu_c \end{pmatrix}, \quad \Sigma = \begin{pmatrix} \Sigma_{(ab)(ab)} & \Sigma_{(ab)c} \\ \Sigma_{c(ab)} & \Sigma_{cc} \end{pmatrix}$$

Then, from part (a), the marginal distribution of x_{ab} is Gaussian:

$$p(x_{ab}) = \mathcal{N}(\mu_{ab}, \Sigma_{(ab)(ab)})$$

Again, from part (a), the conditional distribution of x_a given x_b is Gaussian:

$$p(x_a|x_b) = \mathcal{N}(\mu_a + \Sigma_{ab}\Sigma_{bb}^{-1}(x_b - \mu_b), \Sigma_{aa} - \Sigma_{ab}\Sigma_{bb}^{-1}\Sigma_{ba})$$

Question 2

(a)

Since μ is the upper part of mean vector, then its variance is given by the upper-left part of the covariance matrix, which is Λ^{-1}. Thus, $p(x) = \mathcal{N}(\mu, \Lambda^{-1})$.

(b)

From question 1 (a), the conditional distribution of y given x is Gaussian, and its mean and variance are:

$$\mu_{y|x} = \mu_y + \Sigma_{yx}\Sigma_{xx}^{-1}(x - \mu_x)$$

$$\Sigma_{y|x} = \Sigma_{yy} - \Sigma_{yx}\Sigma_{xx}^{-1}\Sigma_{xy}$$

Substitute the given values into the equations, we have:

$$\mu_{y|x} = A\mu + b + A\Lambda^{-1}\Lambda(x - \mu) = Ax + b$$

$$\Sigma_{y|x} = L^{-1} + A\Lambda^{-1}A^T - A\Lambda^{-1}\Lambda\Lambda^{-1}A^T = L^{-1}$$

Hence, $p(y|x) = \mathcal{N}(Ax + b, L^{-1})$.

Question 3

The log likelihood function is:

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

Taking the derivative with respect to Σ, we have:

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

Setting the derivative to zero, we have:

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

Proof.

Symmetry:

$$\begin{aligned}
\Sigma^T &= \left(\frac{1}{N}\sum_{n=1}^{N}(x_n - \mu)(x_n - \mu)^T\right)^T \\
&= \frac{1}{N}\sum_{n=1}^{N}\left((x_n - \mu)(x_n - \mu)^T\right)^T \\
&= \frac{1}{N}\sum_{n=1}^{N}(x_n - \mu)(x_n - \mu)^T \\
&= \Sigma
\end{aligned}$$

Positive semi-definite: For any vector z, we have:

$$
\begin{aligned}
z^T \Sigma z &= z^T \left(\frac{1}{N} \sum_{n=1}^{N} (x_n - \mu)(x_n - \mu)^T \right) z \\
&= \frac{1}{N} \sum_{n=1}^{N} z^T (x_n - \mu)(x_n - \mu)^T z \\
&= \frac{1}{N} \sum_{n=1}^{N} (z^T (x_n - \mu))^2 \\
&\geq 0
\end{aligned}
$$

□

Question 4

(a)

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

The maximum likelihood estimation of σ^2 is the value that maximizes the log likelihood function:

$$
\frac{\partial}{\partial \sigma_{ML}^2} \frac{1}{N} \sum_{n=1}^{N} \ln p(x_n | \mu, \sigma^2) = 0
$$

Swap the derivative and the sum, we have:

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

Hence, the estimation of σ^2 is:

$$
(\sigma_{ML}^2)^{(N)} = (\sigma_{ML}^2)^{(N-1)} + a_{N-1} \frac{\partial}{\partial (\sigma_{ML}^2)^{(N-1)}} \ln p(x_N | \mu_{ML}, (\sigma_{ML}^2)^{(N-1)})
$$

Let this equal to the first equation, we have:

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

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

(b)

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

Same as part (a), we find the sequential form of Σ_{ML}:

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

Let this equal to the first equation, we have:

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

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

Question 5

From Bayes' theorem, we have:

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

Since both the prior and likelihood are Gaussian, the posterior is also Gaussian. We now focus on the exponential term of the right-hand side of the equation:

$$\begin{aligned}
p(\mu)\prod_{n=1}^{N}p(x_n|\mu) &= K\exp\left(-\frac{1}{2}(\mu-\mu_0)^T\Sigma_0^{-1}(\mu-\mu_0) - \frac{1}{2}\sum_{n=1}^{N}(x_n-\mu)^T\Sigma^{-1}(x_n-\mu)\right) \\
&= K\exp\left(-\frac{1}{2}\mu^T(\Sigma_0^{-1}+N\Sigma^{-1})\mu + \mu^T(\Sigma_0^{-1}\mu_0 + \Sigma^{-1}\sum_{n=1}^{N}x_n) + \text{const}\right)
\end{aligned}$$

where *const* is the terms that do not contain μ.
Thus, the posterior is Gaussian with mean and covariance given by:

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

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

Hence, the posterior is:

$$\begin{aligned}
p(\mu|X) &= \mathcal{N}(\mu_N, \Sigma_N) \\
&= \mathcal{N}\left((\Sigma_0^{-1}+N\Sigma^{-1})^{-1}(\Sigma_0^{-1}\mu_0 + \Sigma^{-1}N\mu_{ML}), (\Sigma_0^{-1}+N\Sigma^{-1})^{-1}\right)
\end{aligned}$$