Machine Learning HW 2

12012727 WANG Duolei

Q1. (a) True

(b) Consider the quadratic form of x_a, x_b:

$$\Delta^2 = \frac{1}{2}\begin{pmatrix} x_a^T - \mu_a^T & x_b^T - \mu_b^T \end{pmatrix}\begin{pmatrix} \Lambda_{aa} & \Lambda_{ab} \\ \Lambda_{ba} & \Lambda_{bb} \end{pmatrix}\begin{pmatrix} x_a - \mu_a \\ x_b - \mu_b \end{pmatrix}$$

And we know that the coefficient between x_a^T and x_a is the $\sum_{a|b}$, the coefficient of x^T is $_{a|b}$. Comparing the expansion of the Δ^2, we get:

$$\begin{cases} \sum_{a|b} = \Lambda_{aa}^{-1} \\ \mu_{a|b} = \mu_a - \Lambda_{aa}^{-1}\Lambda_{ab}(x_b - \mu_b) \end{cases}$$

And consider the inverse of $\sum$:

$$\begin{pmatrix} (A - BD^{-1}C)^{-1} & -(A - BD^{-1}C)^{-1}BD^{-1} \\ -DC(A - BD^{-1}C) & \cdots \end{pmatrix}$$

Thus, we get:

$$\begin{cases} \sum_{a|b} = \sum_{aa} - \sum_{ab}\sum_{bb}^{-1}\sum_{ba} \\ \mu_{a|b} = \mu_a - \sum_{ab}\sum_{bb}^{-1}(x_b - \mu_b) \end{cases}$$

Q2.

1. As z follows normal distribution, x also follows the normal distribution, we need to calculate the coefficient of the quadratic form and the coefficient of x^T:

$$\begin{aligned} \Delta^2 &= \frac{1}{2}\begin{pmatrix} x_a^T - \mu_a^T & x_b^T - \mu_b^T \end{pmatrix}\begin{pmatrix} \Lambda_{aa} & \Lambda_{ab} \\ \Lambda_{ba} & \Lambda_{bb} \end{pmatrix}\begin{pmatrix} x_a - \mu_a \\ x_b - \mu_b \end{pmatrix} \\ &= -\frac{1}{2}x_b^T\Lambda_{bb}x_b + x_b^T[\Lambda_{bb} - \mu_b - \Lambda_{ba}(x_a - \mu_a)] \\ &= -\frac{1}{2}x_b^T\Lambda_{bb}x_b + x_b^T m,\ where\ m := \Lambda_{bb} - \mu_b - \Lambda_{ba}(x_a - \mu_a) \end{aligned}$$

Consider the integral of the x_b part, transform the above equation into quadratic form:

$$-\frac{1}{2}(x_b - \Lambda_{bb}^{-1}m)\Lambda_{bb}(x_b - \Lambda_{bb}^{-1}m)$$

We can get after the integral, the quadratic containing x_a becomes:

$$-\frac{1}{2}x_a^T(\Lambda_{aa} - \Lambda_{ab}\Lambda_{bb}^{-1}\Lambda_{ba})x_a + x_a^T(\Lambda_{aa} - \Lambda_{ab}\Lambda_{bb}^{-1}\Lambda_{ba})\mu_a$$

Comparing the coefficient, we get:

$$\begin{cases} \sum_a = (\Lambda_{aa} - \Lambda_{ab}\Lambda_{ab}^{-1}\Lambda_{ab})^{-1} \\ \mu_a = \sum_a(\Lambda_{aa} - \Lambda_{ab}\Lambda_{ab}^{-1}\Lambda_{ab})^{-1}\mu_a = \mu_a \end{cases}$$

And considet the inverse of $\sum$, we get:

$$\begin{cases} \sum_a = \sum_{aa} = \mathbf{\Lambda}^{-1} \\ \mu_a = \mu_a \end{cases}$$

which implies that $x \sim \mathcal{N}(\mu_a, \mathbf{\Lambda}^{-1})$.

2. Firstly, we can get the inverse Λ of the covariance matrix $\mathbf{R}$:

$$\mathbf{R} = \begin{pmatrix} \mathbf{\Lambda} + \mathbf{A^T L A} & -\mathbf{A^T L} \\ -\mathbf{LA} & \mathbf{L} \end{pmatrix}$$

And according to Q1, we get:

$$y|x \sim \mathcal{N}(y|\mu_{y|x}, \sum y|x)$$

where:

$$\begin{cases} \mu_{y|x} = \mu_y - \mathbf{L^{-1}}(-\mathbf{LA})(x-\mu) = \mathbf{A}\mu + \mathbf{b} + \mathbf{A}x - \mathbf{A}\mu = \mathbf{A}x + \mathbf{b} \\ \sum_{y|x} = \mathbf{L}^{-1} \end{cases}$$

Q3. Derivative the function by A and let it be zero, we get the derivation form:

$$\sum_{k=1}^{N} (A^{-1}(x_n - \mu))(A^{-1}(x_n - \mu))^T - N(A^{-1})^T = 0$$

Thus:

$$[A^{-1}(\sum_{k=1}^{N}(x_n - \mu)(x_n - \mu)^T) - N](A^{-1})^T = 0$$

We get:

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

Q4. (a)

$$\begin{aligned} {\sigma^2_{ML}}^{(N)} &= \frac{1}{N}\sum_{k=1}^{N}(x_n - \mu)^2 \\ &= \frac{1}{N}[(N-1)\sigma^2_{ML}{}^(N-1) + (x_n - \mu)^2] \\ &= {\sigma^2_{ML}}^{(N-1)} + \frac{1}{N}[(x_n - \mu)^2 - {\sigma^2_{ML}}^{(N-1)}] \end{aligned}$$

Using the Robbins-Monro method, we get the expression of z:

$$\frac{\partial \ln p}{\partial(\sigma^2)} = \frac{1}{2\sigma^2}[\frac{(x-\mu)^2}{\sigma^2} - 1]$$

Substitute the form into the formula, and choose $\alpha_N = \frac{2\sigma^4}{N}$:

$${\sigma^2_{ML}}^{(N)} = \sigma_{ML}^{(N-1)} + \frac{2\sigma^4}{N}\frac{1}{2\sigma^2}[\frac{(x-\mu)^2}{\sigma^2} - 1] = \sigma_{ML}^{(N-1)} + \frac{1}{N}[(x_n - \mu)^2 - {\sigma^2_{ML}}^{(N-1)}]$$

(b) Using the symbol A to replace covariance matrix $\sum$.

$$\begin{aligned} A^{(N)} &= \frac{1}{N}\sum_{k=1}^{N}(x_k - \mu)(x_k - \mu)^T \\ &= \frac{1}{N}[(N-1)A^{(N-1)} + (x_N - \mu)(x_N - \mu)^T] \\ &= A^{(N-1)} + \frac{1}{N}[(x_N - \mu)(x_N - \mu)^T - A^{(N-1)}] \end{aligned}$$

As the same, we get:

$$\frac{\ln(p)}{\partial A} = -\frac{A^{-1}}{2}[1 - (x - \mu_{ML})(x - \mu_{ML})^T A^{-1}]$$

Then:

$$A^{(N)} = A^{(N-1)} + \alpha_N \frac{A^{-1}}{2}[(x_N - \mu_{ML})(x_N - \mu_{ML})^T - A^{(N-1)}]$$

Let $\alpha_N = \frac{2A}{N}$, we get the same result.

Q5. According to the formula given by the text:

$$\mu_N = \frac{\sigma^2}{N\sigma_0^2 + \sigma^2}\mu_0 + \frac{N\sigma_0^2}{N\sigma_0^2 + \sigma^2}\mu_{ML},\ \frac{1}{\sigma_N^2} = \frac{1}{\sigma_0^2} + \frac{N}{\sigma^2}$$

Thus:

$$\sigma_N = \frac{\sigma\sigma_0}{\sqrt{\sigma^2 + N\sigma_0^2}}$$

where $\sigma = \sqrt{\sum},\ \sigma_0 = \sqrt{\sum_0}, \mu_{ML} = \frac{1}{N}\sum_{n=1}^{N} x_n$. And $p(\mu|\mathbf{X}) = \mathcal{N}(\mu|\mu_N, \sigma_N^2)$.

We get:

$$p(\mu_N|X) = \frac{1}{\sqrt{2\pi}\sigma_N}\exp\{-\frac{(x-\mu_N)^2}{2\sigma_N^2}\}$$

with $\mu_N = \frac{1}{N}\sum_{n=1}^{N} x_n,\ \sigma_N = \frac{\sigma\sigma_0}{\sqrt{\sigma^2+N\sigma_0^2}}$.