Homework 2
M1522.001300 Probabilistic Graphical Models (2016 Fall)

2015-21259 Hyunwoo Lee

Date: October 6 Thursday

1 Variable Elimination

1.1 Example

Consider the network demonstrated in Figure 1. The chain fule for this newtork asserts that

$$\begin{aligned} P(C,D,I,G,S,L,J,H) &= P(C)P(D|C)P(I)P(G|I,D)P(S|I) \\ &\quad P(L|G)P(J|L,S)P(H|G,J) \\ &= \phi_C(C)\phi_D(D,C)\phi_I(I)\phi_G(G,I,D) \\ &\quad \phi_S(S,I)\phi_L(L,G)\phi_J(J,L,S)\phi_H(H,G,J) \end{aligned}$$

Apply variable elimination algorithm to compute $P(J)$ by using elimination ordering: G, I, S, L, H, C, D. Show VE step by step and fill the Table 1.

The algorithm of the variable elimination is as follows:

1. Pick a variable ordering with the query variable at the end of the list
2. Initialize the active factor list
3. Introduce the evidence by conditioning
4. Repeat for $i = 1$ to n
   (a) Take the next variable X_i from the ordering
   (b) Take all the factors that have X_i as an argument out of the active factor list, and multiply them, then sum over all values of X_i, creating a new factor m_{X_i}
   (c) Put m_{X_i} on the active factor list

First, we select the ordering as G, I, S, L, H, C, D with the query variable J at the end: $\{G, I, S, L, H, C, D, J\}$.

Second, the initialization of the active factor list is as follows:

$$\{P(C), P(D|C), P(I), P(G|I,D), P(S|I), P(L|G), P(J|L,S), P(H|G,J)\}$$

Third, because there is no evidence for query $P(J)$, we can skip the conditioning.

Fourth, now let us see the step by step of the operation.

1. Eliminate G
   The factors related to G is $\{P(G|I,D), P(L|G), P(H|G,J)\}$.
   The variables involved are $\{G, D, I, L, H, J\}$.
   The new term $m_G(I,D,L,H,J)$ is

$$m_G(I, D, L, H, J) = \sum_G P(G|I, D)P(L|G)P(H|G, J)$$

Now, the active factor list becomes

$$\{P(C), P(D|C), P(I), P(S|I), P(J|L, S), m_G(D, I, L, H, J)\}.$$

2. Eliminate I
   The factors related to I is $\{P(I), P(S|I), m_G(D, I, L, H, J)\}$.
   The variables involved are $\{D, S, I, L, H, J\}$.
   The new term $m_I(D, S, L, H, J)$ is

   $$m_I(D, S, L, H, J) = \sum_I P(I)P(S|I)m_G(D, I, L, H, J)$$

   Now, the active factor list becomes

   $$\{P(C), P(D|C), P(J|L, S), P(S|I), m_I(D, S, L, H, J)\}.$$

3. Eliminate S
   The factors related to S is $\{P(J|L, S), m_I(D, S, L, H, J)\}$.
   The variables involved are $\{J, L, S, D, H\}$.
   The new term $m_S(J, L, D, H)$ is

   $$m_S(J, L, D, H) = \sum_I P(J|L, S)m_I(D, S, L, H, J)$$

   Now, the active factor list becomes $\{P(C), P(D|C), m_S(J, L, D, H)\}$.

4. Eliminate L
   The factors related to L is $\{m_S(J, L, D, H)\}$.
   The variables involved are $\{J, L, D, H\}$.
   The new term $m_L(J, D, H)$ is

   $$m_L(J, D, H) = \sum_l m_S(J, L, D, H)$$

   Now, the active factor list becomes $\{P(C), P(D|C), m_L(J, D, H)\}$.

5. Eliminate H
   The factors related to H is $\{m_L(J, D, H)\}$.
   The variables involved are $\{J, D, H\}$.
   The new term $m_H(J, D)$ is

   $$m_H(J, D) = \sum_H m_L(J, D, H)$$

   Now, the active factor list becomes $\{P(C), P(D|C), m_H(J, D)\}$.

6. Eliminate C
   The factors related to C is $\{P(C), P(D|C)\}$.
   The variables involved are $\{C, D\}$.
   The new term $m_C(D)$ is

   $$m_C(D) = \sum_C P(C)P(D|C) = P(D)$$

Now, the active factor list becomes $\{P(D), m_H(J,D)\}$.

7. Eliminate D
   The factors related to D is $\{P(D), m_H(J,D)\}$.
   The variables involved are $\{J, D\}$.
   The new term $m_D(J)$ is
   $$m_D(J) = \sum_D P(D)m_H(J,D) = P(J)$$
   Now, we get $P(J) = \sum_D P(D)m_H(J,D)$.

We can summarize the step by step in the form of the table as follows:

Step	Variable eliminated	Factors used	Variables involved	New factor
1	G	$P(G\|I,D), P(L\|G), P(H\|G,J)$	G,D,I,L,H,J	$m_G(I,D,L,H,J)$
2	I	$P(I), P(S\|I), m_G(D,I,L,H,J)$	D,S,I,L,H,J	$m_L(D,S,L,H,J)$
3	S	$P(J\|L,S), m_I(D,S,L,H,J)$	J,L,S,D,H	$m_S(J,L,D,H)$
4	L	$m_S(J,L,D,H$	J,L,D,H	$m_L(J,D,H)$
5	H	$m_L(J,D,H)$	J,D,H	$m_H(J,D)$
6	C	$P(C), P(D\|C)$	C,D	$m_C(D)$
7	D	$P(D), m_H(J,D)$	J,D	$m_D(J)$

1.2 Variable Elimination in Clique Tree

Consider a chain graphical model with the structure $X_1 - X_2 - \cdots - X_n$, where each X_i take on one of d possible assignments. You can form the following clique tree for this GM: $\mathrm{C}_1 - \mathrm{C}_2, \cdots, \mathrm{C}_{n-1}$, where Scope$[\mathrm{C}_1] = \{X_i, X_{i+1}\}$. You can assume that this clique tree has already been calibrated. Using this clique tree, we can directly obtain $P(X_i, X_{i+1})$. Your goal in this question is to compute $P(X_i, X_j)$, for any $j > i$.

1. Briefly, describe how variable elimination can be used to compute $P(X_i, X_j)$, for some $j > i$, in linear time, given the calibrated clique tree.

   Because we assume the calibrated clique tree, we can see a calibrated clique tree as a distribution.
   So, we have $P(C_i) = \beta_i(C_i)$, where C_i is the clique involving the nodes. Before, finding the value of $P(X_1, X_2, \cdots X_n)$, let us think about $P(X_1, X_2, X_3)$.
   $$P(X_1, X_2, X_3) = P(X_1, X_2)P(X_3|X_2)$$
   And,
   $$P(X_3|X_2) = \frac{P(X_2,X_3)}{P(X_2)} = \frac{\beta(X_2,X_3)}{\sum_{X_2}\beta(X_2,X_3)}$$
   So,
   $$P(X_1, X_2, X_3) = P(X_1, X_2)P(X_3|X_2) = P(X_1, X_2)\frac{\beta(X_2,X_3)}{\sum_{X_2}\beta(X_2,X_3)}$$
   We can generalize it to $P(X_1, X_2, \cdots X_n)$ as,

$$\begin{aligned} P(X_1, X_2, \cdots X_n) &= P(X_1, X_2) \prod_{i=2}^{n-1} P(X_{i+1}|X_i) \\ &= P(X_1, X_2) \prod_{i=2}^{n-1} \frac{\beta(X_i, X_{i+1})}{\sum_{X_{i+1}} \beta(X_i, X_{i+1})} \\ &= \beta(X_1, X_2) \prod_{i=2}^{n-1} \frac{\beta(X_i, X_{i+1})}{\sum_{X_{i+1}} \beta(X_i, X_{i+1})} \end{aligned}$$

Now, we can sum up the unnecessary variables.

$$P(X_1, X_n) = \sum_{X_{n-1}} \sum_{X_{n-2}} \cdots \sum_{X_2} \beta(X_1, X_2) \prod_{i=2}^{n-1} \frac{\beta(X_i, X_{i+1})}{\sum_{X_i} \beta(X_i, X_{i+1})}$$

2. What is the running time of the algorithm in part one? If you wanted to compute $P(X_i, X_j)$ for all n choose 2 choices i and j?

   Let say there needs time d to sum up it by one variable.
   To make the denominator in the each factors, it needs $d \times (n-2) = O(n)$
   For summing up, it needs additional $d \times (n-2) = O(n)$

   Because it depends only on the distance between two nodes, the running time is same.

3. Consider a particular chain $X_1 - X_2 - X_3 - X_4$. Show that by caching $P(X_1, X_3)$, you can compute $P(X_1, X_4)$ more efficiently than directly applying variable elimination as described in part 1.2.1.

   We can get $P(X_1, X_4)$ from $P(X_1, X_3)$ by the following steps.

$$\begin{aligned} P(X_1, X_4) &= \sum_{X_3} P(X_1, X_3) P(X_4|X_3) \\ &= \sum_{X_3} \beta(X_1, X_3) \frac{\beta(X_3, X_4)}{\sum_{X_4} \beta(X_3, X_4)} \end{aligned}$$

4. Using the intuition in part three, design a dynamic programming algorithm (caching partial results) which computes $P(X_i, X_j)$ for all n choose 2 choices of i and j in time asymptotically much lower than the complexity you described in part 1.2.2. What is the asymptotic running time of your algorithm?

   We can generalize the method in the part 3.
   When we want $P(X_i, X_j)$, where $i < j$, first,

$$\begin{aligned} P(X_i, X_{i+2}) &= \sum_{X_{i+1}} P(X_i, X_{i+1}) P(X_{i+2}|X_{i+1}) \\ &= \sum_{X_{i+1}} \beta(X_i, X_{i+1}) \frac{\beta(X_{i+1}, X_{i+2})}{\sum_{X_{i+2}} \beta(X_{i+1}, X_{i+2})} \end{aligned}$$

   Then, let's cache the result and repeat the process until we get $P(X_i, X_j)$.

   Let say there needs time d to sum up it by one variable.
   Then, first it needs d to make denominator in $\frac{\beta(X_i, X_{i+1})}{\sum_{X_i} \beta(X_i, X_{i+1})}$
   And to sum up there needs another d.
   So, for the distance added one from X_1, it needs $2d$.
   That is, to find $P(X_1, X_n)$ from $P(X_1, X_2)$, it need $2d \times n = O(n)$ It depends only on the distance between two nodes, for $P(X_i, X_j)$ where $i < j$, it needs $2d \times (j - i - 1) = O(n)$.

1.3 Chains or trees

Discuss whether true for false: the complexity of variable elimination is the same in graphical models that are chains or trees. If true, discuss why it is true. If false, discuss why it is false.

"True" if the execution is one by one.
Let c be the execution cost of message passing in a clique tree.
In view of message passing as a variable elimination process, each node only needs $2c$ for making the messages.
Mainly, the complexity depends on the number of nodes, when we assume the making of the message is operated on one by one.
But if we can execute parallally, in case of tree, the nodes in the same level (i.e. the same number of path to the root) can be executed parallally which makes the overall execution time faster.
In this case, we can expect $O(logn)$, and the answer is "False".

1.4 Implementation of Variable Elimination

(Solved separately)

2 Exact Inference

2.1 Message passing on a tree

Consider the DGM in Figure 3 which represents the following fictitious biological model. Each G_i represents the genotype of a person: $G_i = 1$ if they have a healthy gene and $G_i = 2$ if they have an unhealthy gene. G_2 and G_3 may inherit the unhealthy gene from their parent G_1. $X_i \in R$ is a continuous measure of blood pressure, which is low if you are healthy and high if you are unhealthy. We define the CPDs as follows:

$$
\begin{aligned}
p(G_1) &= [0.5, 0.5] \\
p(G_2|G_1) &= \begin{pmatrix} 0.9 & 0.1 \\ 0.1 & 0.9 \end{pmatrix} \\
p(G_3|G_1) &= \begin{pmatrix} 0.9 & 0.1 \\ 0.1 & 0.9 \end{pmatrix} \\
p(X_i|G_i = 1) &= \mathcal{N}(X_i|\mu = 50, \sigma^2 = 10) \\
p(X_i|G_i = 2) &= \mathcal{N}(X_i|\mu = 60, \sigma^2 = 10)
\end{aligned}
$$

Before starting to solve the problems, by marginalization, we get

$$
\begin{aligned}
p(G_2) &= \textstyle\sum_{G_1} p(G_1, G_2) \\
&= \textstyle\sum_{G_1} p(G_1)\cdot p(G_2|G_1) \\
&= p(G_2|G_1=1)p(G_1=1) + p(G_2|G_1=2)p(G_1=2) \\
&= \begin{pmatrix}0.9 & 0.1\end{pmatrix} 0.5 + \begin{pmatrix}0.1 & 0.9\end{pmatrix} 0.5 \\
&= \begin{pmatrix}0.5 & 0.5\end{pmatrix} \\
p(G_3) &= \begin{pmatrix}0.5 & 0.5\end{pmatrix} \\
p(X_i) &= \textstyle\sum_{G_i} p(X_i, G_i) \\
&= \textstyle\sum_{G_i} p(X_i|G_i)p(G_i) \\
&= p(X_i|G_i=1)p(G_i=1) + p(X_i|G_i=2)p(G_i=2) \\
&= 0.5\cdot\mathcal{N}(X_i|\mu=50, \sigma^2=10) + 0.5\cdot\mathcal{N}(X_i|\mu=60, \sigma^2=10) \\
&\qquad (\because \text{Every } p(G_i) = \begin{pmatrix}0.5 & 0.5\end{pmatrix})
\end{aligned}
$$

1. Suppose you observe $X_2 = 50$, and X_1 is unobserved. What is the posterior belief on G_1, i.e., $p(G_1|X_2 = 50)$?

   By Bayesian Theorem, $p(G_1|X_2 = 50) = \frac{p(X_2=50|G_1)\cdot p(G_1)}{p(X_2=50)}$

   1) $G_1 = 1$

      By the marginalization,

$$
\begin{aligned}
p(X_2, G_1) &= \textstyle\sum_{G_3}\sum_{G_2}\sum_{X_3}\sum_{X_1} p(G_1, G_2, G_3, X_1, X_2, X_3) \\
&= \textstyle\sum_{G_3}\sum_{G_2}\sum_{X_3}\sum_{X_1} p(G_1)p(X_1|G_1)p(G_2|G_1)p(G_3|G_1)p(X_2|G_2)p(X_3|G_3) \\
&= p(G_1)\textstyle\sum_{G_2} p(G_2|G_1)p(X_2|G_2)
\end{aligned}
$$

$$
\begin{aligned}
p(G_1=1|X_2=50) &= \tfrac{p(X_2=50,G_1=1)}{p(X_2=50)} \\
&= \tfrac{p(G_1=1)\sum_{G_2} p(G_2|G_1=1)p(X_2=50|G_2)}{p(X_2=50)} \\
&= \tfrac{p(G_1=1)(p(G_2=1|G_1=1)p(X_2=50|G_2=1)+p(G_2=2|G_1=1)p(X_2=50|G_2=2))}{p(X_2=50)} \\
&= \tfrac{0.5\cdot 0.9\cdot\mathcal{N}(X_i=50|\mu=50,\sigma^2=10)+0.5\cdot 0.1\cdot\mathcal{N}(X_i=50|\mu=60,\sigma^2=10)}{0.5\cdot\mathcal{N}(X_i=50|\mu=50,\sigma^2=10)+0.5\cdot\mathcal{N}(X_i=50|\mu=60,\sigma^2=10)} \\
&= 0.8946457192609383
\end{aligned}
$$

   2) $G_1 = 2$

$$
\begin{aligned}
p(G_1=1|X_2=50) &= \tfrac{p(G_1=2)(p(G_2=1|G_1=2)p(X_2=50|G_2=1)+p(G_2=2|G_1=2)p(X_2=50|G_2=2))}{p(X_2=50)} \\
&= \tfrac{0.5\cdot 0.1\cdot\mathcal{N}(X_i=50|\mu=50,\sigma^2=10)+0.5\cdot 0.9\cdot\mathcal{N}(X_i=50|\mu=60,\sigma^2=10)}{0.5\cdot\mathcal{N}(X_i=50|\mu=50,\sigma^2=10)+0.5\cdot\mathcal{N}(X_i=50|\mu=60,\sigma^2=10)} \\
&= 0.10535428073947115
\end{aligned}
$$

   Now, we get $P(G_1|X_2 = 50) = \begin{pmatrix}0.8946457192609383 & 0.10535428073947115\end{pmatrix}$

2. Now suppose you observe $X_2 = 50$ and $X_3 = 50$. What is $p(G_1|X_2, X_3)$? Explain your answer intuitively

   By the marginalization,

$$
\begin{aligned}
p(G_1, X_2, X_3) &= \textstyle\sum_{X_1}\sum_{G_2}\sum_{G_3} p(G_1)p(X_1|G_1)p(G_2|G_1)p(G_3|G_1)p(X_2|G_2)p(X_3|G_3) \\
&= \textstyle\sum_{G_2}\sum_{G_3} p(G_1)p(G_2|G_1)p(G_3|G_1)p(X_2|G_2)p(X_3|G_3) \\
&= p(G_1)(p(G_3=1|G_1)p(X_3|G_3=1) + p(G_3=2|G_1)p(X_3|G_3=2)) \\
&\quad \times(p(G_2=1|G_1)p(X_2|G_2=1) + p(G_2=2|G_1)p(X_2|G_3=2))
\end{aligned}
$$

When $G_1 = 1$,

$$
\begin{aligned}
&p(G_1 = 1, X_2, X_3) \\
&\quad = p(G_1 = 1)(p(G_3 = 1|G_1 = 1)p(X_3|G_3 = 1) + p(G_3 = 2|G_1 = 1)p(X_3|G_3 = 2)) \\
&\quad\quad \times(p(G_2 = 1|G_1 = 1)p(X_2|G_2 = 1) + p(G_2 = 2|G_1 = 1)p(X_2|G_3 = 2)) \\
&\quad = 0.5 \cdot (0.9 \cdot \mathcal{N}(X_3|\mu = 50, \sigma^2 = 10) + 0.1 \cdot \mathcal{N}(X_3|\mu = 60, \sigma^2 = 10)) \\
&\quad\quad \times(0.9 \cdot \mathcal{N}(X_2|\mu = 50, \sigma^2 = 10) + 0.1 \cdot \mathcal{N}(X_2|\mu = 60, \sigma^2 = 10))
\end{aligned}
$$

When $G_1 = 2$,

$$
\begin{aligned}
&p(G_1 = 2, X_2, X_3) \\
&\quad = p(G_1 = 2)(p(G_3 = 1|G_1 = 2)p(X_3|G_3 = 1) + p(G_3 = 2|G_1 = 2)p(X_3|G_3 = 2)) \\
&\quad\quad \times(p(G_2 = 1|G_1 = 2)p(X_2|G_2 = 1) + p(G_2 = 2|G_1 = 2)p(X_2|G_3 = 2)) \\
&\quad = 0.5 \cdot (0.1 \cdot \mathcal{N}(X_3|\mu = 50, \sigma^2 = 10) + 0.9 \cdot \mathcal{N}(X_3|\mu = 60, \sigma^2 = 10)) \\
&\quad\quad \times(0.1 \cdot \mathcal{N}(X_2|\mu = 50, \sigma^2 = 10) + 0.9 \cdot \mathcal{N}(X_2|\mu = 60, \sigma^2 = 10))
\end{aligned}
$$

So,

$$
p(G_1, X_2 = 50, X_3 = 50) = \begin{pmatrix} 0.00645543020617 & 0.0000895215074278 \end{pmatrix}
$$

And,

$$
p(X_2, X_3) = \textstyle\sum_{G_1} p(G_1, X_2, X_3)
$$

Therefore,

$$
\begin{aligned}
p(G_1|X_2 = 50, X_3 = 50) &= \tfrac{p(G_1, X_2=50, X_3=50)}{p(X_2=50, X_3=50)} \\
&= \begin{pmatrix} 0.986322052271 & 0.0136779477291 \end{pmatrix}
\end{aligned}
$$

Intuitively, because X_2 and X_3 are healthy, it is natural that we may believe G_1 may be healthy rather than unhealthy.

3. Now suppose $X_2 = 60$, $X_3 = 60$. What is $p(G_1|X_2, X_3)$? Explain your answer intuitively.

$$
\begin{aligned}
p(G_1|X_2 = 60, X_3 = 60) &= \tfrac{p(X_2=60, X_3=60, G_1)}{p(X_2=60, X_3=60)} \\
&= \begin{pmatrix} 0.0136779477291 & 0.986322052271 \end{pmatrix}
\end{aligned}
$$

Intuitively, because X_2 and X_3 are unhealthy, it is natural that we may believe G_1 may be unhealthy rather than healthy.

4. Now suppose $X_2 = 50$, $X_3 = 60$. What is $p(G_1|X_2, X_3)$? Explain your answer intuitively.

$$
\begin{aligned}
p(G_1|X_2 = 50, X_3 = 60) &= \tfrac{p(X_2=50, X_3=60, G_1)}{p(X_2=50, X_3=60)} \\
&= \begin{pmatrix} 0.5 & 0.5 \end{pmatrix}
\end{aligned}
$$

Intuitively, because X_2 is healthy and X_3 is unhealthy, it is hard to infer the status of G_1. The belief may be similar.

2.2 Gaussian Belief Propagation

If we set the model as jointly Gaussian, all marginals and all messages will be Gaussian. The key operations we need are to multiply together two Gaussian factors, and to marginalize out a variable from a joint Gaussian factor. For multiplication, we can use the fact that the product of two Gaussians is Gaussian:

$$\begin{aligned}\mathcal{N}(x|\mu_1, \lambda_1^{-1}) \times \mathcal{N}(x|\mu_2, \lambda_2^{-1}) &= C\mathcal{N}(x|\mu, \lambda^{-1}) \\ \lambda &= \lambda_1 + \lambda_2 \\ \mu &= \lambda^{-1}(\mu_1\lambda_1 + \mu_2\lambda_2)\end{aligned}$$

where

$$C = \sqrt{\frac{\lambda}{\lambda_1\lambda_2}} exp(\tfrac{1}{2}(\lambda_1\mu_1^2(\lambda^{-1}\lambda_1 - 1) + \lambda_2\mu_2^2(\lambda^{-1}\lambda_2 - 1) + 2\lambda^{-1}\lambda_1\lambda_2\mu_1\mu_2))$$

So, in this problem, prove the above equation. (Hint: use completing the square.)

By the definition of the Gaussian (normal) distribution,

$$\mathcal{N}(x|\mu, \sigma^2) = \frac{1}{\sqrt{2\sigma^2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

With the definition, lets find the product of two Gaussians,

$$\mathcal{N}(x|\mu_1, \lambda_1^{-1}) \times \mathcal{N}(x|\mu_2, \lambda_2^{-1}) = \frac{1}{\sqrt{2\lambda_1^{-1}\pi}} e^{-\frac{(x-\mu_1)^2}{2\lambda_1^{-1}}} \times \frac{1}{\sqrt{2\lambda_2^{-1}\pi}} e^{-\frac{(x-\mu_2)}{2\lambda_2^{-1}}}$$

$$= \frac{1}{2\pi\sqrt{\lambda_1^{-1}\lambda_2^{-1}}} e^{-(\frac{(x-\mu_1)^2}{2\lambda_1^{-1}} + \frac{(x-\mu_2)^2}{2\lambda_2^{-1}})}$$

$$= \frac{1}{2\pi\sqrt{\lambda_1^{-1}\lambda_2^{-1}}} e^{-\frac{1}{2}(\frac{(x-\mu_1)^2}{\lambda_1^{-1}} + \frac{(x-\mu_2)^2}{\lambda_2^{-1}})}$$

$$= \frac{1}{2\pi\sqrt{\lambda_1^{-1}\lambda_2^{-1}}} e^{-\frac{1}{2}(\frac{\lambda_2^{-1}(x-\mu_1)^2 + \lambda_1^{-1}(x-\mu_2)^2}{\lambda_1^{-1}\lambda_2^{-1}})}$$

$$= \frac{1}{2\pi\sqrt{\lambda_1^{-1}\lambda_2^{-1}}} e^{-\frac{1}{2}(\frac{(\lambda_1^{-1}+\lambda_2^{-1})x^2 - 2(\lambda_2^{-1}\mu_1 + \lambda_1^{-1}\mu_2)x + \lambda_2^{-1}\mu_1^2 + \lambda_1^{-1}\mu_2^2}{\lambda_1^{-1}\lambda_2^{-1}})}$$

$$= \frac{1}{2\pi\sqrt{\lambda_1^{-1}\lambda_2^{-1}}} e^{-\frac{1}{2}\frac{(\lambda_1^{-1}+\lambda_2^{-1})}{\lambda_1^{-1}\lambda_2^{-1}}(x^2 - 2\frac{(\lambda_2^{-1}\mu_1 + \lambda_1^{-1}\mu_2)}{(\lambda_1^{-1}+\lambda_2^{-1})}x + \frac{\lambda_2^{-1}\mu_1^2 + \lambda_1^{-1}\mu_2^2}{(\lambda_1^{-1}+\lambda_2^{-1})})}$$

$$= \frac{1}{2\pi\sqrt{\lambda_1^{-1}\lambda_2^{-1}}} e^{-\frac{1}{2}(\frac{1}{\lambda_1^{-1}} + \frac{1}{\lambda_2^{-1}})((x - \frac{\lambda_1\mu_1 + \lambda_2\mu_2}{\lambda_1+\lambda_2})^2 - (\frac{\lambda_1\mu_1+\lambda_2\mu_2}{\lambda_1+\lambda_2})^2 + \frac{\lambda_1\mu_1^2+\lambda_2\mu_2^2}{\lambda_1+\lambda_2})}$$

$$= \frac{1}{2\pi\sqrt{\lambda_1^{-1}\lambda_2^{-1}}} e^{-\frac{1}{2}(\lambda_1+\lambda_2)((x - \frac{\lambda_1\mu_1+\lambda_2\mu_2}{\lambda_1+\lambda_2})^2 - \frac{\lambda_1^2\mu_1^2 + 2\lambda_1\mu_1\lambda_2\mu_2 + \lambda_2^2\mu_2^2}{(\lambda_1+\lambda_2)^2} + \frac{(\lambda_1+\lambda_2)(\lambda_1\mu_1^2+\lambda_2\mu_2^2)}{(\lambda_1+\lambda_2)^2})}$$

$$= \frac{1}{2\pi\sqrt{\lambda_1^{-1}\lambda_2^{-1}}} e^{-\frac{1}{2}(\frac{1}{(\lambda_1+\lambda_2)^{-1}})((x - \frac{\lambda_1\mu_1+\lambda_2\mu_2}{\lambda_1+\lambda_2})^2 - \frac{\lambda_1^2\mu_1^2 + 2\lambda_1\mu_1\lambda_2\mu_2 + \lambda_2^2\mu_2^2 - (\lambda_1+\lambda_2)\lambda_1\mu_1^2 - (\lambda_1+\lambda_2)\lambda_2\mu_2^2}{(\lambda_1+\lambda_2)^2})}$$

$$= \frac{1}{\sqrt{2(\lambda_1+\lambda_2)^{-1}\pi}} e^{-\frac{1}{2(\lambda_1+\lambda_2)^{-1}}(x-\frac{\lambda_1\mu_1+\lambda_2\mu_2}{\lambda_1+\lambda_2})^2} C_0 e^{\frac{1}{2}(\frac{\lambda_1^2\mu_1^2+2\lambda_1\mu_1\lambda_2\mu_2+\lambda_2^2\mu_2^2-(\lambda_1+\lambda_2)\lambda_1\mu_1^2-(\lambda_1+\lambda_2)\lambda_2\mu_2^2}{\lambda_1+\lambda_2})}$$

$$= \frac{1}{\sqrt{2\lambda^{-1}\pi}} e^{-\frac{1}{2\lambda_1^{-1}}(x-\lambda^{-1}(\lambda_1\mu_1+\lambda_2\mu_2))^2} C_0 e^{\frac{1}{2}((\lambda_1\mu_1^2(\lambda^{-1}\lambda_1-1)+\lambda_2\mu_2^2(\lambda^{-1}\lambda_2-1)+2\lambda^{-1}\lambda_1\lambda_2\mu_1\mu_2)}$$

$$= C\mathcal{N}(x|\mu, \lambda^{-1})$$

where,

$$\begin{array}{rcl} \lambda & = & \lambda_1 + \lambda_2 \\ \mu & = & \lambda^{-1}(\mu_1\lambda_1 + \mu_2\lambda 2) \\ C & = & C_0 e^{\frac{1}{2}((\lambda_1\mu_1^2(\lambda^{-1}\lambda_1-1)+\lambda_2\mu_2^2(\lambda^{-1}\lambda_2-1)+2\lambda^{-1}\lambda_1\lambda_2\mu_1\mu_2)} \end{array}$$