Probabilistic Machine Learning

Exercise 5

Summer Term 2023

Full Name	Matriculation No
Batuhan Özcömlekci	6300476
Aakarsh Nair	6546577

Eberhards Karls University of Tübingen, June 4, 2023

1(a) Show that

$$N(x; a, A)\, N(x; b, A) = N(a; b, A+B)\, N(x; (A^{-1}+B^{-1})^{-1}(A^{-1}a + B^{-1}b), (A^{-1}+B^{-1})^{-1})$$

We start by showing that their multiplicative factors are equivlent

LHS

$$(2\pi)^{-D/2} \det(A)^{-1/2} * (2\pi)^{-D/2} \det(B)^{-1/2} = (2\pi)^{-D} \det(A)^{-1/2} \det(B)^{-1/2}$$

$$= (2\pi)^{-D} \det(AB)^{-1/2}$$

Usig $\det(A)\det(B) = \det(AB)$

RHS :

$$(2\pi)^{-D/2} \det(A+B)^{-1/2} (2\pi)^{-D/2} \det((A^{-1}+B^{-1})^{-1})^{-1/2}$$

$$= (2\pi)^{-D} \det(A+B)^{-1/2} \det((A^{-1}+B^{-1})^{-1})^{-1/2}$$

$$= (2\pi)^{-D} \det((A+B)(A^{-1}+B^{-1})^{-1})^{-1/2}$$

usig the Wood berry identity $(A^{-1}+B^{-1})^{-1} = A - A(A+B)^{-1}A = A(I-(A+B)^{-1}A)$

$$= (2\pi)^{-D} \left[\det((A+B)\, A\, (I-(A+B)^{-1}A))\right]^{-1/2}$$

$$= (2\pi)^{-D} \left[\det(A+B)\, \det(A)\, \det(I-(A+B)^{-1}A)\right]^{-1/2}$$

$$= (2\pi)^{-D} \det(A)^{-1/2} \left[\det((A+B) - (A+B)(A+B)^{-1}A)\right]^{-1/2}$$

$$= (2\pi)^{-D} \det(A)^{-1/2} \left[\det(A+B-A)\right]^{-1/2}$$

$$= (2\pi)^{-D} \det(A)^{-1/2} \det(B)^{-1/2}$$

$$= (2\pi)^{-D} \det(AB)^{-1/2}$$

Thus the multiplyig exponent in both RHS and LHS are equal.

We now compare the exponent term of LHS & RHS ignoring the $-1/2$ factor for simplicity

$$\text{LHS :} \quad (x-a)^T A^{-1} (x-a) + (x-b)^T B^{-1} (x-b)$$

opening the symmetric multiplicative terms we have.

$$(x-a)^T (A^{-1}x - A^{-1}a) + (x-b)^T (B^{-1}x - B^{-1}b)$$

$$\Rightarrow x^T A^{-1} x - x^T A^{-1} a - a^T A^{-1} x + a^T A^{-1} a + x^T B^{-1} x - x^T B^{-1} b - b^T B^{-1} x + b^T B^{-1} b$$

$\Rightarrow$ combining $x^T \underbrace{\cdot * \cdot}\, x$ terms using multiplicative & distributive laws.

$$\Rightarrow x^T (A^{-1} + B^{-1}) x - x^T (A^{-1}a + B^{-1}b) - (a^T A^{-1} + b^T B^{-1}) x + a^T A^{-1} a + b^T B^{-1} b$$

we know that A, B are positive semi-definite & symmetric as they are covariance matrices. $\Rightarrow (A^{-1})^T = A^{-1}$

$$\Rightarrow \underbrace{x^T (A^{-1} + B^{-1}) x}_{\text{LHS \#1}} - \underbrace{x^T (A^{-1}a + B^{-1}b)}_{\text{LHS \#2}} - \underbrace{(A^{-1}a + B^{-1}b)^T x}_{\text{LHS \#3}} + \underbrace{a^T A^{-1} a}_{\text{LHS \#4}} + \underbrace{b^T B^{-1} b}_{\text{LHS \#5}}$$

We label the terms and argue that they arise equivalently on RHS as well.

RHS : Consider the exponent term ignoring the common $-1/2$ factor of RHS we see. we have.

we consider the exponent arising from $N(a; b, A+B)$, with label RHS#1

RHS#1

$$(a-b)^T (A+B)^{-1} (a-b)$$

$$\Rightarrow (a-b)^T \left[(A+B)^{-1} a - (A+B)^{-1} b \right]$$

$$\Rightarrow \underbrace{a^T (A+B)^{-1} a}_{RHS\#1.1} - \underbrace{a^T (A+B)^{-1} b}_{RHS\#1.2} - \underbrace{b^T (A+B)^{-1} a}_{RHS\#1.3} + \underbrace{b^T (A+B)^{-1} b}_{RHS\#1.4}$$

Next we consider the terms arising in the exponent for

$$N\left(x ; (A^{-1}+B^{-1})^{-1} (A^{-1}a + B^{-1}b), (A^{-1}+B^{-1})^{-1}\right)$$

RHS#2

$$\left(x - (A^{-1}+B^{-1})^{-1} (A^{-1}a + B^{-1}b)\right)^T (A^{-1}+B^{-1}) \left(x - (A^{-1}+B^{-1})^{-1} (A^{-1}a + B^{-1}b)\right)$$

$$\Rightarrow \left(x - (A^{-1}+B^{-1})^{-1} (A^{-1}a + B^{-1}b)\right)^T \left((A^{-1}+B^{-1}) x - (A^{-1}a + B^{-1}b)\right)$$

$$\Rightarrow \overbrace{x^T (A^{-1}+B^{-1}) x}^{RHS\#2.1} - \overbrace{x^T (A^{-1}a + B^{-1}b)}^{RHS\#2.2} - \overbrace{(A^{-1}a + B^{-1}b)^T x}^{RHS\#2.3} +$$

$$+ \underbrace{(A^{-1}a + B^{-1}b)^T \left((A^{-1}+B^{-1})^{-1}\right)^T (A^{-1}a + B^{-1}b)}_{RHS\#2.4}$$

we note that RHS#2.1 = LHS#1

RHS#2.2 = LHS#2

RHS#2.3 = LHS#3

We first simplify RHS # 2.4

$$(A^{-1}a + B^{-1}b)^T \left((A^{-1}+B^{-1})^{-1}\right)^T (A^{-1}a + B^{-1}b)$$

① A^{-1} is symmetric positive definite
② B^{-1} is symmetric positive definite.
③ $(A^{-1}+B^{-1})$ is symmetric positive definite.
④ $\left((A^{-1}+B^{-1})^{-1}\right)^T = (A^{-1}+B^{-1})^{-1}$

$$\Rightarrow (A^{-1}a + B^{-1}b)^T (A^{-1}+B^{-1})^{-1} (A^{-1}a + B^{-1}b)$$

$$\Rightarrow (a^T A^{-1} + b^T B^{-1}) (A^{-1}+B^{-1})^{-1} (A^{-1}a + B^{-1}b)$$

$$\Rightarrow \left[a^T A^{-1} (A^{-1}+B^{-1})^{-1} + b^T B^{-1} (A^{-1}+B^{-1})^{-1}\right] (A^{-1}a + B^{-1}b)$$

Simplifying using the Woodbury identities.

$$(A^{-1}+B^{-1})^{-1} = A - A(A+B)^{-1}A = B - B(A+B)^{-1}B$$

$$\Rightarrow \left[a^T A^{-1}\left(A - A(A+B)^{-1}A\right) + b^T B^{-1}\left(B - B(A+B)^{-1}B\right)\right](A^{-1}a + B^{-1}b)$$

$$= \left[\left(a^T - a^T (A+B)^{-1}A\right) + \left(b^T - b^T (A+B)^{-1}B\right)\right](A^{-1}a + B^{-1}b)$$

$$= \left[\left(a^T (A^{-1}a + B^{-1}b) - a^T (A+B)^{-1} A (A^{-1}a + B^{-1}b) - b^T (A+B)^{-1} B (A^{-1}a + B^{-1}b) + b^T (A^{-1}a + B^{-1}b)\right]$$

Simplifying.

$$= \Big[\overbrace{a^T A^{-1} a}^{\text{RHS } 2.4.1} + \overbrace{a^T B^{-1} b}^{\text{RHS } 2.4.3} - \overbrace{a^T (A+B)^{-1} a}^{\text{RHS } 2.4.5} - \overbrace{a^T (A+B)^{-1} A B^{-1} b}^{\text{RHS } 2.4.7}$$

$$\underbrace{b^T B^{-1} B}_{\text{RHS } 2.4.2} + \underbrace{b^T A^{-1} a}_{\text{RHS } 2.4.4} - \underbrace{b^T (A+B)^{-1} b}_{\text{RHS } 2.4.6} - \underbrace{b^T (A+B)^{-1} B A^{-1} a}_{\text{RHS } 2.4.8}\Big]$$

We note that RHS 2.4.1 = LHS 4

RHS 2.4.2 = LHS 5

We next note that : RHS 1.1 + RHS 2.4.5 = 0 $\{a^T (A+B)^{-1} a\}$

RHS 1.4 + RHS 2.4.6 = 0 $\{b^T (A+B)^{-1} b\}$

We collect the remaining RHS terms and show they sum to zero :

$$\Rightarrow \Big[- \overbrace{a^T (A+B)^{-1} b}^{\text{RHS } 1.2} - \overbrace{b^T (A+B)^{-1} a}^{\text{RHS } 1.3}$$

$$+ \overbrace{a^T B^{-1} b}^{\text{RHS } 2.4.3} - \overbrace{a^T (A+B)^{-1} A B^{-1} b}^{\text{RHS } 2.4.7}$$

$$+ \underbrace{b^T A^{-1} a}_{\text{RHS } 2.4.4} - \underbrace{b^T (A+B)^{-1} B A^{-1} a}_{\text{RHS } 2.4.8} \Big]$$

collecting terms $a^T . * b$ and $b^T . * a$ matching.

$$\Rightarrow a^T \left[B^{-1} - (A+B)^{-1} - (A+B)^{-1} A B^{-1} \right] b$$

$$+ b^T \left[A^{-1} - (A+B)^{-1} - (A+B)^{-1} B A^{-1} \right] a$$

$$\Rightarrow a^T \left[B^{-1} - (A+B)^{-1} \left[I + A B^{-1} \right] \right] b + b^T \left[A^{-1} - (A+B)^{-1} (I + B A^{-1}) \right] a$$

$$\Rightarrow a^T \left[B^{-1} - (A+B)^{-1} (B + A) B^{-1} \right] b + b^T \left[A^{-1} - (A+B)^{-1} (A+B) A^{-1} \right] a$$

$$\Rightarrow a^T \left[B^{-1} - B^{-1} \right] b + b^T \left[A^{-1} - A^{-1} \right] a$$

$$\Rightarrow 0 + 0 = 0$$

Thus we have shown that bothe the exponent and multiplicative factors of LHS and RHS are equivalent.

$\blacksquare$

1(b) Conditionals of Gaussian Random Variables are Gaussian Random Variables. If (x,y) are jointly distributed as.

$$p(x,y) = \mathcal{N}\left(\begin{bmatrix} x \\ y \end{bmatrix}; \begin{bmatrix} \mu_x \\ \mu_y \end{bmatrix}, \begin{bmatrix} \Sigma_{xx} & \Sigma_{xy} \\ \Sigma_{yx} & \Sigma_{yy} \end{bmatrix}\right)$$

then the conditional $p(x|y) := \frac{p(x,y)}{p(y)}$

is given by:

$$p(x|y) = \mathcal{N}\left(x;\ \mu_x + \Sigma_{xy}\Sigma_{yy}^{-1}(y-\mu_y),\ \Sigma_{xx} - \Sigma_{xy}\Sigma_{yy}^{-1}\Sigma_{yx}\right)$$

Answer: without loss of generalility we consider the case of $p(\tilde{x},\tilde{y})$ such that $\tilde{x} = x - \mu_x$
$\tilde{y} = y - \mu_y$

*For the simplicity of notation, $\tilde{x}$ and $\tilde{y}$ are referred as x and y in the calcutions

① We start by computing two equivalent forms of the inverse using different Schur decompositions. to compute the precision matrix of the joint $p(\tilde{x},\tilde{y})$

$$\Lambda_{xy} = \begin{pmatrix} \Sigma_{xx} & \Sigma_{xy} \\ \Sigma_{yx} & \Sigma_{yy} \end{pmatrix}^{-1}$$

For a matrix $M = \begin{pmatrix} A & B \\ C & D \end{pmatrix}$

The first Schur inverse form is given by.

$$M_1^{-1} = \begin{pmatrix} A & B \\ C & D \end{pmatrix}^{-1} = \begin{pmatrix} A^{-1} + A^{-1}B(D-CA^{-1}B)^{-1}CA^{-1} & -A^{-1}B(D-CA^{-1}B)^{-1} \\ -(D-CA^{-1}B)^{-1}CA^{-1} & (D-CA^{-1}B)^{-1} \end{pmatrix}$$

The second Schur inverse form is given by.

$$M_2^{-1} = \begin{pmatrix} A & B \\ C & D \end{pmatrix}^{-1} = \begin{pmatrix} (A-BD^{-1}C)^{-1} & -(A-BD^{-1}C)^{-1}BD^{-1} \\ -D^{-1}C(A-BD^{-1}C)^{-1} & D^{-1}+D^{-1}C(A-BD^{-1}C)^{-1}BD^{-1} \end{pmatrix}$$

Thus the precision matrix Λ_{xy} of the joint can be written as.

$A = \Sigma_{xx}$ $\quad B = \Sigma_{xy}$ $\quad C = \Sigma_{yx}$ $\quad D = \Sigma_{yy}$.

$$M_1^{-1} = \begin{pmatrix} \Sigma_{xx}^{-1} + \Sigma_{xx}^{-1}\Sigma_{xy}(\Sigma_{yy}-\Sigma_{yx}\Sigma_{xx}^{-1}\Sigma_{xy})^{-1}\Sigma_{yx}\Sigma_{xx}^{-1} & -\Sigma_{xx}^{-1}\Sigma_{xy}(\Sigma_{yy}-\Sigma_{yx}\Sigma_{xx}^{-1}\Sigma_{xy})^{-1} \\ -(\Sigma_{yy}-\Sigma_{yx}\Sigma_{xx}^{-1}\Sigma_{xy})^{-1}\Sigma_{yx}\Sigma_{xx}^{-1} & (\Sigma_{yy}-\Sigma_{yx}\Sigma_{xx}^{-1}\Sigma_{xy})^{-1} \end{pmatrix}$$

We use notation $\Lambda_{y|x} = (\Sigma_{yy}-\Sigma_{yx}\Sigma_{xx}^{-1}\Sigma_{xy})^{-1}$

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

as covariance matrices are symmetric

$$\Sigma_{yx} = \Sigma_{xy}$$

$$\Sigma_{yx}^T = \Sigma_{yx}$$

We note $M_1^{-1}[1,1] = \Sigma_{xx}^{-1} + \Sigma_{xx}^{-1}\Sigma_{xy}(\Sigma_{yy} - \Sigma_{yx}\Sigma_{xx}^{-1}\Sigma_{xy})^{-1}\Sigma_{yx}\Sigma_{xx}^{-1}$

the expression is of the form

$$(Z + UWV^T)^{-1} = Z^{-1} - Z^{-1}U(W^{-1} + V^T Z^{-1} U)^{-1} V^T Z^{-1}$$

where $Z^{-1} = -\Sigma_{xx}^{-1}$ $\quad U = \Sigma_{xy}$ $\quad V = \Sigma_{yx}$

$W^{-1} = \Sigma_{yy}$ $\quad W = \Sigma_{yy}^{-1}$

$$-\left[(-\Sigma_{xx})^{-1} - (-\Sigma_{xx})(\Sigma_{xy})\left((\Sigma_{yy}) + (\Sigma_{yx})^T(-\Sigma_{xx})(\Sigma_{xy})\right)^{-1}(\Sigma_{yx})^T(-\Sigma_{xx})\right]$$

$$\Rightarrow -(-\Sigma_{xx}^{-1} + \Sigma_{xy}\Sigma_{yy}^{-1}\Sigma_{yx})^{-1} \Rightarrow \Lambda_{x|y}$$

$$M_1^{-1} = \begin{bmatrix} \Lambda_{x|y} & -\Sigma_{xx}^{-1}\Sigma_{xy}\Lambda_{y|x} \\ -\Lambda_{y|x}\Sigma_{yx}\Sigma_{xx}^{-1} & \Lambda_{y|x} \end{bmatrix}$$

We now compute the alternate form of

$$M_2^{-1} = \begin{pmatrix} A & B \\ C & D \end{pmatrix}^{-1} = \begin{pmatrix} (A - BD^{-1}C)^{-1} & -(A-BD^{-1}C)^{-1}BD^{-1} \\ -(D^{-1}C(A-BD^{-1}C)^{-1} & D^{-1}+D^{-1}C(A-BD^{-1}C^{-1})BD^{-1} \end{pmatrix}$$

$A = \Sigma_{xx}$ $\quad B = \Sigma_{xy}$ $\quad C = \Sigma_{yx}$ $\quad D = \Sigma_{yy}$.

$$M_2^{-1} = \begin{pmatrix} \Sigma_{xx} & \Sigma_{xy} \\ \Sigma_{yx} & \Sigma_{yy} \end{pmatrix}^{-1} = \begin{pmatrix} (\Sigma_{xx} - \Sigma_{xy}\Sigma_{yy}^{-1}\Sigma_{yx})^{-1} & -(\Sigma_{xx} - \Sigma_{xy}\Sigma_{y}^{-1}\Sigma_{yx})\Sigma_{xy}\Sigma_{yy}^{-1} \\ -(\Sigma_{yy}^{-1}\Sigma_{yx}(\Sigma_{xx} - \Sigma_{xy}\Sigma_{yy}^{-1}\Sigma_{yx})^{-1}) & \Sigma_{yy}^{-1} + \Sigma_{yy}^{-1}\Sigma_{yx}(\Sigma_{xx} - \Sigma_{xy}\Sigma_{y}^{-1}\Sigma_{yx})\Sigma_{xy}\Sigma_{yy}^{-1} \end{pmatrix}$$

$$M_2^{-1} = \begin{pmatrix} \Lambda_{x|y} & -\Lambda_{x|y}\Sigma_{xy}\Sigma_{yy}^{-1} \\ -\Sigma_{yy}^{-1}\Sigma_{yx}\Lambda_{x|y} & \Sigma_{yy}^{-1} + \Sigma_{yy}^{-1}\Sigma_{yx}(\Sigma_{xx} - \Sigma_{xy}\Sigma_{yy}^{-1}\Sigma_{yx})^{-1}\Sigma_{xy}\Sigma_{yy}^{-1} \end{pmatrix}$$

again using

$$(Z + UWV^T)^{-1} = Z^{-1} - Z^{-1}U(W^{-1} + V^TZ^{-1}U)^{-1}V^TZ^{-1}$$

using $\quad Z^{-1} = -\Sigma_{yy}^{-1} \quad U = \Sigma_{yx} \quad W^{-1} = \Sigma_{xx} \quad V^T = \Sigma_{xy} \quad U = \Sigma_{yx}$

$$-\Big(\underbrace{(-\Sigma_{yy}^{-1})}_{Z^{-1}} - (-\Sigma_{yy})^{-1}(\Sigma_{yx})\big(\Sigma_{xx} + (\Sigma_{xy})^{T}(-\Sigma_{yy})(\Sigma_{yx})\big)^{-1}(\Sigma_{xy})(-\Sigma_{y})$$

$$\Rightarrow \; -(-\Sigma_{yy} + \Sigma_{yx}\Sigma_{xx}^{-1}\Sigma_{xy})$$

$$\Rightarrow \; \Lambda_{y|x} \quad \text{also} = \Sigma_{yy}^{-1} + \Sigma_{yy}^{-1}\Sigma_{yx}\Lambda_{x}^{-1}\Sigma_{xy}\Sigma_{yy}^{-1}$$

Comparing entries in the inverses we have.

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

② $$-\Lambda_{y|x}\Sigma_{yx}\Sigma_{xx}^{-1} = -\Sigma_{yy}\Sigma_{yx}\Lambda_{x|y}$$

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

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

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

Computing $p(x,y)/p(y)$
exponent

$$\begin{bmatrix} x & y \end{bmatrix}^T \begin{bmatrix} \Lambda_{x|y} & -\Lambda_{x|y}\Sigma_{xy}\Sigma_{yy}^{-1} \\ -\Sigma_{yy}^{-1}\Sigma_{yx}\Lambda_{x|y} & \Lambda_{y|x} \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} - y^T\Sigma_{yy}^{-1}y$$

$$\Rightarrow \begin{bmatrix} x & y \end{bmatrix}^T \begin{bmatrix} \Lambda_{x|y} & -\Lambda_{x|y}\Sigma_{xy}\Sigma_{yy}^{-1} \\ -\Sigma_{yy}^{-1}\Sigma_{yx}\Lambda_{x|y} & \Lambda_{y|x} \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} - y^T\Sigma_{yy}^{-1}y$$

$$\Rightarrow \quad x^T\Lambda_{x|y}x - x^T\Lambda_{x|y}\Sigma_{yy}^{-1}y - y^T\Sigma_{yy}^{-1}\Sigma_{yx}\Lambda_{x|y}x + y^T(\Lambda_{y|x} - \Sigma_{yy}^{-1})y$$

$$p(x|y) = \mathcal{N}(x; \Sigma_{xy}\Sigma_{yy}^{-1}y, \Sigma_{xx} - \Sigma_{xy}\Sigma_{yy}^{-1}\Sigma_{yx})$$

exponent : $(x - \Sigma_{xy}\Sigma_{yy}^{-1}y)^T \Lambda_{x|y} (x - \Sigma_{xy}\Sigma_{yy}^{-1}y)$

$$\Rightarrow (x^T - y^T\Sigma_{yy}^{-1}\Sigma_{xy}) \Lambda_{x|y} (x - \Sigma_{xy}\Sigma_{yy}^{-1}y)$$

$$\Rightarrow x^T\Lambda_{x|y}x - x^T\Lambda_{xy}\Sigma_{xy}\Sigma_{yy}^{-1}y - y^T\Sigma_{yy}^{-1}\Sigma_{xy}\Lambda_{x|y}x + y^T\Sigma_{yy}^{-1}\Sigma_{xy}\Lambda_{x|y}\Sigma_{xy}\Sigma_{yy}^{-1}y$$

Consider coefficient $y^T . * y$

$$\Sigma_{yy}^{-1}\Sigma_{xy}\Lambda_{x|y}\Sigma_{xy}\Sigma_{yy}^{-1}$$

$$(Z + UWV^T)^{-1} = Z^{-1} - Z^{-1}U(W^{-1} + V^TZ^{-1}U)^{-1}V^TZ^{-1}$$

$$\Rightarrow \Lambda_{y|x} \quad \text{also} = \Sigma_{yy}^{-1} + \Sigma_{yy}^{-1}\Sigma_{yx}\Lambda_{x|y}\Sigma_{xy}\Sigma_{yy}^{-1}$$

$$(\Lambda_{y|x} - \Sigma_{yy}^{-1}) = \Sigma_{yy}^{-1}\Sigma_{yx}\Lambda_{x|y}\Sigma_{xy}\Sigma_{yy}^{-1}$$

Thus coefficient for $y^T . * y$ $p(x|y)$ is same as $\frac{p(x,y)}{p(y)}$

thus by comparison all exponent coefficients are the same.

We now compute the determinat

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

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

We can decompose the determinat of the block matrix $|\Sigma_{XY}|$ into a product of determinat of Σ_{yy} and remaining variance.

$$\det\left|\Sigma_{XY}\right| = \det\begin{vmatrix}\Sigma_{xx} & \Sigma_{xy} \\ \Sigma_{yx} & \Sigma_{yy}\end{vmatrix} = \det(\Sigma_{yy})\det(\Sigma_{xx} - \Sigma_{xy}\Sigma_{yy}^{-1}\Sigma_{yx})$$

$$\det|\Sigma_{XY}| = \det(\Sigma_{x|y})\det(\Sigma_{yy})$$

thus we have $\dfrac{\det|\Sigma_{xy})}{\det|\Sigma_{yy})} = \det(\Sigma_{x|y})$

Thus Finally we have:

$$\frac{p(x,y)}{p(y)} = \frac{\dfrac{1}{(2\pi)^{2D/2}\det(\Sigma_{xy})^{\frac{1}{2}}}\exp\left\{-\frac{1}{2}\left|[x\ y]^T\Lambda_{xy}\begin{bmatrix}x\\y\end{bmatrix}\right|\right\}}{\dfrac{1}{(2\pi)^{D/2}}\det(\Sigma_y)^{\frac{1}{2}}\exp\left\{-\frac{1}{2}y^T\Lambda_y y\right\}}$$

where $\dim(x)+\dim(y) = D+D = 2D$

$$= \frac{1}{(2\pi)^{D/2}}\frac{1}{\det(\Sigma_{xy})^{\frac{1}{2}}/\det(\Sigma_y)^{\frac{1}{2}}}\exp\left\{\frac{-1}{2}\left([x\ y]^T\Lambda_{xy}\begin{bmatrix}x\\y\end{bmatrix} - y^T\Lambda_y y\right\}\right.$$

$$= \frac{1}{(2\pi)^{p/2}} \frac{1}{\det(\Sigma_{x|y})^{\frac{1}{2}}} \exp\left\{ (x - \Sigma_{xy}\Sigma_{yy}^{-1}y)^T \Lambda_{x|y} (x - \Sigma_{xy}\Sigma_{yy}^{-1}y) \right\}$$

$$= \mathcal{N}(x ; -\Sigma_{xy}\Sigma_{yy}^{-1}y, \Lambda_{x|y})$$

$$= P(x|y)$$ as per definition.

We know that here x and y actually refers to $\tilde{x}$ and $\tilde{y}$ by our initial substitution*

Resubstitute $x \triangleq x - \mu_x$, $y \triangleq y - \mu_y$, $\Lambda_{x|y} = (\Sigma_{xx} - \Sigma_{xy}\Sigma_{yy}^{-1}\Sigma_{yx})^{-1}$

$$\boxed{\mathcal{N}(x ; \mu_x - \Sigma_{xy}\Sigma_{yy}^{-1}(y - \mu_y), \Sigma_{xx} - \Sigma_{xy}\Sigma_{yy}^{-1}\Sigma_{yx})}$$

Q.E.D.

(c) Finally use the above results to show the following two statements which help interpretting the covariance matrix and its inverse (the precision matrix.

(a) A zero entry in the i,j^{th} position of the covariance matrix Σ implies that x_i and x_j are marginally independent

(b) A zero-entry in the i,j^{th} position of the precision matrix $\Lambda = \Sigma^{-1}$ implies that x_i and x_j are conditionally independent given all other variables x_k for $k \neq i,j$.

(a) Without loss of generality, we can assume that $(i,j) = (1,2)$, as we can permute the variables without changing the distribution.

we have been given that $\Sigma_{12} = 0$ and from the symmetry of covariance matrix we know $\Sigma_{21} = 0$

$$p(x ; \mu, \Sigma) = \frac{1}{(2\pi)^{D/2} |\Sigma|^{1/2}} \exp\left\{ \frac{-1}{2} (x-\mu)^T \Sigma (x-\mu) \right\} \quad ; x \in \mathbb{R}^D$$

consider the matrix $A = \begin{bmatrix} I_2 & O_2 \\ O_2 & O_{D \times 2} \end{bmatrix}$; $A = A^T$

As we know the affine projection $p(z := Ax) = \mathcal{N}(x ; A\mu, A\Sigma A^T)$

thus we have $z = A \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_D \end{bmatrix} = \begin{bmatrix} x_1 \\ x_2 \\ 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix}$

The new mean $\mu' = A\mu = \begin{bmatrix} \mu_x \\ \mu_y \\ 0 \\ 0 \\ \vdots \end{bmatrix}_{D\times 1}$

The new covariance $\Sigma' = \begin{bmatrix} I_2 & O_2 \\ O_2 & O_{D\times 2} \end{bmatrix} \begin{bmatrix} \Sigma_{11} & \Sigma_{12} & \cdots & \Sigma_{1D} \\ \vdots & & \ddots & \\ \Sigma_{D1} & & & \Sigma_{DD} \end{bmatrix} \begin{bmatrix} I_2 & O_2 \\ O_2 & O_{D\times 2} \end{bmatrix}$

$$= \begin{bmatrix} \Sigma_{11} & \Sigma_{12} & 0 & 0 \\ \Sigma_{21} & \Sigma_{22} & 0 & 0 \\ 0 & 0 & 0 & 0 \\ \vdots & \vdots & & \end{bmatrix}$$

Thus the joint distribution of x_1 x_2 is given by

$$p\left(\begin{bmatrix} x_1 \\ x_2 \end{bmatrix}\right) = \mathcal{N}\left(\begin{bmatrix} x_1 \\ x_2 \end{bmatrix}, \begin{bmatrix} \mu_{x_1} \\ \mu_{x_2} \end{bmatrix}, \begin{bmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{bmatrix}\right)$$

as we have $\Sigma_{21} = \Sigma_{12} = 0$

$$p(x_1 | x_2) = \mathcal{N}(x_1 ; \mu_{x_1} + \Sigma_{12}\Sigma_{22}^{-1}(x_2 - \mu_2), \Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21})$$
$$= \mathcal{N}(x_1 ; \mu_{x_1} ; \Sigma_{11}) = p(x_1)$$

and $p(x_2 | x_1) = p(x_2)$ by symmetry.

$$p(x_1, x_2) = p(x_1 | x_2)\, p(x_2)$$
$$= p(x_1)\, p(x_2)$$

thus x_1 and x_2 are marginally independent. ■

In short, choosing $(i,j) = (1,2)$ w.l.o.g. and the covariance matrix $\Sigma = \begin{pmatrix} \Sigma_{1,1} & 0 & \cdots \\ 0 & \Sigma_{2,2} & \vdots \\ \cdots & & \cdots \end{pmatrix}$ means that $\Sigma_{1,2} = \Sigma_{2,1} = 0$

Therefore using the equation from part b yields

$$p(x_1|x_2) = \mathcal{N}(x_1 ; \mu_1 + \Sigma_{12} \Sigma_{2,2}^{-1} (x_2 - \mu_2), \Sigma_{1,1} - \Sigma_{12} \Sigma_{22}^{-1} \Sigma_{21})$$

$$= \mathcal{N}(x_1 ; \mu_1 + 0 \, \Sigma_{2,2}^{-1} (x_2 - \mu_2), \Sigma_{1,1} - 0 \, \Sigma_{2,2}^{-1} 0)$$

$$= \mathcal{N}(x_1 ; \mu_1, \Sigma_{11})$$

$p(x_1|x_2) = p(x_1) \rightarrow$ marginal

and x_1 and x_2 are marginally independent. Since this can be applied to any $i \neq j$ pair (by permuting), we say that the statement in the first bullet point always hold.

1(c) Finally use the above results to show the following two statements

* zero entry in the i,j-th position of precision matrix Σ^{-1} implies that x_i and x_j are conditionaly independent.

Ans: Without loss of generality assume $i=1$ and $j=2$

Let us partition the vector $x = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_D \end{bmatrix}$ into two pars.

$$x_{p_1} = [x_1 \; x_2]^T \qquad x_{p_2} = [x_3 \; x_4 \; \dots \; x_D]^T$$

The joint distribution

$$p(x_{p_1}, x_{p_2}) \propto \exp\left\{ -\frac{1}{2} \left[(x_{p_1}-\mu_{p_1})^T \; (x_{p_2}-\mu_{p_2})^T \right] \begin{bmatrix} \Lambda_{p_1p_1} & \Lambda_{p_1p_2} \\ \Lambda_{p_2p_1} & \Lambda_{p_2p_2} \end{bmatrix} \begin{bmatrix} x_{p_1}-\mu_{p_1} \\ x_{p_2}-\mu_{p_2} \end{bmatrix} \right\}$$

$$\propto \exp\left\{ -\frac{1}{2} \left[(x_{p_1}-\mu_{p_1})^T \Lambda_{p_1p_1} (x_{p_1}-\mu_{p_1}) + (x_{p_2}-\mu_{p_2})^T \Lambda_{p_2p_1} (x_{p_1}-\mu_{p_1}) \right.\right.$$
$$+ (x_{p_1}-\mu_{p_1})^T \Lambda_{p_1p_2} (x_{p_2}-\mu_{p_2})$$
$$\left.\left. + (x_{p_2}-\mu_{p_2})^T \Lambda_{p_2p_2} (x_{p_2}-\mu_{p_2}) \right] \right\}$$

Conditionig on x_{p_2} variables is equivalent to holding then constant.

We futher know from the definition of conditional inde-pendence that:

$X \perp Y \mid Z$ iff there exists functions g, h such that

$$p(x, y \mid z) = g(x, z)\, h(y, z) \quad \forall x, y, z \text{ and } p(z) > 0$$

Thus if we can factorize

$$p(x_{p_1} | x_{p_2}) = g(x_1, x_{p_2})\, h(x_2 | x_{p_2})$$

we will have shown the conditional independence of $x_1 | x_3 x_4 \dots x_p \perp x_2 | x_3, x_4 \dots x_D$

Starting with first term we see it factorizes.

$$(x_{p_1} - \mu_{p_1})^T \begin{bmatrix} \Lambda_{11} & \Lambda_{12} \\ \Lambda_{21} & \Lambda_{22} \end{bmatrix} (x_{p_2} - \mu_p)$$

here we know $\Lambda_{12} = 0$, $\Lambda_{21} = 0$ as we know the entries in the precision matrix are zero.

$$\begin{bmatrix} (x_1 - \mu_1)^T & (x_2 - \mu_2)^T \end{bmatrix} \begin{bmatrix} \Lambda_{11} & 0 \\ 0 & \Lambda_{22} \end{bmatrix} \begin{bmatrix} x_1 - \mu_1 \\ x_2 - \mu_2 \end{bmatrix}$$

$$\Rightarrow (x_1 - \mu_1)^T \Lambda_{11} (x_1 - \mu_1) + (x_2 - \mu_2)^T \Lambda_{22} (x_2 - \mu_2)$$

Let $g_1(x_1) = (x_1 - \mu_1)^T \Lambda_{11} (x - \mu_1)$ and $h_1(x_2) = (x_2 - \mu_2)^T \Lambda_{22}(x_2 - \mu_2)$

then the first term contains only x_1 and second term contains only x_2

The terms $(x_{p_2}-\mu_{p_2})^T \Lambda_{p_2 p_1} (x_{p_1}-\mu_{p_1})$

and $(x_{p_1}-\mu_{p_1})^T \Lambda_{p_1 p_2} (x_{p_2}-\mu_{p_2})$

are linear in x_{p_1} and can thus be

written as separate additive functions of x_1, x_2

$$\Rightarrow \quad g_2(x_2, x_{p_2}) \quad + \quad h_2(x_2, x_{p_2})$$

The term $(x_{p_2}-\mu_{p_2})^T \Lambda_{p_2 p_2} (x_{p_2}-\mu_{p_2})$ does not contain x_{p_1} and can be considered a constat.

$$p(x_{p_1} | x_{p_2}) \propto \exp\left\{-\frac{1}{2}\left[g_1(x_1) + g_2(x_2, x_{p_2}) + h_1(x_2) + h_2(x_2, x_{p_2}) + \text{constat}\right]\right.$$

$$\propto \exp\left[-\frac{1}{2} g(x_1, x_{p_2})\right] \exp\left[-\frac{1}{2} h(x_2, x_{p_2})\right]$$

Thus the conditionon density $p(x_1, x_2 | x_3, x_4 \cdots x_p)$

factorizes as product of two functions.

$$\exp\left[-\frac{1}{2} g(x_1, x_3 x_4 \cdots x_p)\right] \exp\left[-\frac{1}{2} h(x_2 x_3, x_4, \dots x_p)\right]$$

Thus $x_1 \perp x_2 | x_3 x_4 \cdots x_p$ when $\Lambda_{21} = \Lambda_{12} = 0$

precision matrix entries for $(1,2)$ are zero

and this result holds for every $i \neq j$ pair by simply permuting the indices

In short, for a precision matrix $\Lambda = \Sigma^{-1}$ and a zero entry on the $(i,j) = (1,2)$th index of the Λ means that x_i and x_j are condinitionally independent where other variables are x_k with $k \neq i,j$ needs to be proven.

Using the equation from part b,

$$p(x_1 \mid x_2, x_k \neq x_1 x_2) = \mathcal{N}\left(x_1 ; \mu_1 - \Lambda_{1,1}^{-1} (\underbrace{\Lambda_{1,2}}_{0} \cdots \Lambda_{1k}) \begin{pmatrix} x_2 - \mu_2 \\ \vdots \\ x_k - \mu_k \end{pmatrix}, \right.$$

$$\left. \Lambda_{11}^{-1} - (\underbrace{\Lambda_{12}}_{0} \cdots \Lambda_{1k}) \begin{pmatrix} \Lambda_{22} & \overset{0}{\Lambda_{2,1}} \\ \vdots & \vdots \\ \Lambda_{kk} & \Lambda_{k,1} \end{pmatrix} \right)$$

Setting $\Lambda_{1,2} = \Lambda_{2,1} = 0$

$$= \mathcal{N}\left(x_1 ; \mu_1 - \Lambda_{11}^{-1} (\cdots \Lambda_{1,k} \cdots) \begin{pmatrix} \vdots \\ x_k - \mu_k \\ \vdots \end{pmatrix}, \right.$$

$$\left. \Lambda_{11}^{-1} - (\cdots \Lambda_{1,k} \cdots) \begin{pmatrix} \Lambda_{kk} & \Lambda_{k,1} \\ \vdots & \end{pmatrix} \right)$$

$$= p(x_1 \mid x_{k \neq 1,2})$$

Thus, we have proven that the last term does not involve x_2 and x_1 and x_2 are conditionally independent given $\{x_k\}$.

By permuting the indices, this result can also be generalized for arbitrary (i,j) in the statement.