0.1 Joint Probability

0.1.1 Abstract

In this section, we study Gaussian joint distribution.

0.1.2 Given

$$x \backsim N(x|\mu, \Lambda^{-1})$$
$$y|x \backsim N(y|Ax + b, L^{-1})$$

Inference

$$y = Ax + b + \epsilon, \quad \epsilon \backsim N(0, L^{-1}), \quad x \perp \epsilon$$

0.1.3 To solve

$$\begin{cases} p(y) \\ \\ p(x|y) \end{cases}$$

0.2 Derivation

Derive $p(y)$

$$E[y] = AE[x] + b + E[\epsilon] = A\mu + b$$
$$Var[y] = A\Lambda^{-1}A^T$$
$$\therefore y \backsim N(A\mu + b, A\Lambda^{-1}A^T)$$

Derive $p(x|y)$

Construct dist z

Here we construct a distribution:

$$z = \begin{pmatrix} x \\ y \end{pmatrix} \backsim N(\begin{bmatrix} \mu \\ A\mu + b \end{bmatrix}, \begin{bmatrix} \Lambda^{-1} & \Delta \\ \Delta & A\Lambda^{-1}A^T \end{bmatrix})$$

$$
\begin{aligned}
\Delta &= cov(x,y)\\
&= E[(x-E[x])(y-E[y])^T]\\
&= E[(x-\mu)(y-A\mu-b)^T]\\
&= E[(x-\mu)(Ax+b+\epsilon-A\mu-b)^T)]\\
&= E[(x-\mu)(Ax-A\mu+\epsilon)^T]\\
&= E[(x-\mu)(x-\mu)^T A^T+(x-\mu)\epsilon^T]\\
&= E[(x-\mu)(x-\mu)^T]A^T+E[(x-\mu)\epsilon^T]\\
&\because x\perp\epsilon\\
&\therefore = E[(x-\mu)(x-\mu)^T]A^T\\
&= \Lambda^{-1}A^T
\end{aligned}
$$

$$
\therefore z=\begin{pmatrix}x\\y\end{pmatrix}\backsim N(\begin{bmatrix}\mu\\A\mu+b\end{bmatrix},\begin{bmatrix}\Lambda^{-1} & \Lambda^{-1}A^T\\ \Lambda^{-1}A^T & A\Lambda^{-1}A^T\end{bmatrix})
$$

Construct dist $x.y$

let's set

$$
\begin{aligned}
x.y &= x-\Sigma_{xy}\Sigma_{yy}^{-1}y\\
&= x-(\Lambda^{-1}A^T)(A\Lambda^{-1}A^T)^{-1}y\\
&= x-A^{-1}y\\
&= \begin{pmatrix}I & -A^{-1}\end{pmatrix}\begin{pmatrix}x\\y\end{pmatrix}
\end{aligned}
$$

$$
\begin{aligned}
E[x.y] &= E[x]-A^{-1}E[y]\\
&= \mu-A^{-1}(A\mu+b)\\
&= -A^{-1}b
\end{aligned}
$$

$$
\begin{aligned}
Var[x.y] &= \begin{pmatrix}I & -A^{-1}\end{pmatrix}Var[z]\begin{pmatrix}I\\-(A^{-1})^T\end{pmatrix}\\
&= \begin{pmatrix}I & -A^{-1}\end{pmatrix}\begin{pmatrix}\Lambda^{-1} & \Lambda^{-1}A^T\\ \Lambda^{-1}A^T & A\Lambda^{-1}A^T\end{pmatrix}\begin{pmatrix}I\\-(A^{-1})^T\end{pmatrix}\\
&= \begin{pmatrix}\Lambda^{-1}-A^{-1}\Lambda^{-1}A^T & 0\end{pmatrix}\begin{pmatrix}I\\-(A^{-1})^T\end{pmatrix}\\
&= \Lambda^{-1}-A^{-1}\Lambda^{-1}A^T\\
&\therefore x.y\backsim N(-A^{-1}b,\Lambda^{-1}-A^{-1}\Lambda^{-1}A^T)
\end{aligned}
$$

Construct $x|y$

we got

$$
x|y = x.y+A^{-1}y
$$

here, we can see $A^{-1}y$ as constant C.

then:

$$x|y = x.y + C$$

$$E[x|y] = A^{-1}y - A^{-1}b$$

$$Var[x|y] = Var[x.y]$$

$$\therefore x|y \backsim N(A^{-1}y - A^{-1}b, \Lambda^{-1} - A^{-1}\Lambda^{-1}A^T)$$

Now, according to an edge distribution and conditional distribution, we construct a joint distribution to obtain another edge distribution and conditional distribution.