パターン認識と機械学習
第2章 確率分布

橋本龍二

December 28, 2020

1

2

3 ガウス分布

3.1 条件付きガウス分布

$\boldsymbol{x}$ をガウス分布に従う D 次元ベクトルとする。

$$\boldsymbol{x} = \begin{pmatrix} \boldsymbol{x}_a \\ \boldsymbol{x}_b \end{pmatrix} \sim N(\boldsymbol{x}|\boldsymbol{\mu}, \Sigma)$$

ただし、$\boldsymbol{x}_a$ は M 個、$\boldsymbol{x}_b$ は $D-M$ 個の要素。

$$\boldsymbol{\mu} = \begin{pmatrix} \boldsymbol{\mu}_a \\ \boldsymbol{\mu}_b \end{pmatrix}, \quad \Sigma = \begin{pmatrix} \Sigma_{aa} & \Sigma_{ab} \\ \Sigma_{ba} & \Sigma_{bb} \end{pmatrix}$$

ただし、Σ_{aa} と Σ_{bb} は対称だが $\Sigma_{ba} = {}^T\Sigma_{ab}$ である。また、$\Lambda = \Sigma^{-1}$ とすれば、

$$\Lambda = \begin{pmatrix} \Lambda_{aa} & \Lambda_{ab} \\ \Lambda_{ba} & \Lambda_{bb} \end{pmatrix}$$

ただし、Λ_{aa} と Λ_{bb} は対称だが、${}^T\Lambda_{ab} = \Lambda_{ba}$ である。また Σ_{aa} の逆行列は Λ_{aa} ではないので注意。
このように考えれば、$p(\boldsymbol{x})$ とは $\boldsymbol{x}_a$ と $\boldsymbol{x}_b$ の同時分布である。$\boldsymbol{x}_a|\boldsymbol{x}_b$ の条件付

き分布を考える。

$$\begin{aligned}
{}^T(\boldsymbol{x}-\boldsymbol{\mu})\Sigma^{-1}(\boldsymbol{x}-\boldsymbol{\mu}) &= {}^T\begin{pmatrix}\boldsymbol{x}_a-\boldsymbol{\mu}_a\\ \boldsymbol{x}_b-\boldsymbol{\mu}_b\end{pmatrix}\begin{pmatrix}\Lambda_{aa} & \Lambda_{ab}\\ \Lambda_{ba} & \Lambda_{bb}\end{pmatrix}\begin{pmatrix}\boldsymbol{x}_a-\boldsymbol{\mu}_a\\ \boldsymbol{x}_b-\boldsymbol{\mu}_b\end{pmatrix}\\
&= {}^T(\boldsymbol{x}_a-\boldsymbol{\mu}_a)\Lambda_{aa}(\boldsymbol{x}_a-\boldsymbol{\mu}_a)+{}^T(\boldsymbol{x}_a-\boldsymbol{\mu}_b)\Lambda_{ab}(\boldsymbol{x}_b-\boldsymbol{\mu}_b)\\
&\quad +{}^T(\boldsymbol{x}_b-\boldsymbol{\mu}_b)\Lambda_{ba}(\boldsymbol{x}_a-\boldsymbol{\mu}_a)+{}^T(\boldsymbol{x}_b-\boldsymbol{\mu}_b)\Lambda_{bb}(\boldsymbol{x}_b-\boldsymbol{\mu}_b)\\
&= {}^T\boldsymbol{x}_a\Lambda_{aa}\boldsymbol{x}_a-{}^T\boldsymbol{x}_a\Lambda_{aa}\boldsymbol{\mu}_a+{}^T\boldsymbol{x}_a\Lambda_{ab}(\boldsymbol{x}_b-\boldsymbol{\mu}_b)\\
&\quad -{}^T\boldsymbol{\mu}_a\Lambda_{aa}\boldsymbol{x}_a+{}^T(\boldsymbol{x}_b-\boldsymbol{\mu}_b)\Lambda_{ba}\boldsymbol{x}_a+const\\
&= {}^T\boldsymbol{x}_a\Lambda_{aa}\boldsymbol{x}_a-{}^T\boldsymbol{x}\{\Lambda_{aa}\boldsymbol{\mu}_a-\Lambda_{ab}(\boldsymbol{x}_b-\boldsymbol{\mu}_b)\}\\
&\quad -\{{}^T\boldsymbol{\mu}_a\Lambda_{aa}-{}^T(\boldsymbol{x}_b-\boldsymbol{\mu}_b)\Lambda_{ba}\}\boldsymbol{x}_a+const\\
&= {}^T\boldsymbol{x}_aA^{-1}\boldsymbol{x}_a-{}^T\boldsymbol{x}_aB-{}^TB\boldsymbol{x}_a+const\\
&= {}^T(\boldsymbol{x}_a-AB)A^{-1}(\boldsymbol{x}_a-AB)+const
\end{aligned}$$

ただし、

$$A=\Lambda_{aa}^{-1},\ B=\Lambda_{aa}\boldsymbol{\mu}_a-\Lambda_{ab}(\boldsymbol{x}_b-\boldsymbol{\mu}_b)$$

と置いた。したがって、条件付き分布 $p(\boldsymbol{x}_a|\boldsymbol{x}_b)$ の平均と分散をそれぞれ $\boldsymbol{\mu}_{a|b}$, $\Sigma_{a|b}$ とすると、

$$\begin{aligned}
\boldsymbol{\mu_{a|b}} &= \Lambda_{aa}^{-1}\{\Lambda_{aa}\boldsymbol{\mu}_a-\Lambda_{ab}(\boldsymbol{x}_b-\boldsymbol{\mu}_b)\}\\
&= \boldsymbol{\mu}_a-\Lambda_{aa}^{-1}\Lambda_{ab}(\boldsymbol{x}_b-\boldsymbol{\mu}_b)\\
\Sigma_{a|b} &= \Lambda_{aa}^{-1}
\end{aligned}$$

となる。では $\Lambda_{aa}^{-1},\Lambda_{ab}$ を求めよう (このままの方が簡潔なので普通は求めない)。今、

$$\begin{pmatrix}\Sigma_{aa} & \Sigma_{ab}\\ \Sigma_{ba} & \Sigma_{bb}\end{pmatrix}^{-1}=\begin{pmatrix}\Lambda_{aa} & \Lambda_{ab}\\ \Lambda_{ba} & \Lambda_{bb}\end{pmatrix}$$

であるから、

$$\begin{aligned}
\Lambda_{aa} &= (\Sigma_{aa}-\Sigma_{ab}\Sigma_{bb}^{-1}\Sigma_{ba})^{-1}\\
\Lambda_{ab} &= -(\Sigma_{aa}-\Sigma_{ab}\Sigma_{bb}^{-1}\Sigma_{ba})^{-1}\Sigma_{ab}\Sigma_{bb}^{-1}
\end{aligned}$$

より、

$$
\begin{aligned}
\boldsymbol{\mu}_{a|b} &= \boldsymbol{\mu}_a + \Sigma_{ab}\Sigma_{bb}^{-1}(\boldsymbol{x}_b - \boldsymbol{\mu}_b) \\
\Sigma_{a|b} &= \Sigma_{aa} - \Sigma_{ab}\Sigma_{bb}^{-1}\Sigma_{ba}
\end{aligned}
$$

となった。条件付き分布の平均は $\boldsymbol{x}_b$ の線形関数。

3.2 周辺ガウス分布

3.1 の設定のもとで周辺分布 $p(\boldsymbol{x}_a)$ を求める。

$$
\begin{aligned}
p(\boldsymbol{x}_a) &= \int p(\boldsymbol{x}_a, \boldsymbol{x}_b) d\boldsymbol{x}_b \\
&\propto \int \exp\{-\frac{1}{2}\ {}^{T}(\boldsymbol{x} - \boldsymbol{\mu})\Lambda(\boldsymbol{x} - \boldsymbol{\mu})\} d\boldsymbol{x}_b
\end{aligned}
$$

$$
\begin{aligned}
&{}^{T}(\boldsymbol{x} - \boldsymbol{\mu})\Lambda(\boldsymbol{x} - \boldsymbol{\mu}) \\
= \ & {}^{T}(\boldsymbol{x}_a - \boldsymbol{\mu}_a)\Lambda_{aa}(\boldsymbol{x}_a - \boldsymbol{\mu}_a) + {}^{T}(\boldsymbol{x}_a - \boldsymbol{\mu}_b)\Lambda_{ab}(\boldsymbol{x}_b - \boldsymbol{\mu}_b) \\
& + {}^{T}(\boldsymbol{x}_b - \boldsymbol{\mu}_b)\Lambda_{ba}(\boldsymbol{x}_a - \boldsymbol{\mu}_a) + {}^{T}(\boldsymbol{x}_b - \boldsymbol{\mu}_b)\Lambda_{bb}(\boldsymbol{x}_b - \boldsymbol{\mu}_b) \\
= \ & {}^{T}\boldsymbol{x}_b\Lambda_{bb}\boldsymbol{x}_b + {}^{T}\boldsymbol{x}_b\Lambda_{ba}(\boldsymbol{x}_a - \boldsymbol{\mu}_a) - {}^{T}\boldsymbol{x}_b\Lambda_{bb}\boldsymbol{\mu}_b - {}^{T}\boldsymbol{\mu}_b\Lambda_{bb}\boldsymbol{x}_b + {}^{T}(\boldsymbol{x}_a - \boldsymbol{\mu}_a)\Lambda_{ab}\boldsymbol{x}_b \\
& + {}^{T}\boldsymbol{x}_a\Lambda_{aa}\boldsymbol{x}_a - {}^{T}\boldsymbol{\mu}_a\Lambda_{aa}\boldsymbol{x}_a - {}^{T}\boldsymbol{x}_a\Lambda_{aa}\boldsymbol{\mu}_a - {}^{T}\boldsymbol{x}_a\Lambda_{ab}\boldsymbol{\mu}_b - {}^{T}\boldsymbol{\mu}_b\Lambda_{ba}\boldsymbol{x}_a + const \\
= \ & {}^{T}\boldsymbol{x}_b\Lambda_{bb}\boldsymbol{x}_b - {}^{T}\boldsymbol{x}_b\{\Lambda_{bb}\boldsymbol{\mu}_b - \Lambda_{ba}(\boldsymbol{x}_a - \boldsymbol{\mu}_a)\} - \{{}^{T}\boldsymbol{\mu}_b\Lambda_{bb} - {}^{T}(\boldsymbol{x}_a - \boldsymbol{\mu}_a)\Lambda_{ab}\}\boldsymbol{x}_b \\
& + {}^{T}\boldsymbol{x}_a\Lambda_{aa}\boldsymbol{x}_a - {}^{T}\boldsymbol{x}(\Lambda_{aa}\boldsymbol{\mu}_a + \Lambda_{ab}\boldsymbol{\mu}_b) - ({}^{T}\boldsymbol{\mu}_a\Lambda_{aa} + {}^{T}\boldsymbol{\mu}_b\Lambda_{ba})\boldsymbol{x}_a + const \\
= \ & {}^{T}\boldsymbol{x}_b A^{-1}\boldsymbol{x}_b - {}^{T}\boldsymbol{x}_b B - {}^{T}B\boldsymbol{x}_b \\
& + {}^{T}\boldsymbol{x}_a\Lambda_{aa}\boldsymbol{x}_a - {}^{T}\boldsymbol{x}(\Lambda_{aa}\boldsymbol{\mu}_a + \Lambda_{ab}\boldsymbol{\mu}_b) - ({}^{T}\boldsymbol{\mu}_a\Lambda_{aa} + {}^{T}\boldsymbol{\mu}_b\Lambda_{ba})\boldsymbol{x}_a + const \\
= \ & {}^{T}(\boldsymbol{x}_b - AB)A^{-1}(\boldsymbol{x} - AB) - {}^{T}B\ {}^{T}AB \\
& + {}^{T}\boldsymbol{x}_a\Lambda_{aa}\boldsymbol{x}_a - {}^{T}\boldsymbol{x}(\Lambda_{aa}\boldsymbol{\mu}_a + \Lambda_{ab}\boldsymbol{\mu}_b) - ({}^{T}\boldsymbol{\mu}_a\Lambda_{aa} + {}^{T}\boldsymbol{\mu}_b\Lambda_{ba})\boldsymbol{x}_a + const
\end{aligned}
$$

ただし、

$$
A = \Lambda_{bb}^{-1}, \quad B = \Lambda_{bb}\boldsymbol{\mu}_b - \Lambda_{ba}(\boldsymbol{x}_a - \boldsymbol{\mu}_a)
$$

と置いた。ゆえ、

$$p(\boldsymbol{x}_a) \propto \exp[-\frac{1}{2}\{-\ ^{T}B\ ^{T}AB +\ ^{T}\boldsymbol{x}_a\Lambda_{aa}\boldsymbol{x}_a -\ ^{T}\boldsymbol{x}(\Lambda_{aa}\boldsymbol{\mu}_a + \Lambda_{ab}\boldsymbol{\mu}_b) - (\ ^{T}\boldsymbol{\mu}_a\Lambda_{aa} +\ ^{T}\boldsymbol{\mu}_b\Lambda_{ba})\boldsymbol{x}_a\}]$$

右辺の指数部分の中身を変形してまとめると、

$$\begin{aligned}
& ^{T}\boldsymbol{x}_a(-\ ^{T}\Lambda_{ba}\Lambda_{bb}^{-1}\Lambda_{ba} + \Lambda_{aa})\boldsymbol{x}_a \\
& -\ ^{T}\boldsymbol{x}_a(-\ ^{T}\Lambda_{ba}\boldsymbol{\mu}_b -\ ^{T}\Lambda_{ba}\Lambda_{bb}^{-1}\Lambda_{ba}\boldsymbol{\mu}_a + \Lambda_{aa}\boldsymbol{\mu}_a + \Lambda_{ab}\boldsymbol{\mu}_b) \\
& -(\ ^{T}\boldsymbol{\mu}_b\Lambda_{ba}\boldsymbol{x}_a -\ ^{T}\boldsymbol{\mu}_a\ ^{T}\Lambda_{ba}\Lambda_{bb}^{-1}\Lambda_{ba} +\ ^{T}\boldsymbol{\mu}_a\Lambda_{aa} +\ ^{T}\boldsymbol{\mu}_b\Lambda_{ab})\boldsymbol{x}_a \\
= & \ ^{T}\boldsymbol{x}_a B^{-1}\boldsymbol{x}_a -\ ^{T}\boldsymbol{x}_a C -\ ^{C}\boldsymbol{x}_a \\
= & \ ^{T}(\boldsymbol{x}_a - BC)B^{-1}(\boldsymbol{x}_a - BC) + const
\end{aligned}$$

ただし、

$$\begin{aligned}
B &= (\Lambda_{aa} - \Lambda_{ab}\Lambda_{bb}^{-1}\Lambda_{ba})^{-1} \\
C &= (\Lambda_{aa} - \Lambda_{ab}\Lambda_{bb}^{-1}\Lambda_{ba})\boldsymbol{\mu}_a
\end{aligned}$$

したがって、

$$\begin{aligned}
\boldsymbol{\mu}_{a|b} &= BC = \boldsymbol{\mu}_a \\
\Sigma_{a|b} &= B = \Sigma_{aa}
\end{aligned}$$

3.3 ガウス変数に対するベイズの定理

ベクトル $\boldsymbol{x}$ を正規分布に従う確率変数とし、平均が $\boldsymbol{x}$ の線形関数で共分散が $\boldsymbol{x}$ とは独立であるようなガウス条件付き分布 $p(\boldsymbol{y}|\boldsymbol{x})$ を考える。すなわち、

$$\begin{cases} p(\boldsymbol{x}) = N(\boldsymbol{x}|\boldsymbol{\mu}, \Lambda^{-1}) \\ p(\boldsymbol{y}|\boldsymbol{x}) = N(\boldsymbol{y}|A\boldsymbol{x} + \boldsymbol{b}, L^{-1}) \end{cases}$$

同時分布を求めるため、$\boldsymbol{z} = \begin{pmatrix} \boldsymbol{x} \\ \boldsymbol{y} \end{pmatrix}$ とおく。

$$p(\boldsymbol{z}) = p(\boldsymbol{x}, \boldsymbol{y}) = p(\boldsymbol{x})p(\boldsymbol{y})p(\boldsymbol{y}|\boldsymbol{x})$$

なので、対数をとって、

$$
\begin{aligned}
\log p(\boldsymbol{z}) &= \log p(\boldsymbol{x}) + \log p(\boldsymbol{y}|\boldsymbol{x}) \\
&\propto {}^{T}(\boldsymbol{x}-\boldsymbol{\mu})\Lambda(\boldsymbol{x}-\boldsymbol{\mu}) + {}^{T}(\boldsymbol{y}-A\boldsymbol{x}-\boldsymbol{b})L(\boldsymbol{y}-A\boldsymbol{x}-\boldsymbol{b}) + const \\
&= {}^{T}\boldsymbol{x}\Lambda\boldsymbol{x} + {}^{T}\boldsymbol{x}\,{}^{T}ALA\boldsymbol{x} + {}^{T}\boldsymbol{y}L\boldsymbol{y} - {}^{T}\boldsymbol{x}\,{}^{T}AL\boldsymbol{y} - {}^{T}\boldsymbol{y}LA\boldsymbol{x} \\
&\quad - {}^{T}\boldsymbol{x}\Lambda\boldsymbol{\mu} - {}^{T}\boldsymbol{\mu}\Lambda\boldsymbol{x} + {}^{T}\boldsymbol{x}\,{}^{T}AL\boldsymbol{b} + {}^{T}\boldsymbol{b}LA\boldsymbol{x} - {}^{T}\boldsymbol{y}L\boldsymbol{b} - {}^{T}\boldsymbol{b}L\boldsymbol{y} + const \\
&= {}^{T}\begin{pmatrix}\boldsymbol{x}\\ \boldsymbol{y}\end{pmatrix}\begin{pmatrix}\Lambda + {}^{T}ALA & -{}^{T}AL \\ -LA & L\end{pmatrix}\begin{pmatrix}\boldsymbol{x}\\ \boldsymbol{y}\end{pmatrix} \\
&\quad - {}^{T}\begin{pmatrix}\boldsymbol{x}\\ \boldsymbol{y}\end{pmatrix}\begin{pmatrix}\Lambda\boldsymbol{\mu} - {}^{T}AL\boldsymbol{b} \\ L\boldsymbol{b}\end{pmatrix} - {}^{T}\begin{pmatrix}\Lambda\boldsymbol{\mu} - {}^{T}AL\boldsymbol{b} \\ L\boldsymbol{b}\end{pmatrix}\begin{pmatrix}\boldsymbol{x}\\ \boldsymbol{y}\end{pmatrix} + const \\
&= {}^{T}\boldsymbol{z}R\boldsymbol{z} - {}^{T}\boldsymbol{z}B - {}^{T}B\boldsymbol{z} + const \\
&= {}^{T}(\boldsymbol{z} - R^{-1}B)R(\boldsymbol{z} - R^{-1}B) + const
\end{aligned}
$$

より、$\boldsymbol{z}$ のガウス分布の精度行列は、

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

したがって、分散共分散行列は

$$\Sigma_{\boldsymbol{z}} = R^{-1} = \begin{pmatrix}\Lambda^{-1} & \Lambda^{-1}\,{}^{T}A \\ A\Lambda^{-1} & L^{-1} + A\Lambda^{-1}\,{}^{T}A\end{pmatrix}$$

平均は、

$$
\begin{aligned}
\boldsymbol{\mu}_{\boldsymbol{z}} &= \begin{pmatrix}\Lambda^{-1} & \Lambda^{-1}\,{}^{T}A \\ A\Lambda^{-1} & L^{-1} + A\Lambda^{-1}\,{}^{T}A\end{pmatrix}\begin{pmatrix}\Lambda\boldsymbol{\mu} - {}^{T}AL\boldsymbol{b} \\ L\boldsymbol{b}\end{pmatrix} \\
&= \begin{pmatrix}\boldsymbol{\mu} \\ A\boldsymbol{\mu} + \boldsymbol{b}\end{pmatrix}
\end{aligned}
$$