「パターン認識と機械学習」
2.3.3 ガウス変数に対するベイズの定理

小野寺喜行

ガウス周辺分布／条件付き分布

- あるガウス周辺分布$p(\mathbf{x})$
- 平均が$\mathbf{x}$の線形関数、共分散は$\mathbf{x}$とは独立なガウス条件付き分布$p(\mathbf{y}|\mathbf{x})$

$$\Downarrow$$

周辺分布$p(\mathbf{y})$、条件付き分布$p(\mathbf{x}|\mathbf{y})$を求める

ガウス周辺分布／条件付き分布

周辺分布と条件付き分布を以下のようにおく

$$p(\mathbf{x}) = N\left(\mathbf{x} \mid \boldsymbol{\mu}, \boldsymbol{\Lambda}^{-1}\right)$$

$$p(\mathbf{y} \mid \mathbf{x}) = N\left(\mathbf{y} \mid \mathbf{A}\mathbf{x} + \mathbf{b}, \mathbf{L}^{-1}\right)$$

$\boldsymbol{\mu}, \mathbf{A}, \mathbf{b}$：平均に関係したパラメータ　$\boldsymbol{\Lambda}, \mathbf{L}$：精度行列

同時分布を定義

x,**y**上の同時分布で、次のように定義

$$\mathbf{z} = \begin{pmatrix} \mathbf{x} \\ \mathbf{y} \end{pmatrix}$$

同時分布の対数

$$\begin{aligned}\ln p(\mathbf{z}) &= \ln p(\mathbf{x}) + \ln p(\mathbf{y} \mid \mathbf{x}) \\ &= -\frac{1}{2}\left(\mathbf{x} - \boldsymbol{\mu}\right)^{\mathrm{T}} \boldsymbol{\Lambda} \left(\mathbf{x} - \boldsymbol{\mu}\right) \\ &\quad -\frac{1}{2}\left(\mathbf{y} - \mathbf{A}\mathbf{x} - \mathbf{b}\right)^{\mathrm{T}} \mathbf{L} \left(\mathbf{y} - \mathbf{A}\mathbf{x} - \mathbf{b}\right) \\ &\quad + \mathrm{const}\end{aligned}$$

精度行列を求める

同時分布の対数の2次の項

$$-\frac{1}{2}\mathbf{x}^{\mathrm{T}}\left(\boldsymbol{\Lambda}+\mathbf{A}^{\mathrm{T}}\mathbf{L}\mathbf{A}\right)\mathbf{x}-\frac{1}{2}\mathbf{y}^{\mathrm{T}}\mathbf{L}\mathbf{y}+\frac{1}{2}\mathbf{y}^{\mathrm{T}}\mathbf{L}\mathbf{A}\mathbf{x}+\frac{1}{2}\mathbf{x}^{\mathrm{T}}\mathbf{A}^{\mathrm{T}}\mathbf{L}\mathbf{y}$$

$$=-\frac{1}{2}\begin{pmatrix}\mathbf{x}\\ \mathbf{y}\end{pmatrix}^{\mathrm{T}}\begin{pmatrix}\mathbf{A}+\mathbf{A}^{\mathrm{T}}\mathbf{L}\mathbf{A} & -\mathbf{A}^{\mathrm{T}}\mathbf{L}\\ -\mathbf{L}\mathbf{A} & \mathbf{L}\end{pmatrix}\begin{pmatrix}\mathbf{x}\\ \mathbf{y}\end{pmatrix}$$

$$=-\frac{1}{2}\mathbf{x}^{T}\mathbf{R}\mathbf{z}$$

精度行列を求める

$\mathbf{z}$上のガウス分布の精度行列（逆共分散行列）

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

共分散行列

$$\mathrm{cov}[\mathbf{z}] = \mathbf{R}^{-1} = \begin{pmatrix} \mathbf{\Lambda}^{-1} & \mathbf{\Lambda}^{-1}\mathbf{A}^{\mathrm{T}} \\ \mathbf{A}\mathbf{\Lambda}^{-1} & \mathbf{L}^{-1} + \mathbf{A}\mathbf{\Lambda}^{-1}\mathbf{A}^{\mathrm{T}} \end{pmatrix}$$

平均を求める

同時分布の対数の線形の項

$$\mathbf{x}^{\mathrm{T}}\Lambda\boldsymbol{\mu} - \mathbf{x}^{\mathrm{T}}\mathbf{A}^{\mathrm{T}}\mathbf{L}\mathbf{b} + \mathbf{y}^{\mathrm{T}}\mathbf{L}\mathbf{b} = \begin{pmatrix} \mathbf{x} \\ \mathbf{y} \end{pmatrix}^{\mathrm{T}} \begin{pmatrix} \Lambda\boldsymbol{\mu} - \mathbf{A}^{\mathrm{T}}\mathbf{L}\mathbf{b} \\ \mathbf{L}\mathbf{b} \end{pmatrix}$$

zの平均

$$\mathbf{E}[\mathbf{z}] = \mathbf{R}^{-1} \begin{pmatrix} \Lambda\boldsymbol{\mu} - \mathbf{A}^{\mathrm{T}}\mathbf{L}\mathbf{b} \\ \mathbf{L}\mathbf{b} \end{pmatrix} = \begin{pmatrix} \boldsymbol{\mu} \\ \mathbf{A}\boldsymbol{\mu} + \mathbf{b} \end{pmatrix}$$

平均と共分散

$\mathbf{x}$を周辺化した周辺分布$p(\mathbf{y})$の平均と共分散

平均

$$\mathbf{E}[\mathbf{y}] = \mathbf{A}\boldsymbol{\mu} + \mathbf{b}$$

共分散

$$\mathrm{cov}[\mathbf{y}] = \mathbf{L}^{-1} + \mathbf{A}\boldsymbol{\Lambda}^{-1}\mathbf{A}^{\mathrm{T}}$$

平均と共分散

条件付き分布$p(\mathbf{x}|\mathbf{y})$の平均と共分散

平均

$$\mathbf{E}\left[\mathbf{x} \mid \mathbf{y}\right] = \left(\mathbf{\Lambda} + \mathbf{A}^{\mathrm{T}}\mathbf{L}\mathbf{A}\right)^{-1}\left\{\mathbf{A}^{\mathrm{T}}\mathbf{L}\left(\mathbf{y} - \mathbf{b}\right) + \mathbf{\Lambda}\boldsymbol{\mu}\right\}$$

共分散

$$\mathrm{cov}\left[\mathbf{x} \mid \mathbf{y}\right] = \left(\mathbf{\Lambda} + \mathbf{A}^{\mathrm{T}}\mathbf{L}\mathbf{A}\right)^{-1}$$

ガウス分布の周辺分布と条件付き分布

xの周辺ガウス分布と、**x**が与えられたときのyの条件付き分布を以下のようにおく

$$p(\mathbf{x}) = N\left(\mathbf{x} \mid \boldsymbol{\mu}, \boldsymbol{\Lambda}^{-1}\right)$$

$$p(\mathbf{y} \mid \mathbf{x}) = N\left(\mathbf{y} \mid \mathbf{A}\mathbf{x} + \mathbf{b}, \mathbf{L}^{-1}\right)$$

⇩ このとき

$$p(\mathbf{y}) = N\left(\mathbf{y} \mid \mathbf{A}\boldsymbol{\mu} + \mathbf{b}, \mathbf{L}^{-1} + \mathbf{A}\boldsymbol{\Lambda}^{-1}\mathbf{A}^{\mathrm{T}}\right)$$

$$p(\mathbf{x} \mid \mathbf{y}) = N\left(\mathbf{x} \mid \boldsymbol{\Sigma}\left\{\mathbf{A}^{\mathrm{T}}\mathbf{L}(\mathbf{y} - \mathbf{b}) + \boldsymbol{\Lambda}\boldsymbol{\mu}\right\}, \boldsymbol{\Sigma}\right)$$

ただし $\Sigma = \left(\boldsymbol{\Lambda} + \mathbf{A}^{\mathrm{T}}\mathbf{L}\mathbf{A}\right)^{-1}$