統計的機械学習
第十一回　レポート

37-196360　森田涼介

2019 年 7 月 11 日

宿題

観測データ $\mathscr{D}_{1:n} = \{\boldsymbol{x}_i,\, y_i\}_{i=1}^{n}$ $(y_i \in \mathbb{R})$ について，以下を仮定する。

$$y_i = f(\boldsymbol{x}_i) \quad (i = 1,\, \cdots,\, n) \tag{1}$$
$$f \sim \mathscr{G}\mathscr{P}(f|m, \kappa) \tag{2}$$

これより，

$$y_i = f(\boldsymbol{x}_i) \sim \mathscr{N}(m(\boldsymbol{x}_i),\, \kappa(\boldsymbol{x}_i,\, \boldsymbol{x}_i)) \tag{3}$$

である。このとき，未知の入力 $\boldsymbol{x}_*$ に対する $f(\boldsymbol{x})$ の分布は，

$$p(f(\boldsymbol{x}_*)|\mathscr{D}_{1:n}) = \mathscr{N}\left(f(\boldsymbol{x}_*)|\mu_n(\boldsymbol{x}_*), \sigma_n^2(\boldsymbol{x}_*)\right) \tag{4}$$

となる。$\mu_n(\boldsymbol{x}_*)$, $\sigma_n^2(\boldsymbol{x}_*)$ を解析的に求める。

$\boldsymbol{y} \in \mathbb{R}^n$ は平均 $\boldsymbol{\mu}$，分散 $\boldsymbol{\Sigma}$ の正規分布 $\mathscr{N}(\boldsymbol{\mu},\, \boldsymbol{\Sigma})$ に従うとする。また，$\boldsymbol{y}_1 \in \mathbb{R}^l$, $\boldsymbol{y}_2 \in \mathbb{R}^m$, $(l + m = n)$ を用いて，

$$\boldsymbol{y} = \begin{bmatrix} \boldsymbol{y}_1 \\ \boldsymbol{y}_2 \end{bmatrix} \tag{5}$$

と分割して定義する。これに対し，$\boldsymbol{\mu}$, $\boldsymbol{\Sigma}$ も分割して，

$$\boldsymbol{\mu} = \begin{bmatrix} \boldsymbol{\mu}_1 \\ \boldsymbol{\mu}_2 \end{bmatrix} \tag{6}$$
$$\boldsymbol{\Sigma} = \begin{bmatrix} \boldsymbol{\Sigma}_{11} & \boldsymbol{\Sigma}_{12} \\ \boldsymbol{\Sigma}_{21} & \boldsymbol{\Sigma}_{22} \end{bmatrix} \tag{7}$$

とする。$\boldsymbol{y}_1$ が与えられた下での $\boldsymbol{y}_2$ の条件付き確率を次のように仮定する。

$$p(\boldsymbol{y}_2|\boldsymbol{y}_1) = \mathscr{N}(\boldsymbol{y}_2|\hat{\boldsymbol{\mu}}_2,\, \hat{\boldsymbol{\Sigma}}_2) \tag{8}$$

このとき，次式が成立する。

$$\hat{\boldsymbol{\mu}}_2 = \boldsymbol{\mu}_2 + \boldsymbol{\Sigma}_{21}\boldsymbol{\Sigma}_{11}^{-1}(\boldsymbol{y}_1 - \boldsymbol{\mu}_1) \tag{9}$$
$$\hat{\boldsymbol{\Sigma}}_2 = \boldsymbol{\Sigma}_{22} - \boldsymbol{\Sigma}_{21}\boldsymbol{\Sigma}_{11}^{-1}\boldsymbol{\Sigma}_{12} \tag{10}$$

いま，

$$\boldsymbol{y}_{1:n} = \begin{bmatrix} y_1 & \cdots & y_n \end{bmatrix}^{\mathrm{T}} \tag{11}$$
$$\boldsymbol{m}(\boldsymbol{x}_{1:n}) = \begin{bmatrix} m(\boldsymbol{x}_1) & \cdots & m(\boldsymbol{x}_n) \end{bmatrix}^{\mathrm{T}} \tag{12}$$
$$\boldsymbol{K}(\boldsymbol{x}_{1:n}) = \begin{bmatrix} \kappa(\boldsymbol{x}_1,\, \boldsymbol{x}_1) & \cdot & \kappa(\boldsymbol{x}_1,\, \boldsymbol{x}_n) \\ \vdots & \ddots & \vdots \\ \kappa(\boldsymbol{x}_n,\, \boldsymbol{x}_1) & \cdot & \kappa(\boldsymbol{x}_n,\, \boldsymbol{x}_n) \end{bmatrix} \tag{13}$$

と表すと，

$$\boldsymbol{y}_{1:n} = f(\boldsymbol{x}_{1:n}) \sim \mathscr{N}(\boldsymbol{m}(\boldsymbol{x}_{1:n}),\, \boldsymbol{K}(\boldsymbol{x}_{1:n})) \tag{14}$$

であり，また，

$$y_* = f(\boldsymbol{x}_*) \sim \mathcal{N}(m_*,\ \kappa(\boldsymbol{x}_*,\ \boldsymbol{x}_*)) \tag{15}$$

である。このとき，

$$\boldsymbol{\kappa}(\boldsymbol{x}_*,\ \boldsymbol{x}_{1:n}) = \begin{bmatrix}\kappa(\boldsymbol{x}_*,\ \boldsymbol{x}_1) & \cdots & \kappa(\boldsymbol{x}_*,\ \boldsymbol{x}_n)\end{bmatrix}^{\mathrm{T}} \tag{16}$$

とおくと，

$$\begin{bmatrix}\boldsymbol{y}_{1:n} \\ y_*\end{bmatrix} \sim \mathcal{N}\left(\begin{bmatrix}\boldsymbol{m}(\boldsymbol{x}_{1:n}) \\ m(\boldsymbol{x}_*)\end{bmatrix},\ \begin{bmatrix}\boldsymbol{K}(\boldsymbol{x}_{1:n}) & \boldsymbol{\kappa}(\boldsymbol{x}_*,\ \boldsymbol{x}_{1:n}) \\ \boldsymbol{\kappa}(\boldsymbol{x}_*,\ \boldsymbol{x}_{1:n})^{\mathrm{T}} & \kappa(\boldsymbol{x}_*,\ \boldsymbol{x}_*)\end{bmatrix}\right) \tag{17}$$

と表せる。

$$p(y_*|\boldsymbol{y}_{1:n}) = p(f(\boldsymbol{x}_*)|\mathcal{D}_{1:n}) = \mathcal{N}\left(f(\boldsymbol{x}_*)|\mu_n(\boldsymbol{x}_*), \sigma_n^2(\boldsymbol{x}_*)\right) \tag{18}$$

とおき，$\boldsymbol{y}_1$ を $\boldsymbol{y}_{1:n}$ に，$\boldsymbol{y}_2$ を y_* に置き換えて考える。すると，式 (5) – (10) より，次式が成立する。

$$\mu_n(\boldsymbol{x}_*) = m(\boldsymbol{x}_*) + \boldsymbol{\kappa}(\boldsymbol{x}_*,\ \boldsymbol{x}_{1:n})^{\mathrm{T}}\boldsymbol{K}(\boldsymbol{x}_{1:n})^{-1}(\boldsymbol{y}_{1:n} - \boldsymbol{m}(\boldsymbol{x}_{1:n})) \tag{19}$$

$$\sigma_n^2(\boldsymbol{x}_*) = \kappa(\boldsymbol{x}_*,\ \boldsymbol{x}_*) - \boldsymbol{\kappa}(\boldsymbol{x}_*,\ \boldsymbol{x}_{1:n})^{\mathrm{T}}\boldsymbol{K}(\boldsymbol{x}_{1:n})^{-1}\boldsymbol{\kappa}(\boldsymbol{x}_*,\ \boldsymbol{x}_{1:n}) \tag{20}$$