ガウス過程回帰入門

瀬戸 道生（防衛大学校・数学教育室）

Real, Complex and Functional Analysis Seminar
2020 年 12 月 13 日

学習と過学習

学習とは

$$\text{データ（入力）} \xrightarrow{\text{学習}} \text{予測（出力）}$$

- 繰り返し学習することで，予測精度の向上が期待される．

過学習とは

- 予測がデータに依存し過ぎている状態

ガウス過程回帰とは

- 機械学習 (Machine Learning) で非常に有力かつ有名な手法

ガウス過程回帰の使い方

工学上の問題

ある装置 $x \to y$ があり，
データ x を入力したとき，$\hat{z}$ を観測したとする．
このとき，新たなデータ x_* に対する出力 $z(x_*)$ を，
誤差を考慮して

$$z = y + \varepsilon$$

の形の関数で推定したい．

ガウス過程回帰の使い方

工学上の問題

ある装置 $x \to y$ があり，
データ x を入力したとき，$\hat{z}$ を観測したとする．
このとき，新たなデータ x_* に対する出力 $z(x_*)$ を，
誤差を考慮して

$$z = y + \varepsilon$$

の形の関数で推定したい．

ガウス過程回帰の導入

- y, ε が独立に多次元ガウス分布に従うと仮定する．
- このとき，すべての計算がガウス分布の範囲で納まる．

多次元ガウス分布

μ を $\mathbb{R}^n$ のベクトル，Σ を $n \times n$ の正定値行列とする．

$$N(\boldsymbol{x}; \boldsymbol{\mu}, \Sigma) := \frac{1}{(2\pi)^{n/2}(\det \Sigma)^{1/2}} \exp\left(-\frac{1}{2}\langle \Sigma^{-1}(\boldsymbol{x} - \boldsymbol{\mu}), \boldsymbol{x} - \boldsymbol{\mu}\rangle\right)$$

（多次元ガウス分布（の確率密度関数））

- μ は平均ベクトル，Σ は共分散行列とよばれる．

多次元ガウス分布

μ を $\mathbb{R}^n$ のベクトル，Σ を $n \times n$ の正定値行列とする．

$$N(\boldsymbol{x}; \boldsymbol{\mu}, \Sigma) := \frac{1}{(2\pi)^{n/2}(\det \Sigma)^{1/2}} \exp\left(-\frac{1}{2} \langle \Sigma^{-1}(\boldsymbol{x} - \boldsymbol{\mu}), \boldsymbol{x} - \boldsymbol{\mu} \rangle \right)$$

（多次元ガウス分布（の確率密度関数））

- μ は平均ベクトル，Σ は共分散行列とよばれる．

同時分布

$\boldsymbol{y} \sim N(\mathbf{0}, \Sigma)$ かつ $\varepsilon \sim N(\mathbf{0}, \beta^{-1} I)$ のとき（独立性も仮定する），

$$\boldsymbol{z} = \boldsymbol{y} + \varepsilon \sim N(\mathbf{0}, \Sigma + \beta^{-1} I).$$

- データを Σ の中に織り込む方法がある（カーネル法）．

カーネル法

カーネル関数

X：集合，$k : X \times X$ 上の関数

k はカーネル関数 $\overset{def}{\Leftrightarrow}$ (i) and (ii)

(i) $\forall x, y \in X \;\; k(x, y) = k(y, x)$（対称性）

(ii) $\forall n \in \mathbb{N}$, $\forall c_1, \ldots, c_n \in \mathbb{R}$, $\forall x_1, \ldots, x_n \in X$

$$\sum_{i,j=1}^{n} c_i c_j k(x_i, x_j) \geq 0 \quad \text{（半正定値性）}$$

カーネル法

カーネル関数

X : **集合，** $k : X \times X$ **上の関数**

k **はカーネル関数** $\overset{def}{\Leftrightarrow}$ (i) and (ii)

(i) $\forall x, y \in X \ \ k(x,y) = k(y,x)$ **（対称性）**

(ii) $\forall n \in \mathbb{N}, \forall c_1, \ldots, c_n \in \mathbb{R}, \forall x_1, \ldots, x_n \in X$

$$\sum_{i,j=1}^{n} c_i c_j k(x_i, x_j) \geq 0 \quad \text{（半正定値性）}$$

再生核ヒルベルト空間

$k_x := k(\cdot, x)$ **から内積** $\langle k_y, k_x \rangle_{\mathcal{H}_k} = k(x,y)$ **が構成できる．**

カーネルトリック

データ $x_1, \ldots, x_n \in X$ **を** $x_j \mapsto k_{x_j}$ **と変換する．**

ガウス過程回帰１

仮定

$\forall n \in \mathbb{N},\ \forall \boldsymbol{x} = (x_1, \dots, x_n)\ (x_i \neq x_j)$ に対し，

$$K_n = K_n(\boldsymbol{x}) = \begin{pmatrix} k(x_1, x_1) & \cdots & k(x_1, x_n) \\ \vdots & \ddots & \vdots \\ k(x_n, x_1) & \cdots & k(x_n, x_n) \end{pmatrix} > 0 \text{（固有値がすべて 0 より大）}$$

を仮定する．例えば，K がガウスカーネルなら OK

$$k(x, y) = \exp(-\gamma \|x - y\|^2) \quad (\gamma > 0) \quad \text{（ガウスカーネル）}$$

記号

新しいデータ $\boldsymbol{x}_* = (x_{n+1}, \dots, x_{n+m})$ に対し，

$$K_{n+m} = K_{n+m}(\boldsymbol{x}, \boldsymbol{x}_*) = \begin{pmatrix} K_n & \boldsymbol{k} \\ \boldsymbol{k}^\top & K_m \end{pmatrix} \quad \text{と定める．}$$

ガウス過程回帰 2

$\mathbb{R}^n$ の中を多次元ガウス分布に従って動くベクトルを構成する．

1. $\vee\{k_{x_j}\}_{j=1}^n$ の正規直交基底を $\{\varphi_j\}_{j=1}^n$ とする．
2. $\Phi = (\varphi_i(x_j))$ とおくと，

$$k(x_i, x_j) = \langle k_{x_i}, k_{x_j}\rangle_{\mathcal{H}_k} = \sum_{\ell=1}^{n} \varphi_\ell(x_i)\varphi_\ell(x_j) = (\Phi\Phi^\top \text{ の } i, j \text{ 成分}).$$

3. $\boldsymbol{w} \sim N(\boldsymbol{0}, \alpha^{-1} I_n)$ に対し，$y(\boldsymbol{w}, x_i) = \sum_{j=1}^n \varphi_j(x_i) w_j$ と定めると

$$\boldsymbol{y}(\boldsymbol{w}, \boldsymbol{x}) = \begin{pmatrix} y(\boldsymbol{w}, x_1) \\ \vdots \\ y(\boldsymbol{w}, x_n) \end{pmatrix} \sim N(\boldsymbol{0}, \alpha^{-1}\Phi\Phi^\top) = N(\boldsymbol{0}, \alpha^{-1} K_n)$$

が成り立つ．

ガウス過程回帰 3

ここまでのまとめ

$\boldsymbol{y} \sim N(\mathbf{0}, K_n)$ かつ $\varepsilon \sim N(\mathbf{0}, \beta^{-1} I_n)$ のとき，

$$\boldsymbol{z} = \boldsymbol{y} + \varepsilon \sim N(\mathbf{0}, K_n + \beta^{-1} I_n)$$

が成り立つ．

ガウス過程回帰３

ここまでのまとめ

$\boldsymbol{y} \sim N(\mathbf{0}, K_n)$ かつ $\boldsymbol{\varepsilon} \sim N(\mathbf{0}, \beta^{-1} I_n)$ のとき，

$$\boldsymbol{z} = \boldsymbol{y} + \boldsymbol{\varepsilon} \sim N(\mathbf{0}, K_n + \beta^{-1} I_n)$$

が成り立つ．

条件付き分布

$\widehat{\boldsymbol{z}}$ を観測した後に $\boldsymbol{t}$ が従う分布を

$$p(\boldsymbol{t} \mid \widehat{\boldsymbol{z}}) = C \exp\left(-\frac{1}{2}\left\langle (K_{n+m} + \beta^{-1} I_{n+m})^{-1} \begin{pmatrix} \widehat{\boldsymbol{z}} \\ \boldsymbol{t} \end{pmatrix}, \begin{pmatrix} \widehat{\boldsymbol{z}} \\ \boldsymbol{t} \end{pmatrix} \right\rangle\right)$$

と定める．ここで，C は $\boldsymbol{t}$ で積分して 1 になるような定数（付録を参照）．

ガウス過程回帰 4

$$
\begin{aligned}
p(\boldsymbol{t} \mid \widehat{\boldsymbol{z}}) &= C \exp\left(-\frac{1}{2} \left\langle (K_{n+m} + \beta^{-1} I_{n+m})^{-1} \begin{pmatrix} \widehat{\boldsymbol{z}} \\ \boldsymbol{t} \end{pmatrix}, \begin{pmatrix} \widehat{\boldsymbol{z}} \\ \boldsymbol{t} \end{pmatrix} \right\rangle \right) \\
&= C \exp\left(-\frac{1}{2} \left\langle \begin{pmatrix} L_{11} & L_{12} \\ L_{21} & L_{22} \end{pmatrix} \begin{pmatrix} \widehat{\boldsymbol{z}} \\ \boldsymbol{t} \end{pmatrix}, \begin{pmatrix} \widehat{\boldsymbol{z}} \\ \boldsymbol{t} \end{pmatrix} \right\rangle \right) \\
&= \widetilde{C} \exp\left(-\frac{1}{2} \langle L_{22}(\boldsymbol{t} - \boldsymbol{a}), \boldsymbol{t} - \boldsymbol{a} \rangle \right) \qquad \text{（∵平方完成）}
\end{aligned}
$$

ここで，$\boldsymbol{a}$（平均）は

$$
\boldsymbol{a} = \begin{pmatrix} \sum_{j=1}^{n} c_j k_{x_j}(x_{n+1}) \\ \vdots \\ \sum_{j=1}^{n} c_j k_{x_j}(x_{n+m}) \end{pmatrix}, \qquad \begin{pmatrix} c_1 \\ \vdots \\ c_n \end{pmatrix} = (K_n + \beta^{-1} I_n)^{-1} \widehat{\boldsymbol{z}}.
$$

により与えられる．この $\boldsymbol{a}$ を $z(x_*)$ の予測として採用する．

付録（条件付き分布）

$x \in [x_0, x_0+h]$ のときに $y \in A$ となる条件付き確率

$$P(y \in A \mid x \in [x_0, x_0+h]) = \frac{P((x,y) \in [x_0, x_0+h] \times A)}{P((x,y) \in [x_0, x_0+h] \times \mathbb{R})}$$

を分布関数の言葉で書き直せば，

$$\begin{aligned} P(y \in A \mid x \in [x_0, x_0+h]) &= \frac{\int_A \left(\int_{x_0}^{x_0+h} p(x,y)\ dx \right)\ dy}{\int_{\mathbb{R}} \left(\int_{x_0}^{x_0+h} p(x,y)\ dx \right)\ dy} \\ &= \frac{\int_A \left(\frac{1}{h} \int_{x_0}^{x_0+h} p(x,y)\ dx \right)\ dy}{\int_{\mathbb{R}} \left(\frac{1}{h} \int_{x_0}^{x_0+h} p(x,y)\ dx \right)\ dy} \\ &\to \frac{\int_A p(x_0,y)\ dy}{\int_{\mathbb{R}} p(x_0,y)\ dy} \quad (h \to 0). \end{aligned}$$

このようにして，$x = x_0$ のときの条件付き分布（確率密度関数）$Cp(x_0, y)$ $(1/C = \int_{\mathbb{R}} p(x_0,y)\ dy)$ を得る．

CM

瀬戸・畑中・伊吹「機械学習のための関数解析入門」

内容

- 理工系学部３年生以上を対象とした教科書.
- 数学専攻４年生の就職対策に最適.

内田老鶴圃から 2021 年３月発売予定