GP and RKHS

米田亮介

March 17, 2021

---

1 準備

Definition 1.1 (GP)**.** $\{X(t) \mid 0 \leq t \leq T\}$ が確率過程のとき、$0 \leq t \leq T$ から任意に有限個 $t_1, t_2, \ldots t_n$ を選んだときに、$(X(t_1), X(t_2), \ldots, X(t_n))^\top$ が n 次元の正規分布に従うならば $\{X(t)\}$ を**ガウス過程** *(Gaussian process, GP)* という。

f が GP のときに、各 x, x' に対して

$$m(x) = \mathbb{E}[f(x)], \quad k(x, x') = \mathbb{E}[(f(x) - m(x))(f(x') - m(x'))] \tag{1.1}$$

で平均 m と共分散 k を定めて、

$$f \sim \mathcal{GP}(m, k) \tag{1.2}$$

と書くことにする。

GP を用いた回帰を復習する。観測データ $\mathcal{D} = \{(x_1, y_1), \ldots, (x_N, y_N)\}$ について x と y の間に $y = f(x)$ の関係があり、GP$f \sim \mathcal{GP}(m, k)$ から生成されているとする。このとき事後分布は

$$f \mid \mathcal{D} \sim \mathcal{GP}(\bar{m}, \bar{k}) \tag{1.3}$$

でガウス過程になる。ここで、$k_{XX} = \{k(x_i, x_j)\}_{i,j}, k_{xX} = (k(x, x_1), \ldots, k(x, x_N)), k_{Xx'} = k_{x'X}^\top$ とおいて

$$\bar{m}(x) = m(x) + k_{xX} k_{XX}^{-1} (\boldsymbol{y} - m_X), \tag{1.4}$$
$$\bar{k}(x, x') = k(x, x') - k_{xX} k_{XX}^{-1} k_{Xx'} \tag{1.5}$$

である。

Definition 1.2 (RKHS)**.** 集合 $\mathcal{X}$ 上の**再生核ヒルベルト空間** *(reproducing kernel Hilbert spcae, RKHS)* とは、$\mathcal{X}$ 上の関数からなるヒルベルト空間 $\mathcal{H}$ で、任意の $x \in \mathcal{X}$ に対して $k_x \in \mathcal{H}$ があって

$$\langle f, k_x \rangle_{\mathcal{H}} = f(x) \tag{1.6}$$

が任意の $f \in \mathcal{H}$ に対して成り立つことである。$k(y, x) = k_x(y)$ により定まるカーネル k を $\mathcal{H}$ の**再生核**と呼ぶ。

再生核が一意であることと Moore–Aronszajn の定理によりカーネルと RKHS が 1 対 1 に対応することがわかる。

RKHS を導入することで GP 回帰は非常に見通しがよくなる。

Lemma 1.1. 観測データ $\mathcal{D} = \{(x_1, y_1), \dots, (x_N, y_N)\}$ について x と y の間に $y = f(x)$ の関係があり、f が平均 0 の $f \sim \mathcal{GP}(0, k)$ から生成されているとする。このとき、カーネル k に対応する *RKHS* を $\mathcal{H}_k$ とおくと

$$\overline{m} = \arg\min_{f \in \mathcal{H}_k} \|f\|_{\mathcal{H}_k}, \ \textit{subject to } f(x_i) = y_i, \quad i = 1, \dots, N, \tag{1.7}$$

で表される。

これは平均が RKHS$\mathcal{H}_k$ に入ることを意味している。そのためデータにのる関数の性質を決める上でカーネルを選び方は非常に重要である。

2 RKHS の性質

shift invariant なカーネル、すなわち、$k(x, y) = \Psi(x - y)$ なる関数 Ψ がある状況においては、RKHS が定まる。

Theorem 2.1. k を $\mathbb{R}^d$ 上のカーネルで、$k(x, y) = \Psi(x - y), \Psi \in C(\mathbb{R}^d) \cap L^1(\mathbb{R}^d)$ であるとする。このとき、カーネル k の *RKHS* は

$$\mathcal{H}_k = \{f \in L^2(\mathbb{R}^d) \cap C(\mathbb{R}^d) \mid \|f\|_{\mathcal{H}_k} < \infty\} \tag{2.1}$$

であり、$f, g \in \mathcal{H}_k$ に対して内積は

$$\langle f, g \rangle_{\mathcal{H}_k} = \frac{1}{(2\pi)^{d/2}} \int_{\mathbb{R}^d} \frac{\mathcal{F}[f](\omega)\overline{\mathcal{F}[g](\omega)}}{\mathcal{F}[\Psi](\omega)} \mathrm{d}\omega \tag{2.2}$$

である。

ここで関数 f に対する Fourier 変換を

$$\mathcal{F}[f](\omega) = \frac{1}{(2\pi)^{d/2}} \int_{\mathbb{R}^d} f(x) e^{-ix^\intercal \omega} \mathrm{d}x \tag{2.3}$$

で定めた。

具体例を見ていこう。

Example 2.1 (RBF カーネル)**.** *RBF* カーネルは

$$k_\gamma(x, y) = \Psi(x - y) = \exp(-\|x - y\|^2 / \gamma^2) \tag{2.4}$$

で与えられる。ここでガウス関数の *Fourier* 変換はガウス関数であったことを思い出そう。すなわち、

$$\mathcal{F}[\Psi](\omega) = \frac{\gamma^d}{2^{d/2}} \exp(-\gamma^2 \|\omega\|^2 / 4) \tag{2.5}$$

である。よって *RBF* カーネルの *RKHS* は

$$\mathcal{H}_{k_\gamma} = \{f \in L^2(\mathbb{R}^d) \cap C(\mathbb{R}^d) \mid \|f\|_{\mathcal{H}_{k_\gamma}} < \infty\}, \tag{2.6}$$

$$\|f\|^2_{\mathcal{H}_{k_\gamma}} = \frac{1}{\gamma^d \pi^{d/2}} \int_{\mathbb{R}^d} |\mathcal{F}[f](\omega)|^2 \exp(\gamma^2 \|\omega\|^2/4) \mathrm{d}\omega \tag{2.7}$$

である。特に $\|f\|_{\mathcal{H}_{k_\gamma}}$ が収束するとき、$\mathcal{F}[f](\omega)$ は ω は指数関数的に減衰することがわかる。よって *Paley–Wiener* の定理から $f \in \mathcal{H}_{k_\gamma}$ は解析的であることがわかる。これが *RBF* カーネルから生成される *GP* が C^∞ 級である、と言われることの直感的な説明である。*(*より正確には *GP* のサンプルパスは確率 *0* で $\mathcal{H}_{k_\gamma}$ に属することが示せてしまうのでもう少し高度な数学を用いて議論する必要がある。*)*

Example 2.2 (Laplace カーネル)**.** *Laplace* カーネルは

$$k_\alpha(x, y) = \Psi(x - y) = \exp(-\alpha|x - y|) \tag{2.8}$$

で与えられる。*Fourier* 変換を行えば

$$\mathcal{F}[\Psi](\omega) = \sqrt{\frac{2}{\pi}} \frac{\alpha}{\omega^2 + \alpha^2} \tag{2.9}$$

であるから、*RKHS* は

$$\mathcal{H}_{k_\alpha} = \{f \in L^2(\mathbb{R}^d) \cap C(\mathbb{R}^d) \mid \|f\|_{\mathcal{H}_{k_\alpha}} < \infty\}, \tag{2.10}$$

$$\|f\|^2_{\mathcal{H}_{k_\gamma}} = \frac{1}{\pi} \int_{\mathbb{R}^d} |\mathcal{F}[f](\omega)|^2 (\omega^2 + \alpha^2) \mathrm{d}\omega \tag{2.11}$$

である。微分の *Fourier* 変換について $\mathcal{F}[f'](\omega) = -i\omega\mathcal{F}[f](\omega)$ が成り立つこと思い出すと、

$$\|f\|_{\mathcal{H}_{k_\alpha}} < \infty \Leftrightarrow \|f\|_{L^2} < \infty, \|Df\|_{L^2} < \infty \tag{2.12}$$

となるから $\mathcal{H}_{k_\alpha}$ は *Sobolev* 空間 $H^1(\mathbb{R})$ である。

Example 2.3 (Matérn カーネル)**.** 詳しくは書かないが、*Matérn* カーネル $k_{\alpha,h}$ に対応する *RKHS* は *Sobolev* 空間 $W_2^{\alpha+d/2}(\mathcal{X})$ に同型であることがわかる。

3 周期カーネル

周期カーネルは

$$k_\theta(x, y) = \Psi(x - y) = \theta_0 \exp(\theta_1 \cos(x - y)) \tag{3.1}$$

で与えられる。これまでの違いは考える定義域が $\mathbb{R}$ から $\mathbb{S}^1$ に移ったことである。これに伴って RKHS 内の Fourier 変換は Fourier 級数として理解する必要がある。(ここの正確な証明があるのかは知らないです。)

周期関数 f の Fourier 級数展開の第 n 係数を

$$\hat{f}_n = \frac{1}{\sqrt{2\pi}} \int_{\mathbb{S}^1} f(x) e^{-inx} \mathrm{d}x \tag{3.2}$$

で定める。すると周期カーネルの RKHS は

$$\mathcal{H}_{k_\theta} = \{ f \in L^2(\mathbb{S}^1) \cap C(\mathbb{S}^1) \mid \|f\|_{\mathcal{H}_{k_\theta}} < \infty \} \tag{3.3}$$

であり、内積は

$$\langle f, g \rangle_{\mathcal{H}_{k_\theta}} = \frac{1}{\sqrt{2\pi}} \sum_{n \in \mathbb{Z}} \frac{\hat{f}_n \overline{\hat{g}_n}}{\hat{\Psi}_n} \tag{3.4}$$

となるだろう。ノルムに着目してみよう。

$$\|f\|_{\mathcal{H}_{k_\theta}} = \frac{1}{\sqrt{2\pi}} \sum_{n \in \mathbb{Z}} \frac{|\hat{f}_n|^2}{\hat{\Psi}_n} < \infty \tag{3.5}$$

である。特に Ψ は解析的な関数であるから Paley–Wiener の定理によりある $C, \varepsilon > 0$ が存在して十分大きな n で

$$|\hat{\Psi}_n| \leq C e^{-\varepsilon |n|} \tag{3.6}$$

のように Fourier 係数が指数関数的な減衰を見せる。これより $|\hat{f}_n|$ も n によって指数関数的な減衰を見せることがわかる。また Paley–Wiener の定理を使えば $f \in \mathcal{H}_{k_\theta}$ が解析的であることがわかる。(これは全くもって厳密な証明ではなく、直感的な説明を書いているだけです。)

References

[1] M. Kanagawa, P. Hennig, D. Sejdinovic, and B. K. Sriperumbudur, Gaussian Processes and Kernel Methods: A Review on Connections and Equivalences, arXiv: 1807.02582.

[2] カーネル法入門, 福水健次, 朝倉書店, 2010.