# カーネル法

## なぜカーネルを計算することが高次元での内積に相当するのか

Daiki Tanaka

## カーネル

## 再生核ヒルベルト空間

- ヒルベルト空間
- 再生核
- カーネルトリック

# 実カーネル

集合 $\mathcal{X}$ について、実カーネルとは写像 $\mathcal{X} \times \mathcal{X} \longrightarrow \mathbb{R}$ である。
実カーネル：集合から 2 つ要素を渡すと、何かの実数を返す関数

## 定義：コーシー列

ベクトル空間 $X$ の点列 $(x_n)_{n=1}^{\infty}$ が

$$\lim_{n,m\to\infty} \|x_m - x_n\| = 0 \tag{1}$$

である時、コーシー列であるという。

## 定義：バナッハ空間

ベクトル空間 $X$ の任意のコーシー列 $(x_n)_{n=1}^{\infty}$ に対して、

$$\lim_{n\to\infty} \|x_n - x\| = 0 \text{ すなわち } \lim_{n\to\infty} \|x_n\| = x \tag{2}$$

となる $x \in X$ が保証される時、$X$ は完備であると呼ぶ。完備なノルム空間をバナッハ空間と呼ぶ。

## 定義：ヒルベルト空間

完備な内積空間 (内積によって誘導されるノルム空間) をヒルベルト空間と呼ぶ。

ヒルベルト空間：内積が定義されているかつ任意のコーシー列の収束先が中にあるような空間

# 再生性

以下ではヒルベルト空間が実数体 $\mathbb{R}$ 上で定義されているとし、$\mathcal{H}$ で定義される内積を $\langle \cdot, \cdot \rangle_{\mathcal{H}}$ で表す。

## 定義：再生核ヒルベルト空間

集合 $\mathcal{X}$ 上の再生核ヒルベルト空間 (Reproducing Kernrl Hilbert Space, 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 $x \in \mathcal{X}$ に対してある関数 $k_x \in \mathcal{H} : \mathcal{X} \to \mathbb{R}$ が存在して、以下の再生性を満たす物をいう。

$$\langle f, k_x \rangle_{\mathcal{H}} = f(x) \qquad (\forall f \in \mathcal{H} : \mathcal{X} \to \mathbb{R}) \tag{3}$$

まず $x$ が与えられて、$\mathcal{H}$ の中の任意の関数 $f$ に対して、それぞれの $x$ について上記を満たす関数 $k_x$ が存在する、というイメージ

## 再生核

上記の定義の $k_x$ に対して、$k(y, x) = k_x(y)$ によって定まるカーネル $k : \mathcal{X} \times \mathcal{X} \to \mathbb{R}$ を再生核と呼ぶ。

再生核はいろいろな $x$ について $k_x$ を集めたものというイメージ

## 命題 1

$\mathcal{X}$ 上の RKHS の再生核 $k$ は、$\mathcal{X}$ 上の正定値カーネルであり、RKHS の再生核 $k$ はただ一つ存在する。

## 定理 1：Moore-Aronszajn

集合 $\mathcal{X}$ 上の正定値カーネル $k$ に対し、$\mathcal{X}$ 上の RKHS$\mathcal{H}$ で以下の 3 つの条件を満たす物がただ一つ存在する。

(1) 任意の $x \in \mathcal{X}$ に対して $k(\cdot, x) \in \mathcal{H}$

(2) $\mathrm{Span}\{k(\cdot, x) \mid x \in \mathcal{X}\}$ は $\mathcal{H}$ 内で稠密

(3) $k$ は $\mathcal{H}$ の再生核である。すなわち、
$\langle f, k(\cdot, x)\rangle_{\mathcal{H}} = f(x) \qquad (\forall x \in \mathcal{X}, \forall f \in \mathcal{H})$

命題 1 と定理 1 を合わせると、集合 $\mathcal{X}$ 上の正定値カーネルと RKHS は 1 対 1 で対応し、正定値カーネルが RKHS に対応することがわかる。

# カーネルトリック：カーネルと RKHS の関係

集合 $\mathcal{X}$ 上に実正定値カーネル $k$ が与えられ、対応する RKHS(関数を要素としてもつ空間) を $\mathcal{H}_k$ とする。$\mathcal{X}$ から $\mathcal{H}_k$ への写像 $\Phi : \mathcal{X} \to \mathcal{H}_k$ を

$$x \mapsto k(\cdot, x) \tag{4}$$

によって定義する。すると、再生性の定義から、

$$\langle \Phi(x), \Phi(y) \rangle_{\mathcal{H}_k} = (\Phi(x))(y) = k(x, y) \tag{5}$$

これがまさにカーネル法の要となる仕組みになっていて、カーネルトリックと呼ばれる。
カーネル $k$ で計算した値 $(k(x, y))$ は $k$ に対応する RKHS(高次元) で、関数間の内積をとっていることに相当する。関数間の内積とは例えば：

$$\int_{-\infty}^{\infty} f(x)\, g(x)\, dx \tag{6}$$