A Theory on Reproducing Kernel Hilbert Space for Machine Learning

https://github.com/Shena4746/RKHS-for-ML

June 19, 2022

Contents

About This Document

This document collects basic results on reproducing kernel Hilbert space (RKHS) that are useful in the context of machine learning, aimed at readers with a basic knowledge of functional analysis. The content relies heavily on the existing literature listed in the bibliography, in particular[1][2]. [3] [4] [5]. add fukumizu. add jo.

Shena
June 16, 2122

Notations

- $\diamond$ $\mathbb{N}$: The set of natural numbers excluding zero.
- $\diamond$ $\mathbb{R}$: The set of real numbers.
- $\diamond$ $\mathbb{F}$: A scalar field, either $\mathbb{R}$ or $\mathbb{C}$.
- $\diamond$ E: A nonempty set on which typical functions are defined.
- $\diamond$ $\overline{c}$: The complex conjugate of a scalar c.
- $\diamond$ $\operatorname{Re} s$, $\operatorname{Im} s$: The real and imaginary part of a scalar s.
- $\diamond$ $\langle \cdot, \cdot \rangle_H$: The inner product in the inner product space H.
- $\diamond$ $\ker f$: The kernel or null space of f, i.e., the set of all points at which f is zero.
- $\diamond$ ev_x: The evaluation function at x, i.e., the linear functional such that $\mathrm{ev}_x(f) = f(x)$.
- $\diamond$ $H^\star$, $f^\star$: The dual space of H and the adjoint of f.
- $\diamond$ $S^\perp$: The orthogonal complement of a subspace S.
- $\diamond$ $A \perp B$: Every function in A is orthogonal to that in B.
- $\diamond$ $L^p(\Omega, \mathscr{B}, \mu)$: L^p space over a measure space $(\Omega, \mathscr{B}, \mu)$.

1 Introduction

kernel ridge regression as an natural extension of classical ridge regression.

2 Basic Property of RKHS

say something.

2.1 Kernel and Positive Definite Function

Definition 2.1. *(Kernel, Feature Space, Feature Map) A function $K : E \times E \to \mathbb{F}$ is called a kernel if there is a $\mathbb{F}$-Hilbert space H and a mapping $\varphi : E \to H$ such that*

$$K(x, y) = \langle \varphi(y), \varphi(x) \rangle_H$$

for all $x, y \in E$. Such a φ is called a feature map, and H a feature space.

Example 2.2. (Kernel admits many different feature spaces and feature maps) Consider the function $K : \mathbb{R} \times \mathbb{R} \ni (x, y) \mapsto xy \in \mathbb{R}$. Clearly, K is a kernel with a feature map $E \ni x \mapsto x \in \mathbb{R}$ with a feature space $\mathbb{R}$. On the other hand, the map $E \ni x \mapsto (x/\sqrt{2}, x/\sqrt{2}) \in \mathbb{R}^2$ is also a feature map of K. An analogous argument works in order to device a new feature map of an arbitrary kernel. Therefore, neither feature space nor feature map are uniquely determined. $\triangleleft$

Let us start with an almost trivial remark about codomain of kernel.

Proposition 2.3. *($\mathbb{R}$-valued $\mathbb{C}$-kernel admits a $\mathbb{R}$-feature space) Let $K : E \times E \to \mathbb{C}$ be a kernel with a $\mathbb{C}$-feature space H and a feature map $\varphi : E \to H$. Assume $K(x, y) \in \mathbb{R}$ for all $x, y \in E$. Then $H_0 := H$ equipped with an inner product*

$$\langle f, g \rangle_{H_0} := \mathrm{Re} \langle f, g \rangle_{H_0}$$

is an $\mathbb{R}$-Hilbert space, and $\varphi_0 : E \to H_0$ is a feature map of K.

Proof. It is easy to check that $\langle \cdot, \cdot \rangle_{H_0}$ is indeed an inner product. The proof about φ_0 is quite straightforward. $\square$

In the following proposition, we demonstrate a direct method to prove some elementary transformation of kernels is again a kernel.

Remark 2.4. (An review of tensor product of Hilbert spaces) Let H_1 and H_2 be two $\mathbb{F}$-Hilbert spaces of functions on E_1 and E_2, respectively. Consider the vector space $H_1 \bullet H_2$ spanned by the all functions of the form

$$f_1 \otimes f_2 : E_1 \times E_2 \ni (x_1, x_2) \mapsto f_1(x_1) f_2(x_2) \in \mathbb{F},$$

where f_1 and f_2 run through H_1 and H_2, respectively. We can then introduce an inner product on $H_1 \bullet H_2$ by setting

$$\langle \cdot, \cdot \rangle : (H_1 \bullet H_2) \times (H_1 \bullet H_2) \ni (f_1 \otimes f_2, g_1 \otimes g_2) \mapsto \langle f_1, g_1 \rangle_{H_1} \langle f_2, g_2 \rangle_{H_2} .$$

The smallest complete Hilbert space containing the inner product space $H_1 \bullet H_2$ is called the tensor product of H_1 and H_2, and is denoted by $H_1 \otimes H_2$. ◁

Proposition 2.5. *(Product of kernels) Let K_1 be a kernel on E_1 and K_2 be a kernel on E_2. Then $K_1 \cdot K_2$ is a kernel on $E_1 \times E_2$.*

Proof. Let $\varphi_i : E_i \ni x \mapsto H_i$ be a feature map of K_i. It follows from the definition of tensor product space $H_1 \otimes H_2$ that

$$\begin{aligned} K_1(x_1, y_1) \times K_2(x_2, y_2) &= \langle \varphi_1(y_1), \varphi_1(x_1) \rangle_{H_1} \langle \varphi_2(y_2), \varphi_2(x_2) \rangle_{H_2} \\ &= \langle \varphi_1(y_1) \otimes \varphi_2(y_2), \varphi_1(x_1) \otimes \varphi_2(x_2) \rangle_{H_1 \otimes H_2}, \end{aligned}$$

which justifies our claim. □

The direct method is a bit tricky and often turns out to be very hard to apply especially to more complicated situations. The difficulty comes from the requirement that we have to specify an suitable feature space and a feature map, for instance, the tensor product space in the above proposition. So, we now introduce the concept of positive definite function as a convenient tool for investigating the properties of kernel.

Definition 2.6. *(Positive definite function) Let E be a nonempty set. A function $K : E \times E \to \mathbb{C}$ is called positive definite if for any $n \in \mathbb{N}$ and for any $a \in \mathbb{C}^n$ and $x \in E^n$ there holds*

$$\sum_{i=1}^{n} \sum_{j=1}^{n} a_i \overline{a_j} K(x_i, x_j) \geq 0,$$

where $\overline{a_j}$ is the complex conjugate of a_j.

Remark 2.7. As we see later in section2.3, the notion of positive definite function and that of kernel completely coincide. Consequently, all the results stated in the language of positive definite functions turn out to be true for kernels, and vice versa. ◁

As a preparation, we should point out a minor note about positive definite *real* functions. The proof is left to the reader.

Proposition 2.8. *(Characterization of real positive definite function) A real function $K : E \times E \to \mathbb{R}$ is positive definite if and only if it has the following properties:*

(a) K is symmetric.

(b) The defining inequality (2.6) holds for any $\alpha \in \mathbb{R}^n$ instead of $\mathbb{C}^n$.

We are now ready to state the series of basic properties of positive definite functions.

Proposition 2.9. *Every positive definite function $K : E \times E \to \mathbb{C}$ satisfies*

(a) $K(x, x) \geq 0$ *for every* $x \in E$

(b) $K(x, y) = \overline{K(y, x)}$ *for every* $x, y \in E$

(c) $\overline{K}$ *is also positive definite, and conversely*

(d) $|K(x, y)| \leq K(x, x)K(y, y)$ *for every* $x, y \in E$.

Proof. (a) and (c) clearly hold. For $\alpha, \beta \in \mathbb{C}$ and $x, y \in E$, we have

$$g(\alpha, \beta) := |\alpha|^2 K(x, x) + \alpha\overline{\beta}K(x, y) + \overline{\alpha}\beta K(y, x) + |\beta|^2 K(y, y) \geq 0.$$

Choose $\alpha = \beta = 1$ and $\alpha = i$, $\beta = 1$ to get

$$\begin{aligned} K(x, y) + K(y, x) &= g(1, 1) - K(x, x) - K(y, y) =: A \in \mathbb{R} \\ iK(x, y) - iK(y, x) &= g(i, 1) - K(x, x) - K(y, y) =: B \in \mathbb{R}. \end{aligned}$$

Therefore,

$$\begin{aligned} 2K(y, x) &= A + iB \\ 2K(x, y) &= A - iB, \end{aligned}$$

which proves (b). Finally, for $x, y \in E$ with $K(x, y) \neq 0$ and for $r \in \mathbb{R}$, (b) gives

$$0 \geq g(r, K(x, y)) = r^2 K(x, x) + 2r |K(x, y)|^2 + |K(x, y)|^2 K(y, y).$$

As RHS is quadratic in r, it must satisfy

$$|K(x, x)|^4 - |K(x, y)|^2 K(x, x)K(y, y) \leq 0,$$

from which (d) follows. □

Corollary 2.10. *(Kernel is positive definite) A kernel is positive definite.*

Proof. For the case $\mathbb{F} = \mathbb{C}$, we have

$$\sum_{i=1}^{n} \sum_{j=1}^{n} a_i \overline{a_j} \overline{K(x_i, x_j)} = \sum_{i=1}^{n} \sum_{j=1}^{n} a_i \overline{a_j} \langle \varphi(x_i), \varphi(x_j) \rangle = \left\| \sum_{i=1}^{n} a_i \varphi(x_i) \right\|^2 \geq 0,$$

and hence $\overline{K}$ as well as K are positive definite. □

The following propositions are easy to deduce.

Proposition 2.11. *Let $K_n : E \times E \to \mathbb{F}$ be positive definite functions.*

(a) $aK_1 + bK_2$ *is positive definite if* $a, b \geq 0$.

(b) The function defined by the pointwise limit $\lim_{n\to\infty} K_n(x, y)$ *of* $K_n(x, y)$ *is positive definite, provided that limit is well-defined.*

In other words, the set of all positive definite functions is a closed cone.

Proposition 2.12. *Let $K : E \times E \to \mathbb{F}$ be a positive definite function.*

(a) For an arbitrary map $T : E_1 \to E$, the function

$$K_T : E_1 \times E_1 \ni (x, y) \mapsto K(T(x), T(y)) \in \mathbb{F}$$

is also positive definite. In particular, if $E_1 \subset E$, the restriction of K to $E_1 \times E_1$ is positive definite.

(b) For an arbitrary map $S : E \to \mathbb{F}$, the function $E \times E \ni (x, y) \mapsto S(x)K(x, y)\overline{S(y)}$ is also positive definite.

Example 2.13. (Normalized kernel) If $K(x, x) > 0$ for all x, then

$$K_N(x, y) := \frac{K(x, y)}{\sqrt{K(x, x)K(y, y)}}$$

is also a kernel with $|K_N| \leq 1$. K_N is called a *normlized kernel.* $\triangleleft$

2.2 Properties of Reproducing Kernel

Definition 2.14 (Reproducing kernel hilbert space)**.** *Let E be a nonempty set. A function K defined by*

$$K : E \times E \ni (x, y) \mapsto K(x, y) \in \mathbb{F}$$

is called a reproducing kernel of a Hilbert space H of functions on E if it satisfies the following conditions:

(a) $K(\cdot, x) \in H$ *for every* $x \in E$

(b) $\langle f, K(\cdot, x)\rangle_H = f(x)$ *for every* $x \in E$ *and every* $f \in H$.

Such a Hilbert space, associated with its reproducing kernel, is called a reproducing kernel hilbert space, and is denoted by (H, K).

Remark 2.15. (b) is called the *reproducing property*, and the identity is called the *reproducing identity.* ◁

Remark 2.16. (Reproducing kernel is a kernel) Setting $f = K(\cdot, y)$ at (b) yields $\langle K(\cdot, y), K(\cdot, x)\rangle = K(x, y)$. Therefore, every reproducing kernel is in fact a kernel, as the name suggests. ◁

Example 2.17. (Finite dimensional inner product space is a RKHS) Suppose an inner product space $(H, \langle\cdot,\cdot\rangle)$ of functions on E is of finite dimensional. Since every finite dimensional normed vector space is complete, H is an Hilbert space. Given an orthonormal basis $\langle u_1, \ldots, u_d\rangle$ of H,

$$K(x, y) := \sum_{i=1}^{d} u_i(x)u_i(y)$$

is a reproducing kernel of H. In fact, for $f = \sum_{i=1}^{d} a_i u_i(\cdot)$, we have

$$\langle f, K(\cdot, x)\rangle = \sum_{i,j=1}^{d} a_i e_j(x) \langle e_i(\cdot), e_j(\cdot)\rangle = \sum_{i=1}^{d} a_i e_i(x) = f(x)$$

for every $x \in E$. ◁

Theorem 2.18 (Characterization of RKHS)**.** *A Hilbert space H of functions on a nonempty set E admits a reproducing kernel K if and only if all evaluation functionals $\{\mathrm{ev}_x\}_{x\in E}$ are continuous on H.*

Proof. Suppose (H, K) is a RKHS. For $x \in E$ and for $f \in H$ we have

$$|\mathrm{ev}_x(f)| = |f(x)| = |\langle f, K(\cdot, x)\rangle| \le \|f\| \, \|K(\cdot, x)\| \le \|f\| \, K(x, x)^{1/2} \to 0$$

as $\|f\| \to 0$. Thus, ev_x is a continuous linear functional (with norm $K(x, x)^{1/2}$).

Conversely, if $\mathrm{ev}_x : H \ni f \mapsto f(x) \in \mathbb{F}$ is continuous, then, by Riesz's representation theorem, there exists $r_x \in H$ such that

$$\langle f, r_x\rangle = f(x)$$

for every $f \in H$. If this happens for every $x \in E$, then $K(x, y) := r_x(y)$ is a reproducing kernel of H. □

Corollary 2.19. *Every convergent sequence in RKHS converges pointwise to the same limit.*

Proof. $|f_n(x) - f(x)| = |\mathrm{ev}_x(f_n - f)| \to 0$ when $f_n \to f$ in norm by continuity of evaluation functional. □

The following proposition says that there is one-to-one correspondence between RKHS and reproducing kernel.

Proposition 2.20. *(Uniqueness of H and K)*

(a) Let (H, K) be a RKHS. The subspace H_0 spanned by $\{K(\cdot, x)\}_{x \in E}$ is dense in H.

(b) A Hilbert space admits at most one reproducing kernel.

(c) A function $K : E \times E \to \mathbb{F}$ is a reproducing kernel for at most one Hilbert space. In particular, there is at most one RKHS that has H_0 as a dense subspace.

Proof. For density of H_0, observe $f \in H$ fulfills $f \perp H_0$ if and only if

$$\langle f, K(\cdot, x)\rangle_H = f(x) = 0$$

for every $x \in E$, which is the case precisely when $f \equiv 0$. To check (b), suppose K_1 and K_2 qualify as a reproducing kernel of H. By definition,

$$f(x) = \langle f, K_1(\cdot, x)\rangle_H = \langle f, K_2(\cdot, x)\rangle_H$$

for every $x \in E$, and hence

$$\langle f, K_1(\cdot, x) - K_2(\cdot, x)\rangle_H = 0$$

for every $f \in H$ and $x \in E$. From this we conclude $K_1 = K_2$. Finally suppose that (H_1, K) and (H_2, K) are two RKHSs. Pick $f \in H_1$. By (a), there is $\{f_n\} \subset H_0 \subset H_1 \cap H_2$ such that $f_n \to f$ in H_1-norm. Since $\{f_n\}$ is also an Cauchy sequence in H_2, it admits a limit $g \in H_2$. But Corollary2.19 implies $f = g$, and hence $f \in H_2$. We then have

$$\|f\|_{H_1} = \lim_{n\to\infty} \|f_n\|_{H_1} = \lim_{n\to\infty} \|f_n\|_{H_0} = \lim_{n\to\infty} \|f_n\|_{H_2} = \|f\|_{H_2}.$$

Therefore, H_1 is isometrically included in H_2. Symmetry thus shows that both Hilbert spaces coincide. □

Proposition 2.21. *(Representation of RK in terms of evaluation functional) In an arbitrary RKHS (H, K), the reproducing kernel $K : E \times E \to \mathbb{F}$ always fulfills the identity*

$$K(x, y) = \langle \mathrm{ev}_y, \mathrm{ev}_x\rangle_{H^\star}$$

for all $x, y \in E$, where $H^\star$ is the dual space of H.

Proof. It suffices to show that a function K *defined* by the above equation is also a reproducing kernel. Let $I : H^\star \to H$ be the isometric anti-linear surjection, guaranteed by Riesz's Representation Theorem, that assigns to every functional in $H^\star$ the corresponding representor in H, i.e., $g^\star(f) = \langle f, Ig^\star \rangle_H$ for all $f \in H$ and $g \in H^\star$, where $g^\star$ is the adjoint of g. Then we have

$$K(x, y) = \langle \mathrm{ev}_y, \mathrm{ev}_x \rangle_{H^\star} = \langle I\mathrm{ev}_y, I\mathrm{ev}_x \rangle_H = \mathrm{ev}_x (\mathrm{ev}_y) = (I\mathrm{ev}_y)(x),$$

for all $x, y \in E$, and hence $K(\cdot, y) = I\mathrm{ev}_y \in H$. From this it follows that

$$f(y) = \mathrm{ev}_y(f) = \langle f, I\mathrm{ev}_y \rangle_H = \langle f, K(\cdot, y) \rangle$$

for all $y \in E$. Thus, K is a reproducing kernel. □

2.3 Construction of RKHS

Theorem 2.22. *(RKHS generated by inner product space) Let H_0 be the subspace of $\mathbb{F}^E$, equipped with an inner product $\langle \cdot, \cdot \rangle_{H_0}$ with norm $\|\cdot\|_{H_0}$. Then there exists unique RKHS (H, K) that extends H_0 in the sense that*

(a) $H_0 \subset H \subset \mathbb{F}^E$ and the subspace topology of H_0 in H coincides with the topology of $(H_0, \|\cdot\|_{H_0})$

if and only if H_0 satisfies the following conditions:

(b) every evaluation functional ev_x is continuous in $(H_0, \|\cdot\|_{H_0})$

(c) any Cauchy sequence $\{f_n\} \subset H_0$ converging pointwise to 0 converges to 0 also in H_0-norm.

Consequently, H is isomorphic to the completion of H_0, and it consists of pointwise limit of Cauchy sequence in H_0.

Proof. Suppose such an extension H exists. H satisfies (b) by Theorem2.18. Since H is complete, a Cauchy sequence $\{f_n\} \subset H_0$ tends to some f, for which we have

$$f(x) = \mathrm{ev}_x(f) = \lim_{n\to\infty} \mathrm{ev}_x(f_n) = \lim_{n\to\infty} f_n(x) = 0$$

for every $x \in E$. Therefore, f is identically 0.

Conversely, suppose (b)(c) hold. As Proposition2.20 show the uniqueness of such H, we only have to prove the existence. Let X be the Hilbert space derived by the completion of H_0. In general, X consists of equivalent classes of Cauchy sequence in H_0 equipped with the inner product

$$\langle \cdot, \cdot \rangle_X : X \times X \ni ([\{f_n\}], [\{g_n\}]) \mapsto \lim_{n\to\infty} \langle f_n, g_n \rangle_{H_0} \in \mathbb{F}.$$

Let $f = [\{f_n\}]$ be an element in X with a representative Cauchy sequence $\{f_n\} \subset H_0$. It follows from (a) that

$$|f_n(x) - f_m(x)| = |\mathrm{ev}_x(f_n - f_m)| \to 0,$$

when $n, m \to \infty$. As this implies $\{f_n(x)\}$ is a Cauchy sequence for every $x \in E$, we can define a function $f : E \to \mathbb{F}$ by setting

$$f(x) := \lim_{n\to\infty} f_n(x).$$

It is easy to see that f is well-defined, independent of the choice of a representative $\{f_n\}$. We then define a linear mapping

$$I : X \ni [\{f_n\}] \mapsto f \in \mathbb{F}^E.$$

Obviously, $I([\{f\}]) = f$ for $f \in H_0$. Moreover, I is injective; indeed, if $h = [\{f_n\}] \in X$ and $\lim_{n\to\infty} f_n(x) = 0$ for every $x \in E$, then (b) gives $f_n \to 0$ in H_0-norm and therefore $h \equiv 0$ in X, as required. The isomorphism I induces the Hilbert structure on $H := I(X)$, which makes I isometric on H. Clearly, H_0 is dense in H. Finally, we claim that every evaluation functional ev_x is continuous on H. Fix $x \in E$. As ev_x is assumed in (a) to be (uniformly) continuous on H_0, it admits unique continuous extension T_x onto the closure of H_0 in H, that is, onto whole H. For $f \in H$ and for $f_n \in H_0$ with $f_n \to f$ pointwise, we have

$$T_x(f) = \lim_{n\to\infty} \mathrm{ev}_x(f_n) = \lim_{n\to\infty} f_n(x) = f(x).$$

It follows from Theorem2.18 that H admits a reproducing kernel. □

Theorem 2.23. *(Moore-Aronszajn) For an arbitrary positive definite function $K : E \times E \to \mathbb{F}$, there exists unique RKHS H that has K as its reproducing kernel. Moreover, the subspace H_0 spanned by $\{K(\cdot, x)\}_{x\in E}$ is dense in H.*

Proof. Define an inner product $\langle \cdot, \cdot \rangle_{H_0}$ on H_0 by setting

$$\langle f, g \rangle_{H_0} := \sum_{i=1}^{n} \sum_{j=1}^{n} \alpha_i \beta_i K(y_i, x_i),$$

where $f = \sum_{i=1}^n \alpha_i K(\cdot, x_i)$ and $g = \sum_{j=1}^n \alpha_i K(\cdot, y_i)$. Let us observe

$$\sum_{i=1}^n \sum_{j=1}^n \alpha_i \beta_i K(y_i, x_i) = \sum_{i=1}^n \alpha_i \overline{g(x_i)} = \sum_{j=1}^n \overline{\beta_i} f(y_j),$$

and therefore that the value $\langle f, g \rangle_{H_0}$ is determined by solely by f and g, independent of the choice of representing linear combination. Choosing $g = K(\cdot, x)$ yields

$$\langle f, K(\cdot, x) \rangle_{H_0} = \sum_{i=1}^n \alpha_i \overline{h(x_i)} = \sum_{i=1}^n \alpha_i K(x, x_i) = f(x).$$

So, K fulfills the reproducing identity under $\langle \cdot, \cdot \rangle_{H_0}$. In particular,

$$\|K(\cdot, z)\|_{H_0} = \langle K(\cdot, x), K(\cdot, x) \rangle_{H_0} = K(x, x) \geq 0.$$

This establishes the definiteness of $\langle \cdot, \cdot \rangle_{H_0}$; indeed, if $\langle f, f \rangle_{H_0} = 0$, then we have

$$|f(x)| = |\langle f, K(\cdot, x) \rangle| \leq \langle f, f \rangle^{1/2} K(x, x)^{1/2} = 0,$$

for every $x \in E$, implying $f \equiv 0$. We then conclude that $\langle \cdot, \cdot \rangle_{H_0}$ is in fact an inner product on H_0 as the other requirements are easy to check.

We now show that H_0 fulfills the sufficient conditions of Theorem2.22. First, each ev_x is continuous on H_0; in fact, for $f, g \in H_0$,

$$|\mathrm{ev}_x(f) - \mathrm{ev}_x(g)| = \left| \langle f - g, K(\cdot, x) \rangle_{H_0} \right| \leq \|f - g\|_{H_0} K(x, x)^{1/2}$$

for every $x \in E$. To verify the other condition, let $\{f_n\}$ be a Cauchy sequence in H_0 converging pointwise to 0. Let $B > 0$ be an upper bound for $\|f_n\|_{H_0}$. For $\epsilon > 0$ and large $N \in \mathbb{N}$ we have

$$\|f_n - f_N\| < \frac{\epsilon}{B}$$

for all $n \geq N$. We may write

$$f_N = \sum_{i=1}^k K(\cdot, x_i)$$

for some $\alpha_i \in \mathbb{F}$ and $x_i \in E$, and for some fixed k. It then follows that

$$\|f_n\|_{H_0}^2 = \langle f_n - f_N, f_n \rangle_{H_0} + \langle f_N, f_n \rangle_{H_0} \leq \epsilon + \sum_{i=1}^k f(x_i)$$

for $n \geq N$, and hence $\|f_n\| \to 0$ as $n \to 0$. Therefore, there is a RKHS H that has H_0 as a dense subspace. Furthermore, for each $f \in H$ there is $\{f_n\} \subset H_0$ such that $f_n \to f$ pointwise as well as in H-norm, for which we have

$$f(x) = \lim_{n\to\infty} f_n(x) = \lim_{n\to\infty} \langle f_n, K(\cdot, x)\rangle_{H_0} = \langle f, K(\cdot, x)\rangle_H ,$$

for every $x \in E$. Thus, K is a reproducing kernel of H. Uniqueness follows from Proposition2.20. □

Theorem 2.24. *(Characterization of positive definite function) A function $K : E \times E \to \mathbb{F}$ is positive definite (and thus a reproducing kernel of some RKHS) if and only if K is a kernel, that is, if and only if there exists some mapping φ of E into some $\mathbb{F}$-Hilbert space H such that*

$$K(x, y) = \langle \varphi(y), \varphi(x)\rangle_H$$

for all $x, y \in E$.

Proof. If (H, K) is the RKHS generated by positive definite function K, then the canonical feature map $\varphi_K : E \ni x \mapsto K(\cdot, x) \in H$ obviously qualifies. The converse follows from Corollary2.10. □

Remark 2.25. Theorem2.24 implies that RKHS (H, K) is a natural feature space. ◁

Theorem2.24 is a powerful tool to construct a positive definite function as well as to prove a given function is a kernel *if we can find an appropriate feature space.*

Example 2.26. Let us show that $K(x, y) = \min(x, y)$, $x, y \in \mathbb{R}_+$ is positive definite. Let H:=$L^2(\mathbb{R}_+, \mu)$ be the space of all square integrable functions on $\mathbb{R}_+$ with respect to a σ-finite measure μ. It is well-known that H is a Hilbert space with the inner product $\langle f, g\rangle_H := \int_{\mathbb{R}_+} f\overline{g}d\mu$. Then we have

$$K(x, y) = \int_{\mathbb{R}_+} 1_{[0,y]}(t)1_{[0,x]}(t)d\mu(t) = \langle \varphi(y), \varphi(x)\rangle_H ,$$

where $\varphi : E \ni x \mapsto 1_{[0,x]}(\cdot) \in H$ is the feature map, and $1_A(\cdot)$ is the indicator function of A. Therefore, K is positive definite. ◁

The next theorem construct a RKHS as a continuous embedding into a given feature space. We will repeatedly exploit this constructive method later.

Theorem 2.27. *(RKHS generated by feature map) Let $E \neq \emptyset$. Suppose K is a positive definite kernel with a feature space H_0 and a feature map $\varphi_0 : E \to H_0$. Then the Hilbert space*

$$H := \{f : E \to \mathbb{F} \mid \exists w \in H_0 \,:\, f(x) = \langle w, \varphi_0(x) \rangle_{H_0} \ \forall x \in E\}$$

equipped with the norm

$$\|f\|_H := \inf\{\|w\|_{H_0} \,:\, w \in H_0,\ f = \langle w, \varphi_0(\cdot) \rangle_{H_0}\}$$

is the RKHS with the reproducing kernel K, and H and $\|\cdot\|_H$ are determined independent of the choice of feature space H_0 and feature map φ_0. Moreover, the function

$$V : H_0 \ni w \mapsto \langle w, \varphi_0(\cdot) \rangle_{H_0} \in H$$

acts as an isometrical isomorphism on $(\ker V)^{\perp}$.

Proof. In light of Theorem2.23, it suffices to prove that H is RKHS with reproducing kernel K. The property of V are automatically obtained in the process. It is easy to verify that $\|\cdot\|_H$ is actually a norm on H. As $\ker V$ is closed subspace of H_0, we get the orthogonal decomposition $H_0 = \ker V \oplus (\ker V)^{\perp}$. Let $H_1 := (\ker V)^{\perp}$ and let the restriction of V onto H_1 be denoted by V_1. Since every $f \in H$ can be written as $f = V(w_0 + w_1) = V_1 w_1$, with $w_0 \in \ker V$, $w_1 \in H_1$, we see that $V_1 : H_1 \to H$ is bijective. Similarly, we have

$$\begin{aligned}
\|f\|_H &= \inf\{\|w_0 + w_1\|_{H_0}^2 \,:\, w_0 \in \ker V, w_1 \in H_1, w_0 + w_1 \in V^{-1}(\{f\})\} \\
&= \inf\{\|w_0\|_{H_0}^2 + \|w_1\|_{H_0}^2 \,:\, w_0 \in \ker V, w_1 \in H_1, w_0 + w_1 \in V^{-1}(\{f\})\} \\
&= \inf\{\|w_1\|_{H_0}^2 \,:\, w_1 \in H_1, w_1 \in V^{-1}(\{f\})\} \\
&= \left\|V_1^{-1}(f)\right\|_{H_1} \left(:= \left\|V_1^{-1}(f)\right\|_{H_0}\right),
\end{aligned}$$

from which we conclude that $V_1 : H_1 \to H$ is an isometrical isomorphism, as required, and that H is a Hilbert space.

It remains to show that K qualifies as the reproducing kernel. Observe

$$K(\cdot, x) = \langle \varphi_0(x), \varphi_0(\cdot) \rangle_{H_0} = V\varphi_0(x) \in H.$$

Moreover, the fact $\langle w, \varphi_0(x) \rangle_{H_0} = 0$ for all $w \in \ker V$ implies

$$f(x) = \left\langle V_1^{-1} f, \varphi_0(x) \right\rangle_{H_0} = \langle f, V\varphi_0(x) \rangle_H = \langle f, K(\cdot, x) \rangle_H$$

for all $f \in H$ and $x \in E$. $\square$

Remark 2.28. (Infimum in the norm $\|\cdot\|_H$ at (2.27) is attainable) We continue with the notation in the Theorem2.27. The isometric relation $\|f\|_H = \left\|V_1^{-1}(f)\right\|_{H_1}$ clearly shows that the infimum is achievable within the domain of V_1, namely within the subspace $D := (\ker V)^\perp$ of H_0. From this it follows that the infimum of norm $\|f\|_H$ of $f \in H$ is attained at the D-orthogonal-component of $V^{-1}(\{f\})$. ◁

Corollary 2.29. *(RKHS derived by Fourier transformation) Suppose that $r \in L^1(\mathbb{R}^d, \mathscr{B}, dt)$ is a bounded continuous function with $r(t) > 0$ for all t. Then,*

$$H := \{f \in L^2(\mathbb{R}^d, \mathscr{B}, dt) \mid \int_{\mathbb{R}^d} \frac{|\hat{f}|^2}{r} dt < \infty\}$$

is a RKHS with the inner product

$$\langle f, g \rangle_H := \int_{\mathbb{R}^d} \frac{\hat{f} \cdot \overline{\hat{g}}}{r} dt,$$

where $\hat{f}$ stands for the Fourier transformation of f, and with the RK

$$K : \mathbb{R}^d \times \mathbb{R}^d \ni (x, y) \mapsto \int_{\mathbb{R}^d} e^{-\sqrt{-1}(x-y)t} r(t) dt \in \mathbb{C}.$$

Proof. We prove for the case $d = 1$. The general case is proved similarly. Let $H_0 := L^2(\mathbb{R}, \mathscr{B}, r(t)dt)$. It is easy to deduce from Schwartz inequality that $wr \in L^1(\mathbb{R}, dt)$ for every $w \in H_0$. Let φ be a feature map that assigns, to each $x \in \mathbb{R}$, the function $\mathbb{R} \ni t \mapsto e^{-\sqrt{-1}xt} \in H_0$. Theorem2.27 tells us that $H = \{\widehat{wr} \mid w \in H_0\}$ is a RKHS with the RK K. We claim that the map

$$V : H_0 \ni w \mapsto \langle w, \varphi(\cdot) \rangle_{H_0} \in H$$

is injective. Suppose $\langle w, \varphi(x) \rangle_{H_0} = 0$, that is, suppose that Fourier transformation of wr is identically zero. This necessarily leads to $wr = 0$, and hence $w \equiv 0$, as required. It then follows that $V : H_0 \to H$ is an isometrical isomorphism, and consequently that

$$\langle f, g \rangle_H = \left\langle V^{-1}f, V^{-1}g \right\rangle_{H_0} = \int \frac{\hat{f}}{r} \frac{\overline{\hat{g}}}{r} r dt = \int \frac{\hat{f} \cdot \overline{\hat{g}}}{r} dt.$$

In order to justify the stated expression of H, let us note that $wr \in L^2(\mathbb{R}, dt) \cap L^1(\mathbb{R}, dt)$ for $w \in H_0$. We may write $f \in H$ as $f = \widehat{wr}$ for some $w \in H_0$. The well-known result from the theory of Fourier transformation then yields $\hat{f} = wr$, which implies $w \in H_0 \iff \hat{f}/r \in H_0$. □

2.4 Example of Kernels

Example 2.30. (Polynomial kernel)

$$K_P(x, y) : \mathbb{F}^d \times \mathbb{F}^d \ni (x, y) \mapsto (\langle x, y\rangle + c)^m \in \mathbb{F},$$

$c \geq 0$ and $m \in \mathbb{N}$ is a kernel. Its restriction to $\mathbb{R}^d$ is called a real-valued *polynomial kernel (of degree m).* ◁

Example 2.31. (Taylor type kernel) Suppose that a function f defined on a subset D of $\mathbb{F}^d$ admits a Taylor series expansion within an open subset B of D:

$$f(x) = \sum_{n=1}^{\infty} a_n x^n, \quad x \in B.$$

If $a_n \geq 0$ for all n, then

$$K_(x, y) : \mathbb{F}^d \times \mathbb{F}^d \ni (x, y) \mapsto f(\langle x, y\rangle) = \sum_{n=1}^{\infty} a_n \langle x, y\rangle^n \in \mathbb{F}$$

defines a kernel called a *kernel of Taylor type.* The restriction of K to $B \cap \mathbb{R}^d$ is a real-valued kernel. ◁

Example 2.32. (Exponential kernel)

$$K_E : \mathbb{F}^d \times \mathbb{F}^d \ni (x, y) \mapsto \exp(\beta \langle x, y\rangle) \in \mathbb{F},$$

$\beta \geq 0$ is a kernel called an *exponential kernel.* ◁

Example 2.33. (Binomial kernel)

$$K_B : \mathbb{F}^d \times \mathbb{F}^d \ni (x, y) \mapsto (1 - \langle x, y\rangle)^{-\alpha} \in \mathbb{F},$$

$\alpha \geq 0$, is a kernel called a *binomial kernel*, defied on $\{x \in \mathbb{F}^d \mid |\langle x, y\rangle| < 1\}$, since

$$(1 - t)^{-\alpha} = \sum_{n=1}^{\infty} \binom{-\alpha}{n} (-1)^n t^n$$

for all $|t| < 1$. ◁

Example 2.34. (Fourier type kernel) Suppose a function f on $\mathbb{R}$ admits a pointwise Fourier series expansion within $[-2\pi, 2\pi]$:

$$f(x) = \sum_{n=1}^{\infty} a_n \cos(nx).$$

If $a_n \geq 0$ for all n, then

$$K(x, y) : \mathbb{R}^d \times \mathbb{R}^d \ni (x, y) \mapsto \prod_{i=1}^{d} f(x_i - y_i) \in \mathbb{R}$$

defines a kernel on $[0, 2\pi)^d$ called a kernel of Fourier type. To see this, we may assume $d = 1$ without loss of generality. Since $\{a_n\} \in \ell_1$, each term on the expansion

$$K(x, y) = a_0 + \sum_{n=1}^{\infty} a_n \cos(nx)\cos(ny) + \sum_{n=1}^{\infty} a_n \sin(nx)\sin(ny)$$

is a kernel. $\triangleleft$

2.5 Examples of RKHSs

Example 2.35. (RKHS generated by a finite dimensional kernel) Let $K : E \times E \to \mathbb{F}$ be a kernel and suppose that the inner product space spanned by $\{K(\cdot, x)\}_{x \in E}$ is of finite dimensional. Then it is an Hilbert space; call it H. Theorem2.23 says that H is also the RKHS with reproducing kernel K. We construct a RKHS in an Euclidean space that is isomorphic to H. A basis $B := \langle e_1, \ldots, e_d \rangle$ of H, for instance, $\langle K(\cdot, x_1), \ldots, K(\cdot, x_d) \rangle$ for some x_i's, defines a bijective linear mapping

$$A : H \ni f = \sum_{i=1}^{d} a_i e_i \mapsto (a_i) \in \mathbb{F}^d$$

The inner product of H is then given by

$$\langle f, g \rangle_H = \sum_{i,j=1}^{d} a_i \overline{b_j} K(x_j, x_i) = (Af)^\top M_K \overline{Ag},$$

where $M_K := K_{ij} := (K(x_i, x_j))$ is a $d \times d$ (Hermitian) symmetric matrix.

Now assume that M_K is positive definite in the matrix sense, that is, assume that every eigenvalue λ_i of M_K is positive. Let $\langle u_1, \ldots, u_d \rangle$ be an orthonormal basis derived by the spectral decomposition of M_K, for which we have $M_K = \sum_{i=1} \lambda_i u_i \overline{u_i}^\top$. As before, this basis defines a bijective linear mapping

$$B : \mathbb{F}^d \ni a = \sum_{i=1}^{d} \alpha_i u_i \mapsto (\alpha_i) \in \mathbb{F}^d.$$

It then follows that

$$\langle f, g\rangle_H = \sum_{i=1}^d \left(a^\top u_i\right) \overline{u_i b} = \sum_{i=1}^d \lambda_i \alpha_i \overline{\beta_i} = \sum_{i=1}^d \lambda_i (BAf)_i \overline{(BAg)_i}.$$

Thus, the mapping $F(f) := BAf$ is an isomorphism between H and $\mathbb{F}^d$ equipped with the inner product

$$\langle a, b\rangle_K := a^\top M_K^{-1} \overline{b} = \sum_{i=1}^d \frac{\alpha_i \overline{\beta_i}}{\lambda_i}.$$

Moreover, $(H, \langle\cdot,\cdot\rangle_K)$ is a RKHS with RK M_K. In fact, for $a \in \mathbb{F}^d$ and $k_i := \overline{(K_{i1}, \ldots, K_{in})}^\top$, we have

$$\langle a, k_i\rangle_K = a^\top M_K^{-1} \overline{k_i} = a^\top M_K^{-1} (M_K)_i = a_i,$$

as required. ◁

Example 2.36. (RKHS of polynomial a kernel) Since the linear space spanned by a polynomial kernel of degree d (Example2.30) is finite dimensional, Example2.35 tells us that the corresponding RKHS coincides with the space consisting of polynomials of degree at most d. ◁

Example 2.37. (RKHS of a Gaussian RBF kernel) In Corollary2.29, take $r \in L^1(\mathbb{R}^d, dt)$ as

$$r(t) : \mathbb{R}^d \ni t \mapsto \frac{\sigma}{2\pi} \exp\left(-\frac{\sigma^2}{2} \langle t, t\rangle\right) \in \mathbb{R}.$$

The resulting kernel is

$$K(x, y) : \mathbb{R}^d \ni t \mapsto \exp\left(-\frac{1}{2\sigma^2} \|x - y\|^2\right) \in \mathbb{R},$$

called *Gaussian RBF kernel*. The associated RKHS is given by

$$H = \{f \in L^2(\mathbb{R}^d, dt) \mid \int |\hat{f}(t)|^2 \exp\left(\frac{\sigma^2}{2} \langle t, t\rangle\right) dt < \infty\}$$

equipped with the inner product (equivalent to)

$$\langle f, g\rangle = \int \hat{f}(t) \cdot \hat{g}(t) \exp\left(\frac{\sigma^2}{2} \langle t, t\rangle\right) dt$$

◁

Example 2.38. (RKHS of a Laplacian kernel)

◁

3 Reconstruction of RKHS

3.1 Reconstruction via Restriction

Corollary 3.1. *(RKHS of a Restricted Kernel) Let (H, K) be a RKHS of functions on E. Let $\emptyset \neq E_1 \subset E$. The restriction K_1 of K to $E_1 \times E_1$ is the RK of the Hilbert space*

$$H_1 := \{f|_{E_1} \mid f \in H\}$$

equipped with the norm

$$\|f_1\|_{H_1} := \inf\{\|f\|_H \,:\, f \in H,\, f|_{E_1} = f_1\} = \|f_{1,0}\|_H\,,$$

where $f|_A$ stands for the restriction of f to the set A, and $f_{1,0}$ the extension of f_1 over H by zero.

Proof. Define a feature map

$$\varphi : E_1 \ni x \mapsto K(\cdot, x) \in H,$$

and apply Theorem2.27 to see (H_1, K_1) is the RKHS generated by the feature map φ. Moreover, Remark2.28 tells us that the infimum of $\|f_1\|_{H_1}$ is achievable at the orthogonal projection of $f \in H$ satisfying $f|_{E_1} = f_1$ onto $\ker V^{\perp}$, namely, at $f 1_{E_1} \in H$. □

Proposition 3.2. *($\mathbb{R}$-RKHS of a $\mathbb{R}$-valued $\mathbb{C}$-kernel) Let $K : E \times E \to \mathbb{C}$ be a kernel and H its corresponding $\mathbb{C}$-RKHS, and suppose $K(x, y) \in \mathbb{R}$ for all $x, y \in E$.*

(a) The space

$$H_1 := \{f : E \to \mathbb{R} \mid \exists g \in H,\, \operatorname{Re} g = f\}$$

equipped with the norm

$$\|f\|_{H_1} := \inf\{\|g\|_H \,:\, g \in H,\, \operatorname{Re} g = f\}$$

is the $\mathbb{R}$-RKHS of the kernel $K : E \times E \to \mathbb{R}$.

(b) In particular, if $E = \mathbb{C}^d$, then the space

$$H_2 := \{f \in \mathbb{R}^d \to \mathbb{R} \mid \exists g \in \mathbb{C}^d \to \mathbb{C} \,:\, g \in H,\, \operatorname{Re} g|_{\mathbb{R}^d \times \mathbb{R}^d} = f\}$$

equipped with the norm

$$\|f\|_{H_2} := \inf\{\|g\|_H \,:\, g \in H,\, \operatorname{Re} g|_{\mathbb{R}^d} = f\}$$

is the $\mathbb{R}$-RKHS of the restricted kernel $K|_{\mathbb{R}^d \times \mathbb{R}^d}$.

Proof. Proposition2.3 tells us that $H_0 := H$ with an inner product

$$\langle f, g \rangle_{H_0} := \operatorname{Re} \langle f, g \rangle_H$$

is an $\mathbb{R}$-feature space of a $\mathbb{R}$-feature map

$$\varphi : E \ni x \mapsto K(\cdot, x) \in H_0.$$

For all $f \in H_0$ and $x \in E$, we have

$$\begin{aligned} f(x) = \langle f, \varphi(x) \rangle_H &= \operatorname{Re} \langle f, \varphi(x) \rangle_H + \operatorname{Im} \langle f, \varphi(x) \rangle_H \\ &= \langle f, \varphi(x) \rangle_{H_0} + \operatorname{Im} \langle f, \varphi(x) \rangle_H , \end{aligned}$$

which implies $\langle f, \varphi(x) \rangle_{H_0} = \operatorname{Re} f(x)$. Applying Theorem2.27 then proves (a). (b) is an immediate consequence of (a) and Corollary3.1. □

3.2 Reconstruction via Sum and Product

Theorem 3.3. *(Sum of RKHSs) Let (H_1, K_1) and (H_2, K_2) be two $\mathbb{F}$-RKHSs of functions on the common set E. Then $K := K_1 + K_2$ is the RK of*

$$H := H_1 \oplus H_2 := \{f_1 + f_2 \mid f_1 \in H_1,\ f_2 \in H_2\}$$

with the norm

$$\|f\|_H := \min\{\|f_1\|_{H_1} + \|f_2\|_{H_2} \ : \ f = f_1 + f_2,\ f_1 \in H_1,\ f_2 \in H_2\}.$$

Proof. Let F be the Hilbert sum of H_1 and H_2:

$$F := \{(f_1, f_2) \mid f_1 \in H_1,\ f_2 \in H_2\}$$

equipped with an inner product

$$\langle f, g \rangle_F := \langle f_1, g_1 \rangle_{H_1} + \langle f_2, g_2 \rangle_{H_2} .$$

It is easy to see that the map

$$\varphi : E \ni x \mapsto (K_1(\cdot, x), K_2(\cdot, x)) \in F$$

is a feature map of K, and that we have

$$\langle f, \varphi(x) \rangle_F = f_1(x) + f_2(x)$$

for all $f = (f_1, f_2) \in F$ and $x \in E$. Thus, (H, K) is a RKHS by Theorem2.27. For attainability of $\|\cdot\|_H$, see Remark2.28. □

Theorem 3.4. *(Tensor product of RKHSs) Let K_1 and K_2 be $\mathbb{F}$-kernels defined on E_1 and E_2, respectively, and let H_1 and H_2 be the corresponding $\mathbb{F}$-RKHSs. Set $H := H_1 \otimes H_2$.*

(a) Define the product kernel K of K_1 and K_2 via

$$K : (E_1 \times E_2) \times (E_1 \times E_2) \ni ((x_1, x_2), (y_1, y_2)) \mapsto K_1(x_1, y_1) K_2(x_2, y_2) \in \mathbb{F}.$$

Then (H, K) is a RKHS.

(b) Assume $(E :=) E_1 = E_2$. The RKHS the kernel $K_E(x, y) := K_1(x, y) K_2(x, y)$ coincides with $H_E := \{f|_{E \times E} \mid f \in H_1 \otimes H_2\}$.

Proof. (a) It is easy to see that

$$\varphi : E_1 \times E_2 \ni (x_1, x_2) \mapsto (K_1(\cdot, x_1) K_2(\cdot, x_2)) \in H$$

is a feature map of K. Let H' be the RKHS generated by φ (and hence by K). Let H_0 be the subspace of H' spanned by $\{K(\cdot, x)\}_{x \in E_1 \times E_2}$. Note that $H_0 \subset H_1 \bullet H_2$ is dense in H', and that H' is isomorphic to the completion of H_0 and hence to that of $H_1 \bullet H_2$. It thus follows that H' and $H_1 \otimes H_2$ must coincide. $\square$

3.3 Reconstruction via Operator

4 Inheritance from Kernel to RKHS

4.1 Topological Properties of RKHS

4.2 Differentiablity of RKHS

5 Representor Theorem and its Application

References

[1] Alain Berlinet and Christine Thomas-Agnan. *Reproducing Kernel Hilbert Spaces in Probability and Statistics*. Springer, 1st edition, 2004.

[2] Andreas Christmann and Ingo Steinwart. *Support vector machines*. Springer, 1st edition, 2008.

[3] Felipe Cucker and Ding Xuan Zhou. *Learning theory: An approximation theory viewpoint*. Cambridge University Press, 1st edition, 2007.

[4] Vern I. Paulsen and Mrinal Raghupathi. *An Introduction to the Theory of Reproducing Kernel Hilbert Spaces*. Cambridge University Press, 1st edition, 2016.

[5] Bernhard Schlkopf and Alexander J. Smola. *Learning with Kernels: Support Vector Machines, Regularization, Optimization, and Beyond*. The MIT Press, 1st edition, 2001.