A Theory on Reproducing Kernel Hilbert Space for Machine Learning

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

2020/05/14

What's this?

This document collects basic results on Reproducing Kernel Hilbert Space 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]

Overview

Contents

1 Basic Property of RKHS

1.1 Properties of Reproducing Kernel

Definition 1.1 (Reproducing kernel). *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 1.2. (b) is called the textitreproducing property, and the identity is called the textitreproducing identity. ◁

Theorem 1.3 (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 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 1.4. *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. □

Proposition 1.5. *(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 $f \equiv 0$. (b) is an immediate consequence of (a) and Corollary1.4. 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 Corollary1.4 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. Exactly the same argument also works for the last claim. □

Proposition 1.6. *(Representation of RK in terms of evaluation functional) In 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 a mapping $I : H^\star \to H$ 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

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

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. □

Definition 1.7. *(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.

Remark 1.8. Proposition1.6 tells us that every RK is indeed a kernel, and that the map $E \ni x \mapsto \mathrm{ev}_x \in H^\star$ is a feature map with a feature space $H^\star$. Every RKHS (H, K) also admits a more simple feature map φ_K, called a canonical feature map, given by

$$\varphi_K : E \ni x \mapsto K(\cdot, x) \in H.$$

This clearly shows that, given a kernel, neither feature space nor feature map are uniquely determined. ◁

1.2 RKHS of a kernel

Definition 1.9. *(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*

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

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

Proposition 1.10. *(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 (1) holds for any $\alpha \in \mathbb{R}^n$ instead of $\mathbb{C}^n$.

Proposition 1.11. *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 1.12. *(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. $\square$

Theorem 1.13. *(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.

Proof. Suppose such an extension H exists. H satisfies (b) by Theorem1.3. 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. Let H be the set of all functions $f \in \mathbb{F}^E$ for which there exists a Cauchy sequence $\{f_n\} \subset H_0$ converging pointwise to f. Clearly, $H_0 \subset H \subset \mathbb{F}^E$. The rest of proof consists of the following Lemmas. $\square$

Lemma 1.14. *Let $f, g \in H$ and let $\{f_n\}$ and $\{g_n\}$ be two Cauchy sequences in H_0 that converge pointwise to f and g respectively.*

- *(A) The sequence $\langle f_n, g_n \rangle_{H_0}$ is convergent.*
- *(B) The limit $\lim_{n \to \infty} \langle f_n, g_n \rangle_{H_0}$ depends solely on f and g, independent of the choice of f_n and g_n.*
- *(C) $\langle f, g \rangle_H := \lim_{n \to \infty} \langle f_n, g_n \rangle_{H_0}$ is an inner product on H.*

Proof. It follows from the definition of f_n and g_n that

$$\begin{aligned}\left|\langle f_n, g_n\rangle_{H_0} - \langle f_m, g_m\rangle_{H_0}\right| &= |\langle f_n - f_m, g_n\rangle - \langle f_m, g_n - g_m\rangle| \\ &\geq \|g_n\| \, \|f_n - f_m\| + \|f_m\| \, \|g_n - g_m\| \to 0,\end{aligned}$$

which proves (A). In order to verify (B), suppose $\{f'_n\}$ and $\{g'_n\}$ are also such approximating sequences. We then similarly deduce that

$$|\langle f_n, g_n\rangle - \langle f'_n, g'_n\rangle| \leq \|g_n\| \, \|f_n - f'_n\| + \|f'_n\| \, \|g_n - g'_n\| .$$

$\{f_n - f'_n\}$ and $\{g_n - g'_n\}$ are Cauchy sequences tending pointwise to 0. Thus, assumption (c) gives $\|f_n - f'_n\| \to 0$ and $\|g_n - g'_n\| \to 0$. So, (A) and (B) show that $\langle f, g\rangle_H$ is well-defined. Note that if $\langle f, f\rangle_H = 0$, then for every $x \in E$

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

and hence $f \equiv 0$. As the symmetry, positivity, linearity are quite obvious, we conclude that (C) is true. □

Lemma 1.15. *(A) Let $f \in H$ and let $\{f_n\} \subset H_0$ be a Cauchy sequence converging pointwise to f. Then $f_n \to f$ also in H-norm.*

(B) H_0 is dense in H.

Proof. (A): Fix $\epsilon > 0$. Choose $N \in \mathbb{N}$ large enough so that

$$\|f_n - f_m\|_{H_0} < \epsilon$$

for all $n, m > N$. For fixed n, $\{f_n - f_m\}_{m\in\mathbb{N}}$ is a Cauchy sequence converging pointwise to $f_n - f$. Therefore, by definition of $\langle\cdot,\cdot\rangle_H$,

$$\|f - f_n\|_H = \lim_{n\to\infty} \|f_n - f_m\|_{H_0} \leq \epsilon.$$

(B) is obvious from (A). □

Lemma 1.16. *Every evaluation functional* ev_x *is continuous on H.*

Proof. Fix $x \in E$. As a linear functional ev_x is assumed to be continuous on H_0, it admits unique continuous extension T_x onto the closure of H_0 in H, that is, onto whole H, where we use the assumption (a) and Lemma2(B). T_x is also the evaluation functional on H. Indeed, for f and f_n as in Lemma1.15, we have

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

□

Lemma 1.17. *H is a RKHS satisfying (a) in Theorem1.13. Consequently, H is isomorphic to the completion of H_0 and thus unique RKHS that meets the requirement.*

Proof. We first prove that H is actually a RKHS. In light of Theorem1.3 and Lemma1.16, it suffices to show that H is complete. Let $\{f_n\}$ be a Cauchy sequence in H. Let $x \in E$. $\{f_n(x)\}$ is also a Cauchy in $\mathbb{F}$, and hence converges to some $f(x)$. By Lemma1.15, for every $n \in \mathbb{N}$, there is $g_n \in H_0$ such that $\|f_n - g_n\|_H < n^{-1}$. In view of the inequality

$$\|f - f_n\|_H \leq \|f - g_n\|_H + \|g_n - f_n\|_H ,$$

it suffices to prove that $\|f - g_n\|_H \to 0$. To this end, we show that $\{g_n\}$ is a Cauchy sequence converging pointwise to f (and then apply Lemma1.15).

For fixed $x \in E$, we have

$$\begin{aligned} |g_n(x) - f(x)| &\leq |g_n(x) - f_n(x)| + |f_n(x) - f(x)| \\ &= |\mathrm{ev}_x(g_n - f_n)| + |f_n(x) - f(x)| \to 0 \end{aligned}$$

as $n \to \infty$, since ev_x is continuous and $f_n(x) \to f(x)$ pointwise. Moreover,

$$\begin{aligned} \|g_n - g_m\|_{H_0} &= \|g_n - g_m\|_H \\ &\leq \|g_n - f_n\| + \|f_n - f_m\| + \|g_m - f_m\| \\ &= n^{-1} + \|f_n - f_m\| + n^{-1} \to 0 \end{aligned}$$

when $n, m \to \infty$, as required. H is isomorphic to the completion of H_0 since H_0 is dense in H. Uniqueness of (H, K) comes from Proposition1.5. □

Theorem 1.18. *(Moore-Aronszajn) For 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,

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

From this definiteness of $\langle \cdot, \cdot\rangle_{H_0}$ follows; 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 conditions of Theorem1.13. 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 Proposition1.5. □

Theorem 1.19. *(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 Corollary1.12. $\square$

Remark 1.20. Theorem1.19 implies that RKHS (H, K) is a natural feature space. $\triangleleft$

Theorem1.19 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*, as the following example illustrates.

Example 1.21. 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. $\triangleleft$

The next Theorem relates a feature map (therefore a kernel) and the RKHS the corresponding positive definite function generates.

Theorem 1.22. *(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

$$\|f\|_H := \inf\{\|w\|_{H_0} \,:\, w \in H_0,\, f = \langle w, \varphi_0(\cdot) \rangle_{H_0}\} \tag{2}$$

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

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

is an isometric isomorphism on $(\ker V)^{\perp}$.

Proof. In light of Theorem1.18, 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 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 an isomorphism. 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 this equation, we conclude that $V_1 : H_1 \to H$ is an isometric 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 $x \in E$. □

Remark 1.23. (Infimum in the norm $\|\cdot\|_H$ at (2) is attainable) We continue with the notation in The Theorem1.22. 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 $V^{-1}(\{f\})$. ◁

1.3 Basic Properties of Kernel

Proposition 1.24. *($\mathbb{R}$-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.

Proposition 1.25. *Let $K : E \times E \to \mathbb{F}$ be a kernel.*

(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 a kernel. In particular, the restriction of K to $E_1 \times E_1$ is a kernel if $E_1 \subset E$.

(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 an kernel.

2 Reconstruction of RKHS

2.1 Reconstruction via Restriction

Corollary 2.1. *(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

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

where $f|_A$ stands for the restriction of f to the set A, and 1_A is the indicator function of A.

Proof. Define a feature map

$$\varphi : E_1 \ni x \mapsto K(\cdot, x) 1_{E_1}(x) \in H,$$

and apply Theorem1.22 to see (H_1, K_1) is the RKHS generated by the feature map φ. Moreover, Remark1.23 tells us that the infimum of $\|f_1\|_{H_1}$ is achievable at $f \in H$ that should necessarily satisfy

$$f_1 = \langle f, K(\cdot, x) 1_{E_1}(x) \rangle_H \,,$$

that is, at $f = f_1 1 E_1(x)$. ∎

Proposition 2.2. *Let $K : E \times E \to \mathbb{C}$ be a kernel and H its corresponding $\mathbb{C}$-RKHS.*

(a) If $K(x, y) \in \mathbb{R}$ for all $x, y \in E$ then

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

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) Suppose $E = \mathbb{C}^d$ of the kernel $K : E \times E \to \mathbb{R}$. Then

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

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. Proposition1.24 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 Theorem1.22 then proves (a). (b) is an immediate consequence of (a) and Corollary2.1. $\square$

2.2 Reconstruction via Operator

2.3 Reconstruction via Sum and Product

3 Inheritance from Kernel to RKHS

3.1 Measurability of RKHS

3.2 Separability of RKHS

3.3 Continuity of RKHS

4 Mercer Representation

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.