What's this?

This document consists of a collection of basic results about Reproducing Kernel Hilbert Space useful in machine learning context. Nearly every statement and proof is borrowed mainly from [1] [2]. [3] [4] [5]

Overview

1 Basic Property of RKHS

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).

Theorem 1.2 (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

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

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. $\square$

Corollary 1.3. *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. $\square$

Proposition 1.4. *(Uniqueness of H and K)*

(a) 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.

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 Corollary1.3 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. $\square$

Proposition 1.5. *(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

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

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. □

Definition 1.6. *(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 1.7. Proposition1.5 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.

Definition 1.8. *(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 1.9. *(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 $\mathbb{C}^n$.

TEST CAREFULLY !!!!

Corollary 1.10. *Every 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. □

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. □

Theorem 1.12. *(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.

Proof. Suppose such an extension H exists. H satisfies (b) by Theorem1.2. 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. □

Lemma 1.13. *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. $\square$

Lemma 1.14. *(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). $\square$

Lemma 1.15. *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 Lemma1.14, we have

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

$\square$

Lemma 1.16. *H is a RKHS satisfying (a) in Theorem1.12. 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 Theorem1.2 and Lemma1.15, 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 Lemma1.14, 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 Lemma1.14).

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. Finally, suppose H' is also an qualifying RKHS. As H is isomorphic to H' there exists an isometric isomorphism between them. Furthermore, by Lemma1.14 and Lemma1.15, they share the same evaluation functionals $\{\mathrm{ev}_x\}_{x\in E}$ and hence the same reproducing kernel by Proposition1.5. Thus, they must coincide by Proposition1.4. □

Theorem 1.17. *(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 Theorem1.12. 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 Proposition1.4. □

Theorem 1.18. *(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 there exists some mapping φ of E into some 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 K, then $\varphi_K : E \ni x \mapsto K(\cdot, x) \in H$ obviously qualifies. Conversely, if such a φ exists,then for $\alpha \in \mathbb{F}^n$, $x \in E^n$

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

as required. □

Remark 1.19. Theorem1.18 implies that RKHS (H, K) is a natural feature space.

Theorem 1.20. *(RKHS)*

2 Reconstruction of RKHS

3 Inheritance from Kernel to 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.