Reproducing Kernel Hilbert Spaces

Lecturer: Clayton Scott   *Scribe: Hossein Keshavarz*

Disclaimer: *These notes have not been subjected to the usual scrutiny reserved for formal publications. They may be distributed outside this class only with the permission of the Instructor.*

1   Introduction

This note introduces a particular kind of Hilbert space known as a *reproducing kernel Hilbert space (RKHS)*. We will establish connections with kernels, defined previously, and show that kernels and RKHSs are in one-to-one correspondence. This material is largely drawn from Chapter 4 of [1], although some results are presented in a slightly different way to ease digestion.

2   Reproducing Kernel Hilbert Spaces

Before formally stating the definitions and results, let us to mention that throughout this note, we use the term "Hilbert function space over $\mathcal{X}$" to refer to a Hilbert space whose elements are functions $f : \mathcal{X} \mapsto \mathbb{R}$.

Definition 1. *(Reproducing Kernel) Let $\mathcal{F}$ be a Hilbert function space over $\mathcal{X}$. A reproducing kernel of $\mathcal{F}$ is a function $k : \mathcal{X} \times \mathcal{X} \mapsto \mathbb{R}$ which satisfies the following two properties:*

   1. $k(\cdot, x) \in \mathcal{F}$ for any $x \in \mathcal{X}$.

   2. (Reproducing property) For any $f \in \mathcal{F}$ and any $x \in \mathcal{X}$, $f(x) = \langle f, k(\cdot, x) \rangle_{\mathcal{F}}$.

Theorem 1. *Let $\mathcal{F}$ be a Hilbert function space over $\mathcal{X}$. The following are equivalent:*

   1. $\mathcal{F}$ has a reproducing kernel.

   2. For any $x \in \mathcal{X}$, the function $\delta_x : \mathcal{F} \mapsto \mathbb{R}$ defined by $\delta_x(f) = f(x)$ is continuous.

Definition 2. *Let $\mathcal{F}$ be a Hilbert function space over $\mathcal{X}$. We say $\mathcal{F}$ is a* reproducing kernel Hilbert space *over $\mathcal{X}$ if it satisfies the conditions of the previous theorem.*

Proof. $((1) \Rightarrow (2))$. Suppose that $\mathcal{F}$ has a reproducing kernel, $k : \mathcal{X} \times \mathcal{X} \mapsto \mathbb{R}$. We need to show that if $(f_n)_{n=1}^{\infty}$ converges to $f \in \mathcal{F}$, i.e., $\|f_n - f\|_{\mathcal{F}} \to 0$, then for every $x \in \mathcal{X}$, $|\delta_x(f_n) - \delta_x(f)| = |f_n(x) - f(x)| \to 0$. Now

$$\delta_x(f) = f(x) = \langle f, k(\cdot, x) \rangle_{\mathcal{F}} = \left\langle \lim_{n \to \infty} f_n, k(\cdot, x) \right\rangle_{\mathcal{F}}$$

$$\stackrel{(a)}{=} \lim_{n \to \infty} \langle f_n, k(\cdot, x) \rangle_{\mathcal{F}} = \lim_{n \to \infty} f_n(x) = \lim_{n \to \infty} \delta_x(f_n)$$

Note that the identity "(a)" is implied by the continuity of inner product operator of its first argument. Before proving the reverse implication, we state a classical result from functional analysis named the *Riesz representation theorem*.

Theorem 2 (Riesz representation theorem)**.** *Let $\mathcal{F}$ be a Hilbert space and $L : \mathcal{F} \mapsto \mathbb{R}$ be a linear functional on $\mathcal{F}$. Then L is continuous if and only if there exists $\Phi \in \mathcal{F}$ such that $L(f) = \langle f, \Phi \rangle_{\mathcal{F}}$ for any $f \in \mathcal{F}$.*

$((2) \Rightarrow (1))$. Let $x \in \mathcal{X}$. It's trivial to see that δ_x is a linear functional, so due to the assumed continuity of δ_x, Theorem 2 ensures the existence of a $\Phi_x \in \mathcal{F}$ such that $\delta_x (f) = f (x) = \langle f, \Phi_x \rangle_{\mathcal{F}}$. Now define $k (x_2, x_1) := \Phi_{x_1} (x_2) \in \mathbb{R}$. Note that $k (\cdot, x) = \Phi_x \in \mathcal{F}$ (establishing the first property of reproducing kernels). Also,

$$f (x) = \langle f, \Phi_x \rangle_{\mathcal{F}} = \langle f, k (\cdot, x) \rangle_{\mathcal{F}}$$

which shows that the reproducing property holds. $\qquad\square$

The next result establishes that kernels and reproducing kernels are the same.

Theorem 3. *Let $k : \mathcal{X} \times \mathcal{X} \mapsto \mathbb{R}$. Then k is a kernel if and only if k is a reproducing kernel of some RKHS $\mathcal{F}$ over $\mathcal{X}$.*

Proof. $(\Leftarrow)$. We first prove the reverse implication. Let k be the reproducing kernel of $\mathcal{F}$ and define $\Phi : \mathcal{X} \mapsto \mathcal{F}$ by $\Phi (x) = k (\cdot, x)$. Now for any $x' \in \mathcal{X}$, consider the function $f_{x'} = k (\cdot, x')$. Using the reproducing property and symmetry of the inner product, we obtain

$$k (x, x') = f_{x'} (x) = \langle f_{x'}, k (\cdot, x) \rangle_{\mathcal{F}} = \langle k (\cdot, x'), k (\cdot, x) \rangle_{\mathcal{F}}$$
$$= \langle \Phi (x'), \Phi (x) \rangle_{\mathcal{F}} = \langle \Phi (x), \Phi (x') \rangle_{\mathcal{F}}.$$

Hence, k is a kernel.

$(\Rightarrow)$. The following lemma is needed.

Lemma 1. *If k is a kernel, then it has a feature space $\tilde{\mathcal{F}}$ and a feature map $\tilde{\Phi} : \mathcal{X} \mapsto \tilde{\mathcal{F}}$ such that the map $\mathcal{V} : \tilde{\mathcal{F}} \mapsto (\mathcal{X} \mapsto \mathbb{R})$ (the range of $\mathcal{V}$ is the set of functions from $\mathcal{X}$ to $\mathbb{R}$) given by*

$$\mathcal{V} (\omega) = \langle \omega, \tilde{\Phi} (\cdot) \rangle_{\tilde{\mathcal{F}}} \qquad\qquad (1)$$

is injective.

Proof of Lemma 1. Let $\mathcal{F}_0$, Φ_0 be a feature space and feature map associated to k, respectively, i.e., $k (x, x') = \langle \Phi_0 (x), \Phi_0 (x') \rangle_{\mathcal{F}_0}$. Now define

$$\mathcal{F} = \{ f = \langle \omega, \Phi_0 (\cdot) \rangle_{\mathcal{F}_0} \mid \omega \in \mathcal{F}_0 \}$$

It's worthwhile to mention that, since Φ_0 is not known to be onto, then it's possible to have multiple ω's mapping to f. Consider the kernel (nullspace) $\mathcal{V}$, $ker (\mathcal{V}) = \{ \omega : \mathcal{V}\omega = \langle \omega, \Phi_0 (\cdot) \rangle_{\mathcal{F}_0} = 0 \}$, where the 0 refers to the zero function. One can easily show using the continuity of an inner product in its first argument that $ker (\mathcal{V})$ is closed. The projection theorem then implies that $\mathcal{F}_0 = ker (\mathcal{V}) \oplus \tilde{\mathcal{F}}$ ($\oplus$ denotes the direct sum) where

$$\tilde{\mathcal{F}} = (ker (\mathcal{V}))^{\perp} = \{ v \in \mathcal{F}_0 : \langle \omega, v \rangle_{\mathcal{F}_0} = 0 \; \forall \; \omega \in ker (\mathcal{V}) \}$$

$\tilde{\mathcal{F}}$ is closed by a routine argument, and therefore $\tilde{\mathcal{F}}$ is a Hilbert space by inheriting the inner product of $\tilde{\mathcal{F}}$. For any $\omega \in ker (\mathcal{V})$ and any $x \in \mathcal{X}$, the construction of $\mathcal{V}$ leads to $\mathcal{V} (\omega) = \langle \omega, \Phi_0 (x) \rangle_{\mathcal{F}_0} = 0$. Thus, $\Phi_0 (x) \in (ker (\mathcal{V}))^{\perp} = \tilde{\mathcal{F}}$. Hence, $\tilde{\Phi} = \Phi_0$ is a valid feature map on $\mathcal{X} \to \tilde{\mathcal{F}}$. Finally, the restriction $\mathcal{V} |_{\tilde{\mathcal{F}}}: \tilde{\mathcal{F}} \mapsto \mathcal{F}$ is such that $ker (\mathcal{V} |_{\tilde{\mathcal{F}}}) = \{0\}$, and therefore is injective. $\qquad\square$

Let $(\tilde{\Phi}, \tilde{\mathcal{F}})$ be as in Lemma 1. Define

$$\mathcal{F} = \left\{ f = \langle \omega, \tilde{\Phi} (\cdot) \rangle_{\tilde{\mathcal{F}}} \mid \omega \in \tilde{\mathcal{F}} \right\}$$

$\mathcal{F}$ is clearly a space of functions. The linear map $\mathcal{V} : \tilde{\mathcal{F}} \mapsto \mathcal{F}$ defined by identity (1) (linearity of $\mathcal{V}$ is an immediate consequence of linearity an of inner product in its first argument) is bijective (injective and onto) by the constructive definition of $\mathcal{F}$ and therefore $\mathcal{V}$ is invertible. Now, define an inner product on $\mathcal{F}$ by

$$\langle f, f' \rangle_{\mathcal{F}} = \langle \mathcal{V}^{-1} f, \mathcal{V}^{-1} f' \rangle_{\tilde{\mathcal{F}}}.$$

Since $\mathcal{V}$ is a bijective linear map, so is its inverse, and so it's straightforward to check that $\langle \cdot, \cdot \rangle_{\mathcal{F}}$ defines a valid inner product. Therefore $\mathcal{V}$ is also an isometry (distance preserving map). We also need to show that $\mathcal{F}$ is complete. Thus, let $\{f_n\}_{n \in \mathbb{N}}$ be a Cauchy sequence in $\mathcal{F}$. Since $\mathcal{V}$ is an isometry, we have that $\|f_n - f_m\|_{\mathcal{F}} = \|\mathcal{V}^{-1} f_n - \mathcal{V}^{-1} f_m\|_{\tilde{\mathcal{F}}}$ for any $m, n \in \mathbb{N}$. Hence, $(\mathcal{V}^{-1} f_n)_{n \in \mathbb{N}}$ is a Cauchy sequence in the complete space $\tilde{\mathcal{F}}$ which means that it converges to a limit $\tilde{f} \in \tilde{\mathcal{F}}$. Since $\mathcal{V}$ is an isometry it is continuous, and therefore $f_n = \mathcal{V}(\mathcal{V}^{-1} f_n)$ converges to $\mathcal{V}\tilde{f} \in \mathcal{F}$. This shows that $\mathcal{F}$ is complete.

Next, we show that k is a reproducing kernel of $\mathcal{F}$. If $x \in \mathcal{X}$, then

$$k\left(\cdot, x\right) = \langle \tilde{\Phi}\left(x\right), \tilde{\Phi}\left(\cdot\right) \rangle_{\tilde{\mathcal{F}}} = \mathcal{V}\tilde{\Phi}\left(x\right) \in \mathcal{F}$$

This proves the first property of a reproducing kernel. Lastly, for any $f \in \mathcal{F}$ and any $x \in \mathcal{X}$, observe

$$f\left(x\right) = \langle \mathcal{V}^{-1}\left(f\right), \tilde{\Phi}\left(x\right) \rangle_{\tilde{\mathcal{F}}} = \langle f, \mathcal{V}\tilde{\Phi}\left(x\right) \rangle_{\mathcal{F}} = \langle f, k\left(\cdot, x\right) \rangle_{\mathcal{F}}$$

which established the reproducing property. $\square$

3 Kernels and RKHSs are in one-to-one correspondence

The next two results establish that kernels and RKHSs are in one-to-one correspondence.

Theorem 4. *If $\mathcal{F}$ is a RKHS, then it has a unique reproducing kernel.*

The proof is left as an exercise below.

Theorem 5. *Let k be a kernel on $\mathcal{X}$, and let $\mathcal{F}_0$ and $\Phi_0 : \mathcal{X} \mapsto \mathcal{F}_0$ be an associated feature space and feature map, respectively, such that the mapping $\mathcal{V} : \mathcal{F}_0 \mapsto \mathcal{F} = \{f = \langle \omega, \Phi_0\left(\cdot\right) \rangle_{\mathcal{F}_0} \mid \omega \in \mathcal{F}_0\}$ is injective, whose existence is guarenteed by Lemma 1. Then,*

1. *$\mathcal{F}$ is the unique RKHS with k as its reproducing kernel.*

2. *The set $\mathcal{F}_{pre} = \left\{ \sum_{i=1}^{n} \alpha_i k\left(\cdot, x_i\right) \mid n \in \mathbb{N}, \ \alpha_i \in \mathbb{R}, x_i \in \mathcal{X} \right\}$ is dense in $\mathcal{F}$.*

3. *For $f, f' \in \mathcal{F}_{pre}$ with $f = \sum_{i=1}^{n} \alpha_i k\left(\cdot, x_i\right)$ and $f' = \sum_{i=1}^{n'} \alpha'_i k\left(\cdot, x'_i\right)$,*

$$\langle f, f' \rangle_{\mathcal{F}} = \sum_{i=1}^{n} \sum_{j=1}^{n'} \alpha_i \alpha'_j k\left(x_i, x'_j\right).$$

Proof. We already have seen in the proof of Theorem 3 that $\mathcal{F}$ is an RKHS. The proof of the uniqueness is presented after proving the other two parts of the theorem.

Proof of part 2. Let $\tilde{\mathcal{F}}$ be any RKHS with k as its reproducing kernel. Since $k\left(\cdot, x\right) \in \tilde{\mathcal{F}}$ for any $x \in \mathcal{X}$, we see that $\mathcal{F}_{pre} \subseteq \tilde{\mathcal{F}}$. Now, suppose towards a contradiction that $\mathcal{F}_{pre}$ is not dense in $\tilde{\mathcal{F}}$. Then the orthogonal complement of the closure of $\mathcal{F}_{pre}$ has a non-trivial element, i.e. $\left(\overline{\mathcal{F}_{pre}}\right)^{\perp} \neq \{0\}$. Thus, there exists $f \in \left(\overline{\mathcal{F}_{pre}}\right)^{\perp}$ and $x \in \mathcal{X}$ such that $f\left(x\right) \neq 0$. But $k\left(\cdot, x\right) \in \mathcal{F}_{pre}$, and so

$$0 = \langle f, k\left(\cdot, x\right) \rangle_{\mathcal{F}} = f\left(x\right) \neq 0,$$

a contradiction.

Proof of part 3. Let again $\tilde{\mathcal{F}}$ be any RKHS with k as its reproducing kernel. Using the reproducing property and bi-linearity of inner product, we get

$$\langle f, f' \rangle_{\tilde{\mathcal{F}}} = \sum_{i=1}^{n} \sum_{j=1}^{n'} \alpha_i \alpha_j' \langle k\left(\cdot, x_i\right), k\left(\cdot, x_j'\right) \rangle_{\tilde{\mathcal{F}}} = \sum_{i=1}^{n} \sum_{j=1}^{n'} \alpha_i \alpha_j' k\left(x_i, x_j'\right).$$

Proof of part 1. Assume that $\mathcal{F}_1$ and $\mathcal{F}_2$ are two RKHSs with the same reproducing kernel k. Choose an arbitrary $f \in \mathcal{F}_1$. We will show $f \in \mathcal{F}_2$, establishing $\mathcal{F}_1 \subseteq \mathcal{F}_2$. The reverse inclusion is established similarly.

From part 2, there is a sequence $(f_n)_{n \in \mathbb{N}}, f_n \in \mathcal{F}_{pre}$, such that $\|f - f_n\|_{\mathcal{F}_1} \to 0$. By part 3, $\|\cdot\|_{\mathcal{F}_1}$ and $\|\cdot\|_{\mathcal{F}_2}$ coincide on $\mathcal{F}_{pre}$, and so $(f_n)_{n \in \mathbb{N}}$ is Cauchy in $\mathcal{F}_2$. Therefore, $(f_n)_{n \in \mathbb{N}}$ converges to some $g \in \mathcal{F}_2$. Moreover, for any $x \in \mathcal{X}$, we have

$$g(x) = \lim_{n \to \infty} f_n(x) \qquad \text{(point evaluation at } x \text{ is continuous on } \mathcal{F}_2\text{)}$$
$$= f(x), \qquad \text{(point evaluation at } x \text{ is continuous on } \mathcal{F}_1\text{)}$$

so $f = g$. This shows that $\mathcal{F}_1 \subseteq \mathcal{F}_2$, and the reverse inclusion holds similarly, so $\mathcal{F}_1 = \mathcal{F}_2$ as sets.

It remains to show that the inner products agree. Thus, let f, f' be arbitrary functions in $\mathcal{F}_1 = \mathcal{F}_2$, expressed as limits of functions in $\mathcal{F}_{pre}$ as $f = \lim_{n \to \infty} f_n$ and $f' = \lim_{n \to \infty} f_n'$. Appealing to continuity of an inner product in one argument while the other is held fixed,

$$\langle f, f' \rangle_{\mathcal{F}_1} = \lim_{n \to \infty} \langle f_n, f' \rangle_{\mathcal{F}_1} = \lim_{n \to \infty} \lim_{m \to \infty} \langle f_n, f_m' \rangle_{\mathcal{F}_1} = \lim_{n \to \infty} \lim_{m \to \infty} \langle f_n, f_m' \rangle_{\mathcal{F}_{pre}}$$
$$= \lim_{n \to \infty} \lim_{m \to \infty} \langle f_n, f_m' \rangle_{\mathcal{F}_2} = \lim_{n \to \infty} \langle f_n, f' \rangle_{\mathcal{F}_2} = \langle f, f' \rangle_{\mathcal{F}_2}.$$

So the two RKHSs are the same, and thus $\mathcal{F}$ is the unique RKHS with k as its reproducing kernel. $\square$

By the above properties, for any kernel k with associated RKHS $\mathcal{F}$, the feature map $\Phi : \mathcal{X} \to \mathcal{F}$ defined by $\Phi(x) = k(\cdot, x)$ is a valid feature map, and is called the *canonical feature map*.

For additional background on reproducing kernel Hilbert spaces, see [1].

Exercises

1. Show that convergence of functions in an RKHS implies pointwise convergence. That is, if $\|f_n - f\|_{\mathcal{F}} \to 0$ then for all x, $|f_n(x) - f(x)| \to 0$.

2. Is the Hilbert space $L^2(\mathbb{R})$ a RKHS? If so, what is its reproducing kernel? *Hint:* This problem is super easy if you understand the definitions of RKHSs and $L^2(\mathbb{R})$.

3. Show that if $\mathcal{F}$ is a RKHS, then it has a unique reproducing kernel.

4. Elements of an RKHS inherit properties of the associated kernel. Let k be a kernel on a metric space $\mathcal{X}$ and $\mathcal{F}$ its RKHS. We say k is *bounded* if $\sup_{x \in \mathcal{X}} k(x, x) < \infty$. Show that

   (a) if k is bounded, then every $f \in \mathcal{F}$ is a bounded function.

   (b) if k is bounded, then convergence in $\mathcal{F}$ implies convergence in the supremum norm.

   (c) if k is bounded and for all x, $k(\cdot, x)$ is a continuous function, then every $f \in \mathcal{F}$ is bounded and continuous. *Hint:* Use the denseness of $\mathcal{F}_{pre}$ in $\mathcal{F}$ and part (b).

5. Let k be a kernel with feature map $\Phi_0 : \mathcal{X} \to \mathcal{F}_0$, and consider

$$\mathcal{F} = \{f = \langle \omega, \Phi_0(\cdot) \rangle_{\mathcal{F}_0} \mid \omega \in \mathcal{F}_0\}.$$

Show that

(a) Show that $\mathcal{F}$, equipped with the inner product constructed in the Proof of Theorem 3, is the RKHS associated with k, and is therefore independent of the choice of $\Phi_0, \mathcal{F}_0$.

(b) for all $f \in \mathcal{F}$,
$$\|f\|_{\mathcal{F}} = \inf\{\|\omega\|_{\mathcal{F}_0} \,:\, \omega \in \mathcal{V}^{-1}f\}$$
where $\mathcal{V}\omega = \langle\omega, \Phi_0(\cdot)\rangle_{\mathcal{F}_0}$. *Note:* $\mathcal{V}$ is possibly many-to-one.

6. Consider the proof of Theorem 5.2 in [2], available online through the UM library. The last part of the proof (after the validity of the inner product is established) may be flawed. If the proof is flawed, identify the error(s), and if possible, salvage the proof without making major changes to the proof strategy. If the proof is correct, provide additional details to make the argument more transparent.

References

[1] I. Steinwart and A. Christman, *Support Vector Machines*, Springer, 2008.

[2] M. Mohri, A. Rostamizadeh, and A. Talwalkar, *Foundations of Machine Learning*, MIT Press, 2012.