Notes on Reproducing Kernel Hilbert Spaces

Tan Yee Fan

2009 January 16

1 Vector Spaces

In this article, we use $\mathbb{F}$ to denote a field of scalars. Here, $\mathbb{F}$ is taken to mean either the real field $\mathbb{R}$ and the complex field $\mathbb{C}$. Vector spaces need to be defined with respect to $\mathbb{F}$, but for ease of exposition this fact will often be omitted.

Let V be a set. Define an vector addition operator and a scalar multiplication on V that are both closed on V, i.e., for all $x, y \in V$ and for all $a \in \mathbb{F}$, we have $x + y \in V$ and $ax \in V$. Together V is a *vector space* provided that the following conditions are satisfied.

- For all $x, y \in V$, $x + y = y + x$.
- For all $x, y, z \in V$, $(x + y) + z = x + (y + z)$.
- There exists $0 \in V$ such that for all $x \in V$, $0 + x = x = x + 0$.
- For all $x \in V$, there exists $-x \in V$ such that $(-x) + x = 0 = x + (-x)$.
- There exists $1 \in \mathbb{F}$ such that for all $x \in V$, $1x = x$.
- For all $a, b \in \mathbb{F}$ and for all $x \in V$, $a(bx) = (ab)x$.
- For all $a \in \mathbb{F}$ and for all $x, y \in V$, $a(x + y) = ax + ay$.

Two examples of vector spaces are $\mathbb{R}^d$ and $\mathbb{C}^d$ where vector addition and scalar multiplication are the usual operations. As a vector $x = (x_1, x_2, \ldots, x_d) \in \mathbb{F}^d$ is essentially a function that maps an integer $k \in \{1, 2, \ldots, d\}$ to x_k, we can consider a vector as a function. However, in vector spaces, we are not restricted to functions whose domain is discrete, we can also consider functions that are continuous. When we consider the vectors in a vector as a function, we call the vector space a *function space*.

Let V be a vector space. For a set of vectors $S \subseteq V$, we say that a vector y is a *linear combination* of the vectors in S if there exists scalars a_i such that $y = \sum_{i \in S} a_i x_i$. The *span* of S is the set of linear combinations of the vectors in S. Note that S is a subset of the span of S. Two vectors x and y in V are *linearly independent* if y cannot be expressed as a linear combination of x. A set $S \subseteq V$ is said to be *linearly independent* if any two different vectors in S

are linearly independent. A *basis* of V is a set $S \subseteq V$ of vectors that is linearly independent and spans V. For the vector space $\mathbb{R}^d$, the *standard basis* of $\mathbb{R}^d$ is $\{(1, 0, 0, \ldots, 0), (0, 1, 0, \ldots, 0), (0, 0, 1, \ldots, 0), \ldots, (0, 0, 0, \ldots, 1)\}$.

A *transformation* is a function $T : V_1 \to V_2$, where V_1 and V_2 are vector spaces. When $V_1 = V_2$ and both are function spaces, then we call T an *operator*. When V_1 is a function space and V_2 is a scalar, we call T a *functional*. The transformation T becomes a *linear transformation* when the following hold:

- For all $x, y \in V_1$, $T(x + y) = T(x) + T(y)$.
- For all $a \in \mathbb{F}$, $T(ax) = aT(x)$.

When $V_1 = \mathbb{F}^m$ and $V_2 = \mathbb{F}^n$, then $T(x) = Ax$ for some matrix A. When an operator is a linear transformation, we call it a *linear operator*. When a functional is a linear transformation, we call it a *linear functional*.

2 More Vector Spaces

An *inner product space* is a vector space V equipped with a *inner product*. An inner product on V is a function $\langle \cdot, \cdot \rangle : V \times V \to \mathbb{F}$. An inner product over the space V is required to satisfy the following conditions.

- For all $x, y \in V$, $\langle x, y \rangle = \overline{\langle y, x \rangle}$.
- For all $x, y \in V$ and for all $a \in \mathbb{F}$, $\langle ax, y \rangle = a\langle x, y \rangle$.
- For all $x, y, z \in V$, $\langle x + y, z \rangle = \langle x, z \rangle + \langle y, z \rangle$.
- For all $x \in V$, $\langle x, x \rangle \geq 0$.
- For all $x \in V$, $\langle x, x \rangle = 0$ if and only if $x = 0$.

Note that the above conditions imply the following.

- For all $x \in V$, $\langle x, x \rangle \in \mathbb{R}$.
- For all $x, y \in V$ and for all $a \in \mathbb{F}$, $\langle x, ay \rangle = \overline{a}\langle x, y \rangle$.
- For all $x, y, z \in V$, $\langle x, y + z \rangle = \langle x, y \rangle + \langle x, z \rangle$.
- If $\mathbb{F} = \mathbb{R}$, then for all $x, y \in V$, $\langle x, y \rangle = \langle y, x \rangle$.

For example, $\mathbb{R}^d$ together with the usual vector dot product forms an inner product space. For any inner product space, the *Cauchy-Schwarz inequality* holds: $|\langle x, y \rangle|^2 \leq \langle x, x \rangle \langle y, y \rangle$. In an inner product space, two vectors x and y are said to be *orthogonal* if $\langle x, y \rangle = 0$. A set $S \subseteq V$ is an *orthogonal basis* of V if any two different vectors in S are orthogonal, and that S is a basis of V.

A *norm* on a vector space V is a function $\| \cdot \| : V \to \mathbb{R}$ satisfying the following conditions.

- For all $x \in V$, $\|x\| \geq 0$.

- For all $x \in V$, $\|x\| = 0$ if and only if $x = 0$.
- For all $x \in V$ and $a \in \mathbb{R}$, $\|ax\| = |a|\|x\|$.
- For all $x, y \in V$, $\|x + y\| \leq \|x\| + \|y\|$.

A *normed* vector space is a vector space where a norm is defined. A typical example of a norm defined on a vector space is $\|x\| = \sqrt{\langle x, y \rangle}$. For this norm, the Cauchy-Schwarz inequality can be restated as $|\langle x, y \rangle| \leq \|x\|\|y\|$. A vector x is called a *unit vector* if $\|x\| = 1$. Note that for any vector x where $\|x\| \neq 0$, the vector $\frac{x}{\|x\|}$ is a unit vector. For a normed inner product space V, we say that S is an *orthonormal basis* of V if every vector in S is a unit vector and S is an orthogonal basis of V. For example, the standard basis of $\mathbb{R}$ is an orthonormal basis of $\mathbb{R}$.

A *metric space* consists of a set V and a *distance function* $d : V \times V \to \mathbb{R}$ such that d is a *metric* on V, i.e.,

- For all $x, y \in V$, $d(x, y) \geq 0$.
- For all $x, y \in V$, $d(x, y) = 0$ if and only if $x = y$.
- For all $x, y \in V$, $d(x, y) = d(y, x)$.
- For all $x, y, z \in V$, $d(x, z) \leq d(x, y) + d(y, z)$.

Note that a norm induces a metric $d(x, y) = \|x - y\|$, hence all normed vector spaces are metric spaces.

Let V be a metric space with distance d. A sequence $x_1, x_2, x_3, \ldots$ of elements from V is called a *Cauchy sequence* if for all positive real number $\epsilon > 0$, there exists a positive integer $N > 0$ such that for all $m, n > N$, we have $d(x_m, x_n) < \epsilon$. The *limit* of this Cauchy sequence, if it exists, is defined to be $\lim_{n\to\infty} x_n$. A metric space V is *complete* if all Cauchy sequences of V has a limit in V. For example, $\mathbb{R}^d$ with Euclidean distance is also a complete metric set. However, the real open set $(0, \infty)$ is not complete because the limit of the Cauchy sequence $1, \frac{1}{2}, \frac{1}{3}, \ldots$ is 0, which is not in the set.

3 Hilbert Spaces

A *Banach space* is a complete normed vector space. A *Hilbert space* is a Banach space that is also an inner product space. It generalizes Euclidean spaces to infinite dimensions. For example, $\mathbb{R}^d$ with inner product defined as the dot product is a Hilbert space. An example of infinite-dimension Hilbert space is L_2, the set of functions $f : \mathbb{R} \to \mathbb{R}$ such that f is *square integrable* ($\int_{-\infty}^{\infty} |f(x)|^2 dx$ is finite), with $\langle f, g \rangle = \int_{-\infty}^{\infty} f(x)g(x)dx$. Because a Hilbert space is both a inner product space as well as a normed product space, any Hilbert space will have an orthonormal basis.

A *continuous linear functional* $f : V \to \mathbb{F}$ is a linear functional that also satisfies the property that whenever $O \subseteq \mathbb{F}$ is an open set, we have $f^{-1}(O) \subseteq V$

is also an open set. A functional f is said to be *bounded* if there exists a constant $M > 0$ such that for all $x \in V$, $|f(x)| \leq M\|f\|$. Note that a bounded functional is also continuous. The *Riesz representation theorem* states that if $f : H \to \mathbb{F}$ is a continuous linear functional mapping from a Hilbert space H to a scalar, then for all $x \in H$, there exists a unique $y \in H$ such that $f(x) = \langle x, y \rangle$. In reproducing kernel Hilbert spaces, we will study the case when H is a Hilbert function space.

4 Kernels

Let $X \subseteq \mathbb{R}^d$ be an *attribute space* and let H be a *feature space*. A mapping $\Phi : X \to H$ is known as a *feature map*. A function $K : X \times X \to \mathbb{R}$ satisfying $K(x, x') = \langle \Phi(x), \Phi(x') \rangle$ for all $x, x' \in X$ is called a *kernel*. Typically, H is assumed or defined to be a high-dimension or infinite-dimension Euclidean space, but it is not necessary so. Some properties of kernels are as follows.

- If K_1 and K_2 are kernels, then K defined by $K(x, x') = K_1(x, x') + K_2(x, x')$ is also a kernel.
- If K_1 is a kernel and $\alpha > 0$ is a real number, then K defined by $K(x, x') = \alpha K_1(x, x')$ is also a kernel.
- If K_1 and K_2 are kernels, then K defined by $K(x, x') = K_1(x, x')K_2(x, x')$ is also a kernel.

The *Mercer's theorem* gives a way to determine whether an arbitrary symmetric function $K : X \times X \to \mathbb{R}$ is a kernel. Let $x_1, \ldots, x_N$ be a collection of elements of X, and let $\mathbf{K}$ be the Gram matrix defined by $\mathbf{K}_{ij} = K(x_i, x_j)$. Then the theorem states that K can be expressed as an inner product if any Gram matrix $\mathbf{K}$ formed by any collection of elements of X is a positive semi-definite matrix, i.e., a matrix whose eigenvalues are non-negative. This is equivalent to saying that K is positive semi-definite, i.e., for all f, we have $\int \int f(x)K(x, x')f(x')dxdx' \geq 0$.

5 Reproducing Kernel Hilbert Spaces

Let H be a Hilbert function space over X, i.e., a Hilbert space which consists of functions mapping from X to $\mathbb{R}$. For an element $x \in X$, a *Dirac functional* is a function $\delta_x \in H$ such that for all $f \in H$, we have $\delta_x(f) = f(x)$. H is a *reproducing kernel Hilbert space* if all Dirac functionals in H are bounded. Thus, in a reproducing kernel Hilbert space H, the Riesz representation theorem applies, and hence we have $\delta_x(f) = f(x) = \langle f, K_x \rangle$ for some unique $K_x \in H$. Then, the *reproducing kernel* for the reproducing kernel Hilbert space H is defined as $K : X \times X \to \mathbb{R}$ with $K(x, x') = K_x(x')$. The *reproducing property* of this kernel is such that $\langle f, K(\cdot, x') \rangle = f(x')$, and hence $K(x, x') = \langle K(x, \cdot), K(x', \cdot) \rangle$. It can be shown that all reproducing kernels are positive semi-definite.

Note that for a given kernel, its feature map and feature space cannot be uniquely defined. However, one can construct a reproducing kernel Hilbert space that acts as the canonical feature space for the kernel, as well as the canonical feature map. Put another way, any positive semi-definite kernel is a reproducing kernel of some reproducing kernel Hilbert space. We now consider a kernel $K : X \times X \to \mathbb{R}$. Let $S = \{K(x, \cdot) | x \in X\}$, and let V to be the span of S. Then let H be the union of V and the limits of all Cauchy sequences in V. Let f and g be elements of H defined by $f = \sum_i \alpha_i K(x_i, \cdot)$ and $g = \sum_j \beta_j K(x_j, \cdot)$. Then we can define an inner product between f and g, by $\langle f, g \rangle = \sum_i \sum_j \alpha_i \beta_j \langle K(x_i, \cdot), K(x_j, \cdot) \rangle$. Thus we have defined a Hilbert space H, and we see that $\langle K(x, \cdot), K(x', \cdot) \rangle = K(x, x')$, and hence H is a reproducing kernel Hilbert space of K. Using H as the feature space, we can define a feature map $\Phi(x) = K(x, \cdot)$ which satisfies $\langle \Phi(x), \Phi(x') \rangle = K(x, x')$. It can be proven that H is the only reproducing kernel Hilbert space of K.

It has been noted that the feature space and feature map of a kernel is not unique. However, any feature space of a kernel will be isomorphic to the reproducing kernel Hilbert space of the kernel. Here we give another construction of the feature map. For a linear operator T, if a function f and a scalar λ satisfies $T(f) = \lambda f$, then f is an *eigenfunction* of T and λ is an *eigenvalue* of T. For a kernel K, we can find an eigenfunction ϕ and an eigenvalue λ satisfying $\langle K(x, \cdot), \phi \rangle = \lambda \phi$, or $\int K(x, x') \phi(x') dx' = \lambda \phi(x')$. As long as K is positive semi-definite, we can find an infinite sequence of eigenfunctions $\phi_1, \phi_2, \ldots$ and eigenvalues $\lambda_1, \lambda_2, \ldots$ of K, where $\{\phi_1, \phi_2, \ldots\}$ form an orthonormal basis of L_2 and $\lambda_1 \geq \lambda_2 \geq \ldots \geq 0$, such that $K(x, x') = \sum_{i=1}^{\infty} \lambda_i \phi_i(x) \phi_i(x')$. We can then define an inner product on L_2: for $f, g \in L_2$, $\langle f, g \rangle = \sum_{i=1}^{\infty} \frac{\langle f, \phi_i \rangle \langle g, \phi_i \rangle}{\lambda_i}$. Henceforth, we define another feature map for K to be $\Phi(x) = (\sqrt{\lambda_1} \phi_1(x), \sqrt{\lambda_2} \phi_2(x), \ldots)$, which satisfies $\langle \Phi(x), \Phi(x') \rangle = \sum_{i=1}^{\infty} \sqrt{\lambda_i} \phi_i(x) \sqrt{\lambda_i} \phi_i(x') = K(x, x')$ as desired.

References

[Daumé III, 2004] Daumé III, H. (2004). From zero to reproducing kernel hilbert spaces in twelve pages or less. Available at `http://hal3.name/docs/daume04rkhs.pdf`.

[Steinwart et al., 2006] Steinwart, I., Hush, D. R., and Scovel, C. (2006). An explicit description of the reproducing kernel hilbert spaces of gaussian rbf kernels. *IEEE Transactions on Information Theory*, 52(10):4635–4643.