Reproducing Kernel Hilbert Spaces

Laura Meneghetti

28 Novembre 2017

Contents

Introduction

Reproducing Kernel Hilbert Spaces (RKHS) have developed into an important tool in many areas:

- machine learning
- statistics
- complex analysis
- probability
- group representation theory
- theory of integral operators
- perceptron-style algorithm

Fields

Definition (Field)

A **field** $\mathbb{F}$ is a structure $\langle F, +, \cdot, 0, 1\rangle$ that consists of a universe F, an addition operation $(+)$, a multiplication operation $(\cdot)$, an identity for addition 0 and an identity for moltiplication 1. Furthermore, there exists an inverse operation for addition $(-)$ and when elements are $\neq 0$ an inverse operation for moltiplication $(\cdot)^{-1}$. In addition those operations must satisfy $\forall a, b, c \in F$:

- Associative Laws:

$$a + (b + c) = (a + b) + c$$
$$a \cdot (b \cdot c) = (a \cdot b) \cdot c$$

- Commutative Laws:

$$a + b = b + a$$
$$a \cdot b = b \cdot a$$

Definition (Field)

- Distributive Laws:

$$a \cdot (b + c) = a \cdot b + a \cdot c$$
$$(a + b) \cdot c = a \cdot c + b \cdot c$$

- Identity Laws:

$$a + 0 = 0 + a = a$$
$$a \cdot 1 = 1 \cdot a = a$$

- Inverse Laws:

$$a + (-a) = (-a) + a = 0$$
$$a \cdot a^{-1} = a^{-1} \cdot a = 1, \text{ when } a \neq 0$$

Examples (of Fields)

The rational numers $\mathbb{Q}$, the real numbers $\mathbb{R}$, the complex numbers $\mathbb{C}$

Note that the integers $\mathbb{Z}$ are not a field, because they don't have a multiplicative inverse.

Definition (Ordered Field)

An **ordered field** $\mathbb{F}$ is a field with a binary relation $(\leq)$, that is a linear order: for any $a, b, c \in F$, $(\leq)$ must satisfy these properties:

- Reflexive: $a \leq a$
- Antisymmetric: $a \leq b \;\wedge b \leq a \implies a = b$
- Transitive: $a \leq b \;\wedge\; b \leq c \implies a \leq c$

When $a \leq b \wedge a \neq b$, we will write $a < b$.

Examples (of ordered fields)

The rational $\mathbb{Q}$, the reals $\mathbb{R}$ and the complex numbers $\mathbb{C}$ with the usual ordering.

Complete Space

Definition (Metric)

A **metric space** X is an ordered pair (X, d) where X is a set and d is a metric on X, i.e. a function $d : X \times X \to \mathbb{R}$ such that for any $x, y, z \in X$ the following conditions are satisfied:

- $d(x, y) \geq 0$ (non-negative)
- $d(x, y) = 0 \iff x = y$
- $d(x, y) = d(y, x)$ (symmetry)
- $d(x, z) \leq d(x, y) + d(y, z)$ (triangle inequality)

When $X = \mathbb{R}$ we have that $d(x, y) = |x - y|$.

Definition (Complete space)

Given X a metric space. X is a **complete space** if every Cauchy sequence in X is convergent.

Definition (Cauchy sequence)

Given a metric space (X, d), a **Cauchy sequence** is a sequence $\langle x_i \rangle_{i=0}^{\infty}$ such that: $\forall \epsilon > 0, \epsilon \in \mathbb{R} \;\; \exists N \in \mathbb{N}$ such that $d(x_n, x_m) < \epsilon, \; \forall n, m > N$.

Definition (Convergent sequence)

Let (X, d) be a metric space. A sequence $\langle x_i \rangle_{i=0}^{\infty}$ is **convergent** in X if there is a point $x \in X$ such that $\forall \epsilon > 0, \epsilon \in \mathbb{R} \;\; \exists N \in \mathbb{N}$ such that $d(x, x_n) < \epsilon, \; \forall n > N$.

$$\text{Convergent sequence} \implies \text{Cauchy sequence}$$

The converse is not true!! For example: the decimal expansion of $\sqrt{2}$, $1, 1.4, 1.41, 1.414, 1.4142, ..$ is a Cauchy sequence in $\mathbb{Q}$ that converge to $\sqrt{2} \notin \mathbb{Q}$ or the sequence $\left(1 - \frac{1}{n}\right)_{n=1}^{\infty}$ in the open interval $X = (-1, 1)$ is a Cauchy sequence, that converge to $1 \notin X$.

Definition (Complete field)

Given X a field, X is a **complete field** if it is equipped with a metric and is complete with respect to that metric.

Example (Complete ordered field)

An example of a complete ordered field is given by the set of real numbers $\mathbb{R}$. In this case

$$\text{Convergent sequence} \iff \text{Cauchy sequence}$$

In particular $\mathbb{R}$ is the unique (up to isomorphism) complete ordered field, i.e. for any complete ordered field $\mathbb{F}$ we can find an isomorphism ϕ between $\mathbb{F}$ and $\mathbb{R}$.

Note: $\mathbb{Q}$ obviously is not a complete field.

Definition (Isomorphism)

Given two fields $\mathbb{F}$ and $\mathbb{G}$, ϕ is an isomorphism between $\mathbb{F}$ and $\mathbb{G}$ if ϕ is a function $F \to G$ and obeys certain properties:

- Injective (one-to-one): $\forall f, f' \in F$ s.t. $\phi(f) = \phi(f') \implies f = f'$
- Surjective (onto): $\forall g \in G \;\; \exists f \in F$ s.t. $\phi(f) = g$
  The combination of these two properties states that ϕ is a bijection.
- Preservation: ϕ preserves operations: $\phi(a + b) = \phi(a) + \phi(b)$ and $\phi(ab) = \phi(a)\phi(b)$. Then the image of the identities of F must be the identities of G: $\phi(1_F) = 1_G$ and $\phi(0_F) = 0_G$.

For ordered fields we have to add a clause to the preservation statement.

ϕ must preserve relative ordering:
given $f, f' \in F$ we have $f \leq f' \iff \phi(f) \leq \phi(f')$.

Vector spaces

Definition (Vector Space)

Let $\mathbb{F}$ be a field. $\mathbb{V}$ is a vector space over $\mathbb{F}$ if $\mathbb{V}$ is a structure of the form $\langle V, \mathbb{F}, \oplus, \otimes, \ominus, 0_{\mathbb{V}}\rangle$ consisting of a universe V, a vector addition operation $\oplus$, a scalar multiplication operation $\otimes$, a unary additive inverse operation $\ominus$ and an identity element $0_{\mathbb{V}}$. For any $u, v, w \in V$ and $a, b \in F$

- Associative Law: $(u \oplus v) \oplus w = u \oplus (v \oplus w)$
- Commutative Law: $u \oplus v = v \oplus u$
- Inverse Law: $u \oplus (\ominus u) = 0_{\mathbb{V}}$
- Identity Laws:
$$0_{\mathbb{V}} \oplus u = u$$
$$1 \otimes u = u$$
- Distributive Laws:
$$a \otimes (b \otimes u) = (ab) \otimes u$$
$$(a \oplus b) \otimes u = a \otimes u \oplus b \otimes u$$

Examples of Vector Spaces

1. $\mathbb{R}$, which is a vector space over $\mathbb{R}$. Vector addition and multiplication are just addition and multiplication on $\mathbb{R}$.
2. $\mathbb{R}^n$, the space of n-dimensional vectors of real numbers. Addition is defined point-wise and scalar multiplication is defined by multiplying each element in the vector by the scalar (using standard multiplication in $\mathbb{R}$).
3. $\mathbb{R}^{\mathbb{R}}$, the set of functions from $\mathbb{R} \to \mathbb{R}$. Addition and multiplication are defined point-wise by $(fg)(x) = f(x)g(x)$, $(f+g)(x) = f(x) + g(x)$ and $(af)(x) = af(x)$ for $f, g \in \mathbb{R}^{\mathbb{R}}$ and $a \in \mathbb{R}$. This is a vector space over $\mathbb{R}$.
4. $\mathbb{R}^X$, the set of functions from $X \to \mathbb{R}$, where X is a metric space, is a vector space over $\mathbb{R}$.

Examples of Vector Spaces

5. The set of continuos functions from a metric space $X \to \mathbb{R}$, $C(X)$, forms a vector space over $\mathbb{R}$ using the usual definitions of addition and scalar multiplication.

Recall the definition of continuous function.

Definition (Continuous Function)

Let X be a metric space. A function $f : X \to \mathbb{R}$ is **continuous** at a point $x_0 \in X$ if

$$\forall \epsilon > 0 \;\; \exists \delta > 0 \text{ s.t. } d(x, x_0) < \delta \implies |f(x) - f(x_0)| < \epsilon.$$

A function is continuous if it is continuous everywhere.

Banach Spaces

Definition (Banach Spaces)

A **Banach Space** is a complete normed vector space $(\mathbb{V}, \|\cdot\|)$.

Definition (Norm)

A norm is a function on a vector space $\mathbb{V}$ over $\mathbb{R}$ from V to $\mathbb{R}$ satisfying the following properties $\forall u, v \in V$ and $\forall a \in \mathbb{R}$:

- Non-negative: $\|u\| \geq 0$
- Strictly-Positive: $\|u\| = 0 \implies u = 0$
- Homogeneous: $\|au\| = |a|\|u\|$
- Triangle inequality: $\|u + v\| \leq \|u\| + \|v\|$

Given $v \in V$ the norm of v is indicated as $\|v\|$. A vector space endowed with a norm is called a **normed vector space**.

A single vector space could have multiple norms, but if $\mathbb{V}$ is finite there is at most one norm. (All the norms are equivalent to each other).
We have referred to **complete vector spaces**, whose definition is the same of fields, with the exception that the distance metric $d(x, y)$ is replaced by $\|x - y\|$, where $\|\cdot\|$ is a suitable norm.

Definition (Complete vector space)

Given $\mathbb{V}$ a vector space, $\mathbb{V}$ is a **complete vector space** if it is equipped with a norm and is complete with respect to that norm, i.e.
for every Cauchy sequence $\langle x_n \rangle_{n=0}^{\infty}$ in V there exists an element $x \in V$ such that

$$\lim_{n\to\infty} x_n = x$$

or equivalently

$$\lim_{n\to\infty} \|x_n - x\|_{\mathbb{V}} = 0.$$

Examples of Banach Spaces

1. $\mathbb{R}$ is a complete vector space over $\mathbb{R}$ and is also a Banach space. The most common norm for $\mathbb{R}$ is the absolute norm: $\|x\| = |x|$.
2. $\mathbb{R}^n$ is a complete vector space over $\mathbb{R}$ for any $n > 0$. There are several of norms we can define on it:
   - the euclidean-norm or 2-norm:
     $$\|\langle x_i \rangle_{i=1}^n\|_2 = \sqrt{\sum_{i=1}^n x_i^2}$$
   - the p-norm (for $p \geq 1$):
     $$\|\langle x_i \rangle_{i=1}^n\|_p = \left(\sum_{i=1}^n |x_i|^p\right)^{\frac{1}{p}}$$
   - the maximum or the ∞ norm:
     $$\|\langle x_i \rangle_{i=1}^n\|_\infty = \max\{|x_1|, |x_2|, \ldots, |x_n|\}$$

3. Sequences of infinite length, $\mathbb{R}^\infty$, form a Banach space: To ensure completeness, we need to make sure that these sequences don't diverge under summing. For a positive integer p, we define the ℓ_p space as:

$$\ell_p = \left\{ \langle x_i \rangle_{i=0}^{\infty} : \sum_{i=0}^{\infty} |x_i|^p < \infty \right\}$$

We define a norm on ℓ_p by: $\| \langle x_i \rangle_{i=1}^{\infty} \|_p = \left(\sum_{i=1}^{\infty} |x_i|^p \right)^{\frac{1}{p}}$

4. The space of continuous functions from a metric space X to $\mathbb{R}$, $C(X)$ is a Banach space. The natural norm, called the uniform norm or the sup norm, is defined by: $\|f\|_{sup} = \sup_{x \in X} |f(x)|$.

5. The $\mathcal{L}_p$ spaces over functions from $\mathbb{R}^n$ to $\mathbb{R}$ are Banach spaces. $\mathcal{L}_p$ is defined by:

$$\mathcal{L}_p = \left\{ (f : \mathbb{R}^n \to \mathbb{R}) : \int_{-\infty}^{\infty} |f^p(x)| dx < \infty \right\}$$

We define a norm on $\mathcal{L}_p$ by: $\|f\|_{\mathcal{L}_p} = \left(\int_{-\infty}^{\infty} |f^p(x)| dx \right)^{\frac{1}{p}}$

Hilbert Spaces

Definition (Hilbert Space)

A **Hilbert Space** $\mathscr{H}$ is a Banach Space endowed with a dot-product operation, therefore is a Banach Space whose norm derives from an internal product.

Definition (Inner Product)

Given a vector space $\mathbb{H}$ over a field $\mathbb{F}$, an **internal or dot or scalar product** is a map $\langle \cdot, \cdot \rangle : H \times H \to F$, which is a symmetric positive definite bilinear form: $\forall u, v, w \in \mathscr{H}, a \in \mathbb{F}$

- Symmetry: $\langle v, w \rangle = \langle w, v \rangle$
- Linearity with respect first term: $\langle v + w, u \rangle = \langle v, u \rangle + \langle w, u \rangle$
- Linearity with respect second term: $\langle v, w + u \rangle = \langle v, w \rangle + \langle v, u \rangle$
- Associative: $\langle au, v \rangle = a \langle u, v \rangle$
- Positive Definite: $\langle v, v \rangle > 0 \;\; \forall v \neq 0$

Given a complete vector space $\mathbb{V}$ with a dot product $\langle \cdot, \cdot \rangle_{\mathbb{V}}$ we can define a norm on $\mathbb{V}$ by $\|u\| = \sqrt{\langle u, u \rangle}$, thus making this space into a Banach space and therefore into a full Hilbert space.

Note: Not all Banach Spaces can be made into Hilbert Spaces.

Examples (of Hilbert Spaces)

- $\mathbb{R}$ and $\mathbb{R}^n$:
  $\mathbb{R}^n$ is an Hilbert space for the Euclidean norm. The dot-product is defined as $\langle u, v \rangle_{\mathbb{R}^n} = u \cdot v = \sum_{i=1}^n u_i v_i$ and then $\|u\|_2 = \sqrt{u^T u} = \sqrt{\sum_{i=1}^n u_i^2}$.
- Sequences of infinite length, ℓ_2 space. A dot product and a norm are defined similarly to the case of finite length.
- $\mathcal{L}_2$ space: A dot product of functions from $\mathbb{R}^n$ to $\mathbb{R}$ is defined by $\langle f, g \rangle = \int_{-\infty}^{\infty} f(x) g(x) dx$ and then $\|f\|_{\mathcal{L}_2} = \left(\int_{-\infty}^{\infty} |f(x)|^2 dx \right)^{\frac{1}{2}}$.

Note: ℓ_p and $\mathcal{L}_p$ spaces are only Banach Spaces for $p \neq 2$.

Dirac evaluation functional

Definition (Linear Operator)

Consider a function $\Phi : \mathbb{F} \to \mathbb{G}$, with $\mathbb{F}$ and $\mathbb{G}$ vector spaces over $\mathbb{R}$.
Φ is a **linear operator** if

$$\Phi(\alpha_1 f_1 + \alpha_2 f_2) = \alpha_1 \Phi(f1) + \alpha_2 \Phi(f_2), \quad \forall \alpha_1, \alpha_2 \in \mathbb{R}, \forall f_1, f_2 \in \mathbb{F}.$$

Operators with $\mathbb{G} = \mathbb{R}$ are called **functionals.**

Definition (Operator Norm)

The **operator norm** of a linear operator $\Phi : \mathbb{F} \to \mathbb{G}$ is defined as

$$\|\Phi\| = \sup_{f \in \mathbb{F}} \frac{\|\Phi f\|_{\mathbb{G}}}{\|f\|_{\mathbb{F}}}$$

Φ is called a **bounded linear operator** if $\|\Phi\| < \infty$ or equivalently if there is a constant $M > 0, M \in \mathbb{R}$ s.t. $\|\Phi f\|_{\mathbb{G}} \leq M \|f\|_{\mathbb{F}} \quad \forall f \in \mathbb{F}$.

Definition (Dirac Evaluation Functional)

Let $\mathscr{H}$ be the $\mathcal{L}_2(X)$ space of functions from X to $\mathbb{R}$ for some non empty set X. For an element $x \in X$, a **Dirac evaluation functional** at x is a functional δ_x with

$$\begin{aligned} \delta_x \quad : \quad & \mathscr{H} \to \mathbb{R} \qquad\qquad s.t \;\; \delta_x(f) = f(x) \\ & f \mapsto f(x) \end{aligned}$$

In this case, let X be $\mathbb{R}^n$ for some n. δ_x is bounded if there is a constant $M > 0, M \in \mathbb{R}$ s.t.

$$\|\delta_x f\| = |f(x)| \leq M\|f\|_{\mathcal{L}_2} \;\; \forall f \in \mathscr{H} \text{ and } \forall x \in \mathbb{R}^n.$$

Reproducing Kernel Hilbert Spaces

Definition (Reproducing Kernel Hilbert Spaces)

A **Reproducing Kernel Hilbert Space (RKHS)** is an Hilbert Space $\mathscr{H}$ where all the Dirac evaluation functionals in $\mathscr{H}$ are bounded and continuous.

Theorem 1

Let $(\mathbb{F}, \|\cdot\|_{\mathbb{F}})$, $(\mathbb{G}, \|\cdot\|_{\mathbb{G}})$ *be normed vector spaces. If* Φ *is a linear operator from* $\mathbb{F}$ *to* $\mathbb{G}$*, then the following conditions are equivalent:*

- Φ *is a bounded operator*
- Φ *is continuous on* $\mathbb{F}$
- Φ *is continuous at one point of* $\mathbb{F}$

Riesz representation theorem

Theorem 2 (Riesz representation theorem)

If Φ is a bounded linear functional on an Hilbert Space $\mathscr{H}$, then there is a unique u in $\mathscr{H}$ such that

$$\Phi(f) = \langle f, u \rangle_{\mathscr{H}} \quad \forall f \in \mathscr{H}$$

In the case of Dirac evaluation functionals, we get that: for each δ_x there exists a unique $k_x \in \mathscr{H}$ such that

$$\delta_x f = f(x) = \langle f, k_x \rangle_{\mathscr{H}}$$

Reproducing Kernel

Definition (Reproducing Kernel)

Let $\mathscr{H}$ be an Hilbert space of function from X to $\mathbb{R}$ ($\mathcal{L}_2(X)$) with X a non-empty set. A function $\mathcal{K} : X \times X \to \mathbb{R}$ is called a **Reproducing Kernel** of $\mathscr{H}$ if it satisfies:

- $\forall x \in X \quad k_x = \mathcal{K}(\cdot, x) \in \mathscr{H}$, k_x is called **representer at** x.
- $\forall x \in X,\ \forall f \in \mathscr{H} \quad \langle f, \mathcal{K}(\cdot, x)\rangle = f(x)$ (**Reproducing Property**)

Note: $k_x \in \mathscr{H}$ is a function from X to $\mathbb{R}$, s.t. $k_x(y) = \mathcal{K}(x, y)$.
In particular, for any $x, x' \in X$

$$\mathcal{K}(x, x') = \left\langle \mathcal{K}(\cdot, x), \mathcal{K}(\cdot, x')\right\rangle_{\mathscr{H}} = \left\langle k_x, k_{x'}\right\rangle_{\mathscr{H}}$$

where $k_x, k_{x'}$ are respectively the unique representatives of δ_x and $\delta_{x'}$.
$\implies \mathcal{K}$ is symmetric and positive definite.

From the first property of the previous definition we know that $\mathscr{H}$ contains all function of the form $f = \sum_{j=1}^{N} \alpha_j \mathcal{K}(\cdot, x_j)$ if $x_j \in X$ and we have

$$\|f\|_{\mathscr{H}}^2 = \sum_{j=1}^{N} \sum_{i=1}^{N} \alpha_j \alpha_i \langle \mathcal{K}(\cdot, x_j), \mathcal{K}(\cdot, x_i) \rangle_{\mathscr{H}} = \sum_{j=1}^{N} \sum_{i=1}^{N} \alpha_j \alpha_i \mathcal{K}(x_j, x_i)$$

Theorem 3 (Uniqueness)

If it exists, the reproducing kernel for an Hilbert space $\mathscr{H}$ is unique.

Proof: We assume that $\mathscr{H}$ has two reproducing kernels $\mathcal{K}_1$ and $\mathcal{K}_2$. Then

$$\langle f, \mathcal{K}_1(\cdot, x) - \mathcal{K}_2(\cdot, x) \rangle = f(x) - f(x) = 0 \quad \forall f \in \mathscr{H}, \quad \forall x \in X$$

In particular, if we take $f = \mathcal{K}_1(\cdot, x) - \mathcal{K}_2(\cdot, x)$ we obtain

$$\|\mathcal{K}_1(\cdot, x) - \mathcal{K}_2(\cdot, x)\|_{\mathscr{H}}^2 = 0 \quad \forall x \in X \quad \Longrightarrow \quad \mathcal{K}_1 = \mathcal{K}_2$$

Theorem 4 (Reproducing Kernel equivalent to bounded δ_x)

Let $\mathscr{H}$ be a Hilbert space of functions $f : X \to \mathbb{R}$. Then the evaluation operators δ_x are bounded and continuous functionals if and only if $\mathscr{H}$ has a reproducing kernel $\mathcal{K}$.

Proof:

($\Longleftarrow$) $\mathscr{H}$ is an Hilbert space with reproducing kernel $\mathcal{K}$, then

$$\begin{aligned} |\delta_x f| &= |f(x)| = |\langle f, \mathcal{K}(\cdot, x)\rangle_{\mathscr{H}}| \leq \|\mathcal{K}(\cdot, x)\|_{\mathscr{H}} \|f\|_{\mathscr{H}} \\ &= \langle \mathcal{K}(\cdot, x), \mathcal{K}(\cdot, x)\rangle_{\mathscr{H}}^{\frac{1}{2}} \|f\|_{\mathscr{H}} = \mathcal{K}(x, x)^{\frac{1}{2}} \|f\|_{\mathscr{H}} \end{aligned}$$

$\Longrightarrow$ $\delta_x : \mathscr{H} \to \mathbb{R}$ is a bounded linear operator.

($\Longrightarrow$) Assume that $\delta_x : \mathscr{H} \to \mathbb{R}$ is a bounded linear functional.
Riesz representation theorem $\Longrightarrow$ $\exists\, f_{\delta_x} \in \mathscr{H}$ s.t $\delta_x f = \langle f, f_{\delta_x}\rangle_{\mathscr{H}}$ $\forall f \in \mathscr{H}$
Define $\mathcal{K}(x', x) = f_{\delta_x}(x')$ $\forall x, x' \in X$. Then

$$\mathcal{K}(\cdot, x) = f_{\delta_x} \text{ and } \langle f, \mathcal{K}(\cdot, x)\rangle_{\mathscr{H}} = \delta_x f = f(x).$$

$\Longrightarrow$ $\mathcal{K}$ is a reproducing kernel for $\mathscr{H}$.

Theorem 5 (Properties of reproducing kernels)

Suppose $\mathscr{H}$ is a Hilbert space of functions $f : X \to \mathbb{R}$ with reproducing kernel $\mathcal{K}$. Then we have:

(i) $\mathcal{K}(x,y) = \langle \mathcal{K}(\cdot,x), \mathcal{K}(\cdot,y)\rangle_{\mathscr{H}}$ *for* x, y *in* X.

(ii) $\mathcal{K}(x,y) = \mathcal{K}(y,x)$ *for* x, y *in* X.

(iii) *if* $f, f_n \in \mathscr{H}$, $n \in \mathbb{N}$ *are given such that* f_n *converges to* f *in the Hilbert space norm then* f_n *also converges pointwise to* f.

Proof: (i) and (ii) are already been demonstrated
(iii) Given $f, f_n \in \mathscr{H}$ s.t. $\|f - f_n\|_{\mathscr{H}} \xrightarrow{n\to\infty} 0$, we have:

$$|f_n(x) - f(x)| = |\langle f_n - f, \mathcal{K}(\cdot,x)\rangle_{\mathscr{H}}| \leq \|f - f_n\|_{\mathscr{H}} \|\mathcal{K}(\cdot,x)\|_{\mathscr{H}} \xrightarrow{n\to\infty} 0$$

Kernels

Definition (Kernel)

A function $\mathcal{K} : X \times X \to \mathbb{R}$ is called a **kernel** on a non-empty set X if there exists a Hilbert Space (not necessary a RKHS) $\mathscr{H}$ and a map $\Phi : X \to \mathscr{H}$, such that

$$\mathcal{K}(x, y) = \langle \Phi(x), \Phi(y) \rangle_{\mathscr{H}}$$

Φ is called **feature map**, $\mathscr{H}$ is called **feature space.**

Proposition

Every reproducing kernel is a kernel.

This is easy to prove: we take as Φ the function that $x \mapsto \mathcal{K}(\cdot, x) = k_x$, $k_x \in \mathscr{H}$ and so $\mathcal{K}(x, y) = \langle \mathcal{K}(\cdot, x), \mathcal{K}(\cdot, y) \rangle_{\mathscr{H}}$. Then, the RKHS $\mathscr{H}$ is the feature space.

Example of a kernel, that is not a reproducing kernel

Consider $X = \mathbb{R}^2$ and $\mathcal{K}(x, y) = \langle x, y\rangle^2$. We can write $\mathcal{K}(x, y)$ as:

$$\mathcal{K}(x,y) = x_1^2 y_1^2 + x_2^2 y_2^2 + 2x_1x_2y_1y_2 = \begin{pmatrix} x_1^2 & x_2^2 & \sqrt{2}x_1x_2 \end{pmatrix} \begin{pmatrix} y_1^2 \\ y_2^2 \\ \sqrt{2}y_1y_2 \end{pmatrix}$$

$$= \begin{pmatrix} x_1^2 & x_2^2 & x_1x_2 & x_1x_2 \end{pmatrix} \begin{pmatrix} y_1^2 \\ y_2^2 \\ y_1y_2 \\ y_1y_2 \end{pmatrix}$$

Feature maps: $\phi(x) = (x_1^2, x_2^2, \sqrt{2}x_1x_2)$ or $\tilde{\phi}(x) = (x_1^2, x_2^2, x_1x_2, x_1x_2)$ with feature spaces $\mathscr{H} = \mathbb{R}^3$ or $\tilde{\mathscr{H}} = \mathbb{R}^4$. They are not RHKSs since they are not unique! Furthermore, also the feature map is not unique, only the kernel is.

Positive Definiteness

Definition (Positive definite Matrix)

A real symmetric $N \times N$ matrix K is called **positive semi-definite** if its associated quadratic form is non-negative for any coefficient vector $\mathbf{c} = [c_1, \dots, c_N]^T \in \mathbb{R}^N$ i.e.

$$\sum_{i=1}^{N} \sum_{j=1}^{N} c_i c_j K_{ij} \geq 0$$

and it is **positive definite** if the quadratic form is zero only for $c \equiv 0$.

A positive definite kernel $\mathcal{K}$ can be viewed as an infinite dimensional positive definite matrix K. We can give two different notions of positive definiteness for kernels, that are equivalent for continuous kernels.

Definition (Positive definite kernel-1)

A symmetric kernel $\mathcal{K} : X \times X \to \mathbb{R}$ is called **positive (semi-)definite** on X if its associated kernel matrix $(K)_{i,j} = \mathcal{K}(x_i, x_j)_{i,j=1}^{N}$ is positive (semi-)definite for any $N \in \mathbb{N}$ and for any set of distinct points $\{x_1, \ldots, x_N\} \subset X$.

This definition can be generalized using complex coefficients and kernels.

Definition (Positive definite kernel-2)

A symmetric kernel $\mathcal{K} : X \times X \to \mathbb{R}$ is **(integrally) positive semi-definite** if for any $\mathcal{L}_2$ function f we have that:

$$\int_X \int_X f(x)\mathcal{K}(x, x')f(x')dxdx' \geq 0$$

and it is **positive definite** if that integral is zero only for f equal to the zero function.

Theorem 6

Every kernel is a positive semi-definite function.

Proof: For pairwise distinct $x_1, \dots, x_N$ and $\mathbf{c} \in \mathbb{R}^N$ we have

$$\begin{aligned}
\sum_{i=1}^N \sum_{j=1}^N c_i c_j \mathcal{K}(x_i, x_j) &= \sum_{i=1}^N \sum_{j=1}^N c_i c_j \left\langle \phi(x_i), \phi(x_j) \right\rangle_{\mathscr{H}} \\
&= \left\langle \sum_{i=1}^N c_i \phi(x_i), \sum_{j=1}^N c_j \phi(x_j) \right\rangle_{\mathscr{H}} \\
&= \|\sum_{i=1}^N c_i \phi(x_i)\|_{\mathscr{H}}^2 \geq 0
\end{aligned}$$

reproducing kernel $\Longrightarrow$ kernel $\Longrightarrow$ positive semi-definite function

Moore Aronszajn theorem- Creating of RKHS

Theorem 7 (Moore-Aronszajn)

Let $\mathcal{K} : X \times X \to \mathbb{R}$ be a positive definite kernel. There is a unique (up to isomorphism) RKHS $\mathscr{H}_{\mathcal{K}} \in \mathbb{R}^X$ with reproducing kernel $\mathcal{K}$.

Steps of this construction:

1. define a pre-Hilbert space $\mathscr{H}_0$
2. define a dot product on $\mathscr{H}_0$
3. construct $\mathscr{H}_{\mathcal{K}}$ as a completion of $\mathscr{H}_0$
4. define a dot product on $\mathscr{H}_{\mathcal{K}}$

By hypothesis $\mathcal{K}$ is a positive definite kernel. We define a pre-Hilbert space $\mathscr{H}_0$ as follows:

- first we take $S = \{k_x := \mathcal{K}(\cdot, x) : x \in X\}$, where k_x is the function such that $k_x(y) = \mathcal{K}(x, y)$
- we define the $\mathbb{R}$-linear space $\mathscr{H}_0 = span\{k_x : x \in X\}$, set of all linear combinations of elements from S. Each element of $\mathscr{H}_0$ can be written as:

$$\sum_i \alpha_i k_{x_i} = \sum_i \alpha_i \mathcal{K}(\cdot, x_i)$$

and we equip it with the bilinear form

$$\langle f_1, f_2 \rangle_{\mathcal{K}} = \left\langle \sum_{i=1}^{N} \alpha_i \mathcal{K}(\cdot, x_i), \sum_{j=1}^{M} \beta_j \mathcal{K}(\cdot, y_j) \right\rangle_{\mathcal{K}} = \sum_{i=1}^{N} \sum_{j=1}^{M} \alpha_i \beta_j \mathcal{K}(x_i, y_j)$$

Theorem 8

If $\mathcal{K} : X \times X \to \mathbb{R}$ is a symmetric positive definite kernel then $\langle \cdot, \cdot \rangle_{\mathcal{K}}$ defines an inner product on $\mathscr{H}_0$. Furthermore $\mathscr{H}_0$ is a pre-Hilbert space with reproducing kernel $\mathcal{K}$.

Proof: Obviously $\langle \cdot, \cdot \rangle_{\mathcal{K}}$ is bilinear and symmetric. Moreover, if we consider an arbitrary function $f = \sum_{j=1}^{N} \alpha_j \mathcal{K}(\cdot, x_j) \neq 0$ from $\mathscr{H}_0$, we find that

$$\langle f, f \rangle_{\mathcal{K}} = \sum_{j=1}^{N} \sum_{k=1}^{N} \alpha_j \alpha_k \mathcal{K}(x_j, x_k) > 0$$

because $\mathcal{K}$ is positive definite. Finally, we obtain for this f

$$\langle f, \mathcal{K}(\cdot, y) \rangle_{\mathcal{K}} = \sum_{j=1}^{N} \alpha_j \mathcal{K}(x_j, y) = f(y)$$

$\Longrightarrow \quad \mathscr{H}_0$ pre-Hilbert space with reproducing kernel $\mathcal{K}$.

Note: $\mathscr{H}_0$ is NOT necessarily complete!
We can force it to be complete by taking all Cauchy sequences over $\mathscr{H}_0$ and adding their limits. Then, $\mathscr{H}_\mathcal{K}$ is the set of functions $f \in \mathbb{R}^X$ for which there exists an $\mathscr{H}_0$-Cauchy sequence $\{f_n\}$ converging pointwise to f, i.e. consider a Cauchy sequence $\{f_n\}$, fix a point $x \in X$ and evaluate

$$\begin{aligned} |f_n(x) - f_m(x)| &= |\langle \mathcal{K}(x,\cdot), f_n - f_m \rangle| \leq \|\mathcal{K}(x,\cdot)\|_{\mathscr{H}_0} \|f_n - f_m\|_{\mathscr{H}_0} \\ &= \langle \mathcal{K}(x,\cdot), \mathcal{K}(x,\cdot) \rangle_{\mathscr{H}_0}^{\frac{1}{2}} \|f_n - f_m\|_{\mathscr{H}_0} \\ &= \mathcal{K}(x,x)^{\frac{1}{2}} \|f_n - f_m\|_{\mathscr{H}_0} \end{aligned}$$

Then $\{f_n(x)\}$ is a bounded Cauchy sequence in $\mathbb{R}$, which is complete; thus there exists $f(x) = \lim f_n(x)$. Add all such f's to $\mathscr{H}_0$ to obtain $\mathscr{H}_\mathcal{K}$.
Then $\forall x \in X$ we have that

$$f(x) = \lim_{n\to\infty} f_n(x) = \lim_{n\to\infty} \langle f_n, \mathcal{K}(\cdot,x) \rangle_{\mathscr{H}_0} = \langle f, \mathcal{K}(\cdot,x) \rangle_{\mathscr{H}_\mathcal{K}}$$

thus $\mathcal{K}$ is the reproducing kernel of $\mathscr{H}_\mathcal{K}$.

Another characterization of $\mathscr{H}_{\mathcal{K}}$ and of its dot product is given in terms of eigenfunctions of a linear operator associated with the kernel.

$$T_{\mathcal{K}}(q)(x) = \int_X \mathcal{K}(x,y)q(y)dy \quad q \in \mathcal{L}_2(X), \quad x \in X$$

In order to do this, we need to ensure that $\mathcal{K}$, a positive definite kernel, is continuous and doesn't diverge,i.e.

$$\int\int \mathcal{K}^2(x,x')dxdx' < \infty \implies \mathcal{K}(\cdot,x) \in \mathcal{L}_2(X)$$

This property is known as **finite trace.** If $\mathcal{K}$ does not have a finite trace, we can restrict ourselves to a specific subset of the space X and ensure that $\mathcal{K}$ has a finite trace on that subspace.
We, then, need to introduce the concept of eigenfunction, the functions-space equivalent to an eigenvector.

Definition (Eigenfunction)

An **eigenfunction** of a linear operator D defined on some function space is any non-zero function f in that space that, when acted upon by D, is only multiplied by some scaling factor called an eigenvalue:

$$Df = \lambda f \quad \text{for some scalar eigenvalue } \lambda$$

The definition of eigenvectors is the same except for the fact D is a linear operator defined on a vector space and not on a function space. We can, then, associate to this linear operator a matrix M and obtain that an eigenvector of a matrix M is a vector v s.t.

$$Mv = \gamma v \text{ for some scalar } \gamma \text{, called eigenvalue.}$$

Suppose $\mathcal{K}$ is a kernel, then ϕ is an eigenfunction of $T_{\mathcal{K}}$ if:

$$\int \mathcal{K}(x, x')\phi(x')dx' = \lambda\phi(x') \quad \forall x \in X$$

In dot product notation this corresponds to: $\langle \mathcal{K}(x, \cdot), \phi \rangle_{\mathcal{L}_2} = \lambda\phi$.

Theorem 9 (Mercer-Hilbert-Schmit)

If $\mathcal{K}$ is a positive definite kernel (that is continuous with finite trace), then there exists an infinite sequence of eigenfunctions $\langle \phi_i \rangle_{i=0}^{\infty}$ and eigenvalues λ_i of $T_{\mathcal{K}}$, with $\lambda_1 \geq \lambda_2 \geq \dots$, and $\mathcal{K}$ can be written as:

$$\mathcal{K}(x, x') = \sum_{i=0}^{\infty} \lambda_i \phi_i(x)\phi_i(x')$$

This allow as to construct a RKHS $\mathscr{H}_{\mathcal{K}}$ in that way:

$$\mathscr{H}_{\mathcal{K}} := \left\{ f : f = \sum_{i=1}^{\infty} c_i \phi_i \right\}$$

For $f \in \mathcal{L}_2$, we will denote f in terms of its coefficients in the eigenfunctions:

$$f_i = \langle f, \phi_i \rangle_{\mathcal{L}_2} = \int f(x)\phi_i(x)dx$$

It is a basic result of Fourier analysis that such representation exists and is unique. Given all this, we can now define the inner product on $\mathscr{H}_\mathcal{K}$:

$$\langle f, g \rangle_{\mathscr{H}_\mathcal{K}} = \sum_{i=0}^{\infty} \frac{f_i g_i}{\lambda_i}$$

where we used the $\mathscr{H}_\mathcal{K}$-orthogonality $\langle \phi_j, \phi_k \rangle_{\mathscr{H}_\mathcal{K}} = \frac{\delta_{j,k}}{\sqrt{\lambda_j}\sqrt{\lambda_k}}$ of the eigenfunctions.

Note: $\mathcal{K}$ is the reproducing kernel of $\mathscr{H}_\mathcal{K}$ since the eigenfunction expansion of $\mathcal{K}$, given by theorem 9, and the orthogonality of the eigenfunctions imply

$$\begin{aligned}
\langle f, \mathcal{K}(\cdot, x) \rangle_{\mathscr{H}_\mathcal{K}} &= \left\langle \textstyle\sum_{j=1}^{\infty} c_j\phi_j, \sum_{i=1}^{\infty} \lambda_i\phi_i\phi_i(x) \right\rangle_{\mathscr{H}_\mathcal{K}} \\
&= \textstyle\sum_{i=1}^{\infty} \frac{c_i\lambda_i\phi_i(x)}{\lambda_i} = \sum_{i=1}^{\infty} c_i\phi_i(x) = f(x)
\end{aligned}$$

Feature Maps and Feature Spaces

Given a kernel $\mathcal{K}$ there is a feature map associated $\Phi : X \to \mathscr{H}$ s.t.

$$\mathcal{K}(x, x') = \left\langle \Phi(x), \Phi(x') \right\rangle_{\mathscr{H}}$$

i.e. given a kernel $\mathcal{K}$ there exists a function Φ s.t. the evaluation of the kernel at points x and x' is equivalent to taking the dot product between $\Phi(x)$ and $\Phi(x')$ in some (perhaps unknown) Hilbert space.
This enables us to perform the **kernel trick**, in which dot products are replaced by kernel products (i.e. evaluation of kernels), where we transform the inputs into $\mathscr{H}$ using Φ and then we take the dot product as before.
We have seen how to construct a RKHS $\mathscr{H}$ starting from a positive definite kernel and that $\mathscr{H}$ is unique up to isomorphism. This means that Φ is not absolutely unique, but it is as unique as $\mathscr{H}$ is. We now show the two most common construction of Φ (which are more or less equivalent) starting from the RKHS $\mathscr{H}$ constructed.

Note: Φ is injective, since in the construction made we require $\mathcal{K}$ and then the matrix associated to be positive definite.

1. First definition of Φ: $\Phi(x) := \mathcal{K}(x, \cdot)$. $\mathscr{H}$ is the feature space . By the reproducing property of the reproducing kernel we get:

$$\left\langle \Phi(x), \Phi(x') \right\rangle_{\mathscr{H}} = \left\langle \mathcal{K}(x, \cdot), \mathcal{K}(x', \cdot) \right\rangle_{\mathscr{H}} = \mathcal{K}(x, x')$$

   which satisfies our requirements for Φ.

2. Second definition of Φ: we ignore our constructed $\mathscr{H}$ and use ℓ_2 as the feature space. This construction uses the eigenfunctions ϕ_i and eigenvalues λ_i of $T_{\mathcal{K}}$ and defines Φ by:

$$\Phi(x) := \left\langle \sqrt{\lambda_i}\phi_i(x) \right\rangle_{i=0}^{\infty}$$

   We calculate the dot product by:

$$\langle \Phi(x), \Phi(x') \rangle_{\ell_2} = \left\langle \left\langle \sqrt{\lambda_i}\phi_i(x) \right\rangle_i, \left\langle \sqrt{\lambda_j}\phi_j(x') \right\rangle_j \right\rangle_{\ell_2} =$$

$$= \sum_{i=0}^{\infty} \sqrt{\lambda_i}\phi_i(x)\sqrt{\lambda_i}\phi_i(x') = \sum_{i=0}^{\infty} \lambda_i\phi_i(x)\phi_i(x') = \mathcal{K}(x, x')$$

THANK YOU FOR YOUR ATTENTION