Foundations of Reproducing Kernel Hilbert Spaces

Advanced Topics in Machine Learning

D. Sejdinovic, A. Gretton

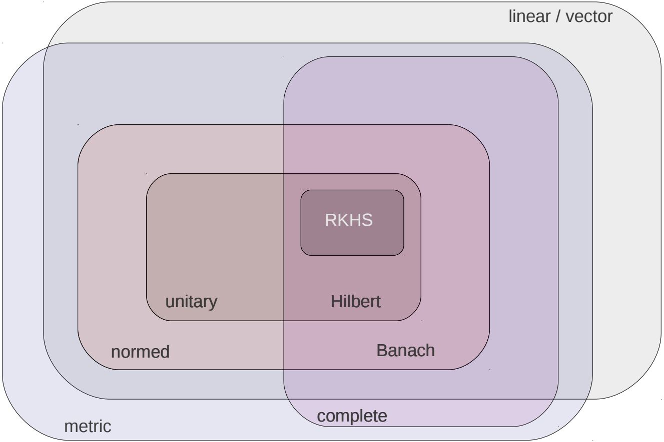
Gatsby Unit
slides and notes are available at www.gatsby.ucl.ac.uk/~dino/teaching

March 11, 2014

Overview

1. Elementary Hilbert space theory
   - Norm. Inner product. Orthogonality
   - Convergence. Complete spaces
   - Linear operators. Riesz representation
2. What is an RKHS?
   - Evaluation functionals view of RKHS
   - Reproducing kernel
   - Inner product between features
   - Positive definite function
   - Moore-Aronszajn Theorem

RKHS: a function space with a very special structure



Outline

1. Elementary Hilbert space theory
   - Norm. Inner product. Orthogonality
   - Convergence. Complete spaces
   - Linear operators. Riesz representation
2. What is an RKHS?
   - Evaluation functionals view of RKHS
   - Reproducing kernel
   - Inner product between features
   - Positive definite function
   - Moore-Aronszajn Theorem

Normed vector space

Definition (Norm)

Let $\mathcal{F}$ be a vector space over the field $\mathbb{R}$ of real numbers (or $\mathbb{C}$). A function $\|\cdot\|_{\mathcal{F}} : \mathcal{F} \to [0, \infty)$ is said to be *a norm* on $\mathcal{F}$ if

1. $\|f\|_{\mathcal{F}} = 0$ if and only if $f = \mathbf{0}$ (*norm separates points*),
2. $\|\lambda f\|_{\mathcal{F}} = |\lambda| \; \|f\|_{\mathcal{F}}$, $\forall \lambda \in \mathbb{R}$, $\forall f \in \mathcal{F}$ (*positive homogeneity*),
3. $\|f + g\|_{\mathcal{F}} \leq \|f\|_{\mathcal{F}} + \|g\|_{\mathcal{F}}$, $\forall f, g \in \mathcal{F}$ (*triangle inequality*).

Normed vector space

Definition (Norm)

Let $\mathcal{F}$ be a vector space over the field $\mathbb{R}$ of real numbers (or $\mathbb{C}$). A function $\|\cdot\|_{\mathcal{F}} : \mathcal{F} \to [0, \infty)$ is said to be *a norm* on $\mathcal{F}$ if

1. $\|f\|_{\mathcal{F}} = 0$ if and only if $f = \mathbf{0}$ (*norm separates points*),
2. $\|\lambda f\|_{\mathcal{F}} = |\lambda| \, \|f\|_{\mathcal{F}}$, $\forall \lambda \in \mathbb{R}$, $\forall f \in \mathcal{F}$ (*positive homogeneity*),
3. $\|f + g\|_{\mathcal{F}} \leq \|f\|_{\mathcal{F}} + \|g\|_{\mathcal{F}}$, $\forall f, g \in \mathcal{F}$ (*triangle inequality*).

In every normed vector space, one can define *a metric* induced by the norm:

$$d(f, g) = \|f - g\|_{\mathcal{F}}.$$

Examples of normed linear spaces

- $(\mathbb{R}, |\cdot|)$, $(\mathbb{C}, |\cdot|)$

Examples of normed linear spaces

- $(\mathbb{R}, |\cdot|)$, $(\mathbb{C}, |\cdot|)$
- $\mathcal{F} = \mathbb{R}^d$: $\|\mathbf{x}\|_p = \left(\sum_{i=1}^d |x_i|^p\right)^{1/p}$, $p \geq 1$ (only quasi-norm for $0 < p < 1$)

Examples of normed linear spaces

- $(\mathbb{R}, |\cdot|)$, $(\mathbb{C}, |\cdot|)$
- $\mathcal{F} = \mathbb{R}^d$: $\|\mathbf{x}\|_p = \left(\sum_{i=1}^d |x_i|^p\right)^{1/p}$, $p \geq 1$ (only quasi-norm for $0 < p < 1$)
  - $p = 1$: Manhattan
  - $p = 2$: Euclidean
  - $p \to \infty$: maximum norm, $\|\mathbf{x}\|_\infty = \max_i |x_i|$

Examples of normed linear spaces

- $(\mathbb{R}, |\cdot|)$, $(\mathbb{C}, |\cdot|)$
- $\mathcal{F} = \mathbb{R}^d$: $\|\mathbf{x}\|_p = \left(\sum_{i=1}^d |x_i|^p\right)^{1/p}$, $p \geq 1$
  (only quasi-norm for $0 < p < 1$)
  - $p = 1$: Manhattan
  - $p = 2$: Euclidean
  - $p \to \infty$: maximum norm, $\|\mathbf{x}\|_\infty = \max_i |x_i|$
- $\mathcal{F} = C[a, b]$: $\|f\|_p = \left(\int_a^b |f(x)|^p \, dx\right)^{1/p}$, $p \geq 1$

Inner product

Definition (Inner product)

Let $\mathcal{F}$ be a vector space over $\mathbb{R}$. A function $\langle \cdot, \cdot \rangle_{\mathcal{F}} : \mathcal{F} \times \mathcal{F} \to \mathbb{R}$ is said to be *an inner product* on $\mathcal{F}$ if

1. $\langle \alpha_1 f_1 + \alpha_2 f_2, g \rangle_{\mathcal{F}} = \alpha_1 \langle f_1, g \rangle_{\mathcal{F}} + \alpha_2 \langle f_2, g \rangle_{\mathcal{F}}$
2. $\langle f, g \rangle_{\mathcal{F}} = \langle g, f \rangle_{\mathcal{F}}$ (**conjugate symmetry** if over $\mathbb{C}$)
3. $\langle f, f \rangle_{\mathcal{F}} \geq 0$ and $\langle f, f \rangle_{\mathcal{F}} = 0$ if and only if $f = 0$.

Inner product

Definition (Inner product)

Let $\mathcal{F}$ be a vector space over $\mathbb{R}$. A function $\langle \cdot, \cdot \rangle_{\mathcal{F}} : \mathcal{F} \times \mathcal{F} \to \mathbb{R}$ is said to be *an inner product* on $\mathcal{F}$ if

1. $\langle \alpha_1 f_1 + \alpha_2 f_2, g \rangle_{\mathcal{F}} = \alpha_1 \langle f_1, g \rangle_{\mathcal{F}} + \alpha_2 \langle f_2, g \rangle_{\mathcal{F}}$
2. $\langle f, g \rangle_{\mathcal{F}} = \langle g, f \rangle_{\mathcal{F}}$ (**conjugate symmetry** if over $\mathbb{C}$)
3. $\langle f, f \rangle_{\mathcal{F}} \geq 0$ and $\langle f, f \rangle_{\mathcal{F}} = 0$ if and only if $f = 0$.

In every inner product vector space, one can define *a norm* induced by the inner product:

$$\|f\|_{\mathcal{F}} = \langle f, f \rangle_{\mathcal{F}}^{1/2}.$$

Examples of inner product

- $\mathcal{F} = \mathbb{R}^d$: $\langle \mathbf{x}, \mathbf{y} \rangle = \sum_{i=1}^{d} x_i y_i$

Examples of inner product

- $\mathcal{F} = \mathbb{R}^d$: $\langle \mathbf{x}, \mathbf{y} \rangle = \sum_{i=1}^d x_i y_i$
- $\mathcal{F} = C[a, b]$: $\langle f, g \rangle = \int_a^b f(x) g(x) dx$

Examples of inner product

- $\mathcal{F} = \mathbb{R}^d$: $\langle \mathbf{x}, \mathbf{y} \rangle = \sum_{i=1}^{d} x_i y_i$
- $\mathcal{F} = C[a,b]$: $\langle f, g \rangle = \int_a^b f(x)g(x)dx$
- $\mathcal{F} = \mathbb{R}^{d \times d}$: $\langle \mathbf{A}, \mathbf{B} \rangle = Tr\left(\mathbf{A}\mathbf{B}^\top\right)$

Examples of inner product

- $\mathcal{F} = \mathbb{R}^d$: $\langle \mathbf{x}, \mathbf{y} \rangle = \sum_{i=1}^{d} x_i y_i$
- $\mathcal{F} = C[a, b]$: $\langle f, g \rangle = \int_a^b f(x) g(x) dx$
- $\mathcal{F} = \mathbb{R}^{d \times d}$: $\langle \mathbf{A}, \mathbf{B} \rangle = Tr\left(\mathbf{A}\mathbf{B}^\top\right)$
- $\mathcal{F}$-set of random variables: $\langle X, Y \rangle = \mathbb{E}[XY]$.

Angles. Orthogonality

Angle θ between $f, g \in \mathcal{F} \setminus \{0\}$ is given by:

$$\cos\theta = \frac{\langle f, g\rangle_{\mathcal{F}}}{\|f\|_{\mathcal{F}}\,\|g\|_{\mathcal{F}}}$$

Angles. Orthogonality

Angle θ between $f, g \in \mathcal{F} \setminus \{0\}$ is given by:

$$\cos\theta = \frac{\langle f, g\rangle_{\mathcal{F}}}{\|f\|_{\mathcal{F}} \|g\|_{\mathcal{F}}}$$

Definition

We say that f is orthogonal to g and write $f \perp g$, if $\langle f, g\rangle_{\mathcal{F}} = 0$. For $M \subset \mathcal{F}$, the orthogonal complement of M is:

$$M^{\perp} := \{g \in \mathcal{F} : f \perp g, \forall f \in M\}.$$

Angles. Orthogonality

Angle θ between $f, g \in \mathcal{F} \setminus \{0\}$ is given by:

$$\cos\theta = \frac{\langle f, g\rangle_{\mathcal{F}}}{\|f\|_{\mathcal{F}} \|g\|_{\mathcal{F}}}$$

Definition

We say that f is orthogonal to g and write $f \perp g$, if $\langle f, g\rangle_{\mathcal{F}} = 0$. For $M \subset \mathcal{F}$, the orthogonal complement of M is:

$$M^{\perp} := \{g \in \mathcal{F} : f \perp g, \forall f \in M\}.$$

- $M^{\perp}$ is a linear subspace of $\mathcal{F}$; $M \cap M^{\perp} = \{0\}$

Key relations in inner product space

- $|\langle f, g\rangle| \leq \|f\| \cdot \|g\|$ (*Cauchy-Schwarz inequality*)
- $2\|f\|^2 + 2\|g\|^2 = \|f+g\|^2 + \|f-g\|^2$ (*the parallelogram law*)
- $4\langle f, g\rangle = \|f+g\|^2 - \|f-g\|^2$ (*the polarization identity*)

Key relations in inner product space

- $|\langle f, g\rangle| \leq \|f\| \cdot \|g\|$ (*Cauchy-Schwarz inequality*)
- $2\|f\|^2 + 2\|g\|^2 = \|f+g\|^2 + \|f-g\|^2$ (*the parallelogram law*)
- $4\langle f, g\rangle = \|f+g\|^2 - \|f-g\|^2$ (*the polarization identity*)
- $f \perp g \implies \|f\|^2 + \|g\|^2 = \|f+g\|^2$ (*Pythagorean theorem*)

Outline

1. Elementary Hilbert space theory
   - Norm. Inner product. Orthogonality
   - Convergence. Complete spaces
   - Linear operators. Riesz representation
2. What is an RKHS?
   - Evaluation functionals view of RKHS
   - Reproducing kernel
   - Inner product between features
   - Positive definite function
   - Moore-Aronszajn Theorem

Cauchy sequence

Definition (Convergent sequence)

A sequence $\{f_n\}_{n=1}^{\infty}$ of elements of a normed vector space $(\mathcal{F}, \|\cdot\|_{\mathcal{F}})$ is said to *converge* to $f \in \mathcal{F}$ if for every $\epsilon > 0$, there exists $N = N(\varepsilon) \in \mathbb{N}$, such that for all $n \geq N$, $\|f_n - f\|_{\mathcal{F}} < \epsilon$.

Cauchy sequence

Definition (Convergent sequence)

A sequence $\{f_n\}_{n=1}^{\infty}$ of elements of a normed vector space $(\mathcal{F}, \|\cdot\|_{\mathcal{F}})$ is said to *converge* to $f \in \mathcal{F}$ if for every $\epsilon > 0$, there exists $N = N(\varepsilon) \in \mathbb{N}$, such that for all $n \geq N$, $\|f_n - f\|_{\mathcal{F}} < \epsilon$.

Definition (Cauchy sequence)

A sequence $\{f_n\}_{n=1}^{\infty}$ of elements of a normed vector space $(\mathcal{F}, \|\cdot\|_{\mathcal{F}})$ is said to be *a Cauchy (fundamental) sequence* if for every $\epsilon > 0$, there exists $N = N(\varepsilon) \in \mathbb{N}$, such that for all $n, m \geq N$, $\|f_n - f_m\|_{\mathcal{F}} < \epsilon$.

Cauchy sequence

Definition (Convergent sequence)

A sequence $\{f_n\}_{n=1}^{\infty}$ of elements of a normed vector space $(\mathcal{F}, \|\cdot\|_{\mathcal{F}})$ is said to *converge* to $f \in \mathcal{F}$ if for every $\epsilon > 0$, there exists $N = N(\varepsilon) \in \mathbb{N}$, such that for all $n \geq N$, $\|f_n - f\|_{\mathcal{F}} < \epsilon$.

Definition (Cauchy sequence)

A sequence $\{f_n\}_{n=1}^{\infty}$ of elements of a normed vector space $(\mathcal{F}, \|\cdot\|_{\mathcal{F}})$ is said to be *a Cauchy (fundamental) sequence* if for every $\epsilon > 0$, there exists $N = N(\varepsilon) \in \mathbb{N}$, such that for all $n, m \geq N$, $\|f_n - f_m\|_{\mathcal{F}} < \epsilon$.

From $\|f_n - f_m\|_{\mathcal{F}} \leq \|f_n - f\|_{\mathcal{F}} + \|f - f_m\|_{\mathcal{F}}$, convergent⇒Cauchy.

Cauchy sequence

Definition (Convergent sequence)

A sequence $\{f_n\}_{n=1}^{\infty}$ of elements of a normed vector space $(\mathcal{F}, \|\cdot\|_{\mathcal{F}})$ is said to *converge* to $f \in \mathcal{F}$ if for every $\epsilon > 0$, there exists $N = N(\varepsilon) \in \mathbb{N}$, such that for all $n \geq N$, $\|f_n - f\|_{\mathcal{F}} < \epsilon$.

Definition (Cauchy sequence)

A sequence $\{f_n\}_{n=1}^{\infty}$ of elements of a normed vector space $(\mathcal{F}, \|\cdot\|_{\mathcal{F}})$ is said to be *a Cauchy (fundamental) sequence* if for every $\epsilon > 0$, there exists $N = N(\varepsilon) \in \mathbb{N}$, such that for all $n, m \geq N$, $\|f_n - f_m\|_{\mathcal{F}} < \epsilon$.

From $\|f_n - f_m\|_{\mathcal{F}} \leq \|f_n - f\|_{\mathcal{F}} + \|f - f_m\|_{\mathcal{F}}$, convergent$\Rightarrow$Cauchy.

Cauchy$\not\Rightarrow$convergent

Examples

Example

1, 1.4, 1.41, 1.414, 1.4142, ... is a Cauchy sequence in $\mathbb{Q}$ which does not converge - because $\sqrt{2} \notin \mathbb{Q}$.

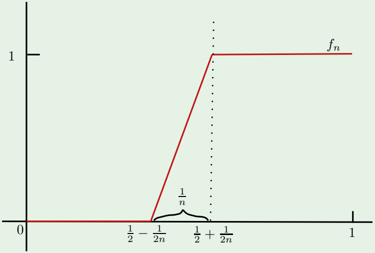
Examples

Example

1, 1.4, 1.41, 1.414, 1.4142, ... is a Cauchy sequence in $\mathbb{Q}$ which does not converge - because $\sqrt{2} \notin \mathbb{Q}$.

Example

$C[0,1]$ with the norm $\|f\|_2 = \left(\int_0^1 |f(x)|^2 \, dx\right)^{1/2}$, a sequence $\{f_n\}$ does not have a continuous limit!



Complete space

Definition (Complete space)

A metric space $\mathcal{F}$ is said to be *complete* if every Cauchy sequence $\{f_n\}_{n=1}^{\infty}$ in $\mathcal{F}$ converges: it has a limit, and this limit is in $\mathcal{F}$.

- i.e., one can find $f \in \mathcal{F}$, s.t. $\lim_{n\to\infty} \|f_n - f\|_{\mathcal{F}} = 0$.

Complete space

Definition (Complete space)

A metric space $\mathcal{F}$ is said to be *complete* if every Cauchy sequence $\{f_n\}_{n=1}^{\infty}$ in $\mathcal{F}$ converges: it has a limit, and this limit is in $\mathcal{F}$.

- i.e., one can find $f \in \mathcal{F}$, s.t. $\lim_{n\to\infty} \|f_n - f\|_{\mathcal{F}} = 0$.
- Complete + norm = Banach

Complete space

Definition (Complete space)

A metric space $\mathcal{F}$ is said to be *complete* if every Cauchy sequence $\{f_n\}_{n=1}^{\infty}$ in $\mathcal{F}$ converges: it has a limit, and this limit is in $\mathcal{F}$.

- i.e., one can find $f \in \mathcal{F}$, s.t. $\lim_{n\to\infty} \|f_n - f\|_{\mathcal{F}} = 0$.
- Complete + norm = Banach
- Complete + inner product = Hilbert

Examples of Hilbert spaces

Example

For an index set A, the space $\ell^2(A)$ of sequences $\{x_\alpha\}_{\alpha \in A}$ of real numbers, satisfying $\sum_{\alpha \in A} |x_\alpha|^2 < \infty$, endowed with the inner product

$$\langle \{x_\alpha\}, \{y_\alpha\} \rangle_{\ell^2(A)} = \sum_{\alpha \in A} x_\alpha y_\alpha$$

is a Hilbert space.

Examples of Hilbert spaces (2)

Example

If ν is a positive measure on $\mathcal{X} \subset \mathbb{R}^d$, then the space

$$L_2(\mathcal{X};\nu) := \left\{ f : \mathcal{X} \to \mathbb{R} \,\middle|\, \|f\|_2 = \left(\int_{\mathcal{X}} |f(x)|^2 d\nu(x) \right)^{1/2} < \infty \right\}$$

is a Hilbert space with inner product

$$\langle f, g \rangle_2 = \int_{\mathcal{X}} f(x) g(x) d\nu(x).$$

Examples of Hilbert spaces (2)

Example

If ν is a positive measure on $\mathcal{X} \subset \mathbb{R}^d$, then the space

$$L_2(\mathcal{X};\nu) := \left\{ f : \mathcal{X} \to \mathbb{R} \;\middle|\; \|f\|_2 = \left(\int_{\mathcal{X}} |f(x)|^2 d\nu(x) \right)^{1/2} < \infty \right\}$$

is a Hilbert space with inner product

$$\langle f, g \rangle_2 = \int_{\mathcal{X}} f(x) g(x) d\nu(x).$$

- Strictly speaking, $L_2(\mathcal{X};\nu)$ is the space of equivalence classes of functions that differ by at most a set of ν-measure zero.

Closed vs. Complete

- **Closed**: $M \subseteq \mathcal{F}$ is closed (in $\mathcal{F}$) if it contains limits of all sequences in M that converge in $\mathcal{F}$
- **Complete**: M is complete (with no reference to a larger space) if all Cauchy sequences in M converge in M

Closed vs. Complete

- **Closed**: $M \subseteq \mathcal{F}$ is closed (in $\mathcal{F}$) if it contains limits of all sequences in M that converge in $\mathcal{F}$
- **Complete**: M is complete (with no reference to a larger space) if all Cauchy sequences in M converge in M
- If M is a **closed subspace** of a Hilbert space $\mathcal{F}$, then:

$$M + M^{\perp} = \left\{ m + m^{\perp} \ : \ m \in M, m^{\perp} \in M^{\perp} \right\} = \mathcal{F}.$$

- In particular, for a closed subspace $M \subsetneq \mathcal{F}$, $M^{\perp}$ contains a non-zero element.

Non-closed subspaces

- Every finite-dimensional subspace of a normed space is **closed**.

Non-closed subspaces

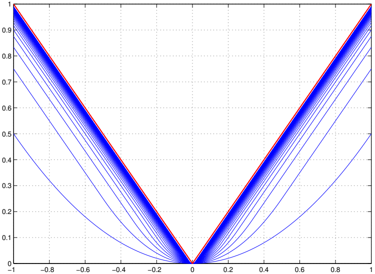
- Every finite-dimensional subspace of a normed space is **closed**.

Example

Let $\mathcal{F} = \{f : [-1,1] \to \mathbb{R}, f \text{ continuous}\}$, with $\|f\|_\infty = \sup |f(x)|$ (Banach space), and $\mathcal{F}^1$ its subspace of **differentiable functions**. *Then $\mathcal{F}^1$ is not closed.*

- Idea: construct a sequence of differentiable functions converging in $\|\cdot\|_\infty$ to $f(x) = |x|$:

$$f_n(x) = \begin{cases} -x - \frac{1}{2n}, & x \le -1/n, \\ \frac{n}{2}x^2, & |x| < 1/n, \\ x - \frac{1}{2n}, & x \ge 1/n. \end{cases}$$



Non-closed subspaces

Example

Let $\mathcal{H}$ be an infinite-dimensional Hilbert space with orthonormal basis $\mathcal{U} = \{u_j\}_{j=1}^{\infty}$. Then span$[\mathcal{U}]$ (finite linear combinations of elements of $\mathcal{U}$) is not closed.

- Take $h = \sum_{j=1}^{\infty} a_j u_j$ with $a_j > 0$ and $\sum_{j=1}^{\infty} a_j^2 < \infty$. Then $h_n = \sum_{j=1}^{n} a_j u_j$ converges to $h \notin span[\mathcal{U}]$.

Non-closed subspaces

Example

Let $\mathcal{H}$ be an infinite-dimensional Hilbert space with orthonormal basis $\mathcal{U} = \{u_j\}_{j=1}^{\infty}$. Then span[$\mathcal{U}$] (finite linear combinations of elements of $\mathcal{U}$) is not closed.

- Take $h = \sum_{j=1}^{\infty} a_j u_j$ with $a_j > 0$ and $\sum_{j=1}^{\infty} a_j^2 < \infty$. Then $h_n = \sum_{j=1}^{n} a_j u_j$ converges to $h \notin span[\mathcal{U}]$.

Recall:

- M closed subspace $\implies$ $M^{\perp}$contains a non-zero element.
- Here: $span[\mathcal{U}]^{\perp} = \{0\}$ (i.e., $span[\mathcal{U}]$ is **dense** in $\mathcal{H}$).

Outline

1. Elementary Hilbert space theory
   - Norm. Inner product. Orthogonality
   - Convergence. Complete spaces
   - Linear operators. Riesz representation
2. What is an RKHS?
   - Evaluation functionals view of RKHS
   - Reproducing kernel
   - Inner product between features
   - Positive definite function
   - Moore-Aronszajn Theorem

Linear operators

Definition (Linear operator)

Consider a function $A : \mathcal{F} \to \mathcal{G}$, where $\mathcal{F}$ and $\mathcal{G}$ are both vector spaces over $\mathbb{R}$. A is said to be a **linear operator** if

$$A(\alpha_1 f_1 + \alpha_2 f_2) = \alpha_1 (Af_1) + \alpha_2 (Af_2) \quad \forall \alpha_1, \alpha_2 \in \mathbb{R},\ f_1, f_2 \in \mathcal{F}.$$

Linear operators

Definition (Linear operator)

Consider a function $A : \mathcal{F} \to \mathcal{G}$, where $\mathcal{F}$ and $\mathcal{G}$ are both vector spaces over $\mathbb{R}$. A is said to be a **linear operator** if

$$A(\alpha_1 f_1 + \alpha_2 f_2) = \alpha_1 (Af_1) + \alpha_2 (Af_2) \quad \forall \alpha_1, \alpha_2 \in \mathbb{R},\ f_1, f_2 \in \mathcal{F}.$$

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

Linear operators

Definition (Linear operator)

Consider a function $A : \mathcal{F} \to \mathcal{G}$, where $\mathcal{F}$ and $\mathcal{G}$ are both vector spaces over $\mathbb{R}$. A is said to be a **linear operator** if

$$A(\alpha_1 f_1 + \alpha_2 f_2) = \alpha_1 (Af_1) + \alpha_2 (Af_2) \quad \forall \alpha_1, \alpha_2 \in \mathbb{R},\ f_1, f_2 \in \mathcal{F}.$$

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

Example

For $g \in \mathcal{F}$, $A_g : \mathcal{F} \to \mathbb{R}$, defined with $A_g f = \langle f, g \rangle_{\mathcal{F}}$ is a linear functional.

$$\begin{aligned} A_g(\alpha_1 f_1 + \alpha_2 f_2) &= \langle \alpha_1 f_1 + \alpha_2 f_2, g \rangle_{\mathcal{F}} \\ &= \alpha_1 \langle f_1, g \rangle_{\mathcal{F}} + \alpha_2 \langle f_2, g \rangle_{\mathcal{F}} \\ &= \alpha_1 A_g f_1 + \alpha_2 A_g f_2. \end{aligned}$$

Boundedness

Definition (Operator norm)

The operator norm of a linear operator $A : \mathcal{F} \to \mathcal{G}$ is defined as

$$\|A\| = \sup_{f \in \mathcal{F}, \|f\|_{\mathcal{F}} \leq 1} \|Af\|_{\mathcal{G}} .$$

If $\|A\| < \infty$, A is called a **bounded linear operator**.

Boundedness

Definition (Operator norm)

The operator norm of a linear operator $A : \mathcal{F} \to \mathcal{G}$ is defined as

$$\|A\| = \sup_{f \in \mathcal{F}, \|f\|_{\mathcal{F}} \leq 1} \|Af\|_{\mathcal{G}} .$$

If $\|A\| < \infty$, A is called a **bounded linear operator**.

$\|A\|$ is the smallest number λ such that the inequality $\|Af\|_{\mathcal{G}} \leq \lambda \|f\|_{\mathcal{F}}$ holds for *every* $f \in \mathcal{F}$.

Boundedness

Definition (Operator norm)

The operator norm of a linear operator $A : \mathcal{F} \to \mathcal{G}$ is defined as

$$\|A\| = \sup_{f \in \mathcal{F}, \|f\|_{\mathcal{F}} \leq 1} \|Af\|_{\mathcal{G}} .$$

If $\|A\| < \infty$, A is called a **bounded linear operator**.

$\|A\|$ is the smallest number λ such that the inequality $\|Af\|_{\mathcal{G}} \leq \lambda \|f\|_{\mathcal{F}}$ holds for *every* $f \in \mathcal{F}$.

bounded operator $\neq$ bounded function

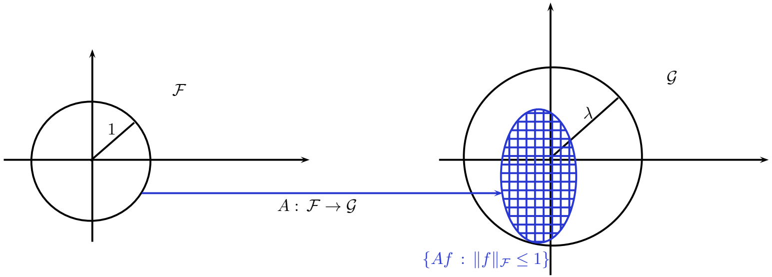
Boundedness

Definition (Operator norm)

The operator norm of a linear operator $A : \mathcal{F} \to \mathcal{G}$ is defined as

$$\|A\| = \sup_{f \in \mathcal{F}, \|f\|_{\mathcal{F}} \leq 1} \|Af\|_{\mathcal{G}} .$$

If $\|A\| < \infty$, A is called a **bounded linear operator**.



Continuity

Definition (Continuity)

Consider a function $A : \mathcal{F} \to \mathcal{G}$, where $\mathcal{F}$ and $\mathcal{G}$ are both normed vector spaces over $\mathbb{R}$. A is said to be **continuous** at $f_0 \in \mathcal{F}$, if for every $\epsilon > 0$, there exists a $\delta = \delta(\epsilon, f_0) > 0$, s.t.

$$\|f - f_0\|_{\mathcal{F}} < \delta \qquad \implies \qquad \|Af - Af_0\|_{\mathcal{G}} < \epsilon.$$

A is said to be **continuous** on $\mathcal{F}$, if it is continuous at every point of $\mathcal{F}$.

Continuity

Definition (Continuity)

Consider a function $A : \mathcal{F} \to \mathcal{G}$, where $\mathcal{F}$ and $\mathcal{G}$ are both normed vector spaces over $\mathbb{R}$. A is said to be **continuous** at $f_0 \in \mathcal{F}$, if for every $\epsilon > 0$, there exists a $\delta = \delta(\epsilon, f_0) > 0$, s.t.

$$\|f - f_0\|_{\mathcal{F}} < \delta \qquad \implies \qquad \|Af - Af_0\|_{\mathcal{G}} < \epsilon.$$

A is said to be **continuous** on $\mathcal{F}$, if it is continuous at every point of $\mathcal{F}$.

Example

For $g \in \mathcal{F}$, $A_g : \mathcal{F} \to \mathbb{R}$, defined with $A_g(f) := \langle f, g \rangle_{\mathcal{F}}$ is continuous on $\mathcal{F}$.

$$|A_g f_1 - A_g f_2| = |\langle f_1 - f_2, g \rangle_{\mathcal{F}}| \leq \|g\|_{\mathcal{F}} \|f_1 - f_2\|_{\mathcal{F}},$$

so can take $\delta = \varepsilon / \|g\|_{\mathcal{F}}$ (Lipschitz).

Summary

- Linear operator $A : \mathcal{F} \to \mathcal{G}$ maps linear subspaces to linear subspaces
  - $Im(A) = A(\mathcal{F})$ is a linear subspace of $\mathcal{G}$.
  - $Null(A) = A^{-1}(\{0\})$ is a linear subspace of $\mathcal{F}$

Summary

- Linear operator $A : \mathcal{F} \to \mathcal{G}$ maps linear subspaces to linear subspaces
  - $Im(A) = A(\mathcal{F})$ is a linear subspace of $\mathcal{G}$.
  - $Null(A) = A^{-1}(\{0\})$ is a linear subspace of $\mathcal{F}$
- Continuous $A : \mathcal{F} \to \mathcal{G}$ maps to closed sets from closed sets
  - If A is also linear, $Null(A) = A^{-1}(\{0\})$ is a **closed linear subspace** of $\mathcal{F}$ because $\{0\}$ is closed in $\mathcal{G}$.

Summary

- Linear operator $A : \mathcal{F} \to \mathcal{G}$ maps linear subspaces to linear subspaces
  - $Im(A) = A(\mathcal{F})$ is a linear subspace of $\mathcal{G}$.
  - $Null(A) = A^{-1}(\{0\})$ is a linear subspace of $\mathcal{F}$
- Continuous $A : \mathcal{F} \to \mathcal{G}$ maps to closed sets from closed sets
  - If A is also linear, $Null(A) = A^{-1}(\{0\})$ is a **closed linear subspace** of $\mathcal{F}$ because $\{0\}$ is closed in $\mathcal{G}$.
- Bounded linear operator $A : \mathcal{F} \to \mathcal{G}$ maps bounded sets to bounded sets

Continuous operator $\equiv$ Bounded operator

Theorem

Let $(\mathcal{F}, \|\cdot\|_{\mathcal{F}})$ and $(\mathcal{G}, \|\cdot\|_{\mathcal{G}})$ be normed linear spaces. If L is a linear operator, then the following three conditions are equivalent:

1. *L is a bounded operator.*
2. *L is continuous on $\mathcal{F}$.*
3. *L is continuous at one point of $\mathcal{F}$.*

Proof

Dual space

Definition (Topological dual)

If $\mathcal{F}$ is a normed space, then the space $\mathcal{F}'$ of *continuous linear* functionals $A : \mathcal{F} \to \mathbb{R}$ is called the topological dual space of $\mathcal{F}$.

Dual space

Definition (Topological dual)

If $\mathcal{F}$ is a normed space, then the space $\mathcal{F}'$ of *continuous linear* functionals $A : \mathcal{F} \to \mathbb{R}$ is called the topological dual space of $\mathcal{F}$.

We have seen that $A_g := \langle \cdot, g \rangle_{\mathcal{F}}$ are continuous linear functionals, so $A_g \in \mathcal{F}'$

Dual space

Definition (Topological dual)

If $\mathcal{F}$ is a normed space, then the space $\mathcal{F}'$ of *continuous linear* functionals $A : \mathcal{F} \to \mathbb{R}$ is called the topological dual space of $\mathcal{F}$.

We have seen that $A_g := \langle \cdot, g \rangle_{\mathcal{F}}$ are continuous linear functionals, so $A_g \in \mathcal{F}'$

Theorem (Riesz representation)

In a Hilbert space $\mathcal{F}$, for every continous linear functional $L \in \mathcal{F}'$, there exists a unique $g \in \mathcal{F}$, such that

$$Lf \equiv \langle f, g \rangle_{\mathcal{F}}.$$

Proof of Riesz representation

Proof.

Existence. Let $L \in \mathcal{F}'$. If $Lf \equiv 0$, then $Lf = \langle f, 0 \rangle_{\mathcal{F}}$, so $g = 0$. Otherwise, $M = Null(L) \subsetneq \mathcal{F}$ is a closed linear subspace of $\mathcal{F}$, so there must exist $h \in M^{\perp}$, with $\|h\|_{\mathcal{F}} = 1$. We claim that we can take $g = (Lh)h$. Indeed, for $f \in \mathcal{F}$, take $u_f = (Lf)h - (Lh)f$. Clearly $u_f \in M$. Thus,

$$
\begin{aligned}
0 &= \langle u_f, h \rangle_{\mathcal{F}} \\
&= \langle (Lf)h - (Lh)f, h \rangle_{\mathcal{F}} \\
&= (Lf)\|h\|_{\mathcal{F}}^2 - (Lh)\langle f, h \rangle_{\mathcal{F}} \\
&= Lf - \langle f, (Lh)h \rangle_{\mathcal{F}}.
\end{aligned}
$$

Uniqueness. If g_1 and g_2 are two representers, then $0 = Lf - Lf = \langle f, g_1 - g_2 \rangle_{\mathcal{F}}\ \forall f$. In particular, $\langle g_1 - g_2, g_1 - g_2 \rangle_{\mathcal{F}} = \|g_1 - g_2\|_{\mathcal{F}}^2 = 0$, so $g_1 = g_2$. □

Orthonormal basis

- orthonormal set $\{u_\alpha\}_{\alpha\in A}$, s.t.

$$\langle u_\alpha, u_\beta\rangle_{\mathcal{F}} = \begin{cases} 1, & \alpha=\beta \\ 0, & \alpha\neq\beta \end{cases}$$

- if also basis, i.e., $\mathcal{F} = span\{u_\alpha\} + span\{u_\beta\} + \cdots$, we define $\hat{f}(\alpha) = \langle f, u_\alpha\rangle_{\mathcal{F}}$

$$\begin{aligned} f &= \sum_{\alpha\in A} \hat{f}(\alpha) u_\alpha \\ \langle f, g\rangle_{\mathcal{F}} &= \sum_{\alpha\in A} \hat{f}(\alpha)\hat{g}(\alpha) \\ &= \left\langle \left\{\hat{f}(\alpha)\right\}, \{\hat{g}(\alpha)\}\right\rangle_{\ell^2(A)} \end{aligned}$$

Isometric isomorphism

Definition (Hilbert space isomorphism)

Two Hilbert spaces $\mathcal{H}$ and $\mathcal{F}$ are said to be *isometrically isomorphic* if there is a **linear bijective map** $U : \mathcal{H} \to \mathcal{F}$, which **preserves the inner product**, i.e., $\langle h_1, h_2 \rangle_{\mathcal{H}} = \langle Uh_1, Uh_2 \rangle_{\mathcal{F}}$.

Isometric isomorphism

Definition (Hilbert space isomorphism)

Two Hilbert spaces $\mathcal{H}$ and $\mathcal{F}$ are said to be *isometrically isomorphic* if there is a **linear bijective map** $U : \mathcal{H} \to \mathcal{F}$, which **preserves the inner product**, i.e., $\langle h_1, h_2 \rangle_{\mathcal{H}} = \langle Uh_1, Uh_2 \rangle_{\mathcal{F}}$.

Riesz representation gives an isomorphism $g \mapsto \langle \cdot, g \rangle_{\mathcal{F}}$ between $\mathcal{F}$ and $\mathcal{F}'$: dual space of a Hilbert space is another (isometrically isomorphic) Hilbert space.

Isometric isomorphism

Definition (Hilbert space isomorphism)

Two Hilbert spaces $\mathcal{H}$ and $\mathcal{F}$ are said to be *isometrically isomorphic* if there is a **linear bijective map** $U : \mathcal{H} \to \mathcal{F}$, which **preserves the inner product**, i.e., $\langle h_1, h_2 \rangle_{\mathcal{H}} = \langle Uh_1, Uh_2 \rangle_{\mathcal{F}}$.

Riesz representation gives an isomorphism $g \mapsto \langle \cdot, g \rangle_{\mathcal{F}}$ between $\mathcal{F}$ and $\mathcal{F}'$: dual space of a Hilbert space is another (isometrically isomorphic) Hilbert space.

Theorem

Every Hilbert space has an orthonormal basis. Thus, all Hilbert spaces are isometrically isomorphic to $\ell^2(A)$, for some set A. We can take $A = \mathbb{N}$ iff Hilbert space is separable.

Summary

Hilbert space:

- is a vector space over $\mathbb{R}$ (or $\mathbb{C}$)

Summary

Hilbert space:

- is a vector space over $\mathbb{R}$ (or $\mathbb{C}$)
- comes equipped with an inner product, a norm and a metric

Summary

Hilbert space:

- is a vector space over $\mathbb{R}$ (or $\mathbb{C}$)
- comes equipped with an inner product, a norm and a metric
- is complete with respect to its metric

Summary

Hilbert space:

- is a vector space over $\mathbb{R}$ (or $\mathbb{C}$)
- comes equipped with an inner product, a norm and a metric
- is complete with respect to its metric
- continuity and boundedness of linear operators are equivalent

Summary

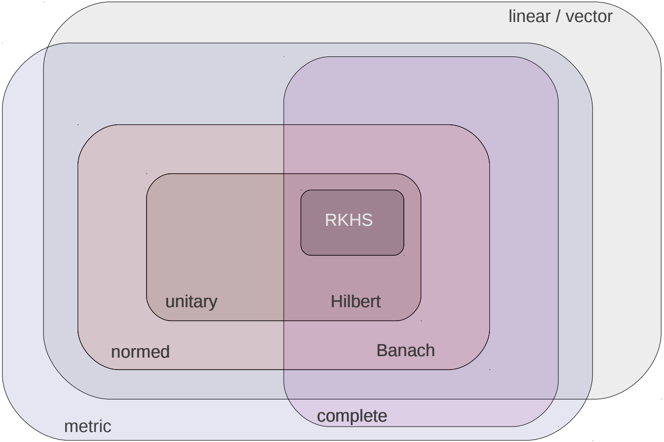
Hilbert space:

- is a vector space over $\mathbb{R}$ (or $\mathbb{C}$)
- comes equipped with an inner product, a norm and a metric
- is complete with respect to its metric
- continuity and boundedness of linear operators are equivalent
- **all** continuous linear functionals arise from the inner product

Outline

1. Elementary Hilbert space theory
   - Norm. Inner product. Orthogonality
   - Convergence. Complete spaces
   - Linear operators. Riesz representation
2. What is an RKHS?
   - Evaluation functionals view of RKHS
   - Reproducing kernel
   - Inner product between features
   - Positive definite function
   - Moore-Aronszajn Theorem

RKHS: a function space with a very special structure



Evaluation functional

Definition (Evaluation functional)

Let $\mathcal{H}$ be a Hilbert space of functions $f : \mathcal{X} \to \mathbb{R}$, defined on a non-empty set $\mathcal{X}$. For a fixed $x \in \mathcal{X}$, map $\delta_x : \mathcal{H} \to \mathbb{R}$, $\delta_x : f \mapsto f(x)$ is called the (Dirac) evaluation functional at x.

Evaluation functional

Definition (Evaluation functional)

Let $\mathcal{H}$ be a Hilbert space of functions $f : \mathcal{X} \to \mathbb{R}$, defined on a non-empty set $\mathcal{X}$. For a fixed $x \in \mathcal{X}$, map $\delta_x : \mathcal{H} \to \mathbb{R}$, $\delta_x : f \mapsto f(x)$ is called the (Dirac) evaluation functional at x.

- Evaluation functional is always linear: For $f, g \in \mathcal{H}$ and $\alpha, \beta \in \mathbb{R}$, $\delta_x(\alpha f + \beta g) = (\alpha f + \beta g)(x) = \alpha f(x) + \beta g(x) = \alpha \delta_x(f) + \beta \delta_x(g)$.

Evaluation functional

Definition (Evaluation functional)

Let $\mathcal{H}$ be a Hilbert space of functions $f : \mathcal{X} \to \mathbb{R}$, defined on a non-empty set $\mathcal{X}$. For a fixed $x \in \mathcal{X}$, map $\delta_x : \mathcal{H} \to \mathbb{R}$, $\delta_x : f \mapsto f(x)$ is called the (Dirac) evaluation functional at x.

- Evaluation functional is always linear: For $f, g \in \mathcal{H}$ and $\alpha, \beta \in \mathbb{R}$, $\delta_x(\alpha f + \beta g) = (\alpha f + \beta g)(x) = \alpha f(x) + \beta g(x) = \alpha \delta_x(f) + \beta \delta_x(g)$.
- But is it continuous?

Discontinuous evaluation

Example

$\mathcal{F}$: the space of polynomials over $[0,1]$, endowed with the L_p norm, i.e.,

$$\|f_1 - f_2\|_p = \left(\int_0^1 |f_1(x) - f_2(x)|^p \, dx\right)^{1/p}.$$

Consider the sequence of functions $\{q_n\}_{n=1}^{\infty}$, where $q_n = x^n$. Then: $\lim_{n\to\infty} \|q_n - 0\|_p = 0$, i.e., $\{q_n\}$ converges to "zero function" in L_p norm, but does not get close to zero function everywhere:

$$1 = \lim_{n\to\infty} \delta_1(q_n) \neq \delta_1(\lim_{n\to\infty} q_n) = 0.$$

Discontinuous evaluation

Example

$\mathcal{F}$: the space of polynomials over $[0,1]$, endowed with the L_p norm, i.e.,

$$\|f_1 - f_2\|_p = \left(\int_0^1 |f_1(x) - f_2(x)|^p \, dx \right)^{1/p}.$$

Consider the sequence of functions $\{q_n\}_{n=1}^{\infty}$, where $q_n = x^n$. Then: $\lim_{n\to\infty} \|q_n - 0\|_p = 0$, i.e., $\{q_n\}$ converges to "zero function" in L_p norm, but does not get close to zero function everywhere:

$$1 = \lim_{n\to\infty} \delta_1(q_n) \neq \delta_1(\lim_{n\to\infty} q_n) = 0.$$

$\delta_1 : f \mapsto f(1)$ is not continuous!

RKHS

Definition (Reproducing kernel Hilbert space)

A Hilbert space $\mathcal{H}$ of functions $f : \mathcal{X} \to \mathbb{R}$, defined on a non-empty set $\mathcal{X}$ is said to be a Reproducing Kernel Hilbert Space (RKHS) if $\delta_x \in \mathcal{H}'$, $\forall x \in \mathcal{X}$.

RKHS

Definition (Reproducing kernel Hilbert space)

A Hilbert space $\mathcal{H}$ of functions $f : \mathcal{X} \to \mathbb{R}$, defined on a non-empty set $\mathcal{X}$ is said to be a Reproducing Kernel Hilbert Space (RKHS) if $\delta_x \in \mathcal{H}'$, $\forall x \in \mathcal{X}$.

Theorem (Norm convergence implies pointwise convergence)

If $\lim_{n\to\infty} \|f_n - f\|_{\mathcal{H}} = 0$, *then* $\lim_{n\to\infty} f_n(x) = f(x)$, $\forall x \in \mathcal{X}$.

RKHS

Definition (Reproducing kernel Hilbert space)

A Hilbert space $\mathcal{H}$ of functions $f : \mathcal{X} \to \mathbb{R}$, defined on a non-empty set $\mathcal{X}$ is said to be a Reproducing Kernel Hilbert Space (RKHS) if $\delta_x \in \mathcal{H}'$, $\forall x \in \mathcal{X}$.

Theorem (Norm convergence implies pointwise convergence)

If $\lim_{n\to\infty} \|f_n - f\|_{\mathcal{H}} = 0$, *then* $\lim_{n\to\infty} f_n(x) = f(x)$, $\forall x \in \mathcal{X}$.

If two functions $f, g \in \mathcal{H}$ are close in the norm of $\mathcal{H}$, then $f(x)$ and $g(x)$ are close for all $x \in \mathcal{X}$

Outline

Will discuss three distinct concepts:

- **reproducing** kernel
- inner product between features (kernel)
- positive definite function

Outline

Will discuss three distinct concepts:

- **reproducing** kernel
- inner product between features (kernel)
- positive definite function

...and then show that they are all equivalent.

Outline

Reproducing kernel

Definition (Reproducing kernel)

Let $\mathcal{H}$ be a Hilbert space of functions $f : \mathcal{X} \to \mathbb{R}$ defined on a non-empty set $\mathcal{X}$. A function $k : \mathcal{X} \times \mathcal{X} \to \mathbb{R}$ is called *a reproducing kernel* of $\mathcal{H}$ if it satisfies

- $\forall x \in \mathcal{X}, \quad k_x = k(\cdot, x) \in \mathcal{H}$,
- $\forall x \in \mathcal{X}, \forall f \in \mathcal{H}, \quad \langle f, k(\cdot, x) \rangle_{\mathcal{H}} = f(x)$ (the reproducing property).

Reproducing kernel

Definition (Reproducing kernel)

Let $\mathcal{H}$ be a Hilbert space of functions $f : \mathcal{X} \to \mathbb{R}$ defined on a non-empty set $\mathcal{X}$. A function $k : \mathcal{X} \times \mathcal{X} \to \mathbb{R}$ is called *a reproducing kernel* of $\mathcal{H}$ if it satisfies

- $\forall x \in \mathcal{X}, \quad k_x = k(\cdot, x) \in \mathcal{H}$,
- $\forall x \in \mathcal{X}, \forall f \in \mathcal{H}, \quad \langle f, k(\cdot, x)\rangle_{\mathcal{H}} = f(x)$ (the reproducing property).

In particular, for any $x, y \in \mathcal{X}$,

$$k(x, y) = \langle k(\cdot, y), k(\cdot, x)\rangle_{\mathcal{H}} = \langle k(\cdot, x), k(\cdot, y)\rangle_{\mathcal{H}}.$$

Reproducing kernel of an RKHS

Theorem

If it exists, reproducing kernel is unique.

Theorem

$\mathcal{H}$ is a reproducing kernel Hilbert space if and only if it has a reproducing kernel.

Outline

1. Elementary Hilbert space theory
   - Norm. Inner product. Orthogonality
   - Convergence. Complete spaces
   - Linear operators. Riesz representation
2. What is an RKHS?
   - Evaluation functionals view of RKHS
   - Reproducing kernel
   - Inner product between features
   - Positive definite function
   - Moore-Aronszajn Theorem

Functions representable as inner products

Definition (Kernel)

A function $k : \mathcal{X} \times \mathcal{X} \to \mathbb{R}$ is called *a kernel* on $\mathcal{X}$ if there exists a Hilbert space (not necessarilly an RKHS) $\mathcal{F}$ and a map $\phi : \mathcal{X} \to \mathcal{F}$, such that $k(x, y) = \langle \phi(x), \phi(y) \rangle_{\mathcal{F}}$.

Functions representable as inner products

Definition (Kernel)

A function $k : \mathcal{X} \times \mathcal{X} \to \mathbb{R}$ is called *a kernel* on $\mathcal{X}$ if there exists a Hilbert space (not necessarilly an RKHS) $\mathcal{F}$ and a map $\phi : \mathcal{X} \to \mathcal{F}$, such that $k(x, y) = \langle \phi(x), \phi(y) \rangle_{\mathcal{F}}$.

- note that we dropped 'reproducing', as $\mathcal{F}$ may not be an RKHS.
- $\phi : \mathcal{X} \to \mathcal{F}$ is called a **feature map**,
- $\mathcal{F}$ is called a **feature space**.

Functions representable as inner products

Definition (Kernel)

A function $k : \mathcal{X} \times \mathcal{X} \to \mathbb{R}$ is called *a kernel* on $\mathcal{X}$ if there exists a Hilbert space (not necessarilly an RKHS) $\mathcal{F}$ and a map $\phi : \mathcal{X} \to \mathcal{F}$, such that $k(x, y) = \langle \phi(x), \phi(y) \rangle_{\mathcal{F}}$.

- note that we dropped 'reproducing', as $\mathcal{F}$ may not be an RKHS.
- $\phi : \mathcal{X} \to \mathcal{F}$ is called a **feature map**,
- $\mathcal{F}$ is called a **feature space**.

Corollary

*Every **reproducing kernel** is a **kernel** (can take $\phi : x \mapsto k(\cdot, x)$, $k(x, y) = \langle k(\cdot, x), k(\cdot, y) \rangle_{\mathcal{H}}$, i.e., RKHS $\mathcal{H}$ is a feature space).*

Non-uniqueness of feature representation

Example

Consider $\mathcal{X} = \mathbb{R}^2$, and $k(x, y) = \langle x, y \rangle^2$

$$
\begin{aligned}
k(x,y) &= x_1^2 y_1^2 + x_2^2 y_2^2 + 2x_1x_2y_1y_2 \\
&= \begin{bmatrix} x_1^2 & x_2^2 & \sqrt{2}x_1x_2 \end{bmatrix} \begin{bmatrix} y_1^2 \\ y_2^2 \\ \sqrt{2}y_1y_2 \end{bmatrix} \\
&= \begin{bmatrix} x_1^2 & x_2^2 & x_1x_2 & x_1x_2 \end{bmatrix} \begin{bmatrix} y_1^2 \\ y_2^2 \\ y_1y_2 \\ y_1y_2 \end{bmatrix}.
\end{aligned}
$$

so we can use the feature maps $\phi(x) = (x_1^2, x_2^2, \sqrt{2}x_1x_2)$ *or* $\tilde{\phi}(x) = \begin{bmatrix} x_1^2 & x_2^2 & x_1x_2 & x_1x_2 \end{bmatrix}$, with feature spaces $\mathcal{H} = \mathbb{R}^3$ *or* $\tilde{\mathcal{H}} = \mathbb{R}^4$.

Non-uniqueness of feature representation

Example

Consider $\mathcal{X} = \mathbb{R}^2$, and $k(x, y) = \langle x, y \rangle^2$

$$\begin{aligned} k(x,y) &= x_1^2 y_1^2 + x_2^2 y_2^2 + 2x_1x_2y_1y_2 \\ &= \left[\begin{array}{ccc} x_1^2 & x_2^2 & \sqrt{2}x_1x_2 \end{array}\right] \left[\begin{array}{c} y_1^2 \\ y_2^2 \\ \sqrt{2}y_1y_2 \end{array}\right] \\ &= \left[\begin{array}{cccc} x_1^2 & x_2^2 & x_1x_2 & x_1x_2 \end{array}\right] \left[\begin{array}{c} y_1^2 \\ y_2^2 \\ y_1y_2 \\ y_1y_2 \end{array}\right]. \end{aligned}$$

so we can use the feature maps $\phi(x) = \left(x_1^2, x_2^2, \sqrt{2}x_1x_2\right)$ *or* $\tilde{\phi}(x) = \left[\begin{array}{cccc} x_1^2 & x_2^2 & x_1x_2 & x_1x_2 \end{array}\right]$, with feature spaces $\mathcal{H} = \mathbb{R}^3$ *or* $\tilde{\mathcal{H}} = \mathbb{R}^4$.

Not RKHS!

Outline

1. Elementary Hilbert space theory
   - Norm. Inner product. Orthogonality
   - Convergence. Complete spaces
   - Linear operators. Riesz representation
2. What is an RKHS?
   - Evaluation functionals view of RKHS
   - Reproducing kernel
   - Inner product between features
   - Positive definite function
   - Moore-Aronszajn Theorem

Positive definite functions

Definition (Positive definite functions)

A symmetric function $h : \mathcal{X} \times \mathcal{X} \to \mathbb{R}$ is positive definite if $\forall n \geq 1$, $\forall (a_1, \ldots a_n) \in \mathbb{R}^n$, $\forall (x_1, \ldots, x_n) \in \mathcal{X}^n$,

$$\sum_{i=1}^{n} \sum_{j=1}^{n} a_i a_j h(x_i, x_j) = \mathbf{a}^\top \mathbf{H} \mathbf{a} \geq 0.$$

The function $h(\cdot, \cdot)$ is *strictly* positive definite if for mutually distinct x_i, the equality holds only when all the a_i are zero.

Kernels are positive definite

Every inner product is a positive definite function, and more generally:

Fact

Every kernel is a positive definite function.

So far

$$\text{reproducing kernel} \implies \text{kernel} \implies \text{positive definite}$$

So far

$$\text{reproducing kernel} \implies \text{kernel} \implies \text{positive definite}$$

Is every positive definite function a reproducing kernel for some RKHS?

Moore-Aronszajn Theorem

Theorem (Moore-Aronszajn - Part I)

Let $k : \mathcal{X} \times \mathcal{X} \to \mathbb{R}$ be positive definite. There is a ***unique RKHS*** *$\mathcal{H} \subset \mathbb{R}^{\mathcal{X}}$ with reproducing kernel k.*

Non-uniqueness of feature representation

Example

Consider $\mathcal{X} = \mathbb{R}^2$, and $k(x, y) = \langle x, y \rangle^2$

$$
\begin{aligned}
k(x, y) &= x_1^2 y_1^2 + x_2^2 y_2^2 + 2 x_1 x_2 y_1 y_2 \\
&= \begin{bmatrix} x_1^2 & x_2^2 & \sqrt{2} x_1 x_2 \end{bmatrix} \begin{bmatrix} y_1^2 \\ y_2^2 \\ \sqrt{2} y_1 y_2 \end{bmatrix} \\
&= \begin{bmatrix} x_1^2 & x_2^2 & x_1 x_2 & x_1 x_2 \end{bmatrix} \begin{bmatrix} y_1^2 \\ y_2^2 \\ y_1 y_2 \\ y_1 y_2 \end{bmatrix}.
\end{aligned}
$$

so we can use the feature maps $\phi(x) = \begin{bmatrix} x_1^2 & x_2^2 & \sqrt{2} x_1 x_2 \end{bmatrix}$ *or* $\tilde{\phi}(x) = \begin{bmatrix} x_1^2 & x_2^2 & x_1 x_2 & x_1 x_2 \end{bmatrix}$, with feature spaces $\mathcal{H} = \mathbb{R}^3$ *or* $\tilde{\mathcal{H}} = \mathbb{R}^4$.

Non-uniqueness of feature representation

Example

Consider $\mathcal{X} = \mathbb{R}^2$, and $k(x, y) = \langle x, y \rangle^2$

$$
\begin{aligned}
k(x, y) &= x_1^2 y_1^2 + x_2^2 y_2^2 + 2 x_1 x_2 y_1 y_2 \\
&= \begin{bmatrix} x_1^2 & x_2^2 & \sqrt{2} x_1 x_2 \end{bmatrix} \begin{bmatrix} y_1^2 \\ y_2^2 \\ \sqrt{2} y_1 y_2 \end{bmatrix} \\
&= \begin{bmatrix} x_1^2 & x_2^2 & x_1 x_2 & x_1 x_2 \end{bmatrix} \begin{bmatrix} y_1^2 \\ y_2^2 \\ y_1 y_2 \\ y_1 y_2 \end{bmatrix}.
\end{aligned}
$$

so we can use the feature maps $\phi(x) = \begin{bmatrix} x_1^2 & x_2^2 & \sqrt{2} x_1 x_2 \end{bmatrix}$ *or* $\tilde{\phi}(x) = \begin{bmatrix} x_1^2 & x_2^2 & x_1 x_2 & x_1 x_2 \end{bmatrix}$, with feature spaces $\mathcal{H} = \mathbb{R}^3$ *or* $\tilde{\mathcal{H}} = \mathbb{R}^4$.

$\mathcal{H}$ and $\tilde{\mathcal{H}}$ are not RKHS - RKHS of k is unique

Non-uniqueness of feature representation

- There are (infinitely) many feature space representations (and we can even work in one or more of them, if it's convenient!)

$$\langle \phi(x), \phi(y) \rangle_{\mathbb{R}^3} = a y_1^2 + b y_2^2 + c\sqrt{2} y_1 y_2 = k_x(y) = \langle k_x, k_y \rangle_{\mathcal{H}_k}$$

$$\phi(x) =$$

$$\begin{bmatrix} a = x_1^2 & b = x_2^2 & c = \sqrt{2} x_1 x_2 \end{bmatrix}$$

Non-uniqueness of feature representation

- There are (infinitely) many feature space representations (and we can even work in one or more of them, if it's convenient!)

$$\langle \phi(x), \phi(y) \rangle_{\mathbb{R}^3} = a y_1^2 + b y_2^2 + c\sqrt{2} y_1 y_2 \quad = \quad k_x(y) = \langle k_x, k_y \rangle_{\mathcal{H}_k}$$

$$\phi(x) = \begin{bmatrix} a = x_1^2 & b = x_2^2 & c = \sqrt{2} x_1 x_2 \end{bmatrix}$$

$$\left\langle \tilde{\phi}(x), \tilde{\phi}(y) \right\rangle_{\mathbb{R}^4} = \tilde{a} y_1^2 + \tilde{b} y_2^2 + \tilde{c} y_1 y_2 + \tilde{d} y_1 y_2 \quad = \quad k_x(y) = \langle k_x, k_y \rangle_{\mathcal{H}_k}$$

$$\tilde{\phi}(x) = \begin{bmatrix} \tilde{a} = x_1^2 & \tilde{b} = x_2^2 & \tilde{c} = x_1 x_2 & \tilde{d} = x_1 x_2 \end{bmatrix}$$

Non-uniqueness of feature representation

- There are (infinitely) many feature space representations (and we can even work in one or more of them, if it's convenient!)

$$\langle \phi(x), \phi(y) \rangle_{\mathbb{R}^3} = ay_1^2 + by_2^2 + c\sqrt{2}y_1y_2 \quad = \quad k_x(y) = \langle k_x, k_y \rangle_{\mathcal{H}_k}$$

$$\phi(x) = \begin{bmatrix} a = x_1^2 & b = x_2^2 & c = \sqrt{2}x_1x_2 \end{bmatrix}$$

$$\left\langle \tilde{\phi}(x), \tilde{\phi}(y) \right\rangle_{\mathbb{R}^4} = \tilde{a}y_1^2 + \tilde{b}y_2^2 + \tilde{c}y_1y_2 + \tilde{d}y_1y_2 \quad = \quad k_x(y) = \langle k_x, k_y \rangle_{\mathcal{H}_k}$$

$$\tilde{\phi}(x) = \begin{bmatrix} \tilde{a} = x_1^2 & \tilde{b} = x_2^2 & \tilde{c} = x_1x_2 & \tilde{d} = x_1x_2 \end{bmatrix}$$

- But what remains unique?

Non-uniqueness of feature representation

- There are (infinitely) many feature space representations (and we can even work in one or more of them, if it's convenient!)

$$\langle \phi(x), \phi(y) \rangle_{\mathbb{R}^3} = a y_1^2 + b y_2^2 + c\sqrt{2} y_1 y_2 \quad = \quad k_x(y) = \langle k_x, k_y \rangle_{\mathcal{H}_k}$$

$$\phi(x) = \begin{bmatrix} a = x_1^2 & b = x_2^2 & c = \sqrt{2} x_1 x_2 \end{bmatrix}$$

$$\left\langle \tilde{\phi}(x), \tilde{\phi}(y) \right\rangle_{\mathbb{R}^4} = \tilde{a} y_1^2 + \tilde{b} y_2^2 + \tilde{c} y_1 y_2 + \tilde{d} y_1 y_2 \quad = \quad k_x(y) = \langle k_x, k_y \rangle_{\mathcal{H}_k}$$

$$\tilde{\phi}(x) = \begin{bmatrix} \tilde{a} = x_1^2 & \tilde{b} = x_2^2 & \tilde{c} = x_1 x_2 & \tilde{d} = x_1 x_2 \end{bmatrix}$$

- But what remains unique?
- Kernel and its RKHS!

Summary

reproducing kernel $\iff$ kernel $\iff$ positive definite

Summary

reproducing kernel $\Longleftrightarrow$ kernel $\Longleftrightarrow$ positive definite

set of all kernels: $\mathbb{R}_+^{\mathcal{X}\times\mathcal{X}}$

$\overset{1-1}{\longleftrightarrow}$

set of all subspaces of $\mathbb{R}^{\mathcal{X}}$ with continuous evaluation: $\mathit{Hilb}(\mathbb{R}^{\mathcal{X}})$

Outline

1. Elementary Hilbert space theory
   - Norm. Inner product. Orthogonality
   - Convergence. Complete spaces
   - Linear operators. Riesz representation
2. What is an RKHS?
   - Evaluation functionals view of RKHS
   - Reproducing kernel
   - Inner product between features
   - Positive definite function
   - Moore-Aronszajn Theorem

Moore-Aronszajn Theorem

Theorem (Moore-Aronszajn - Part I)

Let $k : \mathcal{X} \times \mathcal{X} \to \mathbb{R}$ be positive definite. There is a ***unique RKHS*** *$\mathcal{H} \subset \mathbb{R}^{\mathcal{X}}$ with reproducing kernel k.*

Moore-Aronszajn Theorem (2)

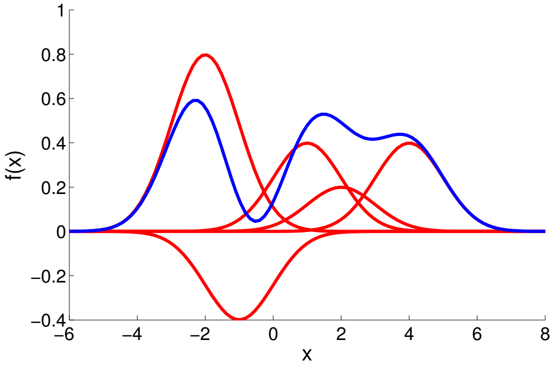
Starting with a positive def. k, construct a **pre-RKHS** $\mathcal{H}_0$ with properties:

1. The evaluation functionals δ_x are continuous on $\mathcal{H}_0$,
2. Any Cauchy sequence f_n in $\mathcal{H}_0$ which converges pointwise to 0 also converges in $\mathcal{H}_0$-norm to 0.

Moore-Aronszajn Theorem (3)

pre-RKHS $\mathcal{H}_0 = span\left\{k(\cdot, x) \,|\, x \in \mathcal{X}\right\}$ will be taken to be the set of functions:

$$f(x) = \sum_{i=1}^{n} \alpha_i k(x_i, x)$$



Moore-Aronszajn Theorem (4)

Theorem (Moore-Aronszajn - Part II)

Space $\mathcal{H}_0 = span\left\{k(\cdot,x)\,|\,x\in\mathcal{X}\right\}$ *is endowed with the inner product*

$$\langle f,g\rangle_{\mathcal{H}_0} = \sum_{i=1}^{n}\sum_{j=1}^{m}\alpha_i\beta_j k(x_i,y_j),$$

where $f=\sum_{i=1}^{n}\alpha_i k(\cdot,x_i)$ *and* $g=\sum_{j=1}^{m}\beta_j k(\cdot,y_j)$*, then* $\mathcal{H}_0$ *is dense in RKHS* $\mathcal{H}$ *of* k*.*

Moore-Aronszajn Theorem (5)

Define $\mathcal{H}$ to be the set of functions $f \in \mathbb{R}^{\mathcal{X}}$ for which there exists a Cauchy sequence $\{f_n\} \in \mathcal{H}_0$ converging **pointwise** to f.

Moore-Aronszajn Theorem (5)

Define $\mathcal{H}$ to be the set of functions $f \in \mathbb{R}^{\mathcal{X}}$ for which there exists a Cauchy sequence $\{f_n\} \in \mathcal{H}_0$ converging **pointwise** to f.

1. We define the inner product between $f, g \in \mathcal{H}$ as the limit of an inner product of the Cauchy sequences $\{f_n\}$, $\{g_n\}$ converging to f and g respectively. Is the inner product well defined, and independent of the sequences used?

Moore-Aronszajn Theorem (5)

Define $\mathcal{H}$ to be the set of functions $f \in \mathbb{R}^{\mathcal{X}}$ for which there exists a Cauchy sequence $\{f_n\} \in \mathcal{H}_0$ converging **pointwise** to f.

1. We define the inner product between $f, g \in \mathcal{H}$ as the limit of an inner product of the Cauchy sequences $\{f_n\}$, $\{g_n\}$ converging to f and g respectively. Is the inner product well defined, and independent of the sequences used?
2. An inner product space must satisfy $\langle f, f \rangle_{\mathcal{H}} = 0$ iff $f = 0$. Is this true when we define the inner product on $\mathcal{H}$ as above?

Moore-Aronszajn Theorem (5)

Define $\mathcal{H}$ to be the set of functions $f \in \mathbb{R}^{\mathcal{X}}$ for which there exists a Cauchy sequence $\{f_n\} \in \mathcal{H}_0$ converging **pointwise** to f.

1. We define the inner product between $f, g \in \mathcal{H}$ as the limit of an inner product of the Cauchy sequences $\{f_n\}$, $\{g_n\}$ converging to f and g respectively. Is the inner product well defined, and independent of the sequences used?
2. An inner product space must satisfy $\langle f, f \rangle_{\mathcal{H}} = 0$ iff $f = 0$. Is this true when we define the inner product on $\mathcal{H}$ as above?
3. Are the evaluation functionals still continuous on $\mathcal{H}$?

Moore-Aronszajn Theorem (5)

Define $\mathcal{H}$ to be the set of functions $f \in \mathbb{R}^{\mathcal{X}}$ for which there exists a Cauchy sequence $\{f_n\} \in \mathcal{H}_0$ converging **pointwise** to f.

1. We define the inner product between $f, g \in \mathcal{H}$ as the limit of an inner product of the Cauchy sequences $\{f_n\}$, $\{g_n\}$ converging to f and g respectively. Is the inner product well defined, and independent of the sequences used?
2. An inner product space must satisfy $\langle f, f \rangle_{\mathcal{H}} = 0$ iff $f = 0$. Is this true when we define the inner product on $\mathcal{H}$ as above?
3. Are the evaluation functionals still continuous on $\mathcal{H}$?
4. Is $\mathcal{H}$ complete (a Hilbert space)?

Summary

reproducing kernel $\iff$ kernel $\iff$ positive definite

Summary

reproducing kernel $\iff$ kernel $\iff$ positive definite

$$\text{all kernels } \mathbb{R}_+^{\mathcal{X}\times\mathcal{X}}$$
$$\overset{1-1}{\longleftrightarrow}$$
$$\text{all function spaces with continuous evaluation } \mathit{Hilb}(\mathbb{R}^{\mathcal{X}})$$

Operations with kernels

Fact (Sum and scaling of kernels)

If k, k_1, and k_2 are kernels on $\mathcal{X}$, and $\alpha \geq 0$ is a scalar, then αk, $k_1 + k_2$ are kernels.

Operations with kernels

Fact (Sum and scaling of kernels)

If k, k_1, and k_2 are kernels on $\mathcal{X}$, and $\alpha \geq 0$ is a scalar, then αk, $k_1 + k_2$ are kernels.

- A difference of kernels is not necessarily a kernel! This is because we cannot have $k_1(x,x) - k_2(x,x) = \langle \phi(x), \phi(x) \rangle_{\mathcal{H}} < 0$.
- This gives the set of all kernels the geometry of a *closed convex cone.*

Operations with kernels

Fact (Sum and scaling of kernels)

If k, k_1, and k_2 are kernels on $\mathcal{X}$, and $\alpha \geq 0$ is a scalar, then αk, $k_1 + k_2$ are kernels.

- A difference of kernels is not necessarily a kernel! This is because we cannot have $k_1(x,x) - k_2(x,x) = \langle \phi(x), \phi(x) \rangle_{\mathcal{H}} < 0$.
- This gives the set of all kernels the geometry of a *closed convex cone.*

$$\mathcal{H}_{k_1+k_2} = \mathcal{H}_{k_1} + \mathcal{H}_{k_2} = \{f_1 + f_2 \, : \, f_1 \in \mathcal{H}_{k_1}, f_2 \in \mathcal{H}_{k_2}\}$$

Operations with kernels (2)

Fact (Product of kernels)

If k_1 and k_2 are kernels on $\mathcal{X}$ and $\mathcal{Y}$, then $k = k_1 \otimes k_2$, given by:

$$k\left((x,y),(x',y')\right) \quad := \quad k_1(x,x')k_2(y,y')$$

is a kernel on $\mathcal{X} \times \mathcal{Y}$. If $\mathcal{X} = \mathcal{Y}$, then $k = k_1 \cdot k_2$, given by:

$$k\left(x,x'\right) \quad := \quad k_1(x,x')k_2(x,x')$$

is a kernel on $\mathcal{X}$.

Operations with kernels (2)

Fact (Product of kernels)

If k_1 and k_2 are kernels on $\mathcal{X}$ and $\mathcal{Y}$, then $k = k_1 \otimes k_2$, given by:

$$k\left((x,y),(x',y')\right) \quad := \quad k_1(x,x')k_2(y,y')$$

is a kernel on $\mathcal{X} \times \mathcal{Y}$. If $\mathcal{X} = \mathcal{Y}$, then $k = k_1 \cdot k_2$, given by:

$$k\left(x,x'\right) \quad := \quad k_1(x,x')k_2(x,x')$$

is a kernel on $\mathcal{X}$.

$$\mathcal{H}_{k_1 \otimes k_2} \cong \mathcal{H}_{k_1} \otimes \mathcal{H}_{k_2}$$

Summary

$$\text{all kernels } \mathbb{R}_+^{\mathcal{X}\times\mathcal{X}}$$

$$\overset{1-1}{\longleftrightarrow}$$

$$\text{all function spaces with continuous evaluation } \mathit{Hilb}(\mathbb{R}^{\mathcal{X}})$$

Summary

all kernels $\mathbb{R}_+^{\mathcal{X}\times\mathcal{X}}$

$$\overset{1-1}{\longleftrightarrow}$$

all function spaces with continuous evaluation $Hilb(\mathbb{R}^{\mathcal{X}})$

bijection between $\mathbb{R}_+^{\mathcal{X}\times\mathcal{X}}$ and $Hilb(\mathbb{R}^{\mathcal{X}})$ preserves geometric structure

Operations with kernels (3)

New kernels from old:

- trivial (linear) kernel on $\mathbb{R}^d$ is $k(x, x') = \langle x, x' \rangle$

Operations with kernels (3)

New kernels from old:

- trivial (linear) kernel on $\mathbb{R}^d$ is $k(x, x') = \langle x, x' \rangle$
- for any $p(t) = a_m t^m + \cdots + a_1 t + a_0$ with $a_i \geq 0$
  $\implies k(x, x') = p(\langle x, x' \rangle)$ is a kernel on $\mathbb{R}^d$

Operations with kernels (3)

New kernels from old:

- trivial (linear) kernel on $\mathbb{R}^d$ is $k(x, x') = \langle x, x' \rangle$
- for any $p(t) = a_m t^m + \cdots + a_1 t + a_0$ with $a_i \geq 0$
  $\implies k(x, x') = p(\langle x, x' \rangle)$ is a kernel on $\mathbb{R}^d$
- **polynomial kernel:** $k(x, x') = (\langle x, x' \rangle + c)^m$, for $c \geq 0$

Operations with kernels (3)

New kernels from old:

- trivial (linear) kernel on $\mathbb{R}^d$ is $k(x, x') = \langle x, x' \rangle$
- for any $p(t) = a_m t^m + \cdots + a_1 t + a_0$ with $a_i \geq 0$
  $\implies k(x, x') = p(\langle x, x' \rangle)$ is a kernel on $\mathbb{R}^d$
- **polynomial kernel:** $k(x, x') = (\langle x, x' \rangle + c)^m$, for $c \geq 0$
- $f(t)$ has Taylor series with non-negative coefficients
  $\implies k(x, x') = f(\langle x, x' \rangle)$ is a kernel on $\mathbb{R}^d$

Operations with kernels (3)

New kernels from old:

- trivial (linear) kernel on $\mathbb{R}^d$ is $k(x, x') = \langle x, x' \rangle$
- for any $p(t) = a_m t^m + \cdots + a_1 t + a_0$ with $a_i \geq 0$
  $\implies k(x, x') = p(\langle x, x' \rangle)$ is a kernel on $\mathbb{R}^d$
- **polynomial kernel:** $k(x, x') = (\langle x, x' \rangle + c)^m$, for $c \geq 0$
- $f(t)$ has Taylor series with non-negative coefficients
  $\implies k(x, x') = f(\langle x, x' \rangle)$ is a kernel on $\mathbb{R}^d$
- **exponential kernel:** $k(x, x') = \exp(\sigma \langle x, x' \rangle)$, for $\sigma > 0$

Operations with kernels (3)

New kernels from old:

- **polynomial kernel:** $k(x, x') = (\langle x, x' \rangle + c)^m$, for $c \geq 0$
- **exponential kernel:** $k(x, x') = \exp(\sigma \langle x, x' \rangle)$, for $\sigma > 0$