Elements of Positive Definite Kernels and Reproducing Kernel Hilbert Spaces

Statistical Data Analysis with Positive Definite Kernels

Kenji Fukumizu

Institute of Statistical Mathematics, ROIS
Department of Statistical Science, Graduate University for Advanced Studies

October 6-10, 2008, Kyushu University

Outline

Definition of positive definite kernel

Definition. Let $\mathcal{X}$ be a set. $k : \mathcal{X} \times \mathcal{X} \to \mathbb{R}$ is a positive definite kernel if $k(x, y) = k(y, x)$ and for every $x_1, \ldots, x_n \in \mathcal{X}$ and $c_1, \ldots, c_n \in \mathbb{R}$

$$\sum_{i,j=1}^{n} c_i c_j k(x_i, x_j) \geq 0,$$

i.e. the symmetric matrix

$$(k(x_i, x_j))_{i,j=1}^{n} = \begin{pmatrix} k(x_1, x_1) & \cdots & k(x_1, x_n) \\ \vdots & \ddots & \vdots \\ k(x_n, x_1) & \cdots & k(x_n, x_) \end{pmatrix}$$

is positive semidefinite.

- The symmetric matrix $(k(x_i, x_j))_{i,j=1}^{n}$ is often called a Gram matrix.

Definition: complex-valued case

Definition. Let $\mathcal{X}$ be a set. $k : \mathcal{X} \times \mathcal{X} \to \mathbb{C}$ is a positive definite kernel if for every $x_1, \ldots, x_n \in \mathcal{X}$ and $c_1, \ldots, c_n \in \mathbb{C}$

$$\sum_{i,j=1}^{n} c_i \overline{c_j} k(x_i, x_j) \geq 0.$$

Remark. The Hermitian property $k(y, x) = \overline{k(x, y)}$ is derived from the positive-definiteness. [Exercise]

Some basic Properties

Fact. Assume $k : \mathcal{X} \times \mathcal{X} \to \mathbb{C}$ is positive definite. Then, for any x, y in $\mathcal{X}$,

1. $k(x, x) \geq 0$.
2. $|k(x, y)|^2 \leq k(x, x)k(y, y)$.

Proof. (1) is obvious. For (2), with the fact $k(y, x) = \overline{k(x, y)}$, the definition of positive definiteness implies that the eigenvalues of the hermitian matrix

$$\begin{pmatrix} k(x,x) & \overline{k(x,y)} \\ k(x,y) & k(y,y) \end{pmatrix}$$

is non-negative, thus, its determinant $k(x, x)k(y, y) - |k(x, y)|^2$ is non-negative. □

Examples

Real valued positive definite kernels on $\mathbb{R}^n$:

- Linear kernel

$$k_0(x, y) = x^T y$$

- Exponential

$$k_E(x, y) = \exp(\beta x^T y) \qquad (\beta > 0)$$

- Gaussian RBF (radial basis function) kernel

$$k_G(x, y) = \exp\left(-\frac{1}{2\sigma^2}\|x - y\|^2\right) \qquad (\sigma > 0)$$

- Laplacian kernel

$$k_L(x, y) = \exp\left(-\alpha \textstyle\sum_{i=1}^n |x_i - y_i|\right) \qquad (\alpha > 0)$$

- Polynomial kernel

$$k_P(x, y) = (x^T y + c)^d \qquad (c \geq 0, d \in \mathbb{N})$$

Operations that Preserve Positive Definiteness I

Proposition 1

If $k_i : \mathcal{X} \times \mathcal{X} \to \mathbb{C}$ ($i = 1, 2, \ldots$) are positive definite kernels, then so are the following:

1. *(positive combination)* $ak_1 + bk_2 \qquad (a, b \geq 0)$.
2. *(product)* $k_1 k_2 \qquad (k_1(x, y)k_2(x, y))$.
3. *(limit)* $lim_{i\to\infty} k_i(x, y)$, *assuming the limit exists.*

Remark. From Proposition 1, the set of all positive definite kernels is a closed (w.r.t. pointwise convergence) convex cone stable under multiplication.

Proof.
(1): Obvious.
(3): The non-negativity in the definition holds also for the limit.

Operations that Preserve Positive Definiteness II

(2): It suffices to show that two Hermitian matrices A and B are positive semidefinite, so is their component-wise product. This is done by the following lemma. □

Definition. For two matrices A and B of the same size, the matrix C with $C_{ij} = A_{ij}B_{ij}$ is called the Hadamard product of A and B.

The Hadamard product of A and B is denoted by $A \odot B$.

Lemma 2

Let A and B be non-negative Hermitian matrices of the same size. Then, $A \odot B$ is also non-negative.

Operations that Preserve Positive Definiteness III

Proof.
Let

$$A = U\Lambda U^*$$

be the eigendecomposition of A, where

$U = (u^1, \dots, u^p)$: a unitary matrix

Λ: diagonal matrix with non-negative entries $(\lambda_1, \dots, \lambda_p)$

$U^* = \overline{U}^T$.

Then, for arbitrary $c_1, \dots, c_p \in \mathbb{C}$,

$$\sum_{i,j=1} c_i \bar{c}_j (A \odot B)_{ij} = \sum_{a=1}^{p} \lambda_a c_i \bar{c}_j u_i^a \bar{u}_j^a B_{ij} = \sum_{a=1}^{p} \lambda_a \xi^{aT} B \overline{\xi^a},$$

where $\xi^a = (c_1 u_1^a, \dots, c_p u_p^a)^T \in \mathbb{C}^p$.

Since $\xi^{aT} B \overline{\xi^a}$ and λ_a are non-negative for each a, so is the sum. □

Basic construction of positive definite kernels I

Proposition 3

Let V be an vector space with an inner product $\langle \cdot, \cdot \rangle$. If we have a map

$$\Phi : \mathcal{X} \to V, \qquad x \mapsto \Phi(x),$$

a positive definite kernel on $\mathcal{X}$ is defined by

$$k(x, y) = \langle \Phi(x), \Phi(y) \rangle.$$

Proof. Let $x_1, \ldots, x_n$ in $\mathcal{X}$ and $c_1, \ldots, c_n \in \mathbb{C}$.

$$\begin{aligned} \textstyle\sum_{i,j=1}^n c_i \overline{c_j} k(x_i, x_j) &= \textstyle\sum_{i,j=1}^n c_i \overline{c_j} \langle \Phi(x_i), \Phi(x_j) \rangle \\ &= \left\langle \textstyle\sum_{i=1}^n c_i \Phi(x_i), \sum_{j=1}^n c_j \Phi(x_j) \right\rangle \\ &= \left\| \textstyle\sum_{i=1}^n c_i \Phi(x_i) \right\|^2 \quad \geq 0. \end{aligned}$$

Basic construction of positive definite kernels II

Proposition 4

Let $k : \mathcal{X} \times \mathcal{X} \to \mathbb{C}$ *be a positive definite kernel and* $f : \mathcal{X} \to \mathbb{C}$ *be an arbitrary function. Then,*

$$\tilde{k}(x, y) = f(x)k(x, y)\overline{f(y)}$$

is positive definite. In particular,

$$f(x)\overline{f(y)}$$

and

$$\frac{k(x, y)}{\sqrt{k(x, x)}\sqrt{k(y, y)}} \qquad (\textit{normalized kernel})$$

are positive definite.

Proof is left as an exercise.

Proofs of positive definiteness of examples

- Linear kernel: Proposition 3
- Exponential:

$$\exp(\beta x^T y) = 1 + \beta x^T y + \frac{\beta^2}{2!}(x^T y)^2 + \frac{\beta^3}{3!}(x^T y)^3 + \cdots$$

  Use Proposition 1.
- Gaussian RBF kernel:

$$\exp\Big(-\frac{1}{2\sigma^2}\|x-y\|^2\Big) = \exp\Big(-\frac{\|x\|^2}{2\sigma^2}\Big)\exp\Big(\frac{x^T y}{\sigma^2}\Big)\exp\Big(-\frac{\|y\|^2}{2\sigma^2}\Big).$$

  Apply Proposition 4.
- Laplacian kernel: The proof is shown later.
- Polynomial kernel: Just sum and product.

$\square$

Vector space with inner product I

Definition. V: vector space over a field $\mathbb{K} = \mathbb{R}$ or $\mathbb{C}$.
V is called an inner product space if it has an inner product (or scalar product, dot product) $(\cdot,\cdot) : V \times V \to \mathbb{K}$ such that for every $x, y, z \in V$

1. (Strong positivity) $(x, x) \geq 0$, and $(x, x) = 0$ if and only if $x = 0$,
2. (Addition) $(x + y, z) = (x, z) + (y, z)$,
3. (Scalar multiplication) $(\alpha x, y) = \alpha(x, y)$ ($\forall \alpha \in \mathbb{K}$),
4. (Hermitian) $(y, x) = \overline{(x, y)}$.

Vector space with inner product II

$(V, (\cdot, \cdot))$: inner product space.

Norm of $x \in V$:

$$\|x\| = (x, x)^{1/2}.$$

Metric between x and y:

$$d(x, y) = \|x - y\|.$$

Theorem 5

Cauchy-Schwarz inequality

$$|(x, y)| \leq \|x\| \|y\|.$$

Remark: Cauchy-Schwarz inequality holds without requiring $\|x\| = 0 \Rightarrow x = 0$.

Hilbert space I

Definition. A vector space with inner product $(\mathcal{H}, (\cdot, \cdot))$ is called Hilbert space if the induced metric is complete, *i.e.* every Cauchy sequence[1] converges to an element in $\mathcal{H}$.

Remark 1:
A Hilbert space may be either finite or infinite dimensional.

Example 1.
$\mathbb{R}^n$ and $\mathbb{C}^n$ are finite dimensional Hilbert space with the ordinary inner product

$$(x, y)_{\mathbb{R}^n} = \sum_{i=1}^n x_i y_i \quad \text{or} \quad (x, y)_{\mathbb{C}^n} = \sum_{i=1}^n x_i \overline{y_i}.$$

[1] A sequence $\{x_n\}_{n=1}^{\infty}$ in a metric space (X, d) is called a Cauchy sequence if $d(x_n, x_m) \to 0$ for $n, m \to \infty$.

Hilbert space II

Example 2. $L^2(\Omega, \mu)$.
Let $(\Omega, \mathcal{B}, \mu)$ is a measure space.

$$\mathcal{L} = \left\{ f : \Omega \to \mathbb{C} \;\middle|\; \int |f|^2 d\mu < \infty \right\}.$$

The inner product on $\mathcal{L}$ is define by

$$(f, g) = \int f \overline{g} d\mu.$$

$L^2(\Omega, \mu)$ is defined by the equivalent classes identifying f and g if their values differ only on a measure-zero set.

- $L^2(\Omega, \mu)$ is complete. [See e.g. [Rud86] for the proof.]
- $L^2(\mathbb{R}^n, dx)$ is infinite dimensional.

Orthogonality

- Orthogonal complement.

  Let $\mathcal{H}$ be a Hilbert space and V be a closed subspace.

$$V^{\perp} := \{x \in \mathcal{H} \mid (x, y) = 0 \text{ for all } y \in V\}$$

  is a closed subspace, and called the orthogonal complement.

- Orthogonal projection.

  Let $\mathcal{H}$ be a Hilbert space and V be a closed subspace. Every $x \in \mathcal{H}$ can be uniquely decomposed

$$x = y + z, \qquad y \in V \quad \text{and} \quad z \in V^{\perp},$$

  that is,

$$\mathcal{H} = V \oplus V^{\perp}.$$

Complete orthonormal system I

- ONS and CONS.
  A subset $\{u_i\}_{i\in I}$ of $\mathcal{H}$ is called an orthonormal system (ONS) if $(u_i, u_j) = \delta_{ij}$ (δ_{ij} is Kronecker's delta).

  A subset $\{u_i\}_{i\in I}$ of $\mathcal{H}$ is called a complete orthonormal system (CONS) if it is ONS and if $(x, u_i) = 0$ ($\forall i \in I$) implies $x = 0$.

  Fact: Any ONS in a Hilbert space can be extended to a CONS.

- Separability
  A Hilbert space is separable if it has a countable CONS.

Assumption

In this course, a Hilbert space is always assumed to be separable.

Complete orthonormal system II

Theorem 6 (Fourier series expansion)

Let $\{u_i\}_{i=1}^{\infty}$ be a CONS of a separable Hilbert space. For each $x \in \mathcal{H}$,

$$x = \sum_{i=1}^{\infty} (x, u_i) u_i, \qquad \textit{(Fourier expansion)}$$

$$\|x\|^2 = \sum_{i=1}^{\infty} |(x, u_i)|^2. \qquad \textit{(Parseval's equality)}$$

Proof omitted.

Example: CONS of $L^2([0\ 2\pi], dx)$

$$u_n(t) = \frac{1}{\sqrt{2\pi}} e^{\sqrt{-1}nt} \qquad (n = 0, 1, 2, \ldots)$$

Then,

$$f(t) = \sum_{n=0}^{\infty} a_n u_n(t)$$

is the (ordinary) Fourier expansion of a periodic function.

Bounded operator I

Let $\mathcal{H}_1$ and $\mathcal{H}_2$ be Hilbert spaces. A linear transform $T : \mathcal{H}_1 \to \mathcal{H}_2$ is often called operator.

Definition. A linear operator $\mathcal{H}_1$ and $\mathcal{H}_2$ is called bounded if

$$\sup_{\|x\|_{\mathcal{H}_1}=1} \|Tx\|_{\mathcal{H}_2} < \infty.$$

The operator norm of a bounded operator T is defined by

$$\|T\| = \sup_{\|x\|_{\mathcal{H}_1}=1} \|Tx\|_{\mathcal{H}_2} = \sup_{x \neq 0} \frac{\|Tx\|_{\mathcal{H}_2}}{\|x\|_{\mathcal{H}_1}}.$$

(Corresponds to the largest singular value of a matrix.)

Fact. If $T : \mathcal{H}_1 \to \mathcal{H}_2$ is bounded,

$$\|Tx\|_{\mathcal{H}_2} \leq \|T\| \|x\|_{\mathcal{H}_1}.$$

Bounded operator II

Proposition 7

A linear operator is bounded if and only if it is continuous.

Proof. Assume $T : \mathcal{H}_1 \to \mathcal{H}_2$ is bounded. Then,

$$\|Tx - Tx_0\| \leq \|T\| \|x - x_0\|$$

means continuity of T.

Assume T is continuous. For any $\varepsilon > 0$, there is $\delta > 0$ such that $\|Tx\| < \varepsilon$ for all $x \in \mathcal{H}_1$ with $\|x\| < 2\delta$.
Then,

$$\sup_{\|x\|=1} \|Tx\| = \sup_{\|x\|=\delta} \frac{1}{\delta} \|Tx\| \leq \frac{\varepsilon}{\delta}.$$

□

Riesz lemma I

Definition. A linear functional is a linear transform from $\mathcal{H}$ to $\mathbb{C}$ (or $\mathbb{R}$).

The vector space of all the bounded (continuous) linear functionals called the dual space of $\mathcal{H}$, and is denoted by $\mathcal{H}^*$.

Theorem 8 (Riesz lemma)

For each $\phi \in \mathcal{H}^*$, there is a unique $y_\phi \in \mathcal{H}$ such that

$$\phi(x) = (x, y_\phi) \qquad (\forall x \in \mathcal{H}).$$

Proof.

Consider the case of $\mathbb{R}$ for simplicity.

$\Leftarrow$) Obvious by Cauchy-Schwartz.

Riesz lemma II

$\Rightarrow$) If $\phi(x) = 0$ for all x, take $y = 0$. Otherwise, let

$$V = \{x \in \mathcal{H} \mid \phi(x) = 0\}.$$

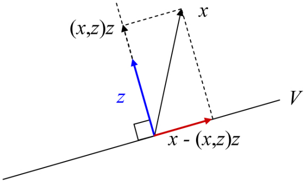
Since ϕ is a bounded linear functional, V is a closed subspace, and $V \neq \mathcal{H}$. Take $z \in V^\perp$ with $\|z\| = 1$. By orthogonal decomposition, for any $x \in \mathcal{H}$,

$$x - (x,z)z \in V.$$

Apply ϕ, then

$$\phi(x) - (x,z)\phi(z) = 0, \qquad \text{i.e.,} \quad \phi(x) = (x, \phi(z)z).$$

Take $y_\phi = \phi(z)z$. □



Reproducing kernel Hilbert space I

Definition.
Let $\mathcal{X}$ be a set. A reproducing kernel Hilbert space (RKHS) (over $\mathcal{X}$) is a Hilbert space $\mathcal{H}$ consisting of functions on $\mathcal{X}$ such that for each $x \in \mathcal{X}$ there is a function $k_x \in \mathcal{H}$ with the property

$$\langle f, k_x \rangle_{\mathcal{H}} = f(x) \qquad (\forall f \in \mathcal{H}) \qquad \text{(reproducing property)}.$$

$k(\cdot, x) := k_x(\cdot)$ is called a reproducing kernel of $\mathcal{H}$.

Fact 1. A reproducing kernel is Hermitian (symmetric).

Proof.

$$k(y,x) = \langle k(\cdot,x), k_y \rangle = \langle k_x, k_y \rangle = \overline{\langle k_y, k_x \rangle} = \overline{\langle k(\cdot,y), k_x \rangle} = \overline{k(x,y)}.$$

□

Fact 2. The reproducing kernel is unique, if exists. [Exercise]

Positive definite kernel and RKHS I

Proposition 9 (RKHS ⇒ positive definite kernel)

The reproducing kernel of a RKHS is positive definite.

Proof.

$$\begin{aligned}\textstyle\sum_{i,j=1}^n c_i \overline{c_j} k(x_i, x_j) &= \textstyle\sum_{i,j=1}^n c_i \overline{c_j} \langle k(\cdot, x_i), k(\cdot, x_j)\rangle \\ &= \textstyle\langle \sum_{i=1}^n c_i k(\cdot, x_i), \sum_{j=1}^n c_j k(\cdot, x_j)\rangle \geq 0\end{aligned}$$

Positive definite kernel and RKHS II

Theorem 10 (positive definite kernel ⇒ RKHS. Moore-Aronszajn)

Let $k : \mathcal{X} \times \mathcal{X} \to \mathbb{C}$ (or $\mathbb{R}$) be a positive definite kernel on a set $\mathcal{X}$. Then, there uniquely exists a RKHS $\mathcal{H}_k$ on $\mathcal{X}$ such that

1. $k(\cdot, x) \in \mathcal{H}_k$ *for every* $x \in \mathcal{X}$,
2. $\mathrm{Span}\{k(\cdot, x) \mid x \in \mathcal{X}\}$ *is dense in* $\mathcal{H}_k$,
3. k *is the reproducing kernel on* $\mathcal{H}_k$, *i.e.*
$$\langle f, k(\cdot, x)_{\mathcal{H}} \rangle = f(x) \qquad (\forall x \in \mathcal{X}, \forall f \in \mathcal{H}_k).$$

Positive definite kernel and RKHS III

One-to-one correspondence between positive definite kernels and RKHS.

$$k \qquad \longleftrightarrow \qquad \mathcal{H}_k$$

- Proposition 9: RKHS $\mapsto$ positive definite kernel k.
- Theorem 10: $k \mapsto \mathcal{H}_k$ (injective).

RKHS as a feature space

If we define

$$\Phi : \mathcal{X} \to \mathcal{H}_k, \quad x \mapsto k(\cdot, x),$$

then,

$$\langle \Phi(x), \Phi(y) \rangle = \langle k(\cdot, x), k(\cdot, y) \rangle = k(x, y).$$

RKHS associated with a positive definite kernel k gives a desired feature space!!

Another characterization

Proposition 11

Let $\mathcal{H}$ be a Hilbert space consisting of functions on a set $\mathcal{X}$. Then, $\mathcal{H}$ is a RKHS if and only if the evaluation map

$$e_x : \mathcal{H} \to \mathbb{K}, \qquad e_x(f) = f(x),$$

is a continuous linear functional for each $x \in \mathcal{X}$.

Proof. Assume $\mathcal{H}$ is a RKHS. The boundedness of e_x is obvious from

$$|e_x(f)| = |\langle f, k_x \rangle| \leq \|k_x\| \|f\|.$$

Conversely, assume the evaluation map is continuous. By Riesz lemma, there is $k_x \in \mathcal{H}$ such that

$$\langle f, k_x \rangle = e_x(f) = f(x),$$

which means $\mathcal{H}$ is a RKHS with k_x a reproducing kernel. □

Some properties of RKHS

The functions in a RKHS are "nice" functions under some conditions.

Proposition 12

Let k be a positive definite kernel on a topological space $\mathcal{X}$, and $\mathcal{H}_k$ be the associated RKHS. If $\mathrm{Re}[k(y,x)]$ is continuous for every $x, y \in \mathcal{X}$, then all the functions in $\mathcal{H}_k$ are continuous.

Proof. Let f be an arbitrary function in $\mathcal{H}_k$.

$$|f(x) - f(y)| = |\langle f, k(\cdot, x) - k(\cdot, y)\rangle| \leq \|f\| \|k(\cdot, x) - k(\cdot, y)\|.$$

The assertion is easy from

$$\|k(\cdot, x) - k(\cdot, y)\|^2 = k(x, x) + k(y, y) - 2\mathrm{Re}[k(x, y)].$$

□

Remark. It is also known ([BTA04]) that if $k(x, y)$ is differentiable, then all the functions in $\mathcal{H}_k$ are differentiable.

c.f. L^2 space contains non-continuous functions.

Proof of Theorem 10

Proof. (Described in $\mathbb{R}$ case.)

- Construction of an inner product space:

$$H_0 := \mathrm{Span}\{k(\cdot, x) \mid x \in \mathcal{X}\}.$$

Define an inner product on H_0:
for $f = \sum_{i=1}^{n} a_i k(\cdot, x_i)$ and $g = \sum_{j=1}^{m} b_j k(\cdot, y_j)$,

$$\langle f, g \rangle := \sum_{i=1}^{n} \sum_{j=1}^{m} a_i b_j k(x_i, y_j).$$

This is independent of the way of representing f and g from the expression

$$\langle f, g \rangle = \sum_{j=1}^{m} b_j f(y_j) = \sum_{i=1}^{n} a_i g(x_i).$$

- Reproducing property on H_0:

$$\langle f, k(\cdot,x)\rangle = \sum_{i=1}^{n} a_i k(x_i, x) = f(x).$$

- Well-defined as an inner product:
It is easy to see $\langle \cdot, \cdot \rangle$ is bilinear form, and

$$\|f\|^2 = \sum_{i,j=1}^{n} a_i a_j k(x_i, x_j) \geq 0$$

by the positive definiteness of f.
If $\|f\| = 0$, from Cauchy-Schwarz inequality,[2]

$$|f(x)| = |\langle f, k(\cdot,x)\rangle| \leq \|f\|\|k(\cdot,x)\| = 0$$

for all $x \in \mathcal{X}$; thus $f = 0$.

[2]Note that Cauchy-Schwarz inequality holds without assuming strong positivity of the inner product.

- Completion:
  Let $\mathcal{H}$ be the completion of H_0.
  - H_0 is dense in $\mathcal{H}$ by the completion.
  - $\mathcal{H}$ is realized by functions:
    Let $\{f_n\}$ be a Cauchy sequence in $\mathcal{H}$. For each $x \in \mathcal{X}$, $\{f_n(x)\}$ is a Cauchy sequence, because

$$|f_n(x) - f_m(x)| = |\langle f_n - f_m, k(\cdot, x)\rangle| \le \|f_n - f_m\| \|k(\cdot, x)\|.$$

    Define $f(x) = \lim_n f_n(x)$.
    This value is the same for equivalent sequences, because $\{f_n\} \sim \{g_n\}$ implies

$$|f_n(x) - g_n(x)| = |\langle f_n - g_n, k(\cdot, x)\rangle| \le \|f_n - g_n\| \|k(\cdot, x)\| \to 0.$$

    Thus, any element $[\{f_n\}]$ in $\mathcal{H}$ can be regarded as a function f on $\mathcal{X}$.

RKHS of polynomial kernel

Polynomial kernel on $\mathbb{R}$:

$$k(x,y) = (xy + c)^d \qquad (c > 0, d \in \mathbb{N}).$$

Proposition 13

$\mathcal{H}_k$ is $d+1$ dimensional vector space with a basis $\{1, x, x^2, \ldots, x^d\}$.

Proof. Omitted. Hint: Use

$$k(x,z) = z^d x^d + \binom{d}{1} c z^{d-1} x^{d-1} + \binom{d}{2} c^2 z^{d-2} x^{d-2} + \cdots + \binom{d}{d-1} c^{d-1} zx + c^d.$$

RKHS as a Hilbertian subspace

- $\mathcal{X}$: set.
- $\mathbb{C}^{\mathcal{X}}$: all functions on $\mathcal{X}$ with the pointwise-convergence topology[3].
- $\mathcal{G} = L^2(\mathcal{T}, \mu)$, where $(\mathcal{T}, \mathcal{B}, \mu)$ is a measure space.
- Suppose

$$H(\cdot; x) \in L^2(\mathcal{T}, \mu) \qquad \text{for all } x \in \mathcal{X}.$$

- Construct a continuous embedding

$$j : L^2(\mathcal{T}, \mu) \to \mathbb{C}^{\mathcal{X}},$$

$$F \mapsto f(x) = \int F(t)\overline{H(t; x)} d\mu(t) = (F, H(\cdot; x))_{\mathcal{G}}.$$

- Assume $\mathrm{Span}\{H(t; x) \mid x \in \mathcal{X}\}$ is dense in $L^2(\mathcal{T}, \mu)$. Then, j is injective.

[3] $f_n \to f \Leftrightarrow f_n(x) \to f(x)$ for every x.

RKHS as a Hilbertian subspace II

- Define $\mathcal{H} := \mathrm{Im} j$.
- Define an inner product on $\mathcal{H}$ by

$$\langle f, g \rangle_{\mathcal{H}} := (F, G)_{\mathcal{G}} \qquad \text{where} \quad f = j(F), g = j(G).$$

- We have $j : L^2(\mathcal{T}, \mu) \cong \mathcal{H}$ (isomorphic) as Hilbert spaces, and

$$\mathcal{H} = \left\{ f \in \mathbb{C}^{\mathcal{X}} \;\middle|\; \exists F \in L^2(\mathcal{T}, \mu), f(x) = \int F(t)\overline{H(t;x)} d\mu(t) \right\}.$$

Proposition 14

$\mathcal{H}$ is a RKHS, and its reproducing kernel is

$$k(x, y) = \langle j(H(\cdot; x)), j(H(\cdot; y)) \rangle_{\mathcal{H}} = \int H(t; x)\overline{H(t; y)} d\mu(t).$$

Proof.

$$f(x) = (F, H(\cdot, x))_{\mathcal{G}} = \langle f, j(H(\cdot, x)) \rangle_{\mathcal{H}}.$$

□

Explicit realization of RKHS by Fourier transform

Special case given by Fourier transform.

- $\mathcal{X} = \mathcal{T} = \mathbb{R}$.
- $\mathcal{G} = L^2(\mathbb{R}, \rho(t)dt)$. $\rho(t)$: continuous, $\rho(t) > 0$, $\int \rho(t)dt < \infty$.
- $H(t;x) = e^{-\sqrt{-1}xt}$.
  Note: $\mathrm{Span}\{H(t;x) \mid x \in \mathcal{X}\}$ is dense $L^2(\mathbb{R}, \rho(t)dt)$.

- Fact.

$$\mathcal{H} = \left\{ f \in L^2(\mathbb{R}, dx) \;\middle|\; \int \frac{|\hat{f}(t)|^2}{\rho(t)} dt < \infty \right\}.$$

$$\langle f, g \rangle_{\mathcal{H}} = \int \frac{\hat{f}(t)\overline{\hat{g}(t)}}{\rho(t)} dt.$$

$$k(x,y) = \int e^{-\sqrt{-1}(x-y)t} \rho(t) dt.^{4}$$

[4] We can directly confirm this a positive definite kernel.

Explicit realization of RKHS by Fourier transform II

Proof. Let $f = j(F)$. By definition,

$$f(x) = \int F(t) e^{\sqrt{-1}tx} \rho(t) dt. \qquad \text{(Fourier transform)}$$

Since $F(t)\rho(t) \in L^1(\mathbb{R}, dt) \cap L^2(\mathbb{R}, dt)$[5], the Fourier isometry of $L^2(\mathbb{R}, dt)$ tells

$$f(x) \in L^2(\mathbb{R}, dx) \quad \text{and} \quad \hat{f}(t) = \frac{1}{2\pi} \int f(x) e^{-\sqrt{-1}xt} dx = F(t)\rho(t).$$

Thus,

$$F(t) = \frac{\hat{f}(t)}{\rho(t)}.$$

By the definition of the inner product, for $f = j(F)$ and $g = j(G)$,

$$\langle f, g \rangle_{\mathcal{H}} = (F, G)_{\mathcal{G}} = \int \frac{\hat{f}(t)}{\rho(t)} \frac{\overline{\hat{g}(t)}}{\rho(t)} \rho(t) dt = \int \frac{\hat{f}(t)\overline{\hat{g}(t)}}{\rho(t)} dt.$$

In addition,

$$F \in L^2(\mathbb{R}, \rho(t)dt) \quad \Leftrightarrow \quad \frac{\hat{f}(t)}{\rho(t)} \in L^2(\mathbb{R}, \rho(t)dt) \quad \Leftrightarrow \quad \int \frac{|\hat{f}(t)|^2}{\rho(t)} dt < \infty.$$

[5] Because $\rho(t)$ is bounded, $F \in L^2(\mathbb{R}, \rho(t)dt)$ means $|F(t)|^2 \rho(t)^2 \in L^1(\mathbb{R}, dt)$

Explicit realization of RKHS by Fourier transform III

Examples.

- Gaussian RBF kernel: $k(x,y) = \exp\{-\frac{1}{2\sigma^2}|x-y|^2\}$.
  - Let $\rho(t) = \frac{1}{2\pi}\exp\{-\frac{\sigma^2}{2}t^2\}$,

    i.e. $$\mathcal{G} = L^2(\mathbb{R}, \frac{1}{2\pi}e^{-\frac{\sigma^2}{2}t^2}dt).$$

  - Reproducing kernel = Gaussian RBF kernel:

$$k(x,y) = \frac{1}{2\pi}\int e^{\sqrt{-1}(x-y)t}e^{-\frac{\sigma^2}{2}t^2}dt = \frac{1}{\sigma}\exp\Big(-\frac{1}{2\sigma^2}(x-y)^2\Big)$$

$$\mathcal{H} = \Big\{ f \in L^2(\mathbb{R}, dx) \;\Big|\; \int |\hat{f}(t)|^2 \exp(\frac{\sigma^2}{2}t^2)dt < \infty \Big\}.$$

$$\langle f, g\rangle = \int \hat{f}(t)\overline{\hat{g}(t)}\exp(\frac{\sigma^2}{2}t^2)dt$$

Explicit realization of RKHS by Fourier transform IV

- Laplacian kernel: $k(x,y) = \exp\{-\beta|x-y|\}$.
  - Let $\rho(t) = \frac{1}{2\pi}\frac{1}{t^2+\beta^2}$,

    i.e. $\mathcal{G} = L^2(\mathbb{R}, \frac{dt}{2\pi(t^2+\beta^2)})$.
  - Reproducing kernel $=$ Laplacian kernel:

$$k(x,y) = \frac{1}{2\pi}\int e^{\sqrt{-1}(x-y)t}\frac{1}{t^2+\beta^2}dt = \frac{1}{2\beta}\exp\Big(-\beta|x-y|\Big)$$

[Note: the Fourier image of $\exp(|x-y|)$ is $\frac{1}{2\pi(t^2+1)}$.]

$$\mathcal{H} = \Big\{f \in L^2(\mathbb{R}, dx) \;\Big|\; \int |\hat{f}(t)|^2(t^2+\beta^2)dt < \infty\Big\}.$$

$$\langle f, g\rangle = \int \hat{f}(t)\overline{\hat{g}(t)}(t^2+\beta^2)dt$$

Summary of Sections 1 and 2

- We would like to use a feature vector $\Phi : \mathcal{X} \to \mathcal{H}$ to incorporate high order moments.
- The inner product in the feature space must be computed efficiently. Ideally,

$$\langle \Phi(x), \Phi(y) \rangle = k(x, y).$$

- To satisfy the above relation, the kernel k must be positive definite.
- A positive definite kernel k defines an associated RKHS, where k is the reproducing kernel;

$$\langle k(\cdot, x), k(\cdot, y) \rangle = k(x, y).$$

- Use a RKHS as a feature space, and $\Phi : x \mapsto k(\cdot, x)$ as the feature map.

References

A good reference on Hilbert (and Banach) space is [Rud86]. A more advanced one on functional analysis is [RS80] among many others. For reproducing kernel Hilbert spaces, the original paper is [Aro50]. Statistical aspects are discussed in [BTA04].

[Aro50] Nachman Aronszajn.
Theory of reproducing kernels.
Transactions of the American Mathematical Society, 68(3):337–404, 1950.

[BTA04] Alain Berlinet and Christine Thomas-Agnan.
Reproducing kernel Hilbert spaces in probability and statistics.
Kluwer Academic Publisher, 2004.

[RS80] Michael Reed and Barry Simon.
Functional Analysis.
Academic Press, 1980.

[Rud86] Walter Rudin.
Real and Complex Analysis (3rd ed.).
McGraw-Hill, 1986.