Basics of RKHS and beyond

TT Nguyen

PRaDA@Deakin

March 21, 2019

Contents

Outline

Overview

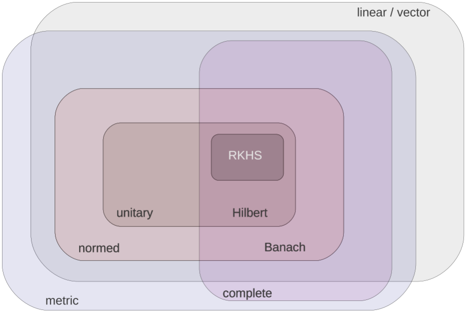

Figure: An overview of spaces [D. Sejdinovic, 2014]. Note that the rkhs part of this slide is heavily based on [D. Sejdinovic, 2014]

Hilbert space in a nutshell

A Hilbert space $\mathcal{H}$ has the following characteristics:

- Vector space over field $\mathbb{R}$ (or $\mathcal{C}$).
- Inner product (thus, norm and metric).
- Complete wrt its metric.
- Boundedness of a linear operator in $\mathcal{H}$ imples an continuous operator.
- All continuous linear functionals defined on $\mathcal{H}$ arise from inner product (Riesz representation).

Dual space of a normed space

Definition ((Topological) dual space)

Let $\mathcal{F}$ be a normed space, then

$$\mathcal{F}' := \{T : \mathcal{F} \to \mathbb{R} | T \text{ is a continuous linear operator (functional)}\}$$

is called the topological dual space of $\mathcal{F}$.

Theorem (Riesz representation)

Let $\mathcal{F}$ be a Hilbert space, then

$$\forall T \in \mathcal{F}', \exists! f \in \mathcal{F}, T(\cdot) = \langle \cdot, f \rangle_{\mathcal{F}}$$

Isometric isomorphism

Definition (Hilbert space isomorphism)

Two Hilbert spaces $\mathcal{H}$ and $\mathcal{F}$ are isometrically isomorphic if there is a linear bijective map $U : \mathcal{H} \to \mathcal{F}$ which preserves the inner product. i.e., $\langle h1, h2\rangle_{\mathcal{H}} = \langle Uh1, Uh2\rangle_{\mathcal{F}}$

Theorem

The dual space $\mathcal{H}'$ of a Hilbert space $\mathcal{H}$ is another Hilbert space. In addition, they are isometrically isomorphic.

Outline

RHKS in a nutshell

- A special space of well-behaved functions in $\mathbb{R}^{\mathcal{X}}$: If two functions are close in norm, they are close point-wise.
- Every RKHS is associated with a unique reproducing kernel. In reverse, every positive definite function $\mathcal{X} \times \mathcal{X} \rightarrow \mathbb{R}$ is associated with a unique RKHS.
- Pre-RKHS $\mathcal{H}_0$ is a useful *representative* of RKHS $\mathcal{H}$ in the sense that in order to prove some property that is preserved over limit in $\mathcal{H}$, we only need to prove it in $\mathcal{H}_0$. This is possible due to that $\mathcal{H}_0$ is *dense* in $\mathcal{H}$.

Dirac evaluation functional

Definition (Evaluation functional)

Let $\mathcal{H} \in \mathbb{R}^{\mathcal{X}}$ be a Hilbert space of functions $f : \mathcal{X} \to \mathbb{R}$, and

$$\forall x \in \mathcal{X}, \delta_x : \mathcal{H} \to \mathbb{R}, f \mapsto f(x)$$

δ_x is called the Dirac evaluation functional at x.

Theorem

$\forall x \in \mathcal{X}, \delta_x$ is linear, but not necessarily continuous; thus, $\{\delta_x : x \in \mathcal{X}\}$ is not necessarily a subspace of the dual space $\mathcal{H}'$.

Reproducing kernel Hilbert space

Definition (Reproducing kernel Hilbert space)

A Hilbert space $\mathcal{H}$ of functionals $f : \mathcal{X} \to \mathbb{R}$ is a reproducing kernel Hilbert space (RKHS) if $\forall x \in \mathcal{X}, \delta_x \in \mathcal{H}'$.

Theorem (Norm convergence implies pointwise convergence)

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

Reproducing kernels

Definition (Kernels)

A function $k : \mathcal{X} \times \mathcal{X} \to \mathbb{R}$ is a kernel on $\mathcal{X}$ iff there exists a Hilbert space $\mathcal{H}$ and a feature map $\phi : \mathcal{X} \to \mathcal{H}$ s.t. $k(x, y) = \langle \phi(x), \phi(y) \rangle_{\mathcal{H}}$.

Definition (Reproducing kerneks of a Hilbert space)

Let $\mathcal{H}$ be a Hilbert space of functions $f : \mathcal{X} \to \mathbb{R}$. A function $k : \mathcal{X} \times \mathcal{X} \to \mathbb{R}$ is a reproducing kernel of $\mathcal{H}$ if:

- $\forall x \in \mathcal{X}, k_x := k(\cdot, x) \in \mathcal{H}$
- $\forall x \in \mathcal{X}, \forall f \in \mathcal{H}, f(x) = \langle f, k_x \rangle_{\mathcal{H}}$

Reproducing kernels (con't)

Theorem

A reproducing kernel is a kernel, but the reverse is not nessarily true.

Theorem

If exists, reproducing kernel is unique.

Theorem

$\mathcal{H}$ is a RKHS iff it has a reproducing kernel.

Moore-Aronszajn Theorem

Theorem (Moore-Aronszajn - Part I)

Let $k : \mathcal{X} \times \mathcal{X} \to \mathbb{R}$ be positive definite, then there exits a unique RKHS $\mathcal{H} \in \mathcal{R}^{\mathcal{X}}$ with reproducing kernel k

Theorem (Moore-Aronszajn - Part II)

The pre-RKHS $\mathcal{H}_0$ is *dense* in a RKHS $\mathcal{H} \in \mathbb{R}^{\mathcal{X}}$ where $\mathcal{H}_0 := span\{k(\cdot, x) : x \in \mathcal{X}\}$ with the inner product defined as:

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

and $f(\cdot) = \sum_{i=1}^{n} \alpha_i k(\cdot, x_i) \in \mathcal{H}_0, g(\cdot) = \sum_{j=1}^{m} \alpha_j k(\cdot, y_i) \in \mathcal{H}_0$

Constructing RKHS from pre-RKHS

Theorem (RKHS from pre-RKHS)

Define $\mathcal{H}$ from $\mathcal{H}_0$ as follows:

- $\mathcal{H} := \{f \in \mathbb{R}^{\mathcal{X}} : \exists$ a Cauchy sequence $\{f_n\} \in \mathcal{H}_0$ converges pointwise to $f\}$
- $\forall f, g \in \mathcal{H}, \langle f, g\rangle_{\mathcal{H}} := \lim_{n\to\infty}\langle f_n, g_n\rangle_{\mathcal{H}_0}$

Then, $\mathcal{H}$ is a well-defined RKHS with reproducing kernel k.

Outline

Lipschitz continuity

- Lipschitz coninuity limits how fast the function can change.
- Every function that has bounded first partial derivatives is Lipschitz [1], i.e., if

$$\left|\frac{\partial f(\boldsymbol{x})}{\partial x_j}\right| < L, \forall \boldsymbol{x} \in D \subset \mathbb{R}^n, \forall j \tag{1}$$

where $0 \leq L \leq \infty$, then

$$\forall \boldsymbol{x}, \boldsymbol{y} \in D, |f(\boldsymbol{x}) - f(\boldsymbol{y})| \leq L\|\boldsymbol{x} - \boldsymbol{y}\|_1. \tag{2}$$

[1] Proof: similar idea to https://math.stackexchange.com/questions/1257553/a-multivariate-function-with-bounded-partial-derivatives-is-lipschitz.

Union bound

Consider the probability space $(\Omega, \mathcal{F}, P)$, and any sequence of events $A_t \in \mathcal{F}$ for $1 \leq t \leq \infty$, we have:

$$P(\{A_t, \forall t\}) \geq 1 - \sum_t P(\bar{A}_t) \tag{3}$$

Bound on the standard normal CDF

Consider $r \sim \mathbb{N}(0,1)$ and $\forall c$, we have

$$P(\{r \leq c\}) \geq 1 - 0.5\exp(-\frac{c^2}{2}). \tag{4}$$

Discretization of a compact space

Consider $D \subset [0, r]^d$, $\tilde{D}$ is a finitely discretized version of D such that for each point $\tilde{\boldsymbol{x}} \in \tilde{D}$, the coordinate along dimension j, $\tilde{x}_j$ is one of the points acquired by dividing $[0, r]$ along dimension j into τ uniformly spaced points. Then we have:

$$\|\boldsymbol{x} - [\boldsymbol{x}]\|_1 \leq d\frac{r}{\tau}, \forall \boldsymbol{x} \in D \tag{5}$$

where $[\boldsymbol{x}]$ is the closest point to $\boldsymbol{x}$ in $\tilde{D}$.

D. Sejdinovic, A. G. (2014).
Foundations of reproducing kernel hilbert spaces (lecture note in advanced topics in machine learning).