High-Dimensional Statistics: A Non-Asymptotic Viewpoint

Chapter 12: Reproducing kernel Hilbert spaces

Notes by Renxiong Liu

Many problems in statistics–among them interpolation, regression and density estimation, as well as nonparametric forms of dimension reduction and testing–involve optimizing over function spaces, where Hilbert spaces enjoys the favor that they could include a broad class of functions and share similar geometric structure with the Euclidean space. In this chapter, we will focus on a special Hilbert space that are defined by the reproducing kernels–reproducing kernel Hilbert spaces (RKHSs).

12.1 Basics of Hilbert spaces

Definition 12.1. Let V be a vector space, we call $\langle \cdot, \cdot \rangle_V : V \times V \to \mathbb{R}$ an inner product over V, if

- $\langle f, g \rangle_V = \langle g, f \rangle_V$, for any $f, g \in V$.
- $\langle f, f \rangle_V \geq 0$, with *equality* if and only if $f = 0 \in V$.
- $\langle \alpha f + \beta g, h \rangle_V = \alpha \langle f, h \rangle_V + \beta \langle g, h \rangle_V$, for any $f, g, h \in V$ and $\alpha, \beta \in \mathbb{R}$.

We call $(V, \langle \cdot, \cdot \rangle_V)$ an **inner product space** if V is a vector space with inner product $\langle \cdot, \cdot \rangle_V$. Given $(V, \langle \cdot, \cdot \rangle_V)$, we define the induced norm over V with the inner product $\langle \cdot, \cdot \rangle_V$: $\|f\|_V = \sqrt{\langle f, f \rangle_V}$, and call $(V, \|\cdot\|_V)$ (or in short V) corresponding **normed space**.

To introduce the notion of Hilbert space, we first need to know the concept of *Cauchy sequence*. A sequence $\{f_n\}_{n=1}^{\infty} \subseteq V$ is call a Cauchy sequence in $(V, \|\cdot\|_V)$, if for any $\epsilon > 0$, there exists $N \in \mathbb{N}$, such that for any $n > N$ and $l \geq 1$, we have that

$$\|f_{n+l} - f_n\|_V < \epsilon \tag{12.1}$$

Now we have the following notion of Hilbert space:

Definition 12.2. An inner product space $(H, \langle \cdot, \cdot \rangle_H)$ is called a Hilbert space, if any Cauchy sequence $\{f_n\}_{n=1}^{\infty} \subseteq H$ will converge to some element in H.

The property that any any Cauchy sequence will converge to some element in vector space is call *completeness*, hence a Hilbert space is a complete inner product space. Some classic examples are given below.

Example 12.1. Given set

$$l^2(N) = \{(\theta_i)_{i=1}^{\infty} \mid \sum_{i=1}^{\infty} \theta_i^2 < \infty\} \tag{12.2}$$

and define the inner product as $\langle s, t \rangle_{l^2(N)} = \sum_{i=1}^{\infty} s_i t_i$, then the sequence space $(l^2(N), \langle \cdot, \cdot \rangle_{l^2 N})$ is a Hilbert space.

Example 12.2. Given set

$$L^2[0,1] = \{f : [0,1] \to \mathbb{R} \mid \|f\|^2_{L^2[0,1]} = \int_0^1 f^2(x)dx < \infty\} \tag{12.3}$$

and define the inner product $\langle f, g\rangle_{L^2[0,1]} = \int_0^1 f(x)g(x)dx$, then space $(L^2[0,1], \langle\cdot,\cdot\rangle_{L^2[0,1]})$ is a Hilbert space. Moreover, denote by $\{e_i\}_{i=1}^\infty$ a set of complete orthonormal basis in $L^2[0,1]$ (existence ensured by Gram-Schimidt orthogonalization), then for any $f \in L^2[0,1]$, we can represent f by $f = \sum_{i=1}^\infty a_i e_i$ for some $\{a_i\}_{i=1}^\infty$, where $a_i = \langle f, e_i\rangle_{L^2[0,1]}$. By Parseval's theorem, it holds that

$$\|f\|_{L^2[0,1]} = \sum_{i=1}^\infty a_i^2 \tag{12.4}$$

Hence $f \in L^2[0,1]$ if and only $\sum_{i=1}^\infty a_i^2 < \infty$, which means $f \leftrightarrow \{a_i\}_{i=1}^\infty$ defines a isomorphism between $L^2[0,1]$ and $l^2(N)$.

Both $l^2(N)$ and $L^2[0,1]$ are separable Hilbert spaces, for which there is a countable dense subset. For separable Hilber space, there must exist a collection of functions that are orthonormal to each other. Throughout this note, we focus on the separable Hilbert space, which enjoys the favor of such existence.

The last results we will introduce in this section is Riesz representer theorem. To see this result, we first introduce the notion of *linear functional*. A linear functional over Hilbert space $(H, \langle\cdot,\cdot\rangle_H)$ is a mapping $L : H \to \mathbb{R}$, such that it's linear. A linear functional is bounded if there exists $M > 0$ such that $|L(f)| < M\|f\|_H$ for any $f \in H$. Riesz representer theorem states that every bounded linear functional can be reprsented in form of the inner product.

Theorem 12.1 (Riesz representer theorem)**.** *Let L be a bounded linear functional over Hilbert space $(H, \langle\cdot,\cdot\rangle)$, then there exists a unique $g \in H$ such that $L(f) = \langle f, g\rangle$ for any $f \in H$. (For such $g \in H$, we call it the representer of functional L.)*

12.2 Reproducing kernel Hilbert spaces

In this section, we discuss the construction of RKHS. We first introduce the notion of positive semidefinite kernel functions.

Definition 12.3 (Positive semidefinite kernel function)**.** We call a symmetric bivariate function $\mathcal{K} : \mathcal{X} \times \mathcal{X} \to \mathbb{R}$ a positive semidefinite kernel function if for any $n \geq 1$ and $\{x_i\}_{i=1}^n \subseteq \mathcal{X}$, $n \times n$ matrix $K = (\mathcal{K}(x_i, x_j))_{1\leq i,j\leq n}$ is positive semidefinite.

Some examples of positive semidefinite kernel function are given as follows.

Example 12.3 (Linear kernels)**.** Let $\mathcal{X} = \mathbb{R}^d$, we define the linear kernel function as $\mathcal{K}(x, x') = \langle x, x'\rangle$. To see $\mathcal{K}$ is a positive semidefinite kernel, note for any $n \geq 1$, $\{x_k\}_{k=1}^n$ and $a \in \mathbb{R}^n$, we have that

$$a^\top K a = \sum_{i=1}^n \sum_{j=1}^n a_i a_j \mathcal{K}(x_i, x_j) = \sum_{i=1}^n \sum_{j=1}^n \langle a_i x_i, a_j x_j\rangle = \|a^\top x\|_2^2 \geq 0 \tag{12.5}$$

Example 12.4 (Gaussian kernels)**.** Let $\mathcal{X} \subseteq \mathbb{R}^d$ be some compact subset of R^d, define $\mathcal{K}(x, z) = \exp(-\frac{1}{2\sigma^2}\|x - z\|_2^2)$, we can show that $\mathcal{K}$ is a positive semidefinite kernel. To see this result, note the characteristic function of $N(0, \sigma^2 I_d)$ is

$$\phi(t) = \mathbb{E}e^{it^\top X} = \exp(-\frac{1}{2}\sigma^2\|t\|^2) \tag{12.6}$$

we have that $\mathcal{K}(x, z) = \phi(x - z) = \mathbb{E}e^{i(x-z)^\top X}$. Hence for any $n \geq 1$, $\{x_k\}_{k=1}^n$ and $a \in \mathbb{R}^n$, we have that

$$a^\top K a = \sum_{i=1}^n \sum_{j=1}^n a_i a_j \mathbb{E}e^{i(x_i - x_j)^\top X} = \mathbb{E}\sum_{i=1}^n \sum_{j=1}^n a_i a_j e^{ix_i^\top X}\overline{e^{ix_j^\top X}} = \mathbb{E}|\sum_{i=1}^n a_i e^{ix_i^\top X}|^2 \geq 0 \tag{12.7}$$