CS281B/Stat241B (Spring 2008) Statistical Learning Theory **Lecture: 7**

Reproducing Kernel Hilbert Spaces

Lecturer: Peter Bartlett *Scribe: Chunhui Gu*

1 Reproducing Kernel Hilbert Spaces

1.1 Hilbert Space and Kernel

An inner product $\langle u, v \rangle$ can be

1. a usual dot product: $\langle u, v \rangle = v'w = \sum_i v_i w_i$
2. a kernel product: $\langle u, v \rangle = k(v, w) = \psi(v)'\psi(w)$ (where $\psi(u)$ may have infinite dimensions)

However, an inner product $\langle \cdot, \cdot \rangle$ must satisfy the following conditions:

1. Symmetry
$$\langle u, v \rangle = \langle v, u \rangle \; \forall u, v \in \mathcal{X}$$
2. Bilinearity
$$\langle \alpha u + \beta v, w \rangle = \alpha \langle u, w \rangle + \beta \langle v, w \rangle \; \forall u, v, w \in \mathcal{X}, \forall \alpha, \beta \in \mathbb{R}$$
3. Positive definiteness
$$\langle u, u \rangle \geq 0, \; \forall u \in \mathcal{X}$$
$$\langle u, u \rangle = 0 \iff u = 0$$

Now we can define the notion of a Hilbert space.

Definition. A *Hilbert Space* is an inner product space that is complete and separable with respect to the norm defined by the inner product.

Examples of Hilbert spaces include:

1. The vector space $\mathbb{R}^n$ with $\langle a, b \rangle = a'b$, the vector dot product of a and b.
2. The space l_2 of square summable sequences, with inner product $\langle x, y \rangle = \sum_{i=1}^{\infty} x_i y_i$
3. The space L_2 of square integrable functions (i.e., $\int_s f(x)^2 dx < \infty$), with inner product $\langle f, g \rangle = \int_s f(x)g(x)dx$

Definition. $k(\cdot, \cdot)$ is a *reproducing kernel* of a Hilbert space $\mathcal{H}$ if $\forall f \in \mathcal{H}, f(x) = \langle k(x, \cdot), f(\cdot) \rangle$.

A Reproducing Kernel Hilbert Space (RKHS) is a Hilbert space H with a reproducing kernel whose span is dense in H. We could equivalently define an RKHS as a Hilbert space of functions with all evaluation functionals bounded and linear.

For instance, the L_2 space is a Hilbert space, but not an RKHS because the delta function which has the reproducing property

$$f(x) = \int_s \delta(x-u)f(u)du$$

does not satisfy the square integrable condition, that is,

$$\int_s \delta(u)^2 du \not< \infty,$$

thus the delta function is not in L_2.

Now let us define a kernel.

Definition. $k : \mathcal{X} \times \mathcal{X} \rightarrow \mathbb{R}$ is a *kernel* if

1. k is symmetric: $k(x,y) = k(y,x)$.
2. k is positive semi-definite, i.e., $\forall x_1, x_2, ..., x_n \in \mathcal{X}$, the "Gram Matrix" K defined by $K_{ij} = k(x_i, x_j)$ is positive semi-definite. (A matrix $M \in \mathbb{R}^{n\times n}$ is positive semi-definite if $\forall a \in \mathbb{R}^n$, $a'Ma \geq 0$.)

Here are some properties of a kernel that are worth noting:

1. $k(x,x) \geq 0$. (Think about the Gram matrix of $n = 1$)
2. $k(u,v) \leq \sqrt{k(u,u)k(v,v)}$. (This is the Cauchy-Schwarz inequality.)

To see why the second property holds, we consider the case when $n = 2$:

Let $a = \begin{bmatrix} k(v,v) \\ -k(u,v) \end{bmatrix}$. The Gram matrix $K = \begin{pmatrix} k(u,u) & k(u,v) \\ k(v,u) & k(v,v) \end{pmatrix} \succeq 0 \iff a'Ka \geq 0$

$\iff [k(v,v)k(u,u) - k(u,v)^2]k(v,v) \geq 0.$

By the first property we know $k(v,v) \geq 0$, so $k(v,v)k(u,u) \geq k(u,v)^2$.

1.2 Build an Reproducing Kernel Hilbert Space (RKHS)

Given a kernel k, define the "reproducing kernel feature map" $\Phi : \mathcal{X} \rightarrow \mathbb{R}^{\mathcal{X}}$ as:

$$\Phi(x) = k(\cdot, x)$$

Consider the vector space:

$$\text{span}(\{\Phi(x) : x \in \mathcal{X}\}) = \{f(\cdot) = \sum_{i=1}^{n} \alpha_i k(\cdot, x_i) : n \in \mathbb{N}, x_i \in \mathcal{X}, \alpha_i \in \mathbb{R}\}$$

For $f = \sum_i \alpha_i k(\cdot, u_i)$ and $g = \sum_i \beta_i k(\cdot, v_i)$, define $\langle f, g \rangle = \sum_{i,j} \alpha_i \beta_j k(u_i, v_j)$.

Note that:

$$\langle f, k(\cdot, x) \rangle = \sum_i \alpha_i k(x, u_i) = f(x), \text{ i.e., } k \text{ has the reproducing property.}$$

We show that $\langle f, g \rangle$ is an inner product by checking the following conditions:

1. Symmetry: $\langle f, g\rangle = \sum_{i,j} \alpha_i\beta_j k(u_i, v_j) = \sum_{i,j} \beta_j\alpha_i k(v_j, u_i) = \langle g, f\rangle$
2. Bilinearity: $\langle f, g\rangle = \sum_i \alpha_i g(u_i) = \sum_j \beta_j f(v_j)$
3. Positive definiteness: $\langle f, f\rangle = \alpha' K\alpha \geq 0$ with equality iff $f = 0$.

From 3 we can also derive:

1. $\langle f, g\rangle^2 \leq \langle f, f\rangle\langle g, g\rangle$

   PROOF. $\forall a \in \mathbb{R}, \langle af + g, af + g\rangle = a^2\langle f, f\rangle + 2a\langle f, g\rangle + \langle g, g\rangle \geq 0$. This implies that the quadratic expression has a non-positive discriminant. Therefore, $\langle f, g\rangle^2 - \langle f, f\rangle\langle g, g\rangle \leq 0$ □
2. $|f(x)|^2 = \langle k(\cdot, x), f\rangle^2 \leq k(x, x)\langle f, f\rangle$, which implies that if $\langle f, f\rangle = 0$ then f is identically zero.

Now we have defined an inner product space $\langle\cdot,\cdot\rangle$. Complete it to give the Hilbert space.

Definition. For a (compact) $\mathcal{X} \subseteq \mathbb{R}^d$, and a Hilbert space $\mathcal{H}$ of functions $f : \mathcal{X} \to \mathbb{R}$, we say $\mathcal{H}$ is a *Reproducing Kernel Hilbert Space* if $\exists k : \mathcal{X} \to \mathbb{R}$, s.t.

1. k has the reproducing property, i.e., $f(x) = \langle f(\cdot), k(\cdot, x)\rangle$
2. k spans $\mathcal{H} = \overline{\text{span}\{k(\cdot, x) : x \in \mathcal{X}\}}$

1.3 Mercer's Theorem

Another way to characterize a symmetric positive semi-definite kernel k is via the Mercer's Theorem.

Theorem 1.1 (Mercer's)**.** Suppose k is a continuous positive semi-definite kernel on a compact set $\mathcal{X}$, and the integral operator $T_k : L_2(\mathcal{X}) \to L_2(\mathcal{X})$ defined by

$$(T_k f)(\cdot) = \int_{\mathcal{X}} k(\cdot, x)f(x)dx$$

is positive semi-definite, that is, $\forall f \in L_2(\mathcal{X})$,

$$\int_{\mathcal{X}} k(u, v)f(u)f(v)dudv \geq 0$$

Then there is an orthonormal basis $\{\psi_i\}$ of $L_2(\mathcal{X})$ consisting of eigenfunctions of T_k such that the corresponding sequence of eigenvalues $\{\lambda_i\}$ are non-negative. The eigenfunctions corresponding to non-zero eigenvalues are continuous on $\mathcal{X}$ and $k(u, v)$ has the representation

$$k(u, v) = \sum_{i=1}^{\infty} \lambda_i\psi_i(u)\psi_i(v)$$

where the convergence is absolute and uniform, that is,

$$\lim_{n\to\infty} \sup_{u,v} |k(u, v) - \sum_{i=1}^{n} \lambda_i\psi_i(u)\psi_i(v)| = 0$$

To take an analogue in the finite case, that is, $\mathcal{X} = \{x_1, \ldots, x_n\}$. Let $K_{ij} = k(x_i, x_j)$, and $f : \mathcal{X} \to \mathbb{R}^n$ with $f_i = f(x_i)$. Then,

$$T_k f = \sum_{i=1}^{n} k(\cdot, x_i) f_i$$

$$\forall f,\ f'Kf \geq 0 \Rightarrow K \succeq 0 \Rightarrow K = \sum \lambda_i v_i v_i'$$

Hence,

$$k(x_i, x_j) = K_{ij} = (V \Lambda V')_{ij} = \sum_{k=1}^{n} \lambda_k v_{ki} v_{kj} = \sum_{k=1}^{n} \lambda_k \psi_k(x_i) \psi_k(x_j) \Rightarrow \psi_k(x_i) = (v_k)_i$$

We summarize several equivalent conditions on continuous, symmetric k defined on compact $\mathcal{X}$:

1. Every Gram matrix is positive semi-definite.
2. T_k is positive semi-definite.
3. k can be expressed as $k(u, v) = \sum_i \lambda_i \psi_i(u) \psi_i(v)$.
4. k is the reproducing kernel of an RKHS of functions on $\mathcal{X}$.