**Math 547: Mathematical Foundations of Statistical Learning Theory Fall 2017**

Lecture 10 — September 13

*Instructor: Stas Minsker* *Scribe: Mose Wintner*

# 10.1 RKHS

Given the feature map $\phi$, we have the kernel

$$K(x_1, x_2) = \langle \phi(x_1), \phi(x_2) \rangle$$

on $\mathbb{S} \times \mathbb{S}$. Given the kernel $K$ that is symmetric and nonnegative, we also have the feature map $\phi$.

## 10.1.1 A different view on RKHS

Let $\mathbb{H}$ be a separable Hilbert space of functions $f : \mathbb{S} \to \mathbb{R}$, where $\mathbb{S}$ is a metric space. Let $\mathbb{H}'$ be the dual space – the space of continuous linear functionals $\mathbb{H} \to \mathbb{R}$. Given $x \in \mathbb{S}$, let $\delta_x : \mathbb{H} \to \mathbb{R}$ be defined by $f \mapsto f(x)$.

**Definition 1.** The RKHS is a Hilbert space such that all point evaluation functionals $\delta_x$ are continuous/bounded for all $x \in \mathbb{S}$.

This implies by the Riesz representation theorem that for all $x \in \mathbb{S}$, there exists $h_x \in \mathbb{H}$ such that

$$\delta_x(f) = \langle h_x, f \rangle$$

for all $f \in \mathbb{H}$. Then we can define

$$K(x, y) = \langle h_x, h_y \rangle,$$

and we see that

$$f(x) = \langle f, K(x, \cdot) \rangle = \langle f, h_x \rangle.$$

## 10.1.2 Another view on RKHS

Let $\mathbb{S}$ be compact, and let $\nu$ be a finite measure on $\mathbb{S}$. Let $K$ be a symmetric positive semidefinite function $\mathbb{S}\times\mathbb{S} \to \mathbb{R}$ continuous in each variable. For $f \in L_2(\nu)$ (i.e. $\int_{\mathbb{S}} f^2(x)\, d\nu(x) < \infty$), consider

$$T_K : L_2(\nu) \to L_2(\nu)$$

defined by

$$(T_K f)(x) = \int_{\mathbb{S}} K(x, y) f(y)\, d\nu(y).$$

Then $T_K$ is a compact operator, so it has eigenvalues $\lambda_j$ and eigenfunctions $\phi_j$ such that

$$(T_K\phi_j)(x) = \sum \lambda_j \phi_j(x).$$

Also $\lambda_j \to 0$ as $j \to \infty$ and the eigenfunctions can be taken to be orthonormal. Given $f = \sum f_j \phi_j$, $T_K(f) = \sum_{j=1}^{\infty} \lambda_j f_j \phi_j$. Then define

$$\langle f_1, f_2 \rangle_{\mathcal{H}_K} = \sum f_{1,j} f_{2,j} / \lambda_j.$$

Then the set of functions in $L_2$ with finite $\mathcal{H}_K$ norm is an RKHS with kernel $K$.

**Mercer's theorem:** $K(x, y) = \sum_j \lambda_j \phi_j(x) \phi_j(y)$.

Take $f$ such that $\sum_j f_j^2 / \lambda_j < \infty$. Then

$$\langle f, K(\cdot, x) \rangle_{\mathcal{H}_K} = \sum_j f_j \phi_j(x) = f(x).$$

Consider the feature map $\phi : \mathbb{S} \to \ell_2$

$$\phi(x) = (\sqrt{\lambda_1}\phi_1(x), \ldots, \sqrt{\lambda_k}\phi_k(x), \ldots).$$

Then $\langle \phi(x), \phi(y) \rangle_{\ell_2} = \sum \lambda_j \phi_j(x) \phi_j(y)$.