Lecture 7: RKHSs revisited

In the last two lectures we showed that RKHSs coincide with Hilbert spaces of functions where pointwise evaluation is a bounded linear functional. We also saw that the sobolev spaces $H^s(\Omega)$ for $s > d/2$ are a concrete example of such spaces.

In this lecture we will give a different construction of RKHSs that is explicit in terms of the kernel.

7.1 PDS Kernels

\cal{X}

Defⁿ Let $\mathcal{X}$ be a set. (Not necessarily a vector space!) A kernel $K : \mathcal{X} \times \mathcal{X} \to \mathbb{R}$ is said to be PDS if for any $X = \{x_1, \ldots, x_m\} \subseteq \mathcal{X}$, the matrix $K(X, X)$ is PDS.

Terminology: we will say A is PDS if $\underline{x}^T A \underline{x} \geq 0$ & strictly PDS if $\underline{x}^T A \underline{x} > 0$ $\forall \underline{x} \neq 0$.

eg Linear Kernel

$K : \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}$ $\quad K(\underline{x}, \underline{x}') = \underline{x}^T \underline{x}'$

Given $X = \{\underline{x}_1, \ldots \underline{x}_n\}$ we have $K(\underline{x}_i, \underline{x}_j) = \underline{x}_i^T \underline{x}_j$
$= \underline{x}_j^T \underline{x}_i = K(\underline{x}_j, \underline{x}_i)$. And

$$\sum_i \sum_j \xi_i \xi_j \cdot K(\underline{x}_i, \underline{x}_j) = \sum_{i,j} \xi_i \cdot \xi_j \cdot \underline{x}_i^T \underline{x}_j$$
$$= \sum_{i=1}^n \sum_{j=1}^n (\xi_i \underline{x}_i)^T (\xi_j \underline{x}_j) = \left\| \sum_i \xi_i \underline{x}_i \right\|_2^2 \geq 0$$

eg Polynomial Kernel

$K : \mathbb{R}^d \times \mathbb{R}^d \longmapsto \mathbb{R}$ $\quad K(\underline{x}, \underline{x}') = (\underline{x}^T \underline{x} + c)^\alpha$

for $c \in \mathbb{R}$ & $\alpha \in \mathbb{N}$.

eg. Gaussian / RBF Kernel

$K : \mathbb{R}^d \times \mathbb{R}^d \longmapsto \mathbb{R}$ $\quad K(\underline{x}, \underline{x}') = \exp(-\gamma \|\underline{x} - \underline{x}'\|^2)$

for $\gamma > 0$.

eg: Sigmoid Kernel

$K : \mathbb{R}^d \times \mathbb{R}^d \longmapsto \mathbb{R}$ $\quad K(\underline{x}, \underline{x}') = \tanh(a \, \underline{x}^T \underline{x}' + b)$

for $a, b > 0$.

7.2 Constructing RKHSs

Now suppose $k: \mathcal{X} \times \mathcal{X} \to \mathbb{R}$ is a PDS kernel & for any $x \in \mathcal{X}$ define the function

$$\varphi_x : \mathcal{X} \to \mathbb{R}, \quad \varphi_x(y) := K(x,y)$$

We now consider the (vector) space of functions

$$H_0 := \left\{ f = \sum_{i=1}^{n} c_i \varphi_{x_i} \;\middle|\; n \in \mathbb{N},\ x_i \in \mathcal{X},\ c_i \in \mathbb{R} \right\}$$

& equip this space with the operation

$$\langle \cdot, \cdot \rangle : H_0 \times H_0 \to \mathbb{R}$$

$$f = \sum_{i=1}^{n} c_i \varphi_{x_i}, \quad g = \sum_{j=1}^{m} b_j \varphi_{y_j}$$

$$\langle f, g \rangle_0 := \sum_{i=1}^{n} \sum_{j=1}^{m} c_i b_j K(x_i, y_j)$$

Observe this readily implies that

$$\langle \varphi_{x_i}, \varphi_{x_j} \rangle_0 = K(x_i, y_j)$$

so that we may write,

$$\langle f, g \rangle_0 := \sum_{j=1}^{m} b_j f(y_j) = \sum_{i=1}^{n} c_i g(x_i). \quad (*)$$

We may now verify that $\langle\cdot,\cdot\rangle$ is sym. & indeed, bi-linear. Also note that (*) implies that $\langle\cdot,\cdot\rangle$ is independent of the particular representation of f, g! Furthermore, since K is PDS, we have

$$\langle f, f\rangle_0 = \sum_{i=1}^{n}\sum_{j=1}^{n} c_i c_j K(x_i, x_j) \geq 0.$$

But this is not yet sufficient to infer that $\langle\cdot,\cdot\rangle_0$ is an inner product since we need to show that $\langle f, f\rangle_0 = 0 \iff f = 0$!

Lemma: Let $\Gamma : \mathcal{X}\times\mathcal{X} \to \mathbb{R}$ be a PDS kernel Then for any $x, x' \in \mathcal{X}$ we have

$$(\Gamma(x,x'))^2 \leq \Gamma(x,x)\Gamma(x',x')$$

Proof in HW.

Now observe that $\langle\cdot,\cdot\rangle : \mathcal{X}\times\mathcal{X} \to \mathbb{R}$ itself is a PDS kernel on $\mathcal{X}$ (we verified this above!) so applying the above lemma with $\Gamma(\cdot,\cdot) \equiv \langle\cdot,\cdot\rangle_0$ we infer that

$$(\langle f, \varphi_x\rangle_0)^2 \leq \langle f, f\rangle_0 \langle \varphi_x, \varphi_x\rangle_0$$

But due to the reproducing property

$$\langle f, \varphi_x \rangle_0 = \langle f, K(x,\cdot) \rangle_0 = f(x)$$

& so we have that

$$|f(x)|^2 \leq \langle f, f \rangle_0 K(x,x) \quad \forall x \in \mathcal{X}$$

Thus, if $\langle f, f \rangle_0 = 0$ then $f = 0$. The converse is trivial to verify & so, $\langle \cdot, \cdot \rangle_0$ is an inner product on H_0 making it a pre-Hilbert space.

By the completion thm for Hilbert spaces we can now complete H_0 wrt $\langle \cdot, \cdot \rangle_0$ to obtain a Hilbert space $(H, \langle \cdot, \cdot \rangle)$.

By the schwartz ineq. we have that

$$\langle f, K(x,\cdot) \rangle_0 \leq \|f\|_0 \cdot \|K(x,\cdot)\|_0$$

so that $f \longmapsto \langle f, K(x,\cdot) \rangle_0$ is continuous. Thus, since H_0 is dense in H we have

$$f(x) = \lim_{n\to\infty} f_n(x) = \lim_{n\to\infty} \langle f_n, K(x,\cdot) \rangle_0$$

$$= \langle f, K(x,\cdot) \rangle \quad \forall x \in \mathcal{X}$$

& so the reproducing property holds on H as well. The space $\mathcal{H}$ is the RKHS associated with the kernel K!

Thm (RKHS)

Let $K: \mathcal{X} \times \mathcal{X} \to \mathbb{R}$ be a PDS Kernel. Then, there exists a Hilbert space H & a mapping $\varphi: \mathcal{X} \to H$ s.t.

$$\forall x, x' \in H \qquad K(x,x') = \langle \varphi(x), \varphi(x') \rangle$$

The map φ is called the feature map of K. Furthermore, H has the reproducing property, ie,

$$\forall f \in H \ \& \ x \in \mathcal{X}, \qquad f(x) = \langle f, K(x, \cdot) \rangle$$

The space H is called the RKHS of K.

Summary

We presented a second construction of RKHSs arising from particular choice of a PDS Kernel.

- Completion of functions of the form

$$f(x) = \sum_{i=1}^{n} c_i K(x_i, x)$$

equipper with inner prod.

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

- The function $\varphi(x) : X \longrightarrow H$ is called a feature map of K if

$$K(x, x') = \langle \varphi(x), \varphi(x') \rangle$$

Notation $\equiv \langle \varphi_x , \varphi_{x'} \rangle$