Some explicit examples of RKHSs

Functions on a finite set

Let $\mathcal{X} = \{a, b, c\}$, a set with three elements. Then, there is a one-to-one correspondence between functions $f : \mathcal{X} \to \mathbb{R}$ and $\mathbb{R}^3$. Indeed, each function f is essentially a vector with three components $(f(a), f(b), f(c))^\top$. So any RKHS is a subspace of $\mathbb{R}^3$ (or $\mathbb{R}^3$) itself, with a possibly non-standard inner product.

1. A kernel function K is now simply 3×3 positive semi-definite symmetric matrix, where $K_{11} = K(a, a)$, $K_{2,3} = K(b, c)$, and so forth. For instance we can consider

$$K = \begin{pmatrix} 1 & 0 & 1 \\ 0 & 2 & 0 \\ 1 & 0 & 3 \end{pmatrix} .$$

The columns of K are precisely k_a, k_b, and k_c. So we have

$$k_a = \begin{pmatrix} 1 \\ 0 \\ 1 \end{pmatrix} \; ; \; k_b = \begin{pmatrix} 0 \\ 2 \\ 0 \end{pmatrix} \; ; \; k_c = \begin{pmatrix} 1 \\ 0 \\ 3 \end{pmatrix} .$$

Note that $\{k_a, k_b, k_c\}$ is a basis for $\mathbb{R}^3$. So in this case the RKHS $\mathcal{H}$ is $\mathbb{R}^3$, but with an unusual inner product.

Let us see everything more explicitly. Fix a vector $v = (v_1, v_2, v_3)^\top \in \mathbb{R}^3$. We can express it as

$$v = \alpha_a k_a + \alpha_b k_b + \alpha_c k_c = K\alpha ,$$

where the vector $\alpha = (\alpha_a, \alpha_b, \alpha_c)^\top$ of real coefficients can be found as $\alpha = K^{-1}v$. Having noticed this, we can now explicitly show what is the inner product in $\mathcal{H}$. Given two vectors, $v = (v_1, v_2, v_3)^\top$ and $w = (w_1, w_2, w_3)$, we define $\alpha = K^{-1}v$ and $\beta = K^{-1}w$, so that $v = K\alpha$ and $w = K\beta$. We have that

$$\begin{aligned} \langle v, w \rangle_{\mathcal{H}} &= \langle \alpha_a k_a + \alpha_b k_b + \alpha_c k_c, \beta_a k_a + \beta_b k_b + \beta_c k_c \rangle \\ &= \alpha_a \beta_a K_{11} + \alpha_a \beta_b K_{12} + \alpha_a \beta_c K_{13} \\ &+ \alpha_b \beta_a K_{21} + \alpha_b \beta_b K_{22} + \alpha_b \beta_c K_{23} \\ &+ \alpha_c \beta_a K_{31} + \alpha_c \beta_b K_{32} + \alpha_c \beta_c K_{33} \\ &= \alpha^\top K \beta = (K^{-1}v)^\top K (K^{-1}w) = v K^{-1} w . \end{aligned}$$

This is the general formula for the induced inner product when K is invertible and $\mathcal{X}$ is finite!

Now let us check that everything makes sense! Consider a function $f \in \mathcal{H}$. We said it is a vector $(f_1, f_2, f_3)^\top$ in $\mathbb{R}^3$, with $f_1 = f(a)$, $f_2 = f(b)$, and $f_3 = f(c)$. Is this consistent with the reproducing property? Let us check it explicitly for a:

$$\langle k_a, f \rangle_{\mathcal{H}} = (k_a^\top K^{-1}) f = \begin{pmatrix} 1 & 0 & 0 \end{pmatrix} \begin{pmatrix} f_1 \\ f_2 \\ f_3 \end{pmatrix} = f_1 = f(a) .$$

2. Let us consider a case where K is not invertible. For instance

$$K = \begin{pmatrix} 1 & 0 & 1 \\ 0 & 2 & 0 \\ 1 & 0 & 1 \end{pmatrix} .$$

Now K has rank 2 and we see that $k_a = k_c$. Hence we have that $\mathcal{H}$ is not anymore the whole $\mathbb{R}^3$, but just the span of $\{k_a, k_b\}$, a two-dimensional subspace of $\mathbb{R}^3$. We have that if $v \in \mathcal{H}$ then again we have unique coefficients α_a and α_b such that $v = \alpha_a k_a + \alpha_b k_b = \tilde{K}\alpha$, with $\alpha = (\alpha_a, \alpha_b)^\top$ and

$$\tilde{K} = \begin{pmatrix} 1 & 0 \\ 0 & 2 \\ 1 & 0 \end{pmatrix} .$$

Now we actually have that $\alpha = K^\dagger v$, where $K^\dagger = (\tilde{K}^\top \tilde{K})^{-1}\tilde{K}^\top$ is a pseudo-inverse (no need to know this, it is just to say that α can be computed given v!). Now, for $v = \tilde{K}\alpha$ and $w = \tilde{K}\beta$, we have that (with a bit of algebraic manipulation)

$$\langle v, w\rangle_{\mathcal{H}} = \alpha^\top K \beta = v^\top \tilde{K}(\tilde{K}^\top K \tilde{K})^{-1}\tilde{K}^\top w .$$

For $f \in \mathcal{H}$, computing everything explicitly, we get that

$$\langle k_a, f\rangle_{\mathcal{H}} = \frac{f_1 + f_3}{2} = \frac{f(a) + f(c)}{2} .$$

At first sight it might seems that something is wrong since we expect $\langle k_a, f\rangle_{\mathcal{H}} = f(a)$. However, in this case $\mathcal{H}$ is precisely the subspace of $\mathbb{R}^3$ of vectors v such that $v_1 = v_3$ (since this is true for both k_a and k_b). So it must be that $f(a) = f(c)$, and so $\langle k_a, f\rangle_{\mathcal{H}} = (f(a) + f(c))/2 = f(a)$.

3. Assume that we are given a feature map $\phi : \mathcal{X} \to \mathbb{R}^4$ such that

$$\phi(a) = (1, 0, 0, 1)\,;\ \phi(b) = (0, 1, 1, 0)\,;\ \phi(c) = (0, 0, 1, 0) .$$

What is the induced RKHS?

We can compute the kernel K taking the inner product between the features (recall that $K(x, x') = \phi(x)^\top \phi(x')$). We find

$$K = \begin{pmatrix} 2 & 0 & 0 \\ 0 & 2 & 1 \\ 0 & 1 & 1 \end{pmatrix} .$$

Since K is invertible, again we will find that $\langle v, w\rangle_{\mathcal{H}} = v^\top K^{-1} w$.

Finite dimensional RKHSs on $\mathbb{R}$

Here we consider the case $\mathcal{X} = \mathbb{R}$.

1. We let $\mathcal{W} = \mathbb{R}^3$ (which of course is a Hilbert space with the usual inner product) and $\phi(x) = (1, \sqrt{2}x, x^2)^\top$. It is clear that $\{\phi(x) : x \in \mathbb{R}\}$ spans the whole $\mathbb{R}^3$, so we will have that $\mathcal{H} = \mathbb{R}^3$. Now, we can compute the kernel

$$K(x, y) = \phi(x)^\top \phi(y) = 1 + 2xy + x^2y^2 = (1 + xy)^2 .$$

The elements of $\mathcal{H}$ will be functions in the form $h_w(x) = w^\top \phi(x) = w_1 + w_2\sqrt{2}x + w_3x^2$. We can then identify h_w with the vector $w = (w_1, w_2, w_3)^\top \in \mathbb{R}^3$. Given two such functions h_w and $h_{w'}$, by definition of inner product in the induced RHKS (from $(\mathcal{W}, \phi)$), we have that

$$\langle h_w, h_w'\rangle_{\mathcal{H}} = w^\top w' = w_1w_1' + w_2w_2' + w_3w_3' .$$

The reproducing property is straightforward. Indeed, fixed a point x_0, we have that k_{x_0} is the function $k_{x_0}(x) = K(x_0, x) = 1 + 2x_0x + x_0^2x^2 = h_{w_0}(x)$, where $w_0 = (1, \sqrt{2}x_0, x_0^2)^\top = \phi(x_0)$. Then, for any w, we have that

$$\langle k_{x_0}, h_w\rangle_{\mathcal{H}} = \phi(x_0)^\top w = w_1 + w_2\sqrt{2}x_0 + w_3x_0^2 = h_w(x) .$$

2. Consider now the kernel

$$K(x, y) = xy\cos(x - y)\,.$$

First let us check that this can be written in the form $K(x, y) = \langle\phi(x), \phi(y)\rangle_{\mathcal{W}}$, for a suitable feature map ϕ valued in some Hilbert space $\mathcal{W}$. To do so, let us expand K. We have

$$K(x, y) = xy(\cos x\cos y + \sin x\sin y) = (x\cos x)(y\cos y) + (x\sin x)(y\sin y)\,.$$

So we can select $\mathcal{W} = \mathbb{R}^2$ and $\phi(x) = (x\cos x, x\sin x)^\top$ and we have $K(x, y) = \phi(x)^\top\phi(y)$. This directly implies that K is a kernel function. Now, we can construct as usual $\mathcal{H}$. We have that it will be isomorphic to $\mathbb{R}^2$ and the functions are in the form $h_w = x(w_1\cos x + w_2\sin x)$, for $w = (w_1, w_2)^\top \in \mathbb{R}^2$. The inner product gives $\langle h_w, h_{w'}\rangle_{\mathcal{H}} = w^\top w' = w_1w_1' + w_2w_2'$. We can check the reproducing property as before. $k_{x_0} = h_{w_0}$, with $w_0 = \phi(x_0)$, and so

$$\langle k_{x_0}, h_w\rangle_{\mathcal{H}} = \phi(x_0)^\top w = x_0(w_1\cos x_0 + w_2\sin x_0) = h_w(x_0)\,.$$

Infinite dimensional RKHSs on $\mathbb{R}$

Again we let $\mathcal{X} = \mathbb{R}$.

1. We let

$$K(x, y) = e^{-(x-y)^2/2}\,.$$

Again, we are looking for some Hilbert space $\mathcal{W}$ and some feature map ϕ such that we can write $K(x, y) = \langle\phi(x), \phi(y)\rangle_{\mathcal{W}}$.

Let us expand our kernel. First we write $K(x, y) = e^{-x^2/2}e^{-y^2/2}e^{xy}$. We can Taylor expand the last term:

$$e^{xy} = \sum_{j=0}^{\infty}\frac{(xy)^j}{j!} = \sum_{j=0}^{\infty}\frac{x^j}{\sqrt{j!}}\frac{y^j}{\sqrt{j!}}\,.$$

So, we get that

$$K(x, y) = \sum_{j=0}^{\infty}\frac{x^je^{-x^2/2}}{\sqrt{j!}}\frac{y^je^{-y^2/2}}{\sqrt{j!}}\,.$$

Now that we have splitted the dependency on x and y, we can find the right $\mathcal{W}$ and ϕ. We let $\mathcal{W} = \ell^2$, the space of squared summable real sequences (namely sequences $w = (w_j)_{j\ge 0}$ such that $\sum_{j=0}^{\infty} w_j^2 < \infty$), where the inner product is $\langle w, w'\rangle_{\mathcal{W}} = \sum_{j=0}^{\infty} w_jw_j'$. Now let us define $\phi(x) = (x^je^{-x^2/2}/\sqrt{j!})_{j\ge 0}$. This is a squared summable sequence, since

$$\sum_{j=0}^{\infty}\frac{x^{2j}e^{-x^2}}{j!} = e^{x^2}e^{-x^2} = 1\,.$$

So $\phi(x) \in \mathcal{W}$, and clearly $\langle\phi(x), \phi(y)\rangle_{\mathcal{W}} = K(x, y)$. Hence, $\mathcal{H}$ is given by the functions in the form

$$h_w(x) = \langle w, \phi(x)\rangle_{\mathcal{W}} = \sum_{j=0}^{\infty}\frac{w_j}{\sqrt{j!}}x^je^{-x^2/2}\,,$$

for $w \in \mathcal{W}$.[1]

The inner product is as usual given by

$$\langle h_w, h_{w'}\rangle_{\mathcal{H}} = \langle w, w'\rangle_{\mathcal{W}} = \sum_{j=0}^{\infty} w_jw_j'\,.$$

---

[1] Note that here we also have that $\mathcal{W}_0 = \overline{\mathrm{Span}(\phi)} = \mathcal{W}$. Indeed, if $w \in \mathcal{W}_0^\perp$, then it must be that $\sum_{j=0}^{\infty}\frac{w_j}{\sqrt{j!}}x^j = 0$ for any x. But the unicity of the Taylor expansion of the function 0 implies that all coefficients are 0, so $w = 0$.

We have that $k_x(x') = e^{-(x-x')^2/2} = \langle \phi(x), \phi(x') \rangle_{\mathcal{W}}$, which of course means that $k_x = h_{\phi(x)}$. This is the reproducing property, as

$$\langle k_x, h_w \rangle_{\mathcal{H}} = \langle \phi(x), w \rangle_{\mathcal{W}} = h_w(x) \,.$$

Note that $\mathcal{H}$ is also the closure (under $\| \cdot \|_{\mathcal{H}}$) of the linear span of $\{k_x \,:\, x \in \mathcal{X}\}$. So the element in $\mathcal{H}$ are (countable) linear combinations of Gaussian bumps, namely we can write every $h \in \mathcal{H}$ as

$$h(x) = \sum_j \alpha_j k_{x_j} = \sum_j \alpha_j e^{-(x-x_j)^2/2} \,.$$

All elements in $\mathcal{H}$ are continuous and smooth, and decay at infinity as $e^{-|x|^2/2}$. Picking a kernel that ensures that functions are regular usually helps with generalisation, preventing overfitting.

2. Consider the kernel

$$K(x,y) = \begin{cases} 1 & \text{if } x = y; \\ 0 & \text{otherwise.} \end{cases}$$

Let $\mathcal{W} = \ell^2(\mathbb{R})$. This is the following Hilbert space. Consider all functions $f : \mathbb{R} \in \mathbb{R}$ and for every f led $D_f = \{x \,:\, f(x) \neq 0\}$. Then

$$\mathcal{W} = \ell^2(\mathbb{R}) = \left\{ f \,:\, D_f \text{ is countable, } \sum_{x \in D_f} f(x)^2 < \infty \right\} .$$

The inner product is defined as

$$\langle f, g \rangle_{\mathcal{W}} = \sum_{x \in D_f \cap D_g} f(x) g(x) = \sum_{x \in \mathbb{R}} f(x) g(x) \,,$$

where we can make sense of the uncountable sum as the terms are 0 whenever $x \notin D_f \cap D_g$.

We let $\phi(x) = \delta_x$, which we define as

$$\delta_x(x') = \begin{cases} 1 & \text{if } x = x'; \\ 0 & \text{otherwise.} \end{cases}$$

Clearly $\phi(x) \in \mathcal{W}$ for all x, and $K(x,y) = \langle \phi(x), \phi(y) \rangle_{\mathcal{W}}$. The set $\{\phi(x) \,:\, x \in \mathcal{X}\}$ generates the whole $\mathcal{W}$. Let us now see that $\mathcal{H} = \mathcal{W}$. Indeed, for any $h \in \mathcal{W}$, we have that

$$\langle \delta_x, h \rangle_{\mathcal{W}} = \sum_{y \in \mathbb{R}} \delta_x(y) h(y) = h(x) \,.$$

So the reproducing property holds directly on $\mathcal{W}$ (which is a space of functions), and hence $\mathcal{W} = \mathcal{H}$.

Note that this kernel treats all points independently (there is no notion of *distance* between points in the kernel), which results in absence of regularity in the functions in the RKHS. Using this kernel will lead to poor generalisation, as it allows to simply fit the data keeping the function equal to 0 on all the rest of the domain.