1.4.2 Reproducing kernel Hilbert spaces

<u>Theorem 5</u> For every kernel k, $\exists$ a feature map ϕ taking values in some inner-product space $\mathcal{H}$, s.t. $k(x,x') = \langle \phi(x), \phi(x') \rangle \quad \forall x, x' \in \mathcal{X}$.

<u>Proof</u> Take $\mathcal{H}$ to be the vector space of functions of the form

$$f(\cdot) = \sum_{i=1}^{n} \alpha_i k(\cdot, x_i) \quad \text{where } n \in \mathbb{N}, \ \alpha \in \mathbb{R}^n \text{ and } x_1, \dots, x_n \in \mathcal{X}.$$

Define feature map: $\phi : \mathcal{X} \to \mathcal{H}$
$x \mapsto k(\cdot, x)$

Define an inner product between f and $g(\cdot) = \sum_{j=1}^{m} \beta_j k(\cdot, x_j')$

$$\langle f, g \rangle = \sum_{i=1}^{n} \sum_{j=1}^{m} \alpha_i \beta_j k(x_i, x_j')$$

This is well-defined. Indeed

$$\langle f, g \rangle = \sum_{i=1}^{n} \alpha_i g(x_i) = \sum_{j=1}^{m} \beta_j f(x_j')$$

The first equality shows that the inner product does not depend on the particular representation of g and the second equality shows the same for f.
$\therefore$ Inner-product is well-defined ✓.

Observe that $\langle k(\cdot, x), f \rangle = \sum_{i=1}^{n} \alpha_i k(x_i, x) = f(x) \qquad (*)$

In particular, $\langle \phi(x), \phi(x') \rangle = \langle k(\cdot, x), k(\cdot, x') \rangle = k(x, x')$ ✓

It remains to show that $\langle \cdot, \cdot \rangle$ is an inner product.
Symmetry ✓
Linearity ✓
Note $\langle f, f \rangle = \sum_{i,j} \alpha_i k(x_i, x_j) \alpha_j \geq 0$ (as k is positive definite) ✓ $\quad (\dagger)$

We want to show that $\langle f, f \rangle = 0 \Rightarrow f(x) = 0 \ \forall x \in \mathcal{X}$

From $(*)$ have $f(x)^2 = \langle k(\cdot, x), f \rangle^2$

$\ulcorner$ If we could use C-S inequality on the RHS, we'd have

$$f(x)^2 \leq \langle k(\cdot, x), k(\cdot, x) \rangle \langle f, f \rangle.$$

Idea: show that $\langle \cdot , \cdot \rangle$ is a kernel and use the C-S inequality for kernels ⌋

Given, $f_1, \dots, f_m \in \mathcal{H}$ and coeffs $\gamma_1, \dots, \gamma_m \in \mathbb{R}$, have

$$\sum_{i,j} \gamma_i \langle f_i, f_j \rangle \gamma_j \overset{\text{lin}}{=} \left\langle \sum_{i=1}^m \gamma_i f_i , \sum_{j=1}^m \gamma_j f_j \right\rangle \overset{(+)}{\geq} 0$$

Thus, we can use the C-S inequality for kernels (Prop 2) to show $\langle f, f \rangle = 0 \Rightarrow f(x) = 0$ ✓ □

<u>Functional analytic facts</u>

Let $(\mathcal{B}, \langle \cdot, \cdot \rangle)$ be an inner product space.

The inner product induces a norm: for $f \in \mathcal{B}$, $\|f\|_{\mathcal{B}}^2 := \langle f, f \rangle$.

A Cauchy sequence $(f_m)_{m=1}^{\infty} \in \mathcal{B}$ satisfies $\|f_m - f_n\|_{\mathcal{B}} \to 0$ as $m, n \to \infty$.

A space where every Cauchy sequence converges to a limit in the space is called <u>complete</u>. A <u>complete</u> inner product space is a <u>Hilbert space</u>.

A subset $V \subseteq \mathcal{B}$ is closed if $\forall (f_m)_{m=1}^{\infty} \in V$ with $f_m \to f \in \mathcal{B}$, have $f \in V$.

<u>Fact</u> If V is a closed subspace of a Hilbert space $\mathcal{B}$ then each $f \in \mathcal{B}$ may be decomposed as $f = u + v$ where $u \in V$, and

$v \in V^{\perp} := \{ w \in \mathcal{B} : \langle w, u \rangle = 0 \ \forall u \in V \}$.

Consider space $\mathcal{H}$ from theorem 5 and let $(f_m)_{m=1}^{\infty}$ be a Cauchy sequence in $\mathcal{H}$.

Then $f_m(x) - f_n(x) = \langle k(\cdot, x), f_m - f_n \rangle$ so by C-S inequality

$$|f_m(x) - f_n(x)| \leq \sqrt{k(x,x)} \, \|f_m - f_n\|_{\mathcal{H}}$$

Thus $(f_m(x))_{m=1}^{\infty}$ is Cauchy for each $x \in \mathcal{X}$, so we may define $f^*(x) := \lim_{m \to \infty} f_m(x)$.

One can show that by adding all such f^* to $\mathcal{H}$ we can complete $\mathcal{H}$, i.e. make it a Hilbert space.

In fact, $\mathcal{H}$ is a reproducing kernel Hilbert space.

<u>Def</u> 2 A Hilbert space $\mathcal{B}$ of fns $\mathcal{X} \to \mathbb{R}$ is a reproducing kernel Hilbert space (RKHS) if $\forall x \in \mathcal{X} \ \exists k_x \in \mathcal{B} : f(x) = \langle k_x, f \rangle \ \forall f \in \mathcal{B}$.

The function $k : \mathcal{X} \times \mathcal{X} \to \mathbb{R}$

$$(x, x') \mapsto \langle k_x, k_{x'} \rangle$$ is known as the <u>reproducing kernel</u> of $\mathcal{B}$.

Note that $\langle k_x, k_{x'} \rangle = k_{x'}(x)$, so $k(\cdot, x') = k_{x'}(\cdot)$.

After adding pointwise limits to $\mathcal{H}$, $\mathcal{H}$ is an RKHS with k as the reproducing kernel.

In fact, $\mathcal{H}$ is the unique RKHS with k as its reproducing kernel.