Kernel methods in machine learning
Homework 1

Thibault de SURREL
thibault.de-surrel@ensta-paris.fr

Exercise 1

1.

Let $n \in \mathbb{N}$, $(a_1, ..., a_n) \in \mathbb{R}^n$ and $(x_1, ..., x_n) \in \mathcal{X}^n$. We have that :

$$\sum_{i=1}^{n}\sum_{j=1}^{n} a_i a_j (\alpha K_1 + \beta K_2)(x_i, x_j) = \underbrace{\alpha}_{>0} \underbrace{\sum_{i=1}^{n}\sum_{j=1}^{n} a_i a_j K_1(x_i, x_j)}_{\geq 0 \text{ as } K_1 \text{ is a p.d kernel}} + \underbrace{\beta}_{>0} \underbrace{\sum_{i=1}^{n}\sum_{j=1}^{n} a_i a_j K_2(x_i, x_j)}_{\geq 0 \text{ as } K_2 \text{ is a p.d kernel}} \geq 0$$

So, $\boxed{\alpha K_1 + \beta K_2 \text{ is positive definite.}}$

2.

Let us denote $K = \alpha K_1 + \beta K_2$. We are going to show that the associated RKHS $\mathcal{H}$ is

$$\mathcal{H} = \alpha \mathcal{H}_1 + \beta \mathcal{H}_2 = \{\alpha f_1 + \beta f_2 \ : \ f_1 \in \mathcal{H}_1, \ f_2 \in \mathcal{H}_2\}$$

with

$$\|f\|_{\mathcal{H}} = \inf\{\sqrt{\alpha \|f_1\|_{\mathcal{H}_1}^2 + \beta \|f_2\|_{\mathcal{H}_2}^2} : \alpha f_1 + \beta f_2 = f\} \tag{1}$$

- First of all, we have, for all $x \in \mathcal{X}$,

$$K(x, \cdot) = \alpha \underbrace{K_1(x, \cdot)}_{\in \mathcal{H}_1} + \beta \underbrace{K_2(x, \cdot)}_{\in \mathcal{H}_2} \in \alpha \mathcal{H}_1 + \beta \mathcal{H}_2.$$

- Let us focus on the topology of $\alpha \mathcal{H}_1 + \beta \mathcal{H}_2$. Let us denote $E = \mathcal{H}_1 \times \mathcal{H}_2$. E is a Hilbert space if we equip it with the following norm :

$$\forall (f_1, f_2) \in E, \ \|(f_1, f_2)\|_E = \sqrt{\alpha \|f_1\|_{\mathcal{H}_1}^2 + \beta \|f_2\|_{\mathcal{H}_2}^2}$$

induced by the following dot product : for $(f_1, f_2), (g_1, g_2) \in E, \langle (f_1, f_2), (g_1, g_2) \rangle_E = \alpha \langle f_1, g_1 \rangle_{\mathcal{H}_1} + \beta \langle f_2, g_2 \rangle_{\mathcal{H}_2}$

Let us consider the canonical surjection s :

$$\begin{array}{rccc} s : & E & \longrightarrow & \alpha \mathcal{H}_1 + \beta \mathcal{H}_2 \\ & (f_1, f_2) & \longmapsto & \alpha f_1 + \beta f_2 \end{array}$$

Let us denote $F = \ker s$. Then, for all $(f_1, f_2) \in F$, $\alpha f_1 + \beta f_2 = 0$. Let us show that F is a closed set. For this, let us consider $(f_n^1, f_n^2)_{n\in\mathbb{N}}$ a sequence of F that converges in E to (f_1, f_2). We want to show that $(f_1, f_2) \in F$. As the sequence $(f_n^1, f_n^2)_{n\in\mathbb{N}}$ converges in E, we have that $(f_n^1)_n$ converges in $\mathcal{H}_1$ to f_1 and $(f_n^2)_n$ converges in $\mathcal{H}_2$ to f_2. So, as convergence in a RKHS implies pointwise convergence, we have that

$$\forall x \in \mathcal{X},\ \alpha f_n^1(x) + \beta f_n^2(x) \underset{n\to\infty}{\longrightarrow} \alpha f_1(x) + \beta f_2(x).$$

But, as $(f_n^1, f_n^2)_{n\in\mathbb{N}}$ is a sequence of F, we have that for all $n \in \mathbb{N}$ and for all $x \in \mathcal{X}, \alpha f_n^1(x) + \beta f_n^2(x) = 0$. Finally, we get that $\alpha f_1(x) + \beta f_2(x) = 0$ for all $x \in \mathcal{X}$. So $(f_1, f_2) \in F$ and F is a closed set.

As F is a closed set of a Hilbert space E, we have

$$E = F \oplus F^\perp \tag{2}$$

If we consider $\tilde{s} := s_{|F^\perp}$, $\tilde{s}$ is a bijection between $F^\perp$ and $\alpha\mathcal{H}_1 + \beta\mathcal{H}_2$. So, we can equip $\alpha\mathcal{H}_1 + \beta\mathcal{H}_2$ with a Hilbertian structure inherited from E. Therefore, $\alpha\mathcal{H}_1 + \beta\mathcal{H}_2$ is a Hilbert space for the norm

$$\forall f \in \alpha\mathcal{H}_1 + \beta\mathcal{H}_2,\ \|f\|_{\alpha\mathcal{H}_1+\beta\mathcal{H}_2} = \|\tilde{s}^{-1}(f)\|_E.$$

Let us now show 1. Let $e = (f_1, f_2) \in E$. From 2, we can decompose e the following way :

$$e = (f_1^F, f_2^F) + \tilde{s}^{-1}(\alpha f_1 + \beta f_2)$$

with $(f_1^F, f_2^F) \in F$. As the sum 2 is orthogonal, we have :

$$\|e\|_E^2 = \|(f_1^F, f_2^F)\|_E^2 + \|\alpha f_1 + \beta f_2\|^2_{\alpha\mathcal{H}_1+\beta\mathcal{H}_2}$$

So, for all function $f = \alpha f_1 + \beta f_2 \in \alpha\mathcal{H}_1 + \beta\mathcal{H}_2$, we have $\|f\|_{\alpha\mathcal{H}_1+\beta\mathcal{H}_2} \leq \|(f_1, f_2)\|_E \leq \sqrt{\alpha\|f_1\|^2_{\mathcal{H}_1} + \beta\|f_2\|^2_{\mathcal{H}_2}}$ with equality if, and only if, $(f_1, f_2) \in F^\perp$. We can deduce 1 :

$$\|f\|_{\alpha\mathcal{H}_1+\beta\mathcal{H}_2} = \inf\{\sqrt{\alpha\|f_1\|^2_{\mathcal{H}_1} + \beta\|f_2\|^2_{\mathcal{H}_2}} : \alpha f_1 + \beta f_2 = f\}$$

- Finally, we show the reproducing property to complete the proof. Let $x \in \mathcal{X}$ and $f \in \alpha\mathcal{H}_1 + \beta\mathcal{H}_2$. We can write $f = \tilde{s}(f_1, f_2)$ where $(f_1, f_2) \in F^\perp$. Moreover, we have that $K(x,\cdot) = \alpha K_1(x,\cdot) + \beta K_2(x,\cdot) = s(K_1(x,\cdot), K_2(x,\cdot))$ and $(K_1(x,\cdot), K_2(x,\cdot)) \in F^\perp$ as $K \neq 0$[1]. So, $K(x,\cdot) = \tilde{s}(K_1(x,\cdot), K_2(x,\cdot))$ and we have

$$\begin{aligned}
\langle f, K_x\rangle_{\alpha\mathcal{H}_1+\beta\mathcal{H}_2} &= \langle (f_1, f_2), (K_1(x,\cdot), K_2(x,\cdot))\rangle_E \\
&= \alpha\langle f_1, K_1(x,\cdot)\rangle_{\mathcal{H}_1} + \beta\langle f_2, K_2(x,\cdot)\rangle_{\mathcal{H}_2} \\
&= \alpha f_1(x) + \beta f_2(x) \text{ using the repoducing proprety in } \mathcal{H}_1 \text{ and in } \mathcal{H}_2 \\
&= f(x)
\end{aligned}$$

Finally, we have

$$\boxed{\begin{aligned} &\mathcal{H} = \alpha\mathcal{H}_1 + \beta\mathcal{H}_2 = \{\alpha f_1 + \beta f_2 :\ f_1 \in \mathcal{H}_1,\ f_2 \in \mathcal{H}_2\} \\ &\text{with} \\ &\|f\|_{\mathcal{H}} = \inf\{\sqrt{\alpha\|f_1\|^2_{\mathcal{H}_1} + \beta\|f_2\|^2_{\mathcal{H}_2}} : \alpha f_1 + \beta f_2 = f\} \end{aligned}}$$

[1] In fact, if $K = 0$, as $\alpha, \beta > 0$ and K_1, K_2 are p.d kernels, that would mean that $K_1 = 0$ and $K_2 = 0$. The RHKS associated is then $\{0\}$, and everything works.

Exercise 2

1.

We need to show two things : the symmetry and the positive-definiteness.

- **Symmetry :** Let $x, y \in \mathcal{X}$. Then, we have, as the dot product is symmetric,

$$K(x,y) = \langle \Psi(x), \Psi(y) \rangle_{\mathcal{F}} = \langle \Psi(y), \Psi(x) \rangle_{\mathcal{F}} = K(y,x)$$

- **Positive-definiteness :** Let $n \in \mathbb{N}$, $(a_1, ..., a_n) \in \mathbb{R}^n$ and $(x_1, ..., x_n) \in \mathcal{X}^n$. We have that

$$\begin{aligned}\sum_{i=1}^n \sum_{j=1}^n a_i a_j K(x_i, x_j) &= \sum_{i=1}^n \sum_{j=1}^n a_i a_j \langle \Psi(x_i), \Psi(x_j) \rangle_{\mathcal{F}} \\ &= \langle \sum_{i=1}^n a_i \Psi(x_i), \sum_{j=1}^n a_j \Psi(x_j) \rangle_{\mathcal{F}} \\ &= \| \sum_{i=1}^n a_i \Psi(x_i) \|^2 \geq 0\end{aligned}$$

So, K is a positive definite kernel on $\mathcal{X}$.

2.

We are going to show that the associated RHKS $\mathcal{H}$ is

$$\mathcal{H} = \{ f \colon x \in \mathcal{X} \mapsto \langle w, \Psi(x) \rangle_{\mathcal{F}}, \ w \in \mathcal{F} \}$$

equipped with the norm

$$\|f\|_{\mathcal{H}} = \inf\{ \|w\|_{\mathcal{F}} \ : \ w \in \mathcal{F}, \ f(x) = \langle w, \Psi(x) \rangle_{\mathcal{F}}, \ \forall x \in \mathcal{X} \}.$$

Let us denote $\tilde{\mathcal{H}} = \{ f \colon x \in \mathcal{X} \mapsto \langle w, \Psi(x) \rangle_{\mathcal{F}}, \ w \in \mathcal{F} \}$. We want to show that $\mathcal{H} = \tilde{\mathcal{H}}$. As in the first exercise, we have several points to verify :

- First of all, we have that, for all $x \in \mathcal{X}$,

$$K(x, \cdot) = \langle \underbrace{\Psi(x)}_{\in \mathcal{F}}, \Psi(\cdot) \rangle_{\mathcal{F}} \in \tilde{\mathcal{H}}.$$

- As in the first exercise, we will consider the surjection :

$$\begin{array}{rccl} s : & \mathcal{F} & \longrightarrow & \tilde{\mathcal{H}} \\ & w & \longmapsto & \langle w, \Psi(\cdot) \rangle_{\mathcal{F}} \end{array}$$

We have that $\ker(s)$ is a closed set. In fact, if $(w_n)_{n \in \mathbb{N}}$ is a sequence of $\ker(s)$ converging in $\mathcal{F}$ to w, then by continuity, we have

$$\forall x \in \mathcal{X}, \ \langle w_n, \Psi(x) \rangle_{\mathcal{F}} \underset{n \to \infty}{\longrightarrow} \langle w, \Psi(x) \rangle_{\mathcal{F}}.$$

As, for all $n \in \mathbb{N}$, $w_n \in \ker(s)$, we have that for all $x \in \mathcal{X}, \langle w_n, \Psi(x)\rangle_{\mathcal{F}} = 0$, so $\langle w, \Psi(x)\rangle_{\mathcal{F}} = 0$ i.e. $w \in \ker(s)$ which is then closed.

As $\mathcal{F}$ is a Hilbert set and $\ker(s)$ is closed, we have

$$\mathcal{F} = \ker(s) \oplus \ker(s)^{\perp}. \tag{3}$$

We can then consider $\tilde{s} = s_{|\ker(s)^{\perp}}$ which is a bijection between $\ker(s)^{\perp}$ and $\tilde{\mathcal{H}}$. So, $\tilde{\mathcal{H}}$ can be equipped with a hilbertian structure thank to the norm :

$$\forall f \in \tilde{\mathcal{H}},\ \|f\|_{\tilde{\mathcal{H}}} = \|\tilde{s}^{-1}(f)\|_{\mathcal{F}}.$$

Using the orthogonal decomposition 3, we have that, for $w \in \mathcal{F}$,

$$w = w_1 + \tilde{s}^{-1}(f_w)$$

where $w_1 \in \ker(s)$ and $f_w \colon x \mapsto \langle w, \Psi(x)\rangle$, we then have

$$\|w\|_{\mathcal{F}}^2 = \|w_1\|_{\mathcal{F}}^2 + \|\tilde{s}^{-1}(f_w)\|_{\mathcal{F}} = \|w_1\|_{\mathcal{F}}^2 + \|f_w\|_{\tilde{\mathcal{H}}}$$

So, we have that, for all $f \colon x \mapsto \langle w, \Psi(x)\rangle \in \tilde{\mathcal{H}}$ where $w \in \mathcal{F}$,

$$\|f\|_{\tilde{\mathcal{H}}} \leq \|w\|_{\mathcal{F}}$$

with equality if, and only if, $w \in \ker(s)^{\perp}$. We then have

$$\forall f \in \tilde{\mathcal{H}},\ \|f\|_{\mathcal{H}} = \inf\{\|w\|_{\mathcal{F}} \ :\ w \in \mathcal{F},\ f(x) = \langle w, \Psi(x)\rangle_{\mathcal{F}},\ \forall x \in \mathcal{X}\}$$

- Finally, we need to show the reproducing property : so $x \in \mathcal{X}$ and $f \in \tilde{\mathcal{H}}$, there exist $w \in \mathcal{F}$ such that $f = \tilde{s}^{-1}(w)$. Moreover, as $K(x,\cdot) \in \tilde{\mathcal{H}}$, there exist $a \in \mathcal{F}$ such that $K(x,\cdot) = \tilde{s}^{-1}(a)$. We then have :

$$\begin{aligned}\langle f, K(x,\cdot)\rangle_{\tilde{\mathcal{H}}} &= \langle w, a\rangle_{\mathcal{F}} \\ &= \langle w, \Psi(x)\rangle_{\mathcal{F}} + \langle w, a - \Psi(x)\rangle_{\mathcal{F}}\end{aligned}$$

As $\langle w, a - \Psi(x)\rangle_{\mathcal{F}} = \langle w, a\rangle_{\mathcal{F}} - \langle w, \Psi(x)\rangle_{\mathcal{F}} = 0$, we have

$$\langle f, K(x,\cdot)\rangle_{\tilde{\mathcal{H}}} = \langle w, \Psi(x)\rangle_{\mathcal{F}} = f(x)$$

Thus we have the reproducing property.

We can then conclude that

$$\boxed{\begin{array}{c}\mathcal{H} = \{f \colon x \in \mathcal{X} \mapsto \langle w, \Psi(x)\rangle_{\mathcal{F}},\ w \in \mathcal{F}\} \\ \text{with} \\ \|f\|_{\mathcal{H}} = \inf\{\|w\|_{\mathcal{F}} \ :\ w \in \mathcal{F},\ f(x) = \langle w, \Psi(x)\rangle_{\mathcal{F}},\ \forall x \in \mathcal{X}\}.\end{array}}$$