Machine learning with kernel methods

Homework n°1

Exercise 1. Kernels

1. $\mathcal{X} = \mathbb{R}_+$, $K(x, x') = \min(x, x')$

- K is obviously symmetric.
- Let $n \in \mathbb{N}^*, (x_1, \dots, x_n) \in \mathbb{R}_+^n, (a_1, \dots a_n) \in \mathbb{R}^n$.

  Note that for $x, x' \geq 0$, $\min(x, x') = \int_0^{\min(x,x')} 1 \, dt = \int_0^{+\infty} \mathbb{1}_{t \leq \min(x,x')} \, dt = \int_0^{+\infty} \mathbb{1}_{t \leq x} \mathbb{1}_{t \leq x'} \, dt$, so:

$$\begin{aligned}
\sum_{i=1}^{n} a_i a_j K(x_i, x_j) = \sum_{i=1}^{n} a_i a_j \int_0^{+\infty} \mathbb{1}_{t \leq x_i} \mathbb{1}_{t \leq x_j} \, dt &= \int_0^{+\infty} \left(\sum_{i=1}^{n} a_i a_j \mathbb{1}_{t \leq x_i} \mathbb{1}_{t \leq x_j} \right) dt \\
&= \int_0^{+\infty} \left(\sum_{i=1}^{n} a_i \mathbb{1}_{t \leq x_i} \right)^2 dt \\
&\geq 0.
\end{aligned}$$

Therefore $\boxed{K \text{ is positive definite}}$.

Remark: We essentially just said that $K(x, x') = \left\langle \mathbb{1}_{[0,x]}, \mathbb{1}_{[0,x']} \right\rangle_{L^2(\mathbb{R})}$, and we can then use the easy implication of Aronszajn's theorem.

2. $\mathcal{X} = \mathbb{R}_+$, $K(x, x') = \max(x, x')$

Take $n = 2, x_1 = 1, x_2 = 2, a_1 = 1, a_2 = -1$, then:

$$\begin{aligned}
\sum_{1 \leq i,j \leq 2} a_i a_j K(x_i, x_j) &= a_1^2 x_1 + a_2^2 x_2 + 2 a_1 a_2 K(x_1, x_2) \\
&= 1 + 2 - 4 \\
&= -1 \\
&< 0,
\end{aligned}$$

so $\boxed{K \text{ is } \textbf{not} \text{ positive definite}}$.

3. $\mathcal{X}$ a set, $f, g : \mathcal{X} \to \mathbb{R}_+$, $K(x,y) = \min(f(x)g(y), f(y)g(x))$.

First note that $K(x,y) = \begin{cases} 0 & \text{if } f(x) = 0 \text{ or } f(y) = 0 \\ f(x)f(y)\min\left(\dfrac{g(x)}{f(x)}, \dfrac{g(y)}{f(y)}\right) & \text{if } f(x) \neq 0 \text{ and } f(y) \neq 0. \end{cases}$

Let $n \in \mathbb{N}^*$, $(x_1, \ldots, x_n) \in \mathcal{X}^n$, $(a_1, \ldots, a_n) \in \mathbb{R}^n$. Then:

$$\begin{aligned}
\sum_{1 \leq i,j \leq n} a_i a_j K(x_i, x_j) &= \sum_{\substack{1 \leq i,j \leq n \\ f(x_i) \neq 0, f(x_j) \neq 0}} a_i a_j K(x_i, x_j) \\
&= \sum_{\substack{1 \leq i,j \leq n \\ f(x_i) \neq 0, f(x_j) \neq 0}} a_i a_j f(x_i) f(x_j) \min\left(\frac{g(x_i)}{f(x_i)}, \frac{g(x_j)}{f(x_j)}\right) \\
&= \sum_{\substack{1 \leq i,j \leq n \\ f(x_i) \neq 0, f(x_j) \neq 0}} b_i b_j \min(y_i, y_j),
\end{aligned}$$

where we set, for $i \in [\![1, n]\!]$ such that $f(x_i) \neq 0$, $b_i = a_i f(x_i)$ and $y_i = \dfrac{g(x_i)}{f(x_i)}$.
Using the fact that min is a positive definite kernel (see question 1.), we deduce that

$$\sum_{1 \leq i,j \leq n} a_i a_j K(x_i, x_j) = \sum_{\substack{1 \leq i,j \leq n \\ f(x_i) \neq 0, f(x_j) \neq 0}} b_i b_j \min(y_i, y_j) \geq 0,$$

so $\boxed{K \text{ is positive definite}}$.

Exercise 2. Non-expansiveness of the Gaussian kernel

For $x \in \mathbb{R}^p$, $\|\varphi(x)\|_{\mathcal{H}}^2 = K(x, x) = 1$.

For $x, x' \in \mathbb{R}^p$, we have:

$$\begin{aligned}\|\varphi(x) - \varphi(x')\|_{\mathcal{H}}^2 &= \|\varphi(x)\|_{\mathcal{H}}^2 + \|\varphi(x')\|_{\mathcal{H}}^2 - 2\langle \varphi(x), \varphi(x')\rangle_{\mathcal{H}} \\ &= 1 + 1 - 2K(x, x') \\ &= 2\left(1 - K(x, x')\right).\end{aligned}$$

Now, we use the fact that $\forall t \in \mathbb{R}, e^t \geq 1 + t$ (exp is convex, so its graph is above its tangents), so $K(x, x') = \exp\left(-\frac{\alpha}{2}\|x - x'\|^2\right) \geq 1 - \frac{\alpha}{2}\|x - x'\|^2$, and:

$$\|\varphi(x) - \varphi(x')\|_{\mathcal{H}}^2 \leq 2\left(1 - \left(1 - \frac{\alpha}{2}\|x - x'\|^2\right)\right) = \alpha\|x - x'\|^2.$$

So $\boxed{\forall x, x' \in \mathbb{R}^p, \|\varphi(x) - \varphi(x')\|_{\mathcal{H}} \leq \sqrt{\alpha}\|x - x'\|}$.

Exercise 3. RKHS

1. For $f_1, g_1 \in \mathcal{H}_1, f_2, g_2 \in \mathcal{H}_2$, we would be tempted to define $\langle f_1 + f_2, g_1 + g_2\rangle_{\mathcal{H}_1+\mathcal{H}_2} = c_1\langle f_1, g_1\rangle_{\mathcal{H}_1} + c_2\langle f_2, g_2\rangle_{\mathcal{H}_2}$ for some constants $c_1, c_2 > 0$. However, the decomposition $f = f_1 + f_2$ ($f_1 \in \mathcal{H}_1, f_2 \in \mathcal{H}_2$) for $f \in \mathcal{H}_1 + \mathcal{H}_2$ is not necessarily unique, so $\langle\cdot,\cdot\rangle_{\mathcal{H}_1+\mathcal{H}_2}$ is ill-defined.

We consider $\mathcal{H}_1 \times \mathcal{H}_2$, which is a Hilbert space when equipped with $\langle (f_1, f_2), (g_1, g_2)\rangle_{\mathcal{H}_1\times\mathcal{H}_2} = \dfrac{1}{\alpha}\langle f_1, g_1\rangle_{\mathcal{H}_1} + \dfrac{1}{\beta}\langle f_2, g_2\rangle_{\mathcal{H}_2}$. Consider the linear mapping

$$\boxed{u : \begin{cases} \mathcal{H}_1 \times \mathcal{H}_2 & \to \quad \mathcal{H}_1 + \mathcal{H}_2 \\ (h_1, h_2) & \mapsto \quad h_1 + h_2. \end{cases}}$$

Its kernel $\ker(u) := u^{-1}(\{0\})$ is closed.

Indeed, if $(f_n, g_n)_{n\geq 0}$ is a sequence in $\ker(u)$ which converges to $(f, g) \in \mathcal{H}_1\times\mathcal{H}_2$, we have, by definition of $\ker(u)$, $\forall n \geq 0, f_n + g_n \overset{(*)}{=} 0$.
And $f_n \xrightarrow[n\to+\infty]{\mathcal{H}_1} f$, $g_n \xrightarrow[n\to+\infty]{\mathcal{H}_2} g$, and the converge in a RKHS implies pointwise convergence.
So for each $x \in \mathcal{X}$, $\lim\limits_{n\to+\infty} f_n(x) = f(x)$, $\lim\limits_{n\to+\infty} g_n(x) = g(x)$, and $(*)$ implies $f(x) + g(x) = 0$. Therefore $f + g = 0$, *i.e.*, $(f, g) \in \ker(u)$.

We thus know that $\mathcal{H}_1 \times \mathcal{H}_2 = \ker(u) \oplus (\ker(u))^\perp$.
Then, the restriction of u to $(\ker(u))^\perp$, which we denote by $\boxed{v = u_{|(\ker(u))^\perp} : (\ker(u))^\perp \to \mathcal{H}_1 + \mathcal{H}_2}$, is an isomorphism.

For $f, g \in \mathcal{H}_1 + \mathcal{H}_2$, define $\boxed{\langle f, g\rangle_{\mathcal{H}_1+\mathcal{H}_2} = \langle v^{-1}(f), v^{-1}(g)\rangle_{\mathcal{H}_1\times\mathcal{H}_2}}$. Since $\mathcal{H}_1 \times \mathcal{H}_2$ is a Hilbert space, and v^{-1} is an isomorphism, $(\mathcal{H}_1 + \mathcal{H}_2, \langle\cdot,\cdot\rangle_{\mathcal{H}_1+\mathcal{H}_2})$ is a Hilbert space.

Let us check that $\mathcal{H}_1 + \mathcal{H}_2$ endowed with the above scalar product is the RKHS of $K := \alpha K_1 + \beta K_2$.

- For $x \in \mathcal{X}$, $K_x = \alpha \underbrace{K_{1,x}}_{\in\mathcal{H}_1} + \beta \underbrace{K_{2,x}}_{\in\mathcal{H}_2} \in \mathcal{H}_1 + \mathcal{H}_2$.
- Let $f \in \mathcal{H}_1 + \mathcal{H}_2$, and let $x \in \mathcal{X}$.
  Define $(f_1, f_2) = v^{-1}(f), (A_x, B_x) = v^{-1}(\alpha K_{1,x} + \beta K_{2,x})$. Then:
  $$\begin{aligned} \langle f, \alpha K_{1,x} + \beta K_{2,x}\rangle_{\mathcal{H}_1+\mathcal{H}_2} &= \langle (f_1, f_2), (A_x, B_x)\rangle_{\mathcal{H}_1\times\mathcal{H}_2} \\ &= \langle (f_1, f_2), (\alpha K_{1,x}, \beta K_{2,x})\rangle_{\mathcal{H}_1\times\mathcal{H}_2} \\ &\quad + \langle (f_1, f_2), (A_x - \alpha K_{1,x}, B_x - \beta K_{2,x})\rangle_{\mathcal{H}_1\times\mathcal{H}_2} \end{aligned}$$
  Then, we notice that:
  $$\begin{aligned} \langle (f_1, f_2), (\alpha K_{1,x}, \beta K_{2,x})\rangle_{\mathcal{H}_1\times\mathcal{H}_2} &= \frac{1}{\alpha}\langle f_1, \alpha K_{1,x}\rangle_{\mathcal{H}_1} + \frac{1}{\beta}\langle f_1, \beta K_{2,x}\rangle_{\mathcal{H}_2} \text{ (definition of } \langle\cdot,\cdot\rangle_{\mathcal{H}_1\times\mathcal{H}_2}) \\ &= f_1(x) + f_2(x) \text{ (reproduction property for } \mathcal{H}_1 \text{ and } \mathcal{H}_2) \\ &= f(x). \end{aligned}$$
  And $u\,(A_x - \alpha K_{1,x}, B_x - \beta K_{2,x}) = (A_x + B_x) - (\alpha K_{1,x} + \beta K_{2,x}) = 0$ because of the definition of A_x, B_x. So $(A_x - \alpha K_{1,x}, B_x - \beta K_{2,x}) \in \ker(u)$, and $(f_1, f_2) \in (\ker(u))^\perp$, and therefore $\langle (f_1, f_2), (A_x - \alpha K_{1,x}, B_x - \beta K_{2,x})\rangle_{\mathcal{H}_1\times\mathcal{H}_2} = 0$.

  Finally, $\langle f, \alpha K_{1,x} + \beta K_{2,x}\rangle_{\mathcal{H}_1+\mathcal{H}_2} = f(x)$, and the reproduction property holds.

So $\mathcal{H}_1 + \mathcal{H}_2$ is the RKHS of $\alpha K_1 + \beta K_2$.

In fact, for $f \in \mathcal{H}_1 + \mathcal{H}_2$, let $f_1 \in \mathcal{H}_1, f_2 \in \mathcal{H}_2$ be such that $f_1 + f_2 = f$.
Then:

$$\begin{aligned}
\|(f_1, f_2)\|^2_{\mathcal{H}_1 \times \mathcal{H}_2} &= \| \underbrace{((f_1, f_2) - v^{-1}(f))}_{\in \ker(u)} + \underbrace{v^{-1}(f)}_{\in (\ker(u))^\perp} \|^2_{\mathcal{H}_1 \times \mathcal{H}_2} \\
&= \|(f_1, f_2) - v^{-1}(f)\|^2_{\mathcal{H}_1 \times \mathcal{H}_2} + \|v^{-1}(f)\|^2_{\mathcal{H}_1 \times \mathcal{H}_2} \\
&\geq \|v^{-1}(f)\|^2_{\mathcal{H}_1 \times \mathcal{H}_2} \\
&= \|f\|^2_{\mathcal{H}_1 + \mathcal{H}_2}.
\end{aligned}$$

Therefore $\boxed{\|f\|_{\mathcal{H}_1 + \mathcal{H}_2} = \inf \{ \|(f_1, f_2)\|_{\mathcal{H}_1 \times \mathcal{H}_2} \mid f_1 \in \mathcal{H}_1, f_2 \in \mathcal{H}_2, f = f_1 + f_2 \}}$ or
$\boxed{\|f\|^2_{\mathcal{H}_1 + \mathcal{H}_2} = \inf \left\{ \frac{1}{\alpha} \|f_1\|^2_{\mathcal{H}_1} + \frac{1}{\beta} \|f_2\|^2_{\mathcal{H}_2} \mid f_1 \in \mathcal{H}_1, f_2 \in \mathcal{H}_2, f = f_1 + f_2 \right\}}$.

2. Let K be a p.d. kernel on $\mathcal{X}$, and $f : \mathcal{X} \to \mathbb{R}$. Let $\mathcal{H}$ be the RKHS with kernel K.

- First, assume that $f \in \mathcal{H}$.
It is clear that $(x, x') \mapsto K(x, x') - \lambda f(x) f(x')$ is symmetric for any λ (due to the symmetry of K).

For $n \in \mathbb{N}^*, (x_1, \ldots, x_n) \in \mathcal{X}^n, (a_1, \ldots, a_n) \in \mathbb{R}^n$:

$$\begin{aligned}
\sum_{1 \leq i,j \leq n} a_i a_j f(x_i) f(x_j) &= \left(\sum_{i=1}^n a_i f(x_i) \right)^2 \\
&= \left(\sum_{i=1}^n a_i \langle f, K_{x_i} \rangle_{\mathcal{H}} \right)^2 \\
&= \left(\left\langle f, \sum_{i=1}^n a_i K_{x_i} \right\rangle_{\mathcal{H}} \right)^2 \\
&\leq \|f\|^2_{\mathcal{H}} \left\| \sum_{i=1}^n a_i K_{x_i} \right\|^2_{\mathcal{H}} \quad \text{(Cauchy-Schwarz)} \\
&= \|f\|^2_{\mathcal{H}} \sum_{1 \leq i,j \leq n} a_i a_j \langle K_{x_i}, K_{x_j} \rangle_{\mathcal{H}} \\
&= \|f\|^2_{\mathcal{H}} \sum_{1 \leq i,j \leq n} a_i a_j K(x_i, x_j).
\end{aligned}$$

This shows that $(x, x') \mapsto K(x, x') - \lambda f(x) f(x')$ is p.d. for $\lambda = 1/\|f\|^2_{\mathcal{H}} > 0$ if $f \neq 0$.
If $f = 0$, then the above kernel (for $\lambda > 0$) is just K, so it is p.d. for any $\lambda > 0$.

- We now assume that $K_1 : (x, x') \mapsto K(x, x') - \lambda f(x) f(x')$ is p.d. for some $\lambda > 0$.
It is easy to check that $K_2 : (x, x') \mapsto \lambda f(x) f(x')$ is also p.d. (the symmetry is obvious, and $\sum_{i,j} a_i a_j \lambda f(x_i) f(x_j) = \left(\sum_i \sqrt{\lambda} a_i f(x_i) \right)^2 \geq 0$).

Moreover, $K = K_1 + K_2$. Using question 1., $\mathcal{H}$ is the sum of the RKHS of K_1, denoted by $\mathcal{H}_1$, and the RKHS of K_2, denoted by $\mathcal{H}_2$: $\mathcal{H} = \mathcal{H}_1 + \mathcal{H}_2$.
And $f \in \mathcal{H}_2$ (if $f \neq 0$, take a such that $f(a) \neq 0$, then $f = \frac{1}{\lambda f(a)} K_{2,a} \in \mathcal{H}_2$; if $f = 0$, it is obvious); and $0 \in \mathcal{H}_1$.
So $f = 0 + f \in \mathcal{H}_1 + \mathcal{H}_2 = \mathcal{H}$.

We have shown that:

$\boxed{f \in \mathcal{H} \text{ if and only if there exists } \lambda > 0 \text{ such that } (x, x') \mapsto K(x, x') - \lambda f(x) f(x') \text{ is p.d.}}$.