KERNEL METHODS IN M.L.

Homework 1 - by DAVID SOTO -

by DAVID SOTO

Exercise 1 : Kernels.

1. We have $\mathcal{X} = \mathbb{R}_+$ and the kernel K defined by :
$\forall (x,y) \in \mathbb{R}_+^2$, $K(x,y) = \min(x,y)$.

* $\forall (x,y) \in \mathbb{R}_+^2$, $K(x,y) = \min(x,y) = \min(y,x) = K(y,x)$
$\Rightarrow$ K is symmetric.

* Now, let's rewrite the kernel K. Indeed, we notice that $\forall (x,y) \in \mathbb{R}_+^2$:

$$K(x,y) = \min(x,y) = \int_{\mathbb{R}_+} \mathbb{1}_{\{t \leq x\}} \mathbb{1}_{\{t \leq y\}} \, dt . \quad (*)$$

From this we can prove the second point of the definition of a p.d. kernel.
Let $N \in \mathbb{N}$, let $(x_1, \dots, x_N) \in \mathbb{R}_+^N$ and let $(a_1, \dots, a_N) \in \mathbb{R}^N$.
We have :

$$A = \sum_{i=1}^N \sum_{j=1}^N a_i a_j K(x_i, x_j) = \sum_{i=1}^N \sum_{j=1}^N a_i a_j \min(x_i, x_j)$$

$$A \overset{(*)}{=} \sum_{i=1}^N \sum_{j=1}^N a_i a_j \int_{\mathbb{R}_+} \mathbb{1}_{\{t \leq x_i\}} \mathbb{1}_{\{t \leq x_j\}} \, dt$$

by the linearity of the integral (since the summations are finite) we have :

$$A = \int_{\mathbb{R}_+} \sum_{i=1}^N \sum_{j=1}^N a_i a_j \mathbb{1}_{\{t \leq x_i\}} \mathbb{1}_{\{t \leq x_j\}} \, dt$$

$$A = \int_{\mathbb{R}_+} \sum_{i=1}^N a_i \mathbb{1}_{\{t \leq x_i\}} \sum_{j=1}^N a_j \mathbb{1}_{\{t \leq x_j\}} \, dt$$

$$A = \int_{\mathbb{R}_+} \left(\sum_{i=1}^N a_i \mathbb{1}_{\{t \leq x_i\}} \right)^2 dt \geq 0$$

That is $\boxed{\sum_{i=1}^{N}\sum_{j=1}^{N} a_i a_j K(x_i, x_j) \geq 0}$.

Hence we can conclude that K is a positive definite kernel.

2. $\mathcal{X} = \mathbb{R}_+$ and we have $\forall (x,y) \in \mathbb{R}_+^2$, $K(x,y) = \max(x,y)$

If we take $N = 2$, $(x_1, x_2) = (5,4) \in \mathbb{R}_+^2$
and $(a_1, a_2) = (1, -1) \in \mathbb{R}^2$, we have:

$$A = \sum_{i=1}^{2}\sum_{j=1}^{2} a_i a_j K(x_i, x_j)$$

$$A = 1 \times \max(5,5) - 1 \times \max(5,4) - 1 \times \max(4,5) + (-1) \times (-1) \times \max(4,4)$$

$$A = \cancel{5} - \cancel{5} - 5 + 4$$

$$A = -1 < 0$$

Thanks to this counterexample we can conclude that the kernel K is not a positive definite kernel.

3. Let $\mathcal{X}$ be a set and $f, g : \mathcal{X} \longrightarrow \mathbb{R}_+$ be two non-negative functions:
$\forall (x,y) \in \mathcal{X}^2$, $K(x,y) = \min(f(x)g(y), f(y)g(x))$

* $\forall (x,y) \in \mathcal{X}^2$,
$K(x,y) = \min(f(x)g(y), f(y)g(x))$
$K(x,y) = \min(f(y)g(x), f(x)g(y))$
$K(x,y) = K(y,x)$
$\Rightarrow$ K is symmetric!

* Now, we notice that $\forall (x,y) \in \mathcal{X}^2$ such that $g(x) > 0$ and $g(y) > 0$, we have:

$$K(x,y) = \min\left(f(x)\,g(y),\ f(y)\,g(x) \right)$$
$$= \min\left(f(x)\,g(y) \times \frac{g(x)}{g(x)},\ f(y)\,g(x) \times \frac{g(y)}{g(y)} \right)$$

$$K(x,y) = g(x)\,g(y)\,\min\left(\frac{f(x)}{g(x)},\ \frac{f(y)}{g(y)} \right) \quad (*)$$

Let $N \in \mathbb{N}$, let $(x_1, \dots, x_N) \in \mathcal{X}^N$ and let $(a_1, \dots, a_N) \in \mathbb{R}^N$.
We have:

$$A = \sum_{i=1}^{N} \sum_{j=1}^{N} a_i a_j K(x_i, x_j)$$

$$A = \sum_{i=1}^{N} \sum_{j=1}^{N} a_i a_j \min\left(f(x_i)\,g(x_j),\ f(x_j)\,g(x_i) \right)$$

$$A = 0 + \sum_{\substack{1 \le i,j \le N \\ g(x_i) > 0 \\ g(x_j) > 0}} a_i a_j\, g(x_i)\, g(x_j) \min\left(\frac{f(x_i)}{g(x_i)},\ \frac{f(x_j)}{g(x_j)} \right)$$

It is patent that $\frac{f(x_i)}{g(x_i)} \geq 0 \ \forall i$ and $\frac{f(x_j)}{g(x_j)} \geq 0 \ \forall j$
(when $g(x_i) > 0$ and $g(x_j) > 0$)

Moreover, we have seen that the kernel $\overline{K} : \mathbb{R}_+^2 \longrightarrow \mathbb{R}$, $(x,y) \longmapsto \min(x,y)$

is a positive definite kernel (see question 1),
Therefore $\overline{K}$ verifies that $\forall\, n \in \mathbb{N}$, $\forall (y_1, \dots, y_n) \in \mathbb{R}_+^n$ and $\forall (b_1, \dots, b_n) \in \mathbb{R}^n$:

$$B = \sum_{i=1}^{n} \sum_{j=1}^{n} b_i b_j \min(y_i, y_j) \geq 0 .$$

If we denote:

$\forall i = 1, \dots, N, \quad b_i = a_i\, g(x_i)$

and $\forall i = 1, \dots, N, \quad y_i = \begin{cases} \frac{f(x_i)}{g(x_i)} & \text{if } g(x_i) > 0 \\ 0 & \text{otherwise} \end{cases} \in \mathbb{R}_+$

We have:

$$A = \sum_{i=1}^{N} \sum_{j=1}^{N} a_i a_j\, K(x_i, x_j)$$

$$A = \sum_{i=1}^{N} \sum_{j=1}^{N} b_i b_j \min(y_i, y_j) \geqslant 0$$

(from question 1!)

That is $A = \sum_{i=1}^{N} \sum_{j=1}^{N} a_i a_j\, K(x_i, x_j) \geqslant 0$

Hence, we can conclude that K is a positive definite kernel

2

Exercice 2: Non-expansiveness of the Gaussian kernel

Setting: We consider the Gaussian kernel $K : \mathbb{R}^p \times \mathbb{R}^p \longrightarrow \mathbb{R}$ such that for all pair of $(x,y) \in \mathbb{R}^p \times \mathbb{R}^p$:

$$K(x,y) = e^{-\frac{\alpha}{2}\|x-y\|^2}$$ (where $\|\,\|$ is the Euclidean norm on $\mathbb{R}^p$).

We call $\mathcal{H}$ the RKHS of K and consider its RKHS mapping $\varphi : \mathbb{R}^p \longrightarrow \mathcal{H}$ such that $\forall (x,y) \in \mathbb{R}^p \times \mathbb{R}^p$

$$K(x,y) = \langle \varphi(x), \varphi(y) \rangle_{\mathcal{H}}.$$

* Let's show that $\|\varphi(x) - \varphi(y)\|_{\mathcal{H}} \leq \sqrt{\alpha}\, \|x-y\|$

Let $(x,y) \in \mathbb{R}^p \times \mathbb{R}^p$, we have:

$$\|\varphi(x) - \varphi(y)\|^2_{\mathcal{H}} = \langle \varphi(x) - \varphi(y), \varphi(x) - \varphi(y) \rangle_{\mathcal{H}}$$

$$= \langle \varphi(x), \varphi(x) - \varphi(y) \rangle_{\mathcal{H}} - \langle \varphi(y), \varphi(x) - \varphi(y) \rangle_{\mathcal{H}}$$

$$= \langle \varphi(x), \varphi(x) \rangle_{\mathcal{H}} - \langle \varphi(x), \varphi(y) \rangle_{\mathcal{H}} - \left(\langle \varphi(y), \varphi(x) \rangle_{\mathcal{H}} - \langle \varphi(y), \varphi(y) \rangle_{\mathcal{H}} \right)$$

$$\|\varphi(x) - \varphi(y)\|^2_{\mathcal{H}} = K(x,x) - K(x,y) - K(y,x) + K(y,y)$$

We notice that $\forall (x,y) \in \mathbb{R}^p \times \mathbb{R}^p$,

$$K(x,y) = \langle \varphi(x), \varphi(y) \rangle_{\mathcal{H}} = \langle \varphi(y), \varphi(x) \rangle_{\mathcal{H}} = K(y,x),$$

i.e **K is symmetric**. Therefore, we have:

$$\|\varphi(x) - \varphi(y)\|^2_{\mathcal{H}} = K(x,x) + K(y,y) - 2K(x,y)$$

$$\|\varphi(x) - \varphi(y)\|^2_{\mathcal{H}} = e^0 + e^0 - 2e^{-\frac{\alpha}{2}\|x-y\|^2}$$

$$\|\varphi(x) - \varphi(y)\|^2_{\mathcal{H}} = 2 - 2e^{-\frac{\alpha}{2}\|x-y\|^2}$$

that is $\|\varphi(x)-\varphi(y)\|_{\mathcal{H}}^2 = 2\left(1-e^{-\frac{\alpha}{2}\|x-y\|^2}\right)$

Now, we use the convexity inequality which states that $\forall u \in \mathbb{R}$, $e^u \geqslant 1+u$.

$$\Rightarrow \quad \forall (x,y) \in \mathbb{R}^p \times \mathbb{R}^p \quad e^{-\frac{\alpha}{2}\|x-y\|^2} \geqslant 1+\left(-\frac{\alpha}{2}\|x-y\|^2\right)$$

$$\Rightarrow \forall (x,y) \in \mathbb{R}^p \times \mathbb{R}^p \quad -e^{-\frac{\alpha}{2}\|x-y\|^2} \leqslant -1+\frac{\alpha}{2}\|x-y\|^2 .$$

Moreover, by noticing that $\forall (x,y) \in \mathbb{R}^p \times \mathbb{R}^p$ $\boxed{-\frac{\alpha}{2}\|x-y\|^2 \leqslant 0}$ (because $\alpha \geqslant 0$), we have:

$\Rightarrow \forall (x,y) \in \mathbb{R}^p \times \mathbb{R}^p$:

$$\|\varphi(x)-\varphi(y)\|_{\mathcal{H}}^2 \leqslant 2\left(1-1+\frac{\alpha}{2}\|x-y\|^2\right)$$

$$\leqslant 2 \times \frac{\alpha}{2}\|x-y\|^2$$

$$\|\varphi(x)-\varphi(y)\|_{\mathcal{H}}^2 \leqslant \alpha\|x-y\|^2$$

$$\Rightarrow \quad \|\varphi(x)-\varphi(y)\|_{\mathcal{H}} \leqslant \sqrt{\alpha}\,\|x-y\| \quad \forall (x,y) \in \mathbb{R}^p \times \mathbb{R}^p$$

$\square$

③

Exercise 3 : Uniqueness of the RKHS :

Let us prove that if $K : \mathcal{X} \times \mathcal{X} \longrightarrow \mathbb{R}$ is a positive definite kernel, then it is the r.k of a UNIQUE RKHS

Let $K : \mathcal{X} \times \mathcal{X} \longrightarrow \mathbb{R}$ be a positive definite kernel.
Then, we know from the course that K is a r.k !
Now let us prove that K is the r.k of a unique RKHS !

* Let us first consider a RKHS $\mathcal{H}$ with r.k K.
By definition of a r.k, we have that:

$$\forall x \in \mathcal{X} \quad , \quad K_x \in \mathcal{H} \; .$$

Therefore the linear space spanned by the functions $K_x : t \longmapsto K(x,t) \quad \forall x \in \mathcal{X}$, which we denote

$$\mathcal{F} = \left\{ \sum_{i=1}^{m} \alpha_i K_{x_i} \; , \; m \in \mathbb{N}, (\alpha_1, \dots, \alpha_m) \in \mathbb{R}^m, (x_1, \dots, x_m) \in \mathcal{X}^m \right\},$$

is clearly a subspace of $\mathcal{H}$.

Let $v \in \mathcal{H}$ be a vector orthogonal to all vectors in $\mathcal{F}$.
In particular we have that $\forall x \in \mathcal{X}$,

$$v(x) \overset{\circledast}{=} \langle v, K_x \rangle_{\mathcal{H}} = 0 \quad (\circledast \text{ reproducing property})$$

that is $v = 0_{\mathcal{H}}$ and since $\mathcal{H}$ is a Hilbert space and $\mathcal{F} \subset \mathcal{H}$ a linear subspace of $\mathcal{H}$ then we can deduce that $\mathcal{F}$ is dense in $\mathcal{H}$.

Moreover the $\mathcal{H}$ norm for functions in $\mathcal{F}$ only depends on the r.k K, because it is given for a function $f = \sum_{i=1}^{m} \alpha_i K_{x_i} \in \mathcal{F}$ by :

$$\| f \|_{\mathcal{H}}^2 = \sum_{i=1}^{m} \sum_{j=1}^{m} \alpha_i \alpha_j \langle K_{x_i}, K_{x_j} \rangle_{\mathcal{H}}$$

$$\|f\|^2_{\mathcal{H}} = \sum_{i=1}^{n} \sum_{j=1}^{n} \alpha_i \alpha_j K(x_i, x_j).$$

* Now we suppose that $\mathcal{H}'$ is also a RKHS with r.k K.

Let us prove that $\mathcal{H} = \mathcal{H}'$.

By using the same arguments as in the first point * we obtain that $\mathcal{F}$ (the linear space spanned by the functions $K_x : t \longmapsto K(x,t)$ $\forall x \in \mathcal{X}$) is dense in $\mathcal{H}'$, and that the $\mathcal{H}'$ norm in $\mathcal{F}$ is given by: $\forall f = \sum_{i=1}^{n} \alpha_i K_{x_i} \in \mathcal{F}$

$$\|f\|^2_{\mathcal{H}'} = \sum_{i=1}^{n} \sum_{j=1}^{n} \alpha_i \alpha_j K(x_i, x_j).$$

We have in particular that $\forall f \in \mathcal{F}, \|f\|_{\mathcal{H}} = \|f\|_{\mathcal{H}'}$.

Let $f \in \mathcal{H}$. By definition of $\mathcal{F}$ being dense in $\mathcal{H}$, $\exists (f_n)_n$ a sequence in $\mathcal{F}$ such that $f_n \xrightarrow[n \to +\infty]{} f$, i.e such that $\|f_n - f\|_{\mathcal{H}} \xrightarrow[n \to +\infty]{} 0$

In particular, we notice that the converging sequence $(f_n)_n$ is a Cauchy sequence for the $\mathcal{H}$ norm, and since $\mathcal{H}$ norm coincides with the $\mathcal{H}'$ norm on $\mathcal{F}$, we can deduce that $(f_n)_n$ is also a Cauchy sequence for the $\mathcal{H}'$ norm and that $(f_n)_n$ converges in $\mathcal{H}'$ to a function $h \in \mathcal{H}'$. Then, by using the "Convergence in a RKHS implies pointwise convergence" Corollary of the course, applied to $\mathcal{H}$ and $\mathcal{H}'$, we have that

$\forall x \in \mathcal{X}, \lim\limits_{n \to +\infty} f_n(x) = f(x) = h(x)$.

i.e, $\forall x \in \mathcal{X}$, $f(x) = h(x) \implies f = h$

and therefore we have $f \in \mathcal{H}'$.

$\Rightarrow \quad \mathcal{H} \subset \mathcal{H}'$

* By using the same arguments (but the other way around!) we obtain $\mathcal{H}' \subset \mathcal{H}$.

$\Rightarrow \quad \mathcal{H} = \mathcal{H}'$, $\mathcal{H}$ and $\mathcal{H}'$ are the same RKHS!

$\Rightarrow$ K is the r.k of a unique RKHS! $\square$