ZEKRI
Oussama

MVA 23/24

Kernel Methods

HW 1

Exercice 1 :

Let $K : \mathcal{X} \times \mathcal{X} \to \mathbb{R}$ be a p.d. kernel, such that

$\forall (x, z) \in \mathcal{X}^2$, $K(x, z) \leq b^2$

First, $\|K_x\|_{\mathcal{H}}^2 = \langle K_x, K_x \rangle_{\mathcal{H}} = K(x, x)$

Then, $\forall f \in B_{\mathcal{H}}(0,1) \subset \mathcal{H}$ and $\forall x \in \mathcal{X}$,

$$f(x) = \langle f, K_x \rangle_{\mathcal{H}}$$

So, $|f(x)| \leq \|f\|_{\mathcal{H}} \, \|K_x\|_{\mathcal{H}}$ (Cauchy-Schwarz)

$\leq 1 \times \sqrt{K(x,x)}$ (because $f \in B_{\mathcal{H}}(0,1)$)

$$|f(x)| \leq b$$

So, by taking the sup, $\boxed{\|f\|_\infty \leq b.}$

Exercice 2:

$\Rightarrow$ First, suppose K is p.d. kernel

- By symmetry of K, $K(x', x) = 1 \Leftrightarrow K(x, x') = 1$
- Suppose that $K(x, x') = K(x', x'') = 1$

We have that, $\forall (a, a', a'') \in \mathbb{R}^3$, $\forall (x, x', x'') \in \mathcal{X}^3$,

$$a^2 K(x,x) + (a')^2 K(x',x') + (a'')^2 K(x'',x'') + 2aa' K(x,x') + 2a'a'' K(x',x'') + 2aa'' K(x,x'') \geq 0$$

But, $K(x,x) = K(x',x') = K(x'',x'') = K(x,x') = K(x',x'') = 1$

So,
$$a^2 + (a')^2 + (a'')^2 + 2(aa' + a'a'') + 2aa'' K(x,x'') \geq 0.$$

if $a' = -a$ and $a'' = a$, we have that

$$-a^2 + 2a^2 K(x, x'') \geq 0.$$

$$\Leftrightarrow \left(2K(x,x'') - 1\right) a^2 \geq 0$$

$$\Leftrightarrow \quad K(x, x'') \geq \frac{1}{2}.$$

$\boxed{\text{It means that } K(x,x'') = 1 \text{ because } K : \mathcal{X}^2 \to \{0;1\}}$

⇐ Then, let us suppose the two assumptions.

Consider the relation $\sim$ defined by $x \sim y \Leftrightarrow K(x,y) = 1$ and let us prove that it is an equivalence relation.

- $\sim$ is reflexive: Because $\forall x \in \mathcal{X}$, $K(x,x) = 1$.
- $\sim$ is symmetric: Because of assumption 1. (This assumption also tells us that K is symmetric!)
- $\sim$ is transitive: Because of assumption 2.

Let $N \in \mathbb{N}$ and consider $(x_1, \ldots, x_N) \in \mathcal{X}^N$ and $(a_1, \ldots, a_N) \in \mathbb{R}^N$.

$$\sum_{i=1}^{N} \sum_{j=1}^{N} a_i a_j K(x_i, x_j) = \sum_{i=1}^{N} a_i^2 + 2 \sum_{(i,j)/x_i \sim x_j} a_i a_j \quad (\text{because } K(x_i,x_j) = 1 \Leftrightarrow x_i \sim x_j)$$

(because $K(x_i,x_i) = 1$)

Let d be the number of equivalence classes and C_k the class number $k \in [\![1;d]\!]$.

For an arbitrary k, if $x \in C_k$, let us note a_x the "associated scalar". For example if $x = x_3$ we have that

$a_x = a_3$.

Thus, we can write that,

$$\sum_{i=1}^{n}\sum_{j=1}^{n} a_i a_j K(x_i, x_j) = \boxed{\sum_{k=1}^{d}\left(\sum_{x\in C_k} a_x\right)^2 \geq 0}$$

Finally, K is a p.d. kernel because it is also symmetric from the first assumption.

Exercice 3:

1) Let K_1 and K_2 be two p.d. kernels on $\mathcal{X}$.

Let $(\alpha, \beta) \in \mathbb{R}_+^2$

Consider $K = \alpha K_1 + \beta K_2$.

- First, $\forall (x, x') \in \mathcal{X}^2$, $K(x, x') = \alpha K_1(x, x') + \beta K_2(x, x')$

$$= \alpha K_1(x', x) + \beta K_2(x', x) = K(x', x)$$

so K is symmetric.

- Then, let $N \in \mathbb{N}$. $\forall (x_1, \dots, x_N) \in \mathcal{X}^N$ and $(a_1, \dots, a_N) \in \mathbb{R}^N$,

we have that,

$$\sum_{i=1}^{N} \sum_{j=1}^{N} a_i a_j K(x_i, x_j) = \alpha \underbrace{\sum_{i=1}^{N} \sum_{j=1}^{N} a_i a_j K_1(x_i, x_j)}_{\geq 0} + \beta \underbrace{\sum_{i=1}^{N} \sum_{j=1}^{N} a_i a_j K_2(x_i, x_j)}_{\geq 0}$$

because K_1 and K_2 are p.d.

$\boxed{\text{So, } K \text{ is also a p.d. kernel}}$

If H_1 and H_2 are the RKHS of K_1 and K_2 respectively,

let us define $H := H_1 + H_2 = \{ h_1 + h_2 \,,\ h_1 \in H_1 \text{ and } h_2 \in H_2 \}$

and
$$\langle \cdot , \cdot \rangle_{H_1 \times H_2} : (H_1 \times H_2)^2 \longrightarrow \mathbb{R}$$
$$\big((f_1, f_2), (g_1, g_2) \big) \longrightarrow \frac{1}{\alpha} \langle f_1, g_1 \rangle_{H_1} + \frac{1}{\beta} \langle f_2, g_2 \rangle_{H_2}$$

an inner product on $H_1 \times H_2$.

Step 1: Show that $\langle \cdot , \rangle_H$ is well defined.

An important thing to do, is to make sure that $H_1 \times H_2 \underset{(*)}{=} H_1 \oplus H_2$, so that we will be able to define an inner product $\langle \cdot , \cdot \rangle_H$ on H.

Define $\tilde{H} = H_1 \cap H_2$.

- If $\tilde{H} = \{0\}$, it is easy to get $(*)$
- If $\tilde{H} \neq \{0\}$, let us define $F = \{ (f, -f), f \in \tilde{H} \}$

We have that, $F \subset H_1 \times H_2$ let us show that F is closed, so that we could write $H_1 \times H_2 = F \oplus F^{\perp}$

Consider a sequence $(f_n) \in F^{\mathbb{N}}$ such that $(f_n, -f_n) \to (f_1, f_2)$

where $(f_1, f_2) \in H_1 \times H_2$

It means that $\begin{cases} f_n \to f_1 \\ -f_n \to f_2 \end{cases}$ i.e. $f_1 = -f_2$

so $(f_1, f_2) \in F$, i.e. F is closed.

As we said earlier, we can now write $H_1 \times H_2 = F \oplus F^{\perp}$

Now, consider $\varphi : \begin{array}{l} H_1 \times H_2 \longrightarrow H_1 + H_2 \\ (f_1, f_2) \longrightarrow f_1 + f_2 \end{array}$ a surjective application

Remark that $F = \varphi^{-1}(\{0\})$

It means that $\varphi_{|F^{\perp}}$ is a bijection.

For $f_{1,2} \in H_1 + H_2$, note $(\varphi_{|F^{\perp}})^{-1}(f_{1,2}) = (f_1, f_2)$

with $(f_1, f_2) \in H_1 \times H_2$

It means that we have a unique decomposition $(f_1, f_2) \in H_1 \times H_2$

for arbitrary $f_{1,2} \in H_1 \times H_2$. i.e. $H_1 \times H_2 = H_1 \oplus H_2$

$$\boxed{\langle f_{1,2}, g_{1,2} \rangle_H = \langle f_1 + f_2, g_1 + g_2 \rangle_{H_1 + H_2} = \langle (f_1, f_2), (g_1, g_2) \rangle_{H_1 \times H_2}}$$

The inner product is well defined.

Step 2: Show that $(H, \langle \cdot, \cdot \rangle_H)$ is the RKHS associated to K.

(i) Let $x \in X$, $K_x(\cdot) := K(x, \cdot)$.

We have that, $K_x = \alpha K_{1,x} + \beta K_{2,x}$, where

$K_{i,x} = K_i(x, \cdot)$, for $i \in \{1, 2\}$

so H contains all functions of the form K_x, $\forall x \in X$.

(ii) Reproducing property: $\forall x \in X$, $K_x = \alpha K_{1,x} + \beta K_{2,x}$

But $\varphi^{-1}_{|F^\perp}(K_x) = (K_x^1, K_x^2)$ (where $K_x^i \in F^\perp$, for $i \in \{1,2\}$)

i.e. $K_x^1 + K_x^2 = \alpha K_{1,x} + \beta K_{2,x}$

i.e. $K_x^1 - \alpha K_{1,x} = -(K_x^2 - \beta K_{2,x})$

so $(K_x^1 - \alpha K_{1,x}, K_x^2 - \beta K_{2,x}) \in F$

So, $\forall f \in H$, $\forall x \in \mathcal{X}$,

$f(x) = f_1(x) + f_2(x)$ which can be defined

as $f = \varphi_{|F^\perp}^{-1}(f_1, f_2)$ by bijectivity of $\varphi_{|F^\perp}$

(here, $(f_1, f_2) \in F^\perp$).

So, $(K_x^1 - \alpha K_{1,x}, K_x^2 - \beta K_{2,x})$ is orthogonal to (f_1, f_2)

$$\text{So, } \left\langle (K_x^1 - \alpha K_{1,x}, K_x^2 - \beta K_{2,x}), (f_1, f_2) \right\rangle_{H_1 \times H_2} \overset{(**)}{=} 0$$

$$\text{But, } \langle f, K_x \rangle_H = \left\langle (f_1, f_2), (K_x^1, K_x^2) \right\rangle_{H_1 \times H_2}$$

$$= \underbrace{\left\langle (f_1, f_2), (K_x^1 - \alpha K_{1,x}, K_x^2 - \beta K_{2,x}) \right\rangle_{H_1 \times H_2}}_{= 0 \text{ of } (**)}$$

$$+ \left\langle (f_1, f_2), (\alpha K_{1,x}, \beta K_{2,x}) \right\rangle_{H_1 \times H_2}$$

$$= \frac{1}{\alpha} \langle f_1, \alpha K_{1,x} \rangle_{H_1} + \frac{1}{\beta} \langle f_2, \beta K_{2,x} \rangle_{H_2}$$

$$= f_1(x) + f_2(x) = f(x)$$

This proves the reproducing property i.e. $(H, \langle \cdot, \cdot \rangle_H)$ is the RKHS of K

2) Let $\mathcal{X}$ a set and $\mathcal{F}$ an Hilbert space.

Let $\Psi : \mathcal{X} \to \mathcal{F}$ and $K : \mathcal{X}^2 \mapsto \mathbb{R}$ such that

$$\forall (x, x') \in \mathcal{X}^2, \ K(x, x') = \langle \Psi(x), \Psi(x') \rangle_{\mathcal{F}}$$

a) K is a p.d. kernel on $\mathcal{X}$

↳ K is symmetric: $\forall (x, x') \in \mathcal{X}^2$, $K(x, x') = \langle \Psi(x), \Psi(x') \rangle_F = \langle \Psi(x'), \Psi(x) \rangle_F = K(x', x)$

↳ Then, let $N \in \mathbb{N}$. $\forall (x_1, \dots, x_N) \in \mathcal{X}^N$ and $(a_1, \dots, a_N) \in \mathbb{R}^N$, we have that,

$$\sum_{i=1}^{N} \sum_{j=1}^{N} a_i a_j K(x_i, x_j) = \sum_{i=1}^{N} \sum_{j=1}^{N} a_i a_j \left\langle \phi(x_i), \phi(x_j) \right\rangle_{\mathcal{F}}$$

$$= \left\langle \sum_{i=1}^{N} a_i \phi(x_i), \sum_{j=1}^{N} a_j \phi(x_j) \right\rangle_{\mathcal{F}}$$

$$= \left\| \sum_{i=1}^{N} a_i \phi(x_i) \right\|_{\mathcal{F}}^2 \geq 0.$$

So K is a p.d. kernel

b) Let us define $H = \{ f_x(t) = \langle \phi(x), \phi(t) \rangle_{\mathcal{F}}, x \in \mathcal{X} \}$

with the inner product $\langle \cdot, \cdot \rangle_H$ defined by:

$$\forall (f_t, f_{t'}) \in H^2, \quad \langle f_t, f_{t'} \rangle_H = \langle \phi(t), \phi(t') \rangle_{\mathcal{F}}$$

- First, H contains all the functions of the form $K_x(\cdot) = \langle \phi(x), \phi(\cdot) \rangle_{\mathcal{F}}$, $\forall x \in \mathcal{X}$. (by definition)

- Then, $\forall x \in \mathcal{X}, \forall x' \in \mathcal{X}, \forall f_x \in H$,

$$f_x(x') = \langle \phi(x), \phi(x') \rangle_{\mathcal{F}} = \langle f_x, K_{x'} \rangle_H$$

$\boxed{\text{So } (H, \langle \cdot, \cdot \rangle_H) \text{ is the RKHS associated to } K.}$

3) Consider K a p.d. kernel on a space $\mathcal{X}$ and $f : \mathcal{X} \mapsto \mathbb{R}$..

$\boxed{\Rightarrow}$ Suppose that $f \in \mathcal{H}$, where $\mathcal{H}$ is the RKHS with kernel K.

- If $f = 0_{\mathcal{H}}$ The result is true because we already know that K is p.d.

- If $f \neq 0_{\mathcal{H}}$ Consider $\lambda = \dfrac{1}{\|f\|_{\mathcal{H}}^2} > 0$

Consider $K'(x, x') = K(x, x') - \dfrac{1}{\|f\|_{\mathcal{H}}^2} f(x) f(x')$

↳ K' is symmetric.

↳ let $N \in \mathbb{N}$, $(a_1, \dots, a_N) \in \mathbb{R}^N$ and $(x_1, \dots, x_N) \in \mathcal{X}^N$

$$\sum_{i,j=1}^{N} a_i a_j K'(x_i, x_j) = \sum_{i,j=1}^{N} a_i a_j K(x_i, x_j) - \lambda \sum_{i,j=1}^{N} a_i a_j f(x_i) f(x_j)$$

$$= \left\langle \sum_{i=1}^{N} a_i K_{x_i}, \sum_{j=1}^{N} a_j K_{x_j} \right\rangle_{\mathcal{H}} - \lambda \left\langle f, \sum_{i=1}^{N} a_i K_{x_i} \right\rangle_{\mathcal{H}} \left\langle f, \sum_{j=1}^{N} a_j K_{x_j} \right\rangle_{\mathcal{H}}$$

But with Cauchy-Schwarz inequality,

$$\sum_{i,j=1}^{N} a_i a_j K'(x_i, x_j) \geq \left\| \sum_{i=1}^{N} a_i K_{x_i} \right\|_H^2 - \lambda \|f\|_H^2 \left\| \sum_{i=1}^{N} a_i K_{x_i} \right\|_H^2$$

$$\geq 0 \quad \text{because} \quad \lambda = \frac{1}{\|f\|_H^2} > 0.$$

$\boxed{\text{So there exists } \lambda > 0 \text{ such that } K(x,x') - \lambda f(x) f(x') \text{ is a p.d kernel}}$

$\Leftarrow$ Consider $K_0(x,x') = f(x) f(x')$, $\forall (x,x') \in \mathcal{X}^2$.

From question 2), we have that K_0 is a p.d. kernel

(In fact, $\mathcal{F} = \mathbb{R}$ so $\langle a, b \rangle_{\mathcal{F}} = a.b$)

Its RKHS $\underline{H_0 \text{ contains in particular } f}$ (also from question 2))

Now, let us remark that $\forall (x,x') \in \mathcal{X}^2$,

$$K(x,x') = K(x,x') - \lambda K_0(x,x') + \lambda K_0(x,x')$$

If we denote $\mathcal{H}'$ the RKHS of the p.d. kernel $K(x,x') - \lambda K_0(x,x')$, we have from question 1 that

$$\mathcal{H} = \mathcal{H}' + \mathcal{H}_0$$

This also means that $\mathcal{H}_0 \subset \mathcal{H}$.

As $f \in \mathcal{H}_0$, we therefore have that $\boxed{f \in \mathcal{H}}$.