# KMML Final homework !

Mahdi Kallel

TOTAL POINTS

## 17.75 / 20

QUESTION 1

### 1 Exercise 1 2 / 3

- **0 pts** Correct
- ✓ - **0.5 pts Error**
- ✓ - **0.5 pts Missing**
- **1 pts** 2 errors

K5 : you can not take x' = -1 (kernel defined on R+). K6 missing

QUESTION 2

### 2 Exercise 2 2.5 / 3

- **0 pts** Correct
- **0.5 pts** Error in "K pd implies (K(x,x')=1 iff K(x',x)=1) "
- **1 pts** Error in "K pd implies (K(x,x')=K(x',x'')=1 => K(x',x'')=1) "
- **1 pts** Proof of the converse (that K is pd if ...) imcomplete
- **1.5 pts** Proof that "K pd implies the two properties" missing
- **1 pts** Lacks clarity
- **2 pts** Incomplete
- ✓ - **0.5 pts Lacks details**
- **1.5 pts** Error

You did not prove that K is symmetric (to prove that it is pd)

QUESTION 3

### 3 3.1 2 / 2

- ✓ - **0 pts Correct**
- **0.5 pts** Lacks information
- **1 pts** Unclear
- **1.5 pts** Not correct
- **2 pts** Not correct

QUESTION 4

### 4 3.2 1 / 1

- ✓ - **0 pts Correct**
- **0.5 pts** Small error
- **1 pts** Not finished
- **1 pts** Error in the proof
- **1 pts** Not solved

QUESTION 5

### 5 3.3 0.5 / 1

- **0 pts** Correct
- ✓ - **0.5 pts Expression of MMD as a function of kernel missing or wrong**
- **0.5 pts** Expression of MMD as a sup missing or wrong

almost good... a "-2" missing though

QUESTION 6

### 6 3.4 2 / 2

- ✓ - **0 pts Correct**
- **2 pts** Not done or wrong
- **1.5 pts** Not solved, but some effort...
- **1 pts** Small error
- **0.5 pts** Small error

QUESTION 7

### 7 4.1 1 / 1

- ✓ - **0 pts Correct**
- **1 pts** wrong argument or missing answer
- **0.25 pts** missing details

QUESTION 8

### 8 4.2 0.75 / 1

- **0 pts** Correct
- **1 pts** major argument is missing
- ✓ - **0.25 pts need to be a bit more rigorous to handle**

**the cases when x_i =0**

- **- 0.5 pts** missing argument
- **- 1 pts** missing answer

QUESTION 9

## 9 4.3 **2 / 2**

✓ **- 0 pts** **Correct**

- **- 0.5 pts** missing minor argument
- **- 1 pts** there is a wrong argument but part of the proof is ok
- **- 2 pts** the proof relies on a wrong statement
- **- 1 pts** missing major argument
- **- 2 pts** missing answer

QUESTION 10

## 10 4.4 **2 / 2**

✓ **- 0 pts** **Correct**

- **- 0.5 pts** missing details or minor mistake
- **- 2 pts** missing or wrong answer
- **- 1 pts** not exactly the right answer
- **- 0 pts** Click here to replace this description.

QUESTION 11

## 11 4.5 **2 / 2**

✓ **- 0 pts** **Correct**

- **- 2 pts** missing or too incomplete answer
- **- 1 pts** inaccurate or incomplete statement
- **- 1.5 pts** the sufficient condition is too strong (only leads to a trivial case)
- **- 0.5 pts** minor details missing

# Kernel Methods HW#5:

Mahdi KALLEL

*Exercice* 1 :

$K_1(x,x') = 2^{x-x'}$ *is not symmetric* : $K(1,0) = 2 \; / \; K(0,1) = 0.5$ *and therefore is not* $p.d$

$K_2(x,x') = 2^{x+x'}$ *is symmetric*, $K(x,x') = 2^{x+x'} = 2^{x'+x} = K(x',x)$

*and* $\forall A, X \in \left(\mathbb{R}^n, \mathbb{R}^n\right) \;\; A^T K A = \sum a_i a_j K(x_i, x_j) = \sum a_i 2^{x_i} 2^{x_j} a_j =$
$\left(a_1 2^{x_1} \; . \; . \; a_n 2^{x_n}\right)^T \left(a_1 2^{x_1} \; . \; . \; a_n 2^{x_n}\right) \geq 0$

$K_3(x,x') = 2^{xx'} = e^{xx' \, log(2)}$,

$\kappa(x,x') = log(2)xx'$ *is* $p.d$ *a* $p.d$ *kernel scaled by the positive factor* $log(2)$.

*We know that if* $K$ *is* $p.d$ *then* $e^K$ *is* $p.d$ *and thus* $K_3$ *is* $p.d$.

$K_4(x,x') = log(1 + xx')$, *if we take* $x = 1$, $x' = 2$. *and* $A = (-2,1)$.

*then* $K = \begin{pmatrix} log(5) & log(3) \\ log(3) & log(2) \end{pmatrix}$ *and* $A^T K A = 4 \, log \, 2 - 4 \, log \, 3 + log \, 5 = log\left(\frac{80}{81}\right) < 0.$

$K_5(x,x') = max(x,x')$ *is not pd for* $A = (-1,1)$, $X = (-1,1)$, *then*
$A^T K_{XX} A = \begin{pmatrix} -1 & 1 \end{pmatrix} \begin{pmatrix} -1 & 1 \\ 1 & 1 \end{pmatrix} \begin{pmatrix} -1 \\ 1 \end{pmatrix} = \begin{pmatrix} -1 & 1 \end{pmatrix} \begin{pmatrix} 2 \\ 0 \end{pmatrix} = -2.$

## 1 Exercise 1 **2 / 3**

- **- 0 pts** Correct
- ✓ **- 0.5 pts** Error
- ✓ **- 0.5 pts** Missing
- **- 1 pts** 2 errors

K5 : you can not take x' = -1 (kernel defined on R+). K6 missing

*Exercice 2:*

$\leftarrow$ *If* $K(x,x') = K(x',x)$ *and* $K(x,x') = K(x',x'') = 1 \implies K(x,x'') = 1$:

*Let's denote by* $x \sim y$ *the relationship* $K(x,y) = 1$*. We now show it's an equivalence relationship:*

$\sim$ *is reflexive since* $K(x,x) = 1 \implies x \sim x$
$\sim$ *is symmetric since if* $K(x,y) = 1$ *then* $K(y,x)$ *meaning* $x \sim y \implies y \sim x$
$\sim$ *is transductive since if* $K(x,y) = K(y,z) = 1$ *then* $K(x,z) = 1$ *meaning if* $x \sim y$ *and* $y \sim z$ *then* $x \sim z$*.*

*Now let's take* $(X, A) \in \left(\mathcal{X}^n, \mathbb{R}^n\right)$ *, thus* $X = (x_1, .. x_n) \; / \; x_i \in \mathcal{X}$

*Now we prove that the order in which we take the elements does not affect the eigenvalues of the Kernel matrix.*

*Now by the Cayley – Hamilton theorem, the eigenvalues are the solutions of the polynomial*

$H(K) = det(XI_n - K)$*. We show that this determinant is independant of the order of the points. We illustrate the reasoning for 3 points. But this holds for any* $n$*.*

*Performing a permutation of two rows or two columns changes the determinant by* $-1$*.*

*And thus performing two of these operations does not change the determinant.*

$$H[K_{(x_1,x_2,x_3)}] = \begin{pmatrix} X - K(x_1,x_1) & -K(x_2,x_2) & -K(x_3,x_3) \\ -K(x_2,x_1) & X - K(x_2,x_2) & -K(x_2,x_3) \\ -K(x_3,x_1) & -K(x_3,x_2) & X - K(x_3,x_3) \end{pmatrix} \implies (\textit{Switch rows } x_1, x_2) \implies$$
$$\begin{pmatrix} -K(x_2,x_1) & X - K(x_2,x_2) & -K(x_2,x_3) \\ X - K(x_1,x_1) & -K(x_2,x_2) & -K(x_3,x_3) \\ -K(x_3,x_1) & -K(x_3,x_2) & X - K(x_3,x_3) \end{pmatrix} \implies (\textit{Switch columuns } x_1, x_2) \implies$$
$$\begin{pmatrix} X - K(x_2,x_2) & -K(x_2,x_1) & -K(x_2,x_3) \\ -K(x_2,x_2) & X - K(x_1,x_1) & -K(x_3,x_3) \\ -K(x_3,x_2) & -K(x_3,x_1) & X - K(x_3,x_3) \end{pmatrix} = H[K_{(x_2,x_1,x_3)}]$$

*And thus changing the order of two elements does not affect the Cayley Hamilton determinant.*

*Therefore it does not affect the eigenvalues of K.*

*Now since the equivalence classes form a partition of* $\mathcal{X}$ *, we can find an ordering of our vector X*

*such that elements in the same equivalence classes are adjacent.*

*Now since for all elements in an equivalence class* $[x]$ *we have* $K(x,y) = 1 \; \forall (y,z) \in [x]^2$

*and if* $x'' \notin [x]$ *then* $K(x',x'') = 0 \; \forall (x',x'') \in \left\{[x], \; \overline{[x]}\right\}$*, the kernel matrix will be block diagonal*

*with each block elements filled by ones.*

$If\ a\ block\ B_i\ has\ size\ |B_i|\ then\ it\ will\ have\ one\ eigenvalue\ of\ |B_i|\ and\ the\ remaining\ eigenvalues\ are\ 0.$

$Thus\ Spec(K) = \cup Spec(B_i) = \cup\{|B_i|, 0\} = \{|B_i|_{B\ is\ block}, 0\}\ which\ elements\ are\ all\ positive.$

$And\ thus\ K\ is\ a\ positive\ definite\ matrix\ whatever\ elements\ (x_1,..., x_n)\ we\ take\ and\ whatever\ the\ ordering.$

########################################################################################

*Exercice* 2 (*suite*) :

$\rightarrow K$ *is pd and* $K(x,x) = 1\ \forall x$

*Since K is p.d then it's associated kernel is symetric* $\implies K(x,y) = K(y,x)$ *and thus*

$K(x,y) = 1 \implies K(y,x) = 1.$

*Let's take* 3 *elements* $x, x', x''$. *and denote by K their kernel matrix.*

*Without loss of generality let's suppose* $K(x,x') = K(x',x'') = 1.$

*With the information we have we know that* $K_{(x,x',x'')}$*looks like this and is symmetric* $(x = K(x,x''))$

$$K = \begin{pmatrix} 1 & 1 & x \\ 1 & 1 & 1 \\ x & 1 & 1 \end{pmatrix}, \ if\ K(x,x'') = 0\ then:\quad det(K) = 0 - 1 + 0 = -1 = \prod_{i=1}^{3} \lambda_i$$

$\implies$ *K has a negative eigenvalue* $\implies$ *K is not p.d.*

*Therefore* $K(x,x'') = 1$. *And thus we conclude our proof.*

## 2 Exercise 2 **2.5 / 3**

- **- 0 pts** Correct
- **- 0.5 pts** Error in "K pd implies (K(x,x')=1 iff K(x',x)=1) "
- **- 1 pts** Error in "K pd implies (K(x,x')=K(x',x'')=1 => K(x',x'')=1) "
- **- 1 pts** Proof of the converse (that K is pd if ...) imcomplete
- **- 1.5 pts** Proof that "K pd implies the two properties" missing
- **- 1 pts** Lacks clarity
- **- 2 pts** Incomplete
- ✓ **- 0.5 pts Lacks details**
- **- 1.5 pts** Error

You did not prove that K is symmetric (to prove that it is pd)

*Exercice* 3.1 :

$L : \mathcal{H} \to \mathbb{R}$

$f \to E_{X\sim P}[f(X)]$

$|L(f)| = |E_{X\sim P}[f(X)]| \leq_{(Jensen,\ |.|\ convex)} E_{X\sim P}[|f(X)|] \leq E_{X\sim P}[| < f, K_x >_{\mathcal{H}}|]$

$\leq E_{X\sim P}[||f||_{\mathcal{H}} * ||K_x||_{\mathcal{H}}] \leq ||f||_{\mathcal{H}} * E_{X\sim P}[||K_x||_{\mathcal{H}}] \leq E_{X\sim P}\left[\sqrt{K(X,X)}\right] * ||f||_{\mathcal{H}}$

*Since K is a bounded operator* $E_{X\sim P}\left[\sqrt{K(X,X)}\right] \leq Max_{\mathcal{X}*\mathcal{X}}(K(x,x')) < \infty$

*And thus L is a linear bounded operator on the hilbert space* $\mathcal{H}$

$\Longrightarrow$ *By Riesz theorem* , $\exists!\, g \in \mathcal{H}\ s.t, \forall f \in \mathcal{H} : L(f) = < f, g > .$

$\Longrightarrow \exists!\, g \in \mathcal{H},\ st.\ \forall y \in \mathcal{X} : L(K_y) = E_{X\sim P}[K_y(X)] = = E_{X\sim P}[ < K_x, K_y > ]$

$= E_{X\sim P}[K(x,y)] = \mu(P)(y) = < K_y, g >$

*This shows that* $\forall y \in \mathcal{H},\ z\ g(y) = \mu(P)(y)$ *and therefore* $\mu = g \in \mathcal{H}$.

### 3 3.1 2 / 2

- ✓ - **0 pts** Correct
- - **0.5 pts** Lacks information
- - **1 pts** Unclear
- - **1.5 pts** Not correct
- - **2 pts** Not correct

*Exercice* 3.2 :

$$\mathbb{E}_{\mathcal{S}}[\ ||\ \mu(P)\ -\ \mu(P_s)||_{\mathcal{H}}]$$

$$\leq\ \mathbb{E}_{x_i\sim\mathcal{S}}[\ Sup_{||f||_{\mathcal{H}}\leq 1}\ < f, \mu(P)\ -\ \mu(P_s) >\ ] \leq \mathbb{E}_{x_i\sim\mathcal{S}}\left[\ Sup_{||f||_{\mathcal{H}}\leq 1}\ \left\{ E_{X\sim P}[f(X)]\ -\ \frac{1}{n}\sum f(x_i)\right\}\right]$$

$$\leq \mathbb{E}_{x_i\sim\mathcal{S}}\left[\ Sup_{||f||_{\mathcal{H}}\leq 1}\ \left\{ E_{\left(\widetilde{X}_i\right)\sim P}\left[\frac{1}{n}\sum f\left(\widetilde{x}_i\right)\right]\ -\ \frac{1}{n}\sum f(x_i)\right\}\right]$$

$$\leq \mathbb{E}_{x_i\sim\mathcal{S}}\left[\ Sup_{||f||_{\mathcal{H}}\leq 1}\ E_{\left(\widetilde{X}_i\right)\sim P}\left[\frac{1}{n}\sum f\left(\widetilde{x}_i\right) - \frac{1}{n}\sum f(x_i)\right]\ \right]$$

[*We show that Sup* $(\mathbb{E})\ \leq \mathbb{E}(\ Sup)$ *below.*)

$$\leq \mathbb{E}_{(X_i)\sim P}\ \mathbb{E}_{\left(\widetilde{X}_i\right)\sim P}\left[\ Sup_{||f||_{\mathcal{H}}\leq 1}\ \left\{\frac{1}{n}\sum\ \left(f\left(\widetilde{x}_i\right)\ -\ f(x_i)\right)\right\}\right]$$

(*The two variables follow the same distribution it does not matter in whic order we take*

$f\left(\widetilde{x}_i\right)$ & $f(x_i)$)

$$\leq \mathbb{E}_{(X_i)\sim P}\ \mathbb{E}_{\left(\widetilde{X}_i\right)\sim P}\mathbb{E}_{\sigma}\left[\ Sup_{||f||_{\mathcal{H}}\leq 1}\ \left\{\frac{1}{n}\sum \sigma_i\ \left(f\left(\widetilde{x}_i\right)\ -\ f(x_i)\right)\right\}\right]$$

$$\leq \mathbb{E}_{(X_i)\sim P}\ \mathbb{E}_{\left(\widetilde{X}_i\right)\sim P}\mathbb{E}_{\sigma}\ Sup_{||f||_{\mathcal{H}}\leq 1}\ \left\{\frac{1}{n}\sum \sigma_i\ \ \left(|f\left(\widetilde{x}_i\right)|\ +\ |f(x_i)|\right)\right\}$$

($X$ *and* $\widetilde{X}$ *are independant and follow the same distribution*)

$$\leq \mathbb{E}_{(X_i)\sim P}\ \mathbb{E}_{\left(\widetilde{X}_i\right)\sim P}\mathbb{E}_{\sigma}\ Sup_{||f||_{\mathcal{H}}\leq 1}\ \left\{\frac{1}{n}\sum \sigma_i\ \ \left(|f\left(\widetilde{x}_i\right)|\ +\ |f(x_i)|\right)\right\}$$

$$\leq 2\mathbb{E}_{(X_i)\sim P}\mathbb{E}_{\sigma}\ \ Sup_{||f||_{\mathcal{H}}\leq 1}\ \left\{\frac{1}{n}\sum \sigma_i\ \ \ |f(x_i)|\right\}$$

$$\leq 2\mathbb{E}_{(X_i)\sim P}\ \mathbb{E}_{\sigma}\ Sup_{||f||_{\mathcal{H}}\leq 1}\ \left\{|\frac{1}{n}\sum \sigma_i\ \ \ f(x_i)|\right\}$$

$$\leq 2\mathbb{E}_{(X_i)\sim P}\ \mathbb{E}_{\sigma}\ Sup_{||f||_{\mathcal{H}}\leq 1}\ \left\{|\frac{1}{n}\sum \sigma_i\ \ \ f(x_i)|\right\}$$

$\leq 2\mathcal{R}_n(||f||_{\mathcal{H}} \leq 1)$ (*from the course*) : $\leq \dfrac{4 * 1\ \sqrt{E\ K(X,X)}}{\sqrt{n}}$

---

*We show that Sup* $(\mathbb{E}) \leq \mathbb{E}(Sup)$ :

$$\forall f \ s.t \ _{||f||_{\mathcal{H}} \leq 1}, \ \frac{1}{n}\sum f\left(\widetilde{x}_i\right) - f(x_i) \ \leq \ Sup_{||f||_{\mathcal{H}} \leq 1} \left(\frac{1}{n}\sum f\left(\widetilde{x}_i\right) - f(x_i)\right)$$

$$\Rightarrow E_{\left(\widetilde{X_i}\right)\sim P}\left[\frac{1}{n}\sum f\left(\widetilde{x}_i\right) - f(x_i)\right] \ \leq E_{\left(\widetilde{X_i}\right)\sim P}\left[\ Sup_{||f||_{\mathcal{H}} \leq 1} \left(\frac{1}{n}\sum f\left(\widetilde{x}_i\right) - f(x_i)\right)\right]$$

$$\Longrightarrow \ Sup_{||f||_{\mathcal{H}} \leq 1} \ E_{\left(\widetilde{X_i}\right)\sim P}\left[\frac{1}{n}\sum f\left(\widetilde{x}_i\right) - f(x_i)\right]$$

$$\leq \ Sup_{||f||_{\mathcal{H}} \leq 1} E_{\left(\widetilde{X_i}\right)\sim P}\left[\ Sup_{||f||_{\mathcal{H}} \leq 1} \left(\frac{1}{n}\sum f\left(\widetilde{x}_i\right) - f(x_i)\right)\right]$$

$$= \ E_{\left(\widetilde{X_i}\right)\sim P}\left[\ Sup_{||f||_{\mathcal{H}} \leq 1} \left(\frac{1}{n}\sum f\left(\widetilde{x}_i\right) - f(x_i)\right)\right]$$

*Therefore we have* : $Sup_{||f||_{\mathcal{H}} \leq 1} \ E_{\left(\widetilde{X_i}\right)\sim P}\left[\frac{1}{n}\sum f\left(\widetilde{x}_i\right) - f(x_i)\right]$

$$\leq E_{\left(\widetilde{X_i}\right)\sim P}\left[\ Sup_{||f||_{\mathcal{H}} \leq 1} \left(\frac{1}{n}\sum f\left(\widetilde{x}_i\right) - f(x_i)\right)\right]$$

---

## 4 3.2 **1 / 1**

- ✓ **- 0 pts** Correct
- **- 0.5 pts** Small error
- **- 1 pts** Not finished
- **- 1 pts** Error in the proof
- **- 1 pts** Not solved

*Exercice* 3.3 :
3) $MMD(S1,\ S2)$ :

$\mu(P_s)$?? *we know that* $\forall f \in H,\ < f, \mu(P_s) > \ = \mathbb{E}_x \sim P_s\ [f(X)]$

*And thus for* $P_s\ = \left(\frac{1}{n}\sum \delta_{xi}\right),\ < f, \mu(P_s) > \ = \frac{1}{n}\sum f(x_i)$

$$||\ \mu(P_{s1}) - \ \mu(P_{S2})||_{\mathcal{H}}^{2} \ = Sup_{f\in\mathcal{H}}\ < \frac{f}{||\ \mu(P_{s1}) - \ \mu(P_{S2})||_{\mathcal{H}}},\ \mu(P_{s1}) - \ \mu(P_{S2}) >^{2} **$$

———————————————————————————————————————

**:

$$< \frac{f}{||\ \mu(P_{s1}) - \ \mu(P_{S2})||_{\mathcal{H}}},\ \mu(P_{s1}) - \ \mu(P_{S2}) > \ \leq \ ||\frac{f}{||\ \mu(P_{s1}) - \ \mu(P_{S2})||_{\mathcal{H}}}||_{\mathcal{H}}\ ||\ \mu(P_{s1}) - \ \mu(P_{S2})||_{\mathcal{H}}$$

*And by cauchy schwarz the maximum is achieved for* $f = \ \alpha\ [\mu(P_{s1}) - \ \mu(P_{S2})]$

———————————————————————————————————————

$$= \ Sup_{||f||_{\mathcal{H}}\leq 1}\ < f,\ \mu(P_{s1}) - \ \mu(P_{S2}) >^{2} = \ Sup_{||f||_{\mathcal{H}}\leq 1} \left\{\frac{1}{n}\sum f(x_i)\ - \ \frac{1}{m}\sum f(y_j)\right\}^{2}$$

$$||\ \mu(P_{s1}) - \ \mu(P_{S2})||_{\mathcal{H}}^{2} \ = ||\ \mu(P_{s1})||_{\mathcal{H}}^{2} \ + ||\ \mu(P_{s2})||_{\mathcal{H}}^{2} \ - 2\ < \mu(P_{s1}), \mu(P_{s2}) >$$

$$< \mu(P), \mu(P') > \ = \ E_{X'\sim P}'\ [\mu(P)(x')\ ] = \ E_{X'\sim P'}\ E_{X\sim P}\ [K(x, x'\ )]$$

$$< \mu(P_{s_1}), \mu(P_{s2}) > \ = E_{X\sim P_{s_1}}\left[\frac{1}{n}\sum_{y_i\in S_2} K(x\ , y_i)\right] = \ \frac{1}{m}\sum_{x_i\in S_1}\left[\frac{1}{n}\sum_{y_i\in S_2} K(x_i\ , y_i)\right]$$

$$= \frac{1}{nm}\sum_{(x_i,y_i)\in S_1.\ S_2} K(x_i, y_i)$$

$$||\ \mu(P_{s1}) - \ \mu(P_{S2})||_{\mathcal{H}}^{2} \ = \frac{1}{n^2}\sum_{(x_i,y_i)\in S_1.\ S_1} K(x_i, y_i)\ + \ \frac{1}{m^2}\sum_{(x_i,y_i)\in S_2.\ S_2} K(x_i, y_i) - \frac{1}{nm}\sum_{(x_i,y_i)\in S_1.\ S_2} K(x_i, y_i)$$

## 5 3.3 0.5 / 1

- **- 0 pts** Correct
- ✓ **- 0.5 pts Expression of MMD as a function of kernel missing or wrong**
- **- 0.5 pts** Expression of MMD as a sup missing or wrong

almost good... a "-2" missing though

*Exercice* 3.4 :

$$MMD(S_1, S_2) = Sup_{||f||_{\mathcal{H}} \leq 1} \left\{ \frac{1}{n} \sum f(x_i) - \frac{1}{m} \sum f(y_j) \right\}^2$$

*We know for* $x \in \mathbb{R}$ *for for* $f(x) = \frac{1}{\sigma} \exp\left( \frac{-x^2}{2\sigma^2} \right) \implies \widehat{f}(\epsilon) = \sqrt{2\pi} \exp\left( \frac{-\sigma^2 \epsilon^2}{2} \right)$

*for* $x \in \mathbb{R}^d \quad f(x) = \sigma^{-d} \exp\left( \frac{-||x||^2}{2\sigma^2} \right) = \prod_i^d \frac{1}{\sigma} \exp\left( \frac{-x_i^2}{2\sigma^2} \right)$

$$\implies \widehat{f}(\epsilon) = \prod_i^d \sqrt{2\pi}\, \sigma \exp\left( \frac{-\sigma^2 \epsilon_i^2}{2} \right) = \left( \sqrt{2\pi} \right)^d \exp\left( \frac{-\sigma^2 ||\epsilon||^2}{2} \right)$$

*For* $K_\sigma(x, y) = \exp\left( \frac{-||x-y||^2}{2\sigma^2} \right)$ *we know that is RKHS is* $\mathcal{H}_\sigma : \left\{ f : \int |\hat{f}(w)|^2 \exp\left( \frac{\sigma^2 ||w||^2}{2} \right) < \infty \right\}$

*For* $\sigma' > \sigma$ *if* $\int |\hat{f}(w)|^2 \exp\left( \frac{\sigma'^2 ||w||^2}{2} \right) < \infty \implies \int |\hat{f}(w)|^2 \exp\left( \frac{\sigma^2 ||w||^2}{2} \right) < \infty$

*And thus* $\mathcal{H}_\sigma' \in \mathcal{H}_\sigma$.

*If we call* $A_1 = \{f : ||f||_{\mathcal{H}_\sigma} \leq 1\} \ / \ A2 = \{f : ||f||_{\mathcal{H}_{\sigma'}} \leq 1\} \ (A_2 \in A_1)$

$$\implies Sup_{f \in A_1} \left\{ \frac{1}{n} \sum f(x_i) - \frac{1}{m} \sum f(y_j) \right\}^2 \geq Sup_{f \in A_2} \left\{ \frac{1}{n} \sum f(x_i) - \frac{1}{m} \sum f(y_j) \right\}^2$$

$$\implies MMD_\sigma(S_1, S_2) \geq MMD_{\sigma'}(S_1, S_2)$$

*And thus for* $K(x, y) = \sigma^{-d} \exp\left( \frac{-||x-y||^2}{2\sigma^2} \right)$ $MMD(S_1, S_2)$ *is a decreasingfunction of* $\sigma$.

## 6 3.4 **2 / 2**

- ✓ **- 0 pts** Correct
- **- 2 pts** Not done or wrong
- **- 1.5 pts** Not solved, but some effort...
- **- 1 pts** Small error
- **- 0.5 pts** Small error

*Exercice 4 :*

1) $K_1(x,x') = \sum a_i < x,x' >^i , a_i \geq 0.$

*We know that* $K(x,x') = < x,x' >$ *is a p.d kernel.*

*And the product of p.d kernels is p.d, and multiplying by* $\alpha \geq 0$ *keeps it p.d*

*Thus* $\forall_i \; a_i < x,x' >^i$ *is p.d*

*Then* $K_1(x,x')$ *is p.d as the (infinite) sum of p.d kernels.*

2) $||x|| \, || x'|| \, \kappa\left(\frac{< x,x' >}{||x|| \, || x'||}\right) =$

*If we take* $K_3(x,x') = ||x|| \, || x'||$ *then* $K_3$ *is p.d since*

$(x_1,.. x_n) \in (S^{p-1})^n \; \forall a \text{ in } \mathbb{R}^n, \quad \sum a_i K_3(x_i,x_j) a_j = \sum a_i \, ||x_i|| \, ||x_j|| \, a_j$

$= a'^T a' \geq 0$ *where* $a' = (a_1||x_1||, \ldots , a_n \, ||x_n||)$

*If we take* $K_1' = \kappa\left(\frac{< x,x' >}{||x|| \, || x'||}\right) = \kappa\left( < \frac{x}{||x||}, \frac{x'}{|| x'||} > \right)$ $(if \; ||x|| \; \& \; || x'|| \neq 0)$

*and* $0 \; if \; ||x|| \; or \; || x'|| = 0$

*then* $K_1'(x,x')$ *can be seen as* $K_1\left(\tilde{x}, \widetilde{x'}\right)$ *where* $\tilde{x}, \widetilde{x'} \in (S^{p-1})^2$, *therefore* $K_1'$ *is p.d*

*And thus* $K_2$ *is the product of* $K_1'$ *and* $K_3$ *which are two p.d kernels and thus it is p.d.*

3) $\kappa(u) = \kappa(1) - \int_u^1 \kappa'(u) du$

*Since* $u \leq 1$ , $\kappa'(u) = \sum_i i \, a_i \, u^i \leq \sum_i i \, a_i = \kappa'(1).$

$\Longrightarrow \int_u^1 \kappa'(u) du \leq \int_u^1 \kappa'(1). du \leq \kappa'(1) \int_u^1 du = \kappa'(1)(1-u) = (1-u)$

$\Longrightarrow \kappa(u) \geq 1 - (1-u) \geq u$

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

$= K(x,x) + K(x',x') - 2 K(x,x') = 2 (1 - \kappa( < x,x' > ) \quad \{x,x' \in S^{p-1}\}$

$$\implies ||\,\phi(x) - \phi(x')||^2_{\mathcal{H}} \leq ||x||^2 + ||\,x'||^2 - \;<x, x'> \;\leq ||\,x - x'||^2$$

7 4.1 **1 / 1**

- ✓ **- 0 pts** Correct
- **- 1 pts** wrong argument or missing answer
- **- 0.25 pts** missing details

*Exercice 4 :*

1) $K_1(x,x') = \sum a_i < x,x' >^i , a_i \geq 0.$

*We know that* $K(x,x') = < x,x' >$ *is a p.d kernel.*

*And the product of p.d kernels is p.d, and multiplying by* $\alpha \geq 0$ *keeps it p.d*

*Thus* $\forall_i\ a_i < x,x' >^i$ *is p.d*

*Then* $K_1(x,x')$ *is p.d as the (infinite) sum of p.d kernels.*

2) $||x||\ ||x'||\ \kappa\left(\frac{< x,x' >}{||x||\ ||x'||}\right) =$

*If we take* $K_3(x,x') = ||x||\ ||x'||$ *then* $K_3$ *is p.d since*

$(x_1,..x_n) \in (\mathbb{S}^{p-1})^n\ \forall a \text{ in } \mathbb{R}^n,\ \sum a_i K_3(x_i,x_j) a_j = \sum a_i\ ||x_i||\ ||x_j||\ a_j$

$= a'^T a' \geq 0$ *where* $a' = (a_1||x_1||, \ldots , a_n\ ||x_n||)$

*If we take* $K_1' = \kappa\left(\frac{< x,x' >}{||x||\ ||x'||}\right) = \kappa\left( < \frac{x}{||x||}, \frac{x'}{||x'||} > \right)$ $(if\ ||x||\ \&\ ||x'|| \neq 0)$

*and* $0$ *if* $||x||$ *or* $||x'|| = 0$

*then* $K_1'(x,x')$ *can be seen as* $K_1(\widetilde{x},\widetilde{x'})$ *where* $\widetilde{x},\widetilde{x'} \in (\mathbb{S}^{p-1})^2$*, therefore* $K_1'$ *is p.d*

*And thus* $K_2$ *is the product of* $K_1'$ *and* $K_3$ *which are two p.d kernels and thus it is p.d.*

3) $\kappa(u) = \kappa(1) - \int_u^1 \kappa'(u)du$

*Since* $u \leq 1$ , $\kappa'(u) = \sum_i i\, a_i\, u^i \leq \sum_i i\, a_i = \kappa'(1).$

$$\Longrightarrow \int_u^1 \kappa'(u)du \leq \int_u^1 \kappa'(1).du \leq \kappa'(1)\int_u^1 du = \kappa'(1)(1-u) = (1-u)$$

$$\Longrightarrow \kappa(u) \geq 1 - (1-u) \geq u$$

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

$$= K(x,x) + K(x',x') - 2\,K(x,x') = 2\,(1 - \kappa( < x,x' > )\quad \left\{x,x' \in \mathbb{S}^{p-1}\right\}$$

$$\implies ||\,\phi(x) - \phi(x')||^2_{\mathcal{H}} \leq ||x||^2 + ||\,x'||^2 - <x, x'> \leq ||\,x - x'||^2$$

8 4.2 **0.75 / 1**

- **- 0 pts** Correct
- **- 1 pts** major argument is missing
- ✓ **- 0.25 pts** need to be a bit more rigorous to handle the cases when x_i =0
- **- 0.5 pts** missing argument
- **- 1 pts** missing answer

*Exercice 4 :*

1) $K_1(x,x') = \sum a_i <x,x'>^i$ , $a_i \geq 0$.

*We know that* $K(x,x') = <x,x'>$ *is a p.d kernel.*

*And the product of p.d kernels is p.d, and multiplying by* $\alpha \geq 0$ *keeps it p.d*

*Thus* $\forall_i\ a_i <x,x'>^i$ *is p.d*

*Then* $K_1(x,x')$ *is p.d as the (infinite) sum of p.d kernels.*

2) $||x||\ ||x'||\ \kappa\left(\frac{<x,x'>}{||x||\ ||x'||}\right) =$

*If we take* $K_3(x,x') = ||x||\ ||x'||$ *then* $K_3$ *is p.d since*

$(x_1,..x_n) \in \left(\mathbb{S}^{p-1}\right)^n\ \forall a \text{ in } \mathbb{R}^n,\ \ \sum a_i K_3(x_i,x_j) a_j = \sum a_i\ ||x_i||\ ||x_j||\ a_j$

$= a'^T a' \geq 0$ *where* $a' = (a_1||x_1||, \ldots, a_n\ ||x_n||)$

*If we take* $K_1' = \kappa\left(\frac{<x,x'>}{||x||\ ||x'||}\right) = \kappa\left(<\frac{x}{||x||}, \frac{x'}{||x'||}>\right)$ $(if\ ||x||\ \&\ ||x'|| \neq 0)$

*and* $0$ *if* $||x||$ *or* $||x'|| = 0$

*then* $K_1'(x,x')$ *can be seen as* $K_1\left(\widetilde{x},\widetilde{x'}\right)$ *where* $\widetilde{x},\widetilde{x'} \in \left(\mathbb{S}^{p-1}\right)^2$, *therefore* $K_1'$ *is p.d*

*And thus* $K_2$ *is the product of* $K_1'$ *and* $K_3$ *which are two p.d kernels and thus it is p.d.*

3) $\kappa(u) = \kappa(1) - \int_u^1 \kappa'(u)du$

*Since* $u \leq 1$ , $\kappa'(u) = \sum_i i\,a_i\,u^i \leq \sum_i i\,a_i = \kappa'(1)$.

$$\Longrightarrow \int_u^1 \kappa'(u)du \leq \int_u^1 \kappa'(1).du \leq \kappa'(1)\int_u^1 du = \kappa'(1)(1-u) = (1-u)$$

$$\Longrightarrow \kappa(u) \geq 1 - (1-u) \geq u$$

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

$$= K(x,x) + K(x',x') - 2\,K(x,x') = 2\,(1 - \kappa(<x,x'>)\ \ \left\{x,x' \in \mathbb{S}^{p-1}\right\}$$

$$\implies ||\,\phi(x) - \phi(x')||^2_{\mathcal{H}} \;\leq\; ||x||^2 + ||\,x'||^2 - \;< x, x' > \;\leq\; ||\,x - x'||^2$$

### 9 4.3 2 / 2

- ✓ **- 0 pts** Correct
- **- 0.5 pts** missing minor argument
- **- 1 pts** there is a wrong argument but part of the proof is ok
- **- 2 pts** the proof relies on a wrong statement
- **- 1 pts** missing major argument
- **- 2 pts** missing answer

*Exercice 4. 4 :*

*If we take the mapping* $\mu(x) = \left(\sqrt{a_i}\, x^{\otimes i}\right)_{i\in\mathbb{N}}$

*Then we have that* $< \mu(x), \mu(y) > \; = \sum a_i \; < x, y >^{\,i}$

*But* $\mu(x)$ *is not really in* $l^2$*as each* $X^{\otimes i}$ *has a different dimension.*

*To go around this we can bring this expression into a suite of reals.*

*We know that* $X^{\otimes i} \in \mathbb{R}^{d^{\,i}}$.

*We create the following functions :*

$$\psi(i) = \mathbb{1}_{\,i > p} + \mathbb{1}_{i > p + p^2} + \mathbb{1}_{\,i > p+p^2+p^3} = \sum_{k}^{\infty} \mathbb{1}_{\,i > \sum_{j=1}^{k} p^j}$$

$\left(\text{to keep the same index } (i) \text{ over the tensor } x^{\otimes i}\right)$

$$\Gamma(i) = i \;\; modulo \; \sum_{j=1}^{\psi(i)} p^j \quad \left(\text{all the indexes of the tensor } x^{\otimes i},\ i \text{ from } 1 \text{ to } P^i\right)$$

*If we take* $\phi(x) = \left(\sqrt{a_{\psi(i)}}\ \left(x^{\otimes\psi(i)}\right)_{\Gamma(i)}\right)_{i\in\mathbb{N}}$ *then we get a suite of reals.*

$$\phi(x) \in l_2 \text{ since } < \phi(x),\ \phi(x) > \; = \sum a_i < x^{\otimes i}, x^{\otimes i} > \; = \sum a_i \; < x, x >^{\,i}$$

$$=_{\{\, x \,\in\, \mathbb{S}^{p-1}\}} \sum a_i = \kappa(1)$$

$$< \phi(x), \phi(y) > \; = \sum_i a_{\psi(i)} \left(x^{\otimes\psi(i)}\right)_{\Gamma(i)} \left(y^{\otimes\psi(i)}\right)_{\Gamma(i)} = \sum_i a_i < x^{\otimes i}, y^{\otimes i} >$$

$$= \sum a_i < x, y >^{\,i} = K(x, y)$$

10 4.4 **2 / 2**

- ✓ **- 0 pts** Correct
- **- 0.5 pts** missing details or minor mistake
- **- 2 pts** missing or wrong answer
- **- 1 pts** not exactly the right answer
- **- 0 pts** Click here to replace this description.

*Exercice* 4.5 :

$$H : \left\{ f_w \quad x \to \ < w, \phi(x) >_{l^2} : w \in l^2 \right\} \ / \ \phi(x) = \left( \sqrt{a_{\psi(i)}} \ \left(x^{\otimes \psi(i)}\right)_{\Gamma(i)} \right)_{i \in \mathbb{N}}$$

$$f_w = \ < w, \phi(x) >_{l^2} = \sum w_i \sqrt{a_{\psi(i)}} \ \left(x^{\otimes \psi(i)}\right)_{\Gamma(i)}$$

$$g_z = \sum b_i < z, x >^{i} = \sum b_i < z^{\otimes i}, x^{\otimes i} > \ = \sum b_{\psi(i)} \ \left(x^{\otimes \psi(i)}\right)_{\Gamma(i)} \left(z^{\otimes \psi(i)}\right)_{\Gamma(i)}$$

*We want* $g(z)$ *to be written as* $f_w$ :

$$\sum b_{\psi(i)} \ \left(x^{\otimes \psi(i)}\right)_{\Gamma(i)} \left(z^{\otimes \psi(i)}\right)_{\Gamma(i)} = \sum w_i \sqrt{a_{\psi}(i)} \ \left(x^{\otimes \psi(i)}\right)_{\Gamma(i)}.$$

*We can make the terms of the sums equal to get equality* :

$$\frac{b_{\psi(i)}}{\sqrt{a_{\psi}(i)}} \ \left(z^{\otimes \psi(i)}\right)_{\Gamma(i)} = w_i \quad \left\{ (w)_i \ \in l^2 \right\}$$

*A sufficient condition is* : $\dfrac{b_{\psi(i)}}{\sqrt{a_{\psi}(i)}} \in l^2$ *which can be rewritten as* : $\dfrac{b_i}{\sqrt{a_i}} \in l^2$,

*this requires all* $a_i > 0$.

*If* $(U)_i = \left( \dfrac{b_{\psi(i)}}{\sqrt{a_{\psi}(i)}} \right)_{i \in N}$ *in* $l^2$ *then* $(V)_i = \left( \dfrac{b_{\psi(i)}}{\sqrt{a_{\psi}(i)}} \ \left(z^{\otimes \psi(i)}\right)_{\Gamma(i)} \right)_{i \in N} \in l_2$ *because* :

$$||V||_{l^2} = \sum \frac{b^2{}_i}{a_i} < z^{\otimes i}, z^{\otimes i} > \ = \sum \frac{b^2{}_i}{a_i} < z, z >^{i} = \sum \frac{b^2{}_i}{a_i} = ||U||_{l^2} \quad \left\{ z \in \mathcal{S}^{P-1} \right\}$$

*Now* $g_z(x)$ *can be written as* $< V, \phi(x) >$ *where* $V \in l^2$. *And thus* $g_z \in \mathcal{H}$.

11 4.5 **2 / 2**

✓ **- 0 pts** Correct

**- 2 pts** missing or too incomplete answer

**- 1 pts** inaccurate or incomplete statement

**- 1.5 pts** the sufficient condition is too strong (only leads to a trivial case)

**- 0.5 pts** minor details missing