MVA "Kernel methods in machine learning" Homework 1

Souhaib Attaiki

March 2019

Exercise 1. Kernels

Question 1 on a:

$$\begin{aligned} K(x,y) &= cos(x-y) \\ &= cos(x)cos(y) + sin(x)sin(y) \\ &= < \phi(x), \phi(y) >_{\mathbb{R}^2} \end{aligned}$$

avec $\phi(x) = (cos(x), sin(x))$, un mapping de $\mathcal{X}$ vers l'espace de Hilbert $\mathbb{R}^2$, donc en appliquant le théorème d'Aronszajn, K est un kernel défini positif.

Question 2 on a, $\forall x, y \in \mathcal{X}$:

$$\begin{aligned} K(x,y) &= \frac{1}{1 - x^\top y} \\ &= \sum_{k=0}^{\infty} (x^\top y)^k \\ &= \lim_{n \to \infty} \sum_{k=0}^{n} (x^\top y)^k \\ &= \lim_{n \to \infty} K_n(x,y) \end{aligned}$$

parce que: $|x^\top y| \leq ||x||_2 ||y||_2 \leq 1$ (en utilisant Cauchy-Schwartz). Or, chaque K_n est une somme et produit de kernel linéaire, Donc K_n est un kernel défini positif. Donc K est un kernel défini positif comme limite simple de suite de kernels définis positifs.

Question 3 On définit, pour $A \in \mathcal{A}, \phi(A) = \mathbb{1}_A - \mathbb{E}(A) \in \mathbb{L}^2(\Omega, \mathcal{A}, \mathcal{P})$ (L'ensemble des variables aléatoires de carré intégrable) Donc, $\forall A, B \in \mathcal{A}$, on a:

$$\begin{aligned} K(A,B) &= P(A \cap B) - P(A)P(B) \\ &= \mathbb{E}(\mathbb{1}_A \mathbb{1}_B) - \mathbb{E}(\mathbb{1}_A)\mathbb{E}(\mathbb{1}_B) \\ &= \mathbb{E}[(\mathbb{1}_A - \mathbb{E}(A)).(\mathbb{1}_B - \mathbb{E}(B))] \\ &= \mathbb{E}[\phi(A)\phi(B)] \\ &= < \phi(A)|\phi(B) >_{\mathbb{L}^2(\Omega,\mathcal{A},\mathcal{P})} \end{aligned}$$

donc en appliquant le théorème d'Aronszajn, K est un kernel défini positif.

Question 4 Soient $x, y \in \mathcal{X}$, on a:

$$K(x,y) = min(f(x)g(y), f(y)g(x)) = g(x)g(y)min(\frac{f(x)}{g(x)}, \frac{f(y)}{g(y)}) = K_1(x,y)K_2(x,y)$$

avec:

$$K_1(x,y) = g(x)g(y) = < \phi_1(x), \phi_1(y) >_{\mathbb{R}}$$

avec $\phi_1(x) = g(x)$, donc selon le théorème d'Aronszajn, K_1 est un kernel défini positif. et:

$$K_2(x,y) = min(\frac{f(x)}{g(x)}, \frac{f(y)}{g(y)}) = \int_0^{+\infty} \mathbb{1}_{[0,\frac{f(x)}{g(x)}]} \mathbb{1}_{[0,\frac{f(y)}{g(y)}]} = < \phi_2(x), \phi_2(y) >_{\mathbb{L}^2(\mathbb{R}^+)}$$

avec $\phi_2(x) = \mathbb{1}_{[0,\frac{f(x)}{g(x)}]} \in \mathbb{L}^2(\mathbb{R}^+)$, donc selon le théorème d'Aronszajn, K_2 est un kernel défini positif. Donc K est kernel défini positif comme produit de deux kernel définis positifs.

Question 5 On note $N = |E|$. Par abus de notation, on peut désigne par $A \in \mathcal{X}$ aussi le vecteur $A \in \mathbb{R}^N$ tel que $A_i = \mathbb{1}_A(x_i)$, le i-eme élément de E. soient $A, B \in \mathcal{X}$, on a:

$$\begin{aligned} K(A,B) &= \frac{|A \cap B|}{|A \cup B|} \\ &= \frac{\sum_{x \in E} \mathbb{1}_A(x)\mathbb{1}_B(x)}{\sum_{x \in E} \mathbb{1}_E(x) - (\mathbb{1}_E(x) - \mathbb{1}_A(x))(\mathbb{1}_E(x) - \mathbb{1}_B(x))} \\ &= A^\top B \times \frac{1}{N} \frac{1}{1 - \frac{(\mathbb{1}-A)^\top}{\sqrt{N}} \frac{\mathbb{1}-B}{\sqrt{N}}} \text{ avec } \mathbb{1} = (1,1,...,1) \\ &= \frac{1}{N} < \phi(A)|\phi(B) >_{\mathbb{R}^N} \times \sum_{k=0}^{\infty} \left(\left(\frac{\mathbb{1}-A}{\sqrt{N}} \right)^\top \frac{\mathbb{1}-B}{\sqrt{N}} \right)^k \\ &= K_1(A,B)K_2(A,B) \end{aligned}$$

avec: $K_1(A,B) = \frac{1}{N} < \phi(A)|\phi(B) >_{\mathbb{R}^N}$ et $\phi : \mathcal{X} \to \mathbb{R}^N, A \to A$, donc K_1 et un kernel défini positif comme résultat du théorème d'Aronszajn et produit d'un kernel et une constante positive. Et:

$$K_2(A,B) = \sum_{k=0}^{\infty} \left(\left(\frac{\mathbb{1}-A}{\sqrt{N}} \right)^\top \frac{\mathbb{1}-B}{\sqrt{N}} \right)^k = \lim_{n \to \infty} \sum_{k=0}^{n} \left(\left(\frac{\mathbb{1}-A}{\sqrt{N}} \right)^\top \frac{\mathbb{1}-B}{\sqrt{N}} \right)^k = \lim_{n \to \infty} K_n(A,B)$$

Or, chaque K_n est une somme et produit du kernel $K^{'}(A,B) = \left(\frac{\mathbb{1}-A}{\sqrt{N}}\right)^{\top} \frac{\mathbb{1}-B}{\sqrt{N}} =< \phi^{'}(A), \phi^{'}(B) >_{\mathbb{R}^N}$ qui est un kernel selon le théorème d'Aronszajn, Donc K_n est un kernel défini positif. Donc K_2 est un kernel défini positif comme limite simple de suite de kernels définis positifs.

D'ou K est un kernel défini positif comme produit de deux kernel définis positifs (K_1 et K_2).

Exercise 2. RKHS

Question 1 Montrons que $\alpha K_1 + \beta K_2$ est un kernel positif défini. La symétrie est évidente et découle de la symétrie de K_1 et de K_2. Et on a, $\forall n \in \mathbb{N}, \{a_i\}_{1\leq i\leq n} \in \mathbb{R}, \{x_i\}_{1\leq i\leq n} \in \mathcal{X}$,

$$\begin{aligned} \sum_{i,j=1}^{n} a_i a_j K(a_i, a_j) &= \sum_{i,j=1}^{n} a_i a_j \left(\alpha K_1(a_i,a_j) + \beta K_2(a_i,a_j)\right) \\ &= \alpha \sum_{i,j=1}^{n} a_i a_j K_1(a_i,a_j) + \beta \sum_{i,j=1}^{n} a_i a_j K_2(a_i,a_j) \geq 0 \end{aligned}$$

parce que $\alpha, \beta \geq 0$ et K_1 et K_2 sont des kernels défini positif. Donc $\alpha K_1 + \beta K_2$ est un kernel défini positif.

Soient H_1 (resp. H_2) le RKHS de K_1 (resp. K_2). Montrons que $H = \alpha H_1 + \beta H_2 = \{\alpha f + \beta g / f \in H_1, g \in H_2\}$ est le RKHS de $\alpha K_1 + \beta K_2$.

Soit l'espace $E = H_1 \times H_2$ qui est muni du produit scalaire $< (f_1, g_1), (f_2, g_2) >_E = \alpha < f_1, f_2 >_{H_1} + \beta < g_1, g_2 >_{H_1}$ est un espace hilbertien (résultat classique).

On a l'application s

$$\begin{aligned} s :& E \to H \\ &(f,g) \to \alpha f + \beta g \end{aligned}$$

est une surjection entre E et H. Considérons $N = s^{-1}(\{0\})$. N est un ensemble fermé. En effet, soit $(f_n, -f_n)$ une suite d'éléments de N qui converge dans E vers (f,g). Par définition de la norme, on a (f_n) converge dans H_1 vers f et $(-f_n)$ converge dans H_2 vers g. Puisque la convergence dans le RKHS implique convergence ponctuel, on a $f = -g$.et donc $(f,g) \in N$, Donc N est fermé. Donc $E = N \oplus N^{\top}$. Donc $\hat{s}$, la restriction de s à $N^{\top}$ est une bijection. Donc H est un espace de Hilbert pour le produit scalaire qu'on définit.

Et on a $\forall x \in \mathcal{X}, \alpha K_1(x,.) + \beta K_2(x,.) \in H$. Il reste donc juste à vérifier la propriété de reproduction.

Soit $x \in \mathcal{X}$ et $f \in H$. On a $f = \hat{s}(f_1, f_2)$ et $K_x = \hat{s}(A_x, B_x)$ où $K_x = \alpha K_{1,x} + \beta K_{2,x}$ et (f_1, f_2) et (A_x, B_x) sont des éléments de $N^{\top}$. On a:

$$< f | K_x >_H =< (f_1, f_2) | (A_x, B_x) >_E =< (f_1, f_2) | (K_{1,x}, K_{2,x}) >_E - < (f_1, f_2) | (A_x - K_{1,x}, B_x - K_{2,x}) >_E$$

et puisque $(A_x - K_{1,x}, B_x - K_{2,x})$ est orthogonal à chaque élément de $N^{\top}$, on conclut que

$$< f | K_x >_H =< (f_1, f_2) | (K_{1,x}, K_{2,x}) >_E = \alpha f(x) + \beta g(x) = f(x)$$

d'ou le résultat. On a montré que $H = \alpha K_1 + \beta K_2$ muni du produit scalaire $< f | g >_H =< \hat{s}^{-1}(f) | \hat{s}^{-1}(g) >_E$ est un RKHS de $\alpha K_1 + \beta K_2$.

Question 2 Le kernel K vérifie les conditions du théorème d'Aronszajn, donc K est un kernel défini positif.

Exercise 3. Sobolev spaces

Question 1 Montrons que $\mathcal{H} = \{f : [0,1] \to \mathbb{R}, \text{absolument continue}, f^{'} \in L^2([0,1]), f(0) = 0\}$ muni de la forme bilinéaire $< f, g >_{\mathcal{H}} = \int_0^1 f'(u)g'(u)du$, est un RKHS, et que $K(x,y) = min(x,y)$ est son kernel reproduisant.

- $\mathcal{H}$ est un espace de pré-hilbertien. En effet, $\mathcal{H}$ est un espace vectoriel de fonctions, et la forme bilinéaire $< f, g >_{\mathcal{H}}$ est symétrique, et positive (évident). De plus, on a

  $$\forall f \in \mathcal{H}, \forall x \in [0,1] : f(x) = f(0) + \int_0^x f^{'}(u)du = \int_0^x f^{'}(u)du$$

  car f est absolument continue et $f(0) = 0$. donc:

  $$|f(x)| = |\int_0^x f^{'}(u)du| \leq \sqrt{x}\left(\int_0^1 f^{'}(u)^2 du\right)^{\frac{1}{2}} = \sqrt{x} < f, f >_{\mathcal{H}}^{\frac{1}{2}} \tag{1}$$

  donc $< f, f >_{\mathcal{H}} = 0 \Rightarrow f(x) = 0, \forall x \in [0,1] \Rightarrow f = 0$, donc $< .,. >_{\mathcal{H}}$ est un produit scalaire, donc $\mathcal{H}$ est un espace pré-hilbertien.

- $\mathcal{H}$ est complet. En effet, soit $(f_n)_{n\in\mathbb{N}}$ est suite de Cauchy dans $\mathcal{H}$. Par définition de la norme dans $\mathcal{H}$ ($||f||_{\mathcal{H}} = ||f^{'}||_{L^2}$), $(f_n^{'})_{n\in\mathbb{N}}$ est une suite de Cauchy dans $L^2[0,1]$, donc converge vers une limite $g \in L^2[0,1]$ car $L^2[0,1]$ est complet. Or, en utilisant l'équation 1, on a que $(f(x)_n)_{n\in\mathbb{N}}$ est une suite de Cauchy dans $\mathbb{R}, \forall x \in [0,1]$, donc elle converge vers $f(x)$ car $\mathbb{R}$ est complet. Et on a:

  $$f(x) = \lim_n f_n(x) = \lim_n \int_0^x f_n^{'}(u)du = \int_0^x g(u)du \tag{2}$$

  en utilisant le théorème de convergence dominée, la domination est assuré par le fait que les fonctions sont continue sur un compact, donc elles sont bornées.

  L'équation 2 montre que f est absolument continue, que $f^{'} = g$ presque partout, de plus, $f^{'} \in L^2[0,1]$. Finalement, on a: $f(0) = \lim_n f_n(0) = 0$, ainsi $f \in \mathcal{H}$, et

  $$\lim_n ||f_n - f||_{\mathcal{H}} = \lim_n ||f_n^{'} - g||_{L^2[0,1]} = 0$$

  donc $\lim_n f_n = f$. Donc $\mathcal{H}$ est complet, d'où le résultat.

- On a $\forall x, y \in [0,1]$:

  $$K_x(y) = K(x,y) = min(x,y) = \begin{cases} y & \text{si } 0 \leq y \leq x \\ x & \text{si } x \leq y \leq 1 \end{cases}$$

  Donc K_x est différentiable presque partout, sa dérivé $\in L^2[0,1]$, et $K_x(0) = 0$ pour tout $x \in [0,1]$, donc $K_x \in \mathcal{H}, \forall x \in [0,1]$.

- Soit x $\in [0,1]$, et $f \in \mathcal{H}$, on a:

$$< f, K_x >_{\mathcal{H}} = \int_0^1 f^{'}(u)K_x^{'}(u)du = \int_0^x f^{'}(u)du = f(x)$$

  donc $\forall f \in \mathcal{H}, \forall x \in [0,1] :< f, K_x >_{\mathcal{H}} = f(x)$.

D'ou $\mathcal{H}$ est un RKHS, et son kernel reproduisant est $K(x,y) = min(x,y)$.

Question 2 Montrons que $\mathcal{H} = \{f : [0,1] \to \mathbb{R}, \text{absolument continue}, f^{'} \in L^2([0,1]), f(0) = f(1) = 0\}$ muni de la forme bilinéaire $< f, g >_{\mathcal{H}} = \int_0^1 f'(u)g'(u)du$, est un RKHS, et que $K(x,y) = min(x,y) - xy$ est son kernel reproduisant.

- On a $\mathcal{H}$ est une espace pré-hilbertien (démonstration similaire au cas précédent)

- La démonstration de $\mathcal{H}$ est complet est similaire au cas précédent, il suffit d'ajouter le fait que $f(1) = \lim_n f_n(1) = 0$.

- On a $\forall x, y \in [0,1]$:

$$K_x(y) = K(x,y) = min(x,y) - xy = \begin{cases} y - xy & \text{si } 0 \leq y \leq x \\ x(1-y) & \text{si } x \leq y \leq 1 \end{cases}$$

  Donc K_x est différentiable presque partout, sa dérivé $\in L^2[0,1]$, et $K_x(0) = K_x(1) = 0$ pour tout $x \in [0,1]$, donc $K_x \in \mathcal{H}, \forall x \in [0,1]$.

- Soit x $\in [0,1]$, et $f \in \mathcal{H}$, on a:

$$\begin{aligned} < f, K_x >_{\mathcal{H}} &= \int_0^1 f^{'}(u)K_x^{'}(u)du \\ &= \int_0^x f^{'}(u)(1-x)du + \int_x^1 f^{'}(u)(-x)du \\ &= (1-x)(f(x) - f(0)) - x(f(1) - f(x)) \\ &= (1 - x + x)f(x) \text{ (car } f(0) = f(1) = 0 \text{)} \\ &= f(x) \end{aligned}$$

  donc $\forall f \in \mathcal{H}, \forall x \in [0,1] :< f, K_x >_{\mathcal{H}} = f(x)$.

- Pour trouver la forme du kernel, on résout l'équation suivante:

$$f(x) =< f, K_x >= \int_0^1 f^{'}(u)K_x^{'}(u)du$$

  sous contrainte que $K_x(0) = K_x(1) = 0$. On trouve que le kernel de la question précédente permet de reproduire f, et pour forcer $K_x(1) = 0$, on ajoute le terme $-xy$.

D'ou $\mathcal{H}$ est un RKHS, et son kernel reproduisant est $K(x,y) = min(x,y) - xy$.

Question 3 Montrons que $\mathcal{H} = \{f : [0,1] \to \mathbb{R}, \text{absolument continue}, f^{'} \in L^2([0,1]), f(0) = f(1) = 0\}$ muni de la forme bilinéaire $< f, g >_{\mathcal{H}} = \int_0^1 (f'(u)g'(u) + f(u)g(u))\, du$, est un RKHS, et que $K(x,y) = \frac{min(sinh(x)sinh(1-y), sinh(1-x)sinh(y))}{sinh(1)}$ est son kernel reproduisant.

- Pour montrer que $\mathcal{H}$ est un espace pré-hilbertien, il suffit de montrer que $< f, g >_{\mathcal{H}}$ est un produit scalaire (le reste au similaire qu'avant). On a $< f, g >_{\mathcal{H}}$ est symétrique et positive (évident). Et comme avant, on a:

$$\begin{aligned} |f(x)| &= |\int_0^x f^{'}(u)du| \\ &\leq \sqrt{x}\left(\int_0^1 f^{'}(u)^2 du\right)^{\frac{1}{2}} \\ &\leq \sqrt{x}\left(\int_0^1 \left(f^{'}(u)^2 + f(u)^2\right) du\right)^{\frac{1}{2}} \\ &= \sqrt{x} < f, f >_{\mathcal{H}}^{\frac{1}{2}} \end{aligned} \tag{3}$$

  donc $< f, f >_{\mathcal{H}} = 0 \Rightarrow f(x) = 0, \forall x \in [0,1] \Rightarrow f = 0$, donc $< ., . >_{\mathcal{H}}$ est un produit scalaire, donc $\mathcal{H}$ est un espace pré-hilbertien.

- $\mathcal{H}$ est complet. En effet, soit $(f_n)_{n\in\mathbb{N}}$ est suite de Cauchy dans $\mathcal{H}$. On a

$$||f||_{\mathcal{H}} = \left(\int_0^1 \left(f^{'}(u)^2 + f(u)^2\right) du\right)^{\frac{1}{2}} = \left(||f^{'}||^2_{\mathbb{L}^2} + ||f||^2_{\mathbb{L}^2}\right)^{\frac{1}{2}}$$

  donc $||f^{'}||^2_{\mathbb{L}^2} \leq ||f||_{\mathcal{H}}$ et $||f||^2_{\mathbb{L}^2} \leq ||f||_{\mathcal{H}}$, donc $(f_n)_{n\in\mathbb{N}}$ est une suite de Cauchy dans $L^2[0,1]$, et converge vers une limite $f \in L^2[0,1]$, et $(f_n^{'})_{n\in\mathbb{N}}$ est une suite de Cauchy dans $L^2[0,1]$, et converge vers une limite $g \in L^2[0,1]$. Par ailleurs, la convergence dans L^2 implique la convergence dans $\mathcal{D}^{'}[0,1]$ (l'espace vectoriel des distributions sur $[0,1]$), donc

$$f_n \to f \text{ dans } \mathcal{D}^{'}$$

  et

$$f_n^{'} \to g \text{ dans } \mathcal{D}^{'}$$

  De la première assertion, on tire que

$$f_n^{'} \to f^{'} \text{ dans } \mathcal{D}^{'}$$

  Par unicité de la limite dans $\mathcal{D}^{'}[0,1]$, on trouve que $f^{'} = g \in L^2$. Et comme $f_n^{'} \to f^{'}$ dans L^2, il s'en suit que $f_n \to f$ dans $\mathcal{H}$.

  En outre, en utilisant l'inéquation $|f(x)| \leq ||f||_{\mathcal{H}}, \forall x \in [0,1]$, on a que $(f(x)_n)_{n\in\mathbb{N}}$ est une suite de Cauchy dans $\mathbb{R}, \forall x \in [0,1]$, donc elle converge vers $h(x)$ car $\mathbb{R}$ est complet. Et on a:

$$h(x) = \lim_n f_n(x) = \lim_n \int_0^x f_n^{'}(u)du = \int_0^x g(u)du \tag{4}$$

  en utilisant le théorème de convergence dominée comme avant.

  L'équation 4 montre que $h^{'} = g = f^{'}$ presque partout, donc $f(x) - f(0) = h(x) - h(0)$. De plus, on a: $h(0) = \lim_n f_n(0) = 0$, donc $f(0) = 0$, de même pour $f(1) = 0$.

- On a $\forall x, y \in [0,1]$:

$$K_x(y) = K(x,y) = \begin{cases} \frac{sinh(1-x)sinh(y)}{sinh(1)} & \text{si } 0 \leq y \leq x \leq 1 \\ \frac{sinh(x)sinh(1-y)}{sinh(1)} & \text{si } 0 \leq x \leq y \leq 1 \end{cases}$$

  Donc K_x est différentiable presque partout, sa dérivé $\in L^2[0,1]$, et $K_x(0) = K_x(1) = 0$ pour tout $x \in [0,1]$, donc $K_x \in \mathcal{H}, \forall x \in [0,1]$.

- Soit x $\in [0,1]$, et $f \in \mathcal{H}$, on a:

$$\begin{aligned}
\int_0^1 f^{'}(u)K_x^{'}(u)du &= \int_0^x f^{'}(u)\frac{sinh(1-x)cosh(u)}{sinh(1)}du - \int_x^1 f^{'}(u)\frac{sinh(x)cosh(1-u)}{sinh(1)}du \\
&= \frac{sinh(1-x)}{sinh(1)}\left([f(u)cosh(u)]_0^x - \int_0^x f(u)sinh(u)du\right) \\
&- \frac{sinh(x)}{sinh(1)}\left([f(u)cosh(1-u)]_x^1 - \int_x^1 f(u)sinh(1-u)du\right) \\
&= \frac{f(x)sinh(1-x)cosh(x)}{sinh(1)} + \frac{f(x)sinh(x)cosh(1-x)}{sinh(1)} - \int_0^1 f(u)K_x(u)du
\end{aligned}$$

  Donc

$$\begin{aligned}
< f, K_x >_{\mathcal{H}} &= f(x)\left[\frac{sinh(1-x)cosh(x)}{sinh(1)} + \frac{sinh(x)cosh(1-x)}{sinh(1)}\right] \\
&= f(x)
\end{aligned}$$

  donc $\forall f \in \mathcal{H}, \forall x \in [0,1] :< f, K_x >_{\mathcal{H}} = f(x)$.

- Pour trouver la forme du kernel, on résout l'équation suivante:

$$f(x) = \int_0^1 f^{'}(u)K^{'}(x,u)du + \int_0^1 f(u)K(x,u)du$$

  en intergrant par partie, en supposant une discontuité en u = x

  et en prenant en compte f(0)=f(1)=0

$$= \int_0^1 f(u)K(x,u)du - \int_0^1 f(u)K^{''}(x,u)du + f(x)(K(x,x^-) - K(x,x^+))$$

  Donc en cherche à résoudre l'équation différentielle suivante: $K^{''}(x,u) = K(x,u), \forall x, u$, avec les conditions suivantes: $\frac{\partial K(x,u)}{\partial u}_{|u=x^-} = \alpha(x)$, $\frac{\partial K(x,u)}{\partial u}_{|u=x^+} = \beta(x)$, $K_x(0) = K_x(1) = 0$, et $\alpha(x) - \beta(x) = 1$. En résolvant cette équation, on trouve la forme du kernel mentionnée plus haut.

Exercise 4. Duality

Question a On veut résoudre le problème d'optimisation suivant:

$$\min_{f \in \mathcal{H}_K} \frac{1}{n}\sum_{i=1}^{n} \ell_{y_i}(f(x_i)) \text{ such that } ||f||_{\mathcal{H}_K} \leq B \tag{5}$$

Son lagrangien s'écrit sous la forme suivante:

$$\mathcal{L}(f,\Lambda) = \frac{1}{n}\sum_{i=1}^{n} \ell_{y_i}(f(x_i)) + \Lambda(||f||_{\mathcal{H}_K} - B)$$

Puisque ℓ_y est convexe, alors le problème 5 est un problème d'optimisation convexe, et pour lequel la dualité forte est vérifié en utilisant les conditions de Slater (évident), donc le problème 5 est équivalent au problème suivant:

$$\min_{f\in\mathcal{H}_K} \frac{1}{n}\sum_{i=1}^{n} \ell_{y_i}(f(x_i)) + \lambda(||f||^2_{\mathcal{H}_K} - B^2)$$

qui est équivalent aussi à (parce qu'on veut minimiser sur f):

$$\min_{f\in\mathcal{H}_K} \frac{1}{n}\sum_{i=1}^{n} \ell_{y_i}(f(x_i)) + \lambda||f||^2_{\mathcal{H}_K} \tag{6}$$

avec $\lambda \geq 0$, la solution optimale du problème dual de 5. Le problème 6 vérifie les conditions du "Representer theorem", donc en utilisant ce dernier, on peut écrire que toute solution du problème s'écrit sous la forme:

$$\hat{f}(x) = \sum_{i=1}^{n} \alpha_i \boldsymbol{K}(x_i, x_j)$$

avec $\boldsymbol{K}$ la matrice de Gram. Et on a:

$$\left(\hat{f}(x_1), ..., \hat{f}(x_n)\right)^\top = \boldsymbol{K}\boldsymbol{\alpha}$$

et

$$||\hat{f}||^2_{\mathcal{H}} = \boldsymbol{\alpha}^\top \boldsymbol{K}\boldsymbol{\alpha}$$

Donc le problème 6 est équivalent à:

$$\min_{\alpha\in\mathbb{R}^n} \frac{1}{n}\sum_{i=1}^{n} \ell_{y_i}(\boldsymbol{K}\boldsymbol{\alpha}_i) + \lambda\boldsymbol{\alpha}^\top\boldsymbol{K}\boldsymbol{\alpha}$$

et donc à:

$$\min_{\alpha\in\mathbb{R}^n} R(\boldsymbol{K}\boldsymbol{\alpha}) + \lambda\boldsymbol{\alpha}^\top\boldsymbol{K}\boldsymbol{\alpha} \tag{7}$$

avec $R(u) = \frac{1}{n}\sum_{i=1}^{n} \ell_{y_i}(u_i)$.

Question b On a d'abord que pour $c > 0$:

$$\sup_{x\in\mathbb{R}^n} x^\top u - cf(x) = \sup_{x\in\mathbb{R}^n} c\left(x^\top \frac{u}{c} - f(x)\right) = c \sup_{x\in\mathbb{R}^n} x^\top \frac{u}{c} - f(x) = cf^*(\frac{u}{c})$$

et on a:

$$
\begin{aligned}
R^*(u) &= \sup_{x\in\mathbb{R}^n} x^\top u - R(x) \\
&= \sup_{x\in\mathbb{R}^n} x^\top u - \frac{1}{n}\sum_{i=1}^n \ell_{y_i}(u_i) \\
&= \sup_{x\in\mathbb{R}^n} \sum_{i=1}^n x_i u_i - \frac{1}{n}\ell_{y_i}(u_i) \\
&= \sum_{i=1}^n \sup_{x_i\in\mathbb{R}} x_i u_i - \frac{1}{n}\ell_{y_i}(u_i) \text{ (Car les différents termes sont indépendants)} \\
&= \sum_{i=1}^n \frac{1}{n}\ell^*_{y_i}(nu_i)
\end{aligned}
$$

D'où le résultat.

Question c Le lagrangien du problème:

$$
\min_{\alpha,u\in\mathbb{R}^n} R(u) + \lambda\alpha^\top K\alpha \text{ such that } u = K\alpha \tag{8}
$$

s'écrit de la façon suivante:

$$
\mathcal{L}(\alpha,u,\mu) = R(u) + \lambda\alpha^\top K\alpha + \mu^\top(K\alpha - u)
$$

la fonction dual s'écrit donc sous la forme:

$$
\begin{aligned}
g(\mu) &= \inf_{\alpha,u}\mathcal{L} \\
&= \inf_{\alpha,u} R(u) + \lambda\alpha^\top K\alpha + \mu^\top(K\alpha - u) \\
&= \inf_{\alpha,u} \big(\, R(u) - \mu^\top u\big) + \big(\lambda\alpha^\top K\alpha + \mu^\top K\alpha\big) \\
&= \inf_{u} \big(\, R(u) - \mu^\top u\big) + \inf_{\alpha} \big(\lambda\alpha^\top K\alpha + \mu^\top K\alpha\big) \\
&= -R^*(\mu) + \inf_{\alpha} h(\alpha) \\
&= -R^*(\mu) - \frac{\mu^\top K\mu}{4\lambda}
\end{aligned}
$$

avec $\inf_\alpha h(\alpha)$ est obtenu en prenant la dérivé de h égal à zéro, et on trouve que: $\alpha = \frac{-\mu}{2\lambda}$. Donc le problème dual s'écrit sous la forme:

$$
\max_{\mu\in\mathbb{R}^n} -R^*(\mu) - \frac{\mu^\top K\mu}{4\lambda}
$$

On peut trouver une solution du problème 8 en utilisant la relation qu'on a trouvé lors de la dérivation de h: $\alpha^* = \frac{-\mu^*}{2\lambda}$.

Question d

- Pour la fonction de coût logistique $\ell_y(u) = \log(1 + e^{-yu})$, on a:

$$
\begin{aligned}
\ell_y^*(u) &= \sup_{x\in\mathbb{R}} xu - \ell_y(x) \\
&= \sup_{x\in\mathbb{R}} xu - \log(1 + e^{-yx}) \\
&= \sup_{x\in\mathbb{R}} \log(e^{xu}) - \log(1 + e^{-yx}) \\
&= \sup_{x\in\mathbb{R}} \log(\frac{e^{xu}}{1 + e^{-yx}}) \\
&= \sup_{x\in\mathbb{R}} \log(\frac{1}{e^{-xu} + e^{-(y+u)x}}) \\
&= -\inf_{x\in\mathbb{R}} \log(e^{-xu} + e^{-(y+u)x}) \\
&= -\log(\inf_{x\in\mathbb{R}} e^{-xu} + e^{-(y+u)x}) \text{ parce que log est croissant} \\
&= -\log(\inf_{x\in\mathbb{R}} e^{xy(-uy)} + e^{xy(-1-uy)}) \text{ parce que } y^2 = 1 \\
&= -\log(\inf_{t\in\mathbb{R}^+} t^a + t^{a-1}) \text{ avec } t = e^{xy} \geq 0 \text{ et } a = -uy \\
&= -\log(\begin{cases} \frac{1}{a^a(1-a)^{1-a}} & \text{si } 0 < a < 1 \\ 0 & \text{sinon} \end{cases}) \text{ en utilisant (*) en bas} \\
&= \begin{cases} a\log(a) + (1-a)\log(1-a) & \text{si } 0 < a < 1 \\ +\infty & \text{sinon} \end{cases}
\end{aligned}
$$

avec (*)

$$
\inf_{x\geq 0} y(x) = x^a + x^{a-1} = \begin{cases} \frac{1}{a^a(1-a)^{1-a}} & \text{si } 0 < a < 1 \\ 0 & \text{sinon} \end{cases}
$$

Ce résultat est facile à démontrer, il suffit juste de séparer les cas et voir que $y(x) \geq 0, \forall x \geq 0$. Pour $a \leq 0$ et donc $a - 1 \leq 0$, on a $y \xrightarrow[x\to+\infty]{} 0$. Pour $a \geq 1$ et donc $a - 1 \geq 0$, on a $y \xrightarrow[x\to 0]{} 0$. Et pour $0 < x < 1$, en cherchant le point d'annulation de la dérivée, on trouve qu'il est égal à $x = \frac{1-a}{a}$, ce qui donne le résultat.

Le problème dual s'écrit donc sous la forme:

$$
\max_{u\in\mathbb{R}^n} -\frac{1}{n}\sum_{i=1}^{n} (-nu_iy_i\log(-nu_iy_i) + (1 + nu_iy_i)\log(1 + nu_iy_i)) - \frac{u^\top Ku}{4\lambda}
$$

sous contrainte que $\frac{-1}{n} < u_iy_i < 0, i = 0, ..., n$

- Pour le squared hinge loss $\ell_y(u) = max(0, 1-yu)^2$, on a:

$$
\begin{aligned}
\ell_y^*(u) &= \sup_{x\in\mathbb{R}} xu - \ell_y(x) \\
&= \sup_{x\in\mathbb{R}} xu - max(0, 1-yx)^2 \\
&= \begin{cases} xu - (1-yx)^2 & \text{si } yx \leq 1 \\ xu & \text{si } yx \geq 1 \end{cases} \\
&= \begin{cases} \xrightarrow[x\to\infty]{} \infty \text{ si } u > 0 & \text{si } u \geq 0 \text{ et } y = 1 \\ -u + \frac{u^2}{4} \text{ (par derivation et séparation des cas)} & \text{si } u \geq 0 \text{ et } y = -1 \\ u + \frac{u^2}{4} \text{ (par derivation et séparation des cas)} & \text{si } u \leq 0 \text{ et } y = 1 \\ \xrightarrow[x\to-\infty]{} \infty \text{ si } u < 0 & \text{si } u \leq 0 \text{ et } y = -1 \end{cases} \\
&= \begin{cases} yu + \frac{u^2}{4} & \text{si } yu \leq 0 \\ +\infty & \text{si } yu > 0 \end{cases}
\end{aligned}
$$

Donc le problème dual s'écrit donc sous la forme:

$$\max_{u\in\mathbb{R}^n} -\frac{1}{n}\sum_{i=1}^{n}\left(ny_iu_i + \frac{(nu_i)^2}{4}\right) - \frac{u^\top K u}{4\lambda}$$

sous contrainte que $u_iy_i \leq 0, i = 0, ..., n$