Homework

Exercise 1. Show that the following kernels are positive definite:

1. On $\mathcal{X} = \mathbb{R}$:
$$\forall x, y \in \mathbb{R}, \qquad K(x,y) = \cos(x-y).$$
2. On $\mathcal{X} = \{x \in \mathbb{R}^p : \|x\|_2 < 1\}$:
$$\forall x, y \in \mathcal{X}, \qquad K(x,y) = 1/\left(1 - x^\top y\right).$$
3. Given a probability space $(\Omega, \mathcal{A}, \mathbb{P})$, on $\mathcal{X} = \mathbb{R}$:
$$\forall x, y \in \mathbb{R}, \qquad K(A,B) = \mathbb{P}(A \cap B) - \mathbb{P}(A)\mathbb{P}(B).$$
4. Let $\mathcal{X}$ be a set and $f, g : \mathcal{X} \mapsto \mathbb{R}_+$ two non-negative functions:
$$\forall x, y \in \mathcal{X}, \qquad K(x,y) = \min\{f(x)g(y), f(y)g(x)\}.$$
5. Given a non-empty finite set E, on $\mathcal{X} = P(E) = \{A : A \subset E\}$:
$$\forall A, B \subset E \qquad K(A,B) = \frac{|A \cap B|}{|A \cup B|}$$
where $|F|$ denotes the cardinality of F, and with the convention $\frac{0}{0} = 0$.

Solution 1. Let $n \in \mathbb{N}$, $x_1, \ldots, x_n \in \mathcal{X}$ (or $A_1, \ldots, A_n \in \mathcal{X}$ depending of the context) and $c_1, \ldots, c_n \in \mathbb{R}$.

1. First of all, by using the evenness of the cosine function, we get immediately that K is symmetric. Then by using Ptolemy's identity
$$\cos(x-y) = \cos(x)\cos(y) + \sin(x)\sin(y),$$
we get
$$\begin{aligned}
\sum_{1 \le i,j \le n} a_i a_j K(x_i, x_j) &= \sum_{1 \le i,j \le n} a_i a_j \cos(x_i - x_j) \\
&= \sum_{1 \le i,j \le n} a_i a_j \left\{\cos(x_i)\cos(x_j) + \sin(x_i)\sin(x_j)\right\} \\
&= \sum_{1 \le i,j \le n} a_i a_j \cos(x_i)\cos(x_j) + \sum_{1 \le i,j \le n} a_i a_j \sin(x_i)\sin(x_j) \\
&= \left(\sum_{1 \le i \le n} a_i \cos(x_i)\right)^2 + \left(\sum_{1 \le i \le n} a_i \sin(x_i)\right)^2 \ge 0
\end{aligned}$$
2. By writing K as following
$$K(x,y) = \lim_{n \to +\infty} \sum_{k=0}^{n} \left\langle x, y \right\rangle_{\mathbb{R}^p}^k$$
and using the stability of positive definite kernels by multiplication, finite sum and limit, we get the desired result.

3. We observe immediately that K is symmetric. By introducing indicator random variable $\mathbb{1}_A : \Omega \to \{0,1\}, \omega \mapsto 1$ *if and only if* $\omega \in A$ for $A \in \mathcal{A}$, we can rewrite K as

$$\begin{aligned} K(A,B) &= \mathbb{E}[\mathbb{1}_{A\cap B}] - \mathbb{E}[\mathbb{1}_A]\cdot\mathbb{E}[\mathbb{1}_B] \\ &= \mathbb{E}[\mathbb{1}_A\cdot\mathbb{1}_B] - \mathbb{E}[\mathbb{1}_A]\cdot\mathbb{E}[\mathbb{1}_B] \\ &= \mathrm{Cov}[\mathbb{1}_A\,,\,\mathbb{1}_B]. \end{aligned}$$

We conclude by the bilinearity of the covariance function:

$$\sum_{1\le i,j\le n} a_i a_j K\left(A_i,A_j\right) = \sum_{1\le i,j\le n} a_i a_j \mathrm{Cov}\left[\mathbb{1}_{A_i}\,,\,\mathbb{1}_{A_j}\right] = \mathrm{Cov}\left[\sum_{1\le i\le n} a_i \mathbb{1}_{A_i}\,,\,\sum_{1\le j\le n} a_j \mathbb{1}_{A_j}\right] = \mathrm{Var}\left[\sum_{1\le i\le n} a_i \mathbb{1}_{A_i}\right] \ge 0.$$

4. Again, K is clearly symmetric. Using the same notation as above

$$\begin{aligned} \sum_{1\le i,j\le n} \alpha_i\alpha_j K(x_i,x_j) &= \sum_{1\le i,j\le n} \alpha_i\alpha_j \int_{\mathbb{R}_+} \mathbb{1}_{u\le f(x_i)g(x_j)}\cdot\mathbb{1}_{u\le f(x_j)g(x_i)}\,\mathrm{d}u \\ &= \sum_{1\le i,j\le n:\ f(x_i)f(x_j)>0} \alpha_i\alpha_j \int_{\mathbb{R}_+} \mathbb{1}_{u\le f(x_i)g(x_j)}\cdot\mathbb{1}_{u\le f(x_j)g(x_i)}\,\mathrm{d}u \\ &= \sum_{\{1\le i,j\le n:\ f(x_i)f(x_j)>0\}} \alpha_i\alpha_j \int_{\mathbb{R}_+} \mathbb{1}_{t\le\frac{g(x_j)}{f(x_j)}}\cdot\mathbb{1}_{t\le\frac{g(x_i)}{f(x_i)}} f(x_i)f(x_j)\,\mathrm{d}t \qquad u = \left(f(x_i)f(x_j)\right)t \\ &= \int_{\mathbb{R}_+}\left[\sum_{i:\ f(x_i)>0} \alpha_i \mathbb{1}_{t\le\frac{g(x_i)}{f(x_i)}} f(x_i)\right]^2 \mathrm{d}t \ge 0 \end{aligned}$$

5. Let's consider $\mathbb{P}$ the uniform probability measure on the measurable space $(E,\mathcal{X})$ defined by

$$\forall A\in\mathcal{X}, \qquad \mathbb{P}(A) = \frac{|A|}{|E|}.$$

From the definition above,$\mathbb{P}$ is well defined since E is a non empty set by hypothesis. Using the suggested convention, we can rewrite K as

$$K(A,B) = \frac{\mathbb{P}(A\cap B)}{\mathbb{P}(A\cup B)} = \frac{\mathbb{P}(A\cap B)}{1-\mathbb{P}\left(A^C\cap B^C\right)} = \mathbb{P}(A\cap B)\sum_{n=0}^{+\infty}\mathbb{P}\left(A^C\cap B^C\right)^n.$$

Using as above the stability of positive definite kernels by multiplication, finite sum and limit, it is enough to show that

$$K_1(A,B) = \mathbb{P}(A\cap B) \quad \text{and } K_2(A,B) = \mathbb{P}\left(A^C\cap B^C\right)$$

are positive definite kernels. First, let's notice that K_1 and K_2 are *both* symmetric. Introducing as in Q2 the indicator functions, we get

$$K_1(A,B) = \mathbb{E}[\mathbb{1}_A\cdot\mathbb{1}_B]$$

then,

$$\sum_{1\le i,j\le n} a_i a_j K_1\left(A_i,A_j\right) = \sum_{1\le i,j\le n} a_i a_j \mathbb{E}\left[\mathbb{1}_{A_i}\cdot\mathbb{1}_{A_j}\right] = \mathbb{E}\left[\sum_{1\le i,j\le n} a_i a_j \mathbb{1}_{A_i}\cdot\mathbb{1}_{A_j}\right] = \mathbb{E}\left[\left(\sum_{1\le i,j\le n} a_i \mathbb{1}_{A_i}\right)^2\right] \ge 0.$$

The same reasoning applies for K_2.

Exercise 2.

1. Let K_1 and K_2 be two positive definite kernels on a set $\mathcal{X}$, and α, β two positive scalars. Show that $\alpha K_1 + \beta K_2$ is positive definite, and describe its RKHS.
2. Let $\mathcal{X}$ be a set and $\mathcal{F}$ be a Hilbert space. Let $\Psi : \mathcal{X} \to \mathcal{F}$ and $K : \mathcal{X} \times \mathcal{X} \to \mathbb{R}$ be:

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

Show that K is a positive definite kernel on $\mathcal{X}$, and describe its RKHS.

Solution 2.

1. Let's denote by $K := \alpha K_1 + \beta K_2$. We notice that since K_1 and K_2 are both symmetric, so is K. Then, for all $n \in \mathbb{N}$, $\{a_i\}_{i=1}^n \subset \mathbb{R}$ and $\{x_i\}_{i=1}^n \subset \mathcal{X}$, we have

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

since $\alpha, \beta \ge 0$ and K_i are definite positive kernels.
Let's eliminate now the case $\alpha\beta = 0$. We denote by $\mathcal{H}_i$ the (unique) RKHS associated with K_i.

- If $\alpha = \beta = 0$, then K is the null kernel and its RKHS is reduce to the null vector.
- If only one of scalars α, β is null. Suppose by symmetry that $\alpha > 0$ and $\beta = 0$. The RKHS of K consists of functions

$$x \in \mathbb{R}^d \mapsto f(x) = \sum_i a_i K(x_i, x) = \alpha \sum_i a_i K_1(x_i, x) = \sum_i \tilde{a}_i K_1(x_i, x)$$

with $\tilde{a}_i = \alpha \cdot a_i \in \mathbb{R}$. We can see in this case that $H = H_1$ which does not contradict the uniqueness of the RKHS associated with K_1 since we also need equality of norms to conclude that H_1 and H are equal, which is, of course, not the case. In fact, by keeping the same notations as above:

$$\|f\|_{\mathcal{H}}^2 = \sum_{i,j} a_i a_j K\left(x_i, x_j\right) = \alpha \sum_{i,j} a_i a_j K_1\left(x_i, x_j\right) = \alpha \|f\|_{\mathcal{H}_1}^2.$$

- Finally, let's assume that $\alpha, \beta > 0$. The RKHS of K consists of functions

$$\begin{aligned}x \in \mathbb{R}^d \mapsto f(x) &= \sum_i a_i K(x_i, x) \\ &= \alpha \sum_i a_i K_1(x_i, x) + \beta \sum_i a_i K_2(x_i, x) \\ &\in \{\alpha f_1 + \beta f_2 : (f_1, f_2) \in \mathcal{H}_1 \times \mathcal{H}_2\} = \mathcal{H}_1 + \mathcal{H}_2 \qquad (1)\end{aligned}$$

where the last equality comes from the fact that $f_1 \in \mathcal{H}_1 \mapsto \alpha f_1 \in \mathcal{H}_1$ (*resp* $f_2 \in \mathcal{H}_2 \mapsto \alpha f_2 \in \mathcal{H}_2$) is bijective for $\alpha, \beta > 0$. By denoting $\mathcal{H}$ the RKHS of K, we have then $\mathcal{H} \subset \mathcal{H}_1 + \mathcal{H}_2$.
By using the same notations as above, we would like to have

$$\begin{aligned}\forall f \in \mathcal{H}, \quad \|f\|_{\mathcal{H}}^2 &= \sum_{i,j} a_i a_j K\left(x_i, x_j\right) \\ &= \alpha \sum_{i,j} a_i a_j K_1\left(x_i, x_j\right) + \beta \sum_{i,j} a_i a_j K_1\left(x_i, x_j\right) \\ &= \frac{1}{\alpha} \|f_1\|^2 + \frac{1}{\beta} \|f_2\|^2 \qquad (2)\end{aligned}$$

where $f_1 = \alpha \sum_i a_i K_1(x_i, \cdot) \in \mathcal{H}_1$, $f_2 = \alpha \sum_i a_i K_2(x_i, \cdot) \in \mathcal{H}_2$ and $f = f_1 + f_2$. The main problem here to define the expression above as our norm in $\mathcal{H}$ is the dependance between f and (f_1, f_2). How can we assure to have the same value for $\|f\|_{\mathcal{H}}$ for another pair (f_1', f_2').
Similarly, we would like to define $\langle \cdot, \cdot \rangle_{\mathcal{H}}$ as

$$\forall f = f_1 + f_2,\ g = g_1 + g_2 \in \mathcal{H} \qquad \left\langle f, g \right\rangle_{\mathcal{H}} = \frac{1}{\alpha}\left\langle f_1, g_1 \right\rangle_{\mathcal{H}_1} + \frac{1}{\beta}\left\langle f_2, g_2 \right\rangle_{\mathcal{H}_2} \tag{3}$$

but it is ill-defined for the same reason as above. To force unicity of the pair (f_1, f_2) for every $f \in \mathcal{H}$, let's introduce $\Psi : \mathcal{H}_1 \times \mathcal{H}_2 \to \mathcal{H}_1 + \mathcal{H}_2$, $(f_1, f_2) \mapsto f_1 + f_2$ where the Hilbert space $\mathcal{H}_1 \times \mathcal{H}_2$ is endowed with the norm (2). Ψ is clearly a linear surjection. To make it injective let's consider its restriction ψ to $(\operatorname{Ker}\Psi)^{\perp}$ [1] which is clearly bijective. Now we have that

$$\forall f \in \mathcal{H}, \quad \|f\|_{\mathcal{H}} = \frac{1}{\alpha}\|f_1\|^2 + \frac{1}{\beta}\|f_2\|^2 \tag{4}$$

is well defined with $(f_1, f_2) = \psi^{-1}(f)$ and the same for the inner product (3) defined on $\mathcal{H}_1 + \mathcal{H}_2$. Let's show that the reproducing property is satisfied. Let $x \in \mathcal{X}$, $f, K_x \in \mathcal{H}_1 + \mathcal{H}_2$. Denoting by f_1, f_2 the image of f from ψ^{-1}, we have

$$\begin{aligned}
\left\langle f, K_x \right\rangle_{H_1+H_2} &= \left\langle f_1 + f_2, \alpha K_1(x, \cdot) \right\rangle_{H_1+H_2} + \left\langle f_1 + f_2, \beta K_2(x, \cdot) \right\rangle_{H_1+H_2} \\
&= \frac{1}{\alpha}\left\langle f_1, \alpha K_1(x, \cdot) \right\rangle_{\mathcal{H}_1} + \frac{1}{\beta}\left\langle f_2, 0 \right\rangle_{\mathcal{H}_2} + \frac{1}{\alpha}\left\langle f_1, 0 \right\rangle_{\mathcal{H}_1} + \frac{1}{\beta}\left\langle f_2, \beta K_2(x, \cdot) \right\rangle_{\mathcal{H}_2} \\
&= \left\langle f_1, K_1(x, \cdot) \right\rangle_{\mathcal{H}_1} + \left\langle f_2, K_2(x, \cdot) \right\rangle_{\mathcal{H}_2} \\
&= f_1(x) + f_2(x) = f(x)
\end{aligned}$$

using the reproducing property in H_1 and H_2 which concludes the proof.

2. First of all, using Aronszajn's theorem, it is clear that K is a positive definite kernel on the set $\mathcal{X}$. Let's describe its RKHS denoted by $\mathcal{H}$.
   - $\mathcal{H}$ consists of functions:

$$\begin{aligned}
\mathbf{x} \in \mathcal{X} \mapsto f(\mathbf{x}) &= \sum_i a_i K(x_i, \mathbf{x}) \\
&= \sum_i a_i \left\langle \psi(x_i), \psi(\mathbf{x}) \right\rangle_{\mathcal{F}} \\
&= \left\langle \sum_i a_i \psi(x_i), \psi(\mathbf{x}) \right\rangle_{\mathcal{F}} \\
&= \left\langle \mathbf{w}, \psi(\mathbf{x}) \right\rangle_{\mathcal{F}}
\end{aligned}$$

with $\mathbf{w} = \sum_i a_i \psi(x_i) \in \overline{\operatorname{span}\{\psi(\mathcal{X})\}}$. Let's explicit its inner product. Let $f_{\mathbf{w}_1}$ and $g_{\mathbf{w}_2}$ be two functions in $\mathcal{H}$ [2], where

$$\mathbf{w}_1 = \sum_i a_i \psi(x_i) \quad \text{and} \quad \mathbf{w}_2 = \sum_j b_j \psi(y_j)$$

1. Note that $\mathcal{H}_1 \times \mathcal{H}_2$ endowed with the norme (2) is a Hilbert space. Therefore every closed subspace of $\mathcal{H}_1 \times \mathcal{H}_2$ admits an orthogonal complement which is easy to verify for $\operatorname{Ker}\Psi$ using the fact that convergence in RKHS implies ponctual convergence.

2. *i.e.* $f_{\mathbf{w}_1}(\mathbf{x}) = \left\langle \mathbf{w}_1, \psi(\mathbf{x}) \right\rangle_{\mathcal{F}}$ and the similar form for g.

with a_i,b_j live in $\mathbb{R}$ and x_i, y_j in $\mathcal{X}$. We have successively

$$\begin{aligned}\left\langle f_{\mathbf{w}_1}, g_{\mathbf{w}_2}\right\rangle_{\mathcal{H}} &= \left\langle \sum_i a_i K(x_i,\cdot), \sum_j b_j\, K(y_j,\cdot)\right\rangle_{\mathcal{F}} \\ &= \sum_{i,j} \alpha_i \beta_j\, K\left(x_i, y_j\right) \\ &= \sum_{i,j} \alpha_i \beta_j \left\langle \psi(x_i), \psi(y_j)\right\rangle_{\mathcal{F}} = \left\langle \mathbf{w}_1, \mathbf{w}_2\right\rangle_{\mathcal{F}}.\end{aligned}$$

It follows from the inner product that

$$\|f_{\mathbf{w}}\|_{\mathcal{H}} = \|\mathbf{w}\|_{\mathcal{F}}\,.$$

Exercise 3.

1. Let
$$\mathcal{H} = \{f : [0,1] \to \mathbb{R},\ \text{absolutely continuous},\ f' \in L^2([0,1]), f(0) = 0\}$$
endowed with the bilinear form
$$\forall f, g \in \mathcal{H}, \qquad \langle f, g\rangle_{\mathcal{H}} = \int_0^1 f'(u) g'(u) du.$$
Show that $\mathcal{H}$ is an RKHS, and compute its reproducing kernel.
2. Same question when
$$\mathcal{H} = \{f : [0,1] \to \mathbb{R},\ \text{absolutely continuous},\ f' \in L^2([0,1]), f(0) = f(1) = 0\}$$
3. Same question, when $\mathcal{H}$ is endowed with the bilinear from:
$$\forall f, g \in \mathcal{H}, \qquad \langle f, g\rangle_{\mathcal{H}} = \int_0^1 (f(u)g(u) + f'(u)g'(u))\, du$$

Solution 3.

1. Let's show that $(\mathcal{H}, \langle\cdot,\cdot\rangle_{\mathcal{H}})$ is an Hilbert space. Let $f, g \in \mathcal{H}$ and $\lambda \in \mathbb{R}$. We have
   - $\lambda f + g$ is also continuous on the compact $[0;1]$, which is equivalent thanks to the Heine-Cantor theorem to $\lambda f + g$ absolutely continuous.
   - $(\lambda f + g)(0) = \lambda f(0) + g(0) = 0$.
   - For all $x \in [0;1]$, $\left|\lambda f'(x) + g'(x)\right| \leq |\lambda| \left|f'(x)\right| + \left|g'(x)\right| \in L^2([0,1])$.
   - $\langle f, f\rangle_{\mathcal{H}} = \int_0^1 (f'(u))^2\, \mathrm{d}u \geq 0$.
   - Suppose we have $\left\langle f, f\right\rangle_{\mathcal{H}} = 0$. By using the Cauchy-Schwartz inequality, we have for all $x \in [0;1]$ and $g \in \mathcal{H}$:
   $$\left(\int_0^x g(u) f'(u)\, \mathrm{d}u\right)^2 \leq \int_0^x (g(u))^2\, \mathrm{d}u \underbrace{\int_0^x (f'(u))^2\, \mathrm{d}u}_{=0}. \tag{5}$$

   By taking $g = 2f$, we get for all $x \in [0;1]$:
   $$0 = \int_0^x 2f(u) f'(u)\, \mathrm{d}u = f(x)^2 - f(0)^2 = f(x)^2,$$
   which shows that $(\mathcal{H}, \langle\cdot,\cdot\rangle_{\mathcal{H}})$ is a pre-Hilbert space.

- Let $\{f_n\}_{n\in\mathbb{N}}$ be a Cauchy sequence in $\mathcal{H}$. We have then $\{f_n\}_{n\in\mathbb{N}}$ a Cauchy sequence in $L^2([0;1])$ endowed with its standard inner product[3]. Let's denote by g its limit and define f:

$$x \in [0;1] \mapsto f(x) := \int_0^x g(u)\,\mathrm{d}u.$$

We have f continuous (defined by an integral) on $[0;1]$ so absolutely continuous since $[0;1]$ is compact, $f(0) = 0$ and $f' = g \in L^2([0;1])$. So $f \in \mathcal{H}$ and we have:

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

- Let's try to find its reproducing kernel K. If K exists, K should verify, for all $f \in \mathcal{H}$, $x \in [0;1]$:

$$K_x(\cdot) \in \mathcal{H} \quad \text{and} \quad \langle f, K_x\rangle_{\mathcal{H}} = \int_0^1 f'(u)K_x'(u)du = f(x).$$

Since $f(x) = \int_0^x f'(u)\,\mathrm{d}u$, it seems natural to propose a candidate for K_x such that

$$K_x'(u) = \begin{cases} 1 \text{ if } u \leq x \\ 0 \text{ otherwise.} \end{cases}$$

Since we also should have $K_x(0) = 0$ and $K_x(\cdot)$ continuous, $K_x(u) = \min(u,x)$ is a good candidate since it verifies the reproducing property and for all $x \in [0;1]$, we have $\mathbf{u} \mapsto \min(\mathbf{u}, x) \in \mathcal{H}$. This shows that $\mathcal{H}$ is a RKHS and we conclude using the uniqueness of the reproducing kernel.

2. By adding the condition $f(1) = 0$ for $f \in \mathcal{H}$, it is quite clear that $(\mathcal{H}, \langle\cdot,\cdot\rangle_{\mathcal{H}})$ is still a pre-Hilbert space. Let's show that its completeness.
By using the same reasoning as above, the only thing that needs to be demonstrated is the fact that the f we've found as limit for $\|\cdot\|_{\mathcal{H}}$, verifies the new condition $f(1) = 0$ which can be easily checked as follow:

$$\forall x \in [0;1], \quad |f_n(x) - f(x)| \leq \int_0^x \left|f_n' - g\right| \leq \left\|f_n' - g\right\|_{L^2[0,1]} \to 0$$

when $n \to +\infty$. Applying this relation with $x = 1$ gives the required result.
Let's try to find its reproducing kernel.
By analogy with above, we will search kernels among functions of the form:

$$K_x'(u) = \begin{cases} \alpha \text{ if } u \leq x \\ \beta \text{ otherwise.} \end{cases}$$

for $\alpha, \beta \in \mathbb{R}$. Since we want $K_x(\cdot)$ continuous, this is equivalent to setting

$$K_x(u) = \begin{cases} \alpha \cdot u \text{ if } u \leq x \\ \beta\cdot(u - x) + \alpha\cdot x \text{ otherwise.} \end{cases}$$

We have two unknown parameters so we need to find two equations to recover them. The first equation ones comes from the condition on $u = 1$: $K_x(1) = 0$, which is equivalent to

$$\beta\cdot(1-x) + \alpha\cdot x = 0.$$

The second equation comes from the reproducing kernel property:

$$\forall f \in \mathcal{H}, \forall x \in [0;1] \qquad f(x) = \langle f, K_x\rangle_{\mathcal{H}} \qquad \text{which is equivalent to} \qquad f(x) = (\alpha - \beta)f(x).$$

We then have to solve the following linear system

$$\begin{cases} (1-x)\cdot\beta + x\cdot\alpha = 0 \\ \qquad\quad -\beta + \quad \alpha = 1 \end{cases}$$

3. This is due to the fact that for all $f \in \mathcal{H}$, $\|f\|_{\mathcal{H}} = \left\|f'\right\|_{L^2([0;1])}$.

which gives $\alpha = 1 - x$ and $\beta = -x$ as unique solution. As above, we have by construction that

$$K_x(u) = \begin{cases} (1-x) \cdot u \text{ if } u \leq x \\ -x \cdot (u-x) + (1-x) \cdot x \text{ otherwise.} \end{cases}$$

which can be simply written as

$$K_x(u) = \min(x\,,\,u) - x \cdot u \tag{6}$$

verifies the reproducing property and for all $x \in [0;1]$, $\mathbf{u} \in [0;1] \mapsto K_x(\mathbf{u}) \in \mathcal{H}$. We conclude as above.

3. Let's proceed as Q1.
   - We have already the fact that $\mathcal{H}$ is a vector space.
   - $\langle f, f \rangle_{\mathcal{H}} = \int_0^1 (f(u))^2 \, \mathrm{d}u + \int_0^1 (f'(u))^2 \, \mathrm{d}u \geq 0$.
   - Let's assume $\langle f, f \rangle_{\mathcal{H}}^2 = 0$. We would necessarily have

     $$\int_0^1 (f'(u))^2 \, \mathrm{d}u = 0,$$

     which implies (shown in Q1) that $f = 0$.
   - Suppose $\{f_n\}_{n \in \mathbb{N}}$ is a Cauchy sequence in $\mathcal{H}$. As in Q1, since we have:

     $$\forall f \in \mathcal{H}, \qquad \left\| f' \right\|_{L^2([0;1])} = \int_0^1 (f'(u))^2 \, \mathrm{d}u \leq \|h\|_{\mathcal{H}}^2,$$

     it follows from the completeness of $L^2([0;1])$ endowed with its usual inner that $\{f_n'\}_{n \in \mathbb{N}}$ converge in $L^2([0;1])$ to a limit g. Let's define

     $$x \in [0;1] \mapsto f(x) := \int_0^x g(u) \, \mathrm{d}u. \tag{7}$$

     We have $f' = g$ and

     $$\|f - f_n\|_{\mathcal{H}}^2 = \underbrace{\int_0^1 (f - f_n)^2}_{:=A_n} + \underbrace{\int_0^1 (g - f_n')}_{:=B_n}. \tag{8}$$

     • B_n it is clear from the convergence of f_n' to g in $L^2([0;1])$ that $B_n \to 0$.

     • A_n let's show that we have uniform convergence of $f - f_n$ to 0. Using the Cauchy-Schwartz inequality yields to

     $$\begin{aligned} \forall x \in [0;1], \quad |f(x) - f_n(x)|^2 &\leq \left| \int_0^x g(u) - f_n'(u) \, \mathrm{d}u \right|^2 \\ &\leq \left(\int_0^x (g(u) - f_n'(u))^2 \, \mathrm{d}u \right) \left(\int_0^x (1)^2 \, \mathrm{d}u \right) \\ &\leq \left\| g(u) - f_n' \right\|_{L^2([0;1])}. \end{aligned}$$

     Since the right term tends to zero, taking the supremum to the right yields to the desired result. Since we are working on a compact with a uniform convergence, we can now switch the limit and the integral and

     $$\lim A_n = \lim \int_0^1 (f - f_n)^2 = \int_0^1 \lim (f - f_n)^2 = 0. \tag{9}$$

   - Let's try to find its reproducing kernel. Let $x \in \mathcal{X} = [0;1]$. We want to find $K_x \in \mathcal{H}$ such that

     $$\forall f \in \mathcal{H}, \quad f(x) = \langle f, K_x \rangle_{\mathcal{H}} = \int_0^1 f(u) K_x(u) \, \mathrm{d}u + \int_0^1 f'(u) K_x'(u) \, \mathrm{d}u. \tag{10}$$

Let $G_x(u) := \int_1^u K_x(t)\,\mathrm{d}t$ be the primitive of K_x equal to zero in $u = 1$. Using an integration by part in (10) leads to

$$\forall f \in \mathcal{H}, \qquad f(x) = \int_0^1 f'(u)\left(K_x'(u) - G_x(u)\right)\mathrm{d}u. \tag{11}$$

As in Q1, we would like $K_x' - G_x$ to satisfies

$$(K_x' - G_x)(u) = \begin{cases} 1 \text{ if } u \le x \\ 0 \text{ otherwise.} \end{cases} \tag{12}$$

Derivating the equation (12) yields to the following linear second order differential equation

$$K_x'' = K_x$$

on domains $[0;x]$ and $[x;1]$ which gives us the following form for K_x:

$$K_x(u) = \begin{cases} A\cdot\mathbf{e}^{-u} + B\cdot\mathbf{e}^{u} \text{ if } u \le x \\ C\cdot\mathbf{e}^{-u} + D\cdot\mathbf{e}^{u} \text{ otherwise.} \end{cases}$$

with A, B, C and D parameters depending on x. Since we have 4 parameters, we need to find 4 equations.

(a) Using the initial condition on K_x: $K_x(0) = 0$, we get

$$A + B = 0. \tag{13}$$

(b) Using the continuity of K_x in $u = x$, leads to

$$A\cdot\mathbf{e}^{-x} + B\cdot\mathbf{e}^{x} = C\cdot\mathbf{e}^{-x} + D\cdot\mathbf{e}^{x}. \tag{14}$$

(c) For $1 \ge u \ge x$, using (12), we have

$$G_x(u) = \int_1^u C\cdot\mathbf{e}^{-t} + D\cdot\mathbf{e}^{t}\,\mathrm{d}t = -C\cdot\mathbf{e}^{-u} + D\cdot\mathbf{e}^{u} = K_x'(u)$$

i.e.

$$C\cdot\mathbf{e}^{-1} - D\cdot\mathbf{e}^{1} = 0. \tag{15}$$

(d) For $0 \le u \le x$, using (12) again, we get

$$G_x(u) = \int_1^x C\cdot\mathbf{e}^{-t} + D\cdot\mathbf{e}^{t}\,\mathrm{d}t + \int_x^u A\cdot\mathbf{e}^{-t} + B\cdot\mathbf{e}^{t}\,\mathrm{d}t = -A\cdot\mathbf{e}^{-u} + B\cdot\mathbf{e}^{u} - 1 = K_x'(u) - 1$$

which is equivalent to

$$A\cdot\mathbf{e}^{-x} - B\cdot\mathbf{e}^{x} + C\cdot(\mathbf{e}^{-1} - \mathbf{e}^{-x}) + D\cdot(\mathbf{e}^{x} - \mathbf{e}^{1}) = -1. \tag{16}$$

We can summarize these 4 equations in the following 4×4 linear system :

$$\begin{pmatrix} 1 & 1 & 0 & 0 \\ \mathbf{e}^{-x} & \mathbf{e}^{x} & -\mathbf{e}^{-x} & -\mathbf{e}^{x} \\ 0 & 0 & \mathbf{e}^{-1} & -\mathbf{e}^{1} \\ \mathbf{e}^{-x} & -\mathbf{e}^{x} & -(\mathbf{e}^{-x} - \mathbf{e}^{-1}) & (\mathbf{e}^{x} - \mathbf{e}^{1}) \end{pmatrix} \begin{pmatrix} A \\ B \\ C \\ D \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \\ -1 \end{pmatrix} \tag{17}$$

By substituting $B = -A$ and $D = \mathbf{e}^{-2}C$ thanks to equations (13) and (15) respectively, we get the following 2×2 linear system

$$\begin{cases} (\mathbf{e}^{-x} - \mathbf{e}^{x})\cdot A - (\mathbf{e}^{x-2} + \mathbf{e}^{-x})\cdot C = 0 \\ (\mathbf{e}^{-x} + \mathbf{e}^{x})\cdot A + (\mathbf{e}^{x-2} - \mathbf{e}^{-x})\cdot C = -1 \end{cases}$$

A small computation gives the determinant of the system

$$\det = \begin{vmatrix} (\mathbf{e}^{-x} - \mathbf{e}^{x}) & -(\mathbf{e}^{x-2} + \mathbf{e}^{-x}) \\ (\mathbf{e}^{-x} + \mathbf{e}^{x}) & (\mathbf{e}^{x-2} - \mathbf{e}^{-x}) \end{vmatrix} = 4\mathbf{e}^{-1}\,\mathrm{ch}(1),$$

and using the Cramer formulas we get

$$A = \frac{\begin{vmatrix} 0 & -(\mathbf{e}^{x-2} + \mathbf{e}^{-x}) \\ -1 & (\mathbf{e}^{x-2} - \mathbf{e}^{-x}) \end{vmatrix}}{4\mathbf{e}^{-1}\,\mathrm{ch}(1)} = \frac{-(\mathbf{e}^{-x} + \mathbf{e}^{x-2})}{4\mathbf{e}^{-1}\,\mathrm{ch}(1)} = -\frac{\mathrm{ch}(x-1)}{2\,\mathrm{ch}(1)} \quad \text{which implies} \quad B = \frac{\mathrm{ch}(x-1)}{2\,\mathrm{ch}(1)}.$$

$$C = \frac{\begin{vmatrix} (\mathbf{e}^{-x} - \mathbf{e}^{x}) & 0 \\ (\mathbf{e}^{-x} + \mathbf{e}^{x}) & -1 \end{vmatrix}}{4\mathbf{e}^{-1}\,\mathrm{ch}(1)} = \frac{2\,\mathrm{sh}(x)}{4\mathbf{e}^{-1}\,\mathrm{ch}(1)} = \frac{\mathrm{sh}(x)}{2\mathbf{e}^{-1}\,\mathrm{ch}(1)} \quad \text{which implies} \quad D = \frac{\mathbf{e}^{-1}\,\mathrm{sh}(x)}{2\,\mathrm{ch}(1)}.$$

Getting back to $K_x(\cdot)$ gives

$$K_x(u) = \begin{cases} \frac{\mathrm{ch}(x-1)}{2\,\mathrm{ch}(1)} \cdot (-\mathbf{e}^{-u} + \mathbf{e}^{u}) \text{ if } u \le x \\ \frac{\mathrm{sh}(x)}{2\,\mathrm{ch}(1)} \cdot (\mathbf{e}^{-u+1} + \mathbf{e}^{u-1}) \text{ otherwise.} \end{cases}$$

which is equivalent to

$$K_x(u) = \begin{cases} \frac{1}{\mathrm{ch}(1)}\,\mathrm{ch}(x-1)\cdot\mathrm{sh}(u) \text{ if } 0 \le u \le x \\ \frac{1}{\mathrm{ch}(1)}\,\mathrm{sh}(x)\cdot\mathrm{ch}(u-1) \text{ if } x \le u \le 1 \end{cases} \tag{18}$$

which verifies all conditions to be a reproducing kernel in $\mathcal{H}$. We conclude using the unicity of the kernel.

Exercise 4. Let $(x_1, y_1), \dots, (x_n, y_n)$ a training set of examples where $x_i \in \mathcal{X}$, a space endowed with a positive definite kernel K, and $y_i \in \{-1, 1\}$, for $i = 1, \dots, n$. $\mathcal{H}_K$ denotes the RKHS of the kernel K. We want to learn a function $f : \mathcal{X} \to \mathbb{R}$ by solving the following optimization problem:

$$\min_{f \in \mathcal{H}_K} \frac{1}{n} \sum_{i=1}^{n} \ell_{y_i}(f(x_i)) \quad \text{such that} \quad \|f\|_{\mathcal{H}_K} \le B \tag{19}$$

where ℓ_y is a convex loss functions (for $y \in \{-1, 1\}$) and $B > 0$ is a parameter.

1. Show that there exists $\lambda \ge 0$ such that the solution to problem (19) can be found by solving the following problem:

$$\min_{\alpha \in \mathbb{R}^n} R(K\alpha) + \lambda \alpha^\top K \alpha \tag{20}$$

where K is the $n \times n$ Gram matrix and $R : \mathbb{R}^n \to \mathbb{R}$ should be explicited.

2. Compute the Fenchel-Legendre transform[4] R^* of R in terms of the Fenchel-Legendre transform ℓ_y^* of ℓ_y.

3. Adding the slack variable $u = K\alpha$, the problem (19) can be written as a constrained optimization problem:

$$\min_{\alpha \in \mathbb{R}^n, u \in \mathbb{R}^n} R(u) + \lambda \alpha^\top K \alpha \quad \text{such that} \quad u = K\alpha. \tag{21}$$

Express the dual problem of (21) in terms of R^*, and explain how a solution to (21) can be found from a solution to the dual problem.

4. For any function $f : \mathbb{R}^n \to \mathbb{R}$, the *Fenchel-Legendre transform* (or *convex conjugate*) of f is the function $f^* : \mathbb{R}^n \to \mathbb{R}$ defined by

$$f^*(u) = \sup_{x \in \mathbb{R}^N} x^T u - f(x)$$

.

4. Explicit the dual problem for the logistic and squared hinge loss:

$$\ell_y(u) = \log\left(1 + e^{-yu}\right)$$
$$\ell_y(u) = \max(0, 1 - yu)^2.$$

Solution 4.

1. Since $f \mapsto \ell_y(\delta_{x_i}[f])$ is convex by composition of a linear function with a convex one and $f \mapsto \|f\|_{\mathcal{H}_K} - B$ is also convex by sum of convex functions, (19) can be seen as a standard **convex optimization problem** equivalent to

$$\min_{f \in \mathcal{H}_K} \frac{1}{n} \sum_{i=1}^{n} \ell_{y_i}(f(x_i)) \quad \text{such that} \quad \|f\|^2_{\mathcal{H}_K} \leq B^2.$$

Since $B > 0$, it is easy to see that $f = 0 \in \operatorname{int} \mathcal{H}_K$ [5] defines a strictly feasible point ($0 = \|f\|_{\mathcal{H}_K} < B$) which guarantees strong duality thanks to Slater's constraint qualifications.
Let's introduce respectively the Lagrangian, dual function and dual problem of the problem above

$$L(f, \lambda) = \frac{1}{n} \sum_{i=1}^{n} \ell_{y_i}(f(x_i)) + \lambda\left(\|f\|^2_{\mathcal{H}_K} - B^2\right) \qquad g(\lambda) = \min_{f \in \mathcal{H}_K} L(f, \lambda) \qquad \max_{\lambda \geq 0} g(\lambda)$$

By denoting f^*, λ^* the corresponding optimas, we have by strong duality

$$\min_{\left\{f \in \mathcal{H}_K : \|f\|^2_{\mathcal{H}_K} \leq B^2\right\}} \frac{1}{n} \sum_{i=1}^{n} \ell_{y_i}(f(x_i)) = L(f^*, \lambda^*) = \min_{f \in \mathcal{H}_K} \left\{ \frac{1}{n} \sum_{i=1}^{n} \ell_{y_i}(f(x_i)) + \lambda^*\left(\|f\|^2_{\mathcal{H}_K} - B^2\right) \right\} = g(\lambda^*).$$

We would like to apply the representer theorem to the right optimization problem but we cannot guarantee that the objective function is strictly increasing in $\|f\|_{H_K}$ as λ^* could be equal to zero. By considering $\mathcal{F}_K = \operatorname{span}\left(K_x : x \in \{x_1, \ldots, x_n\}\right) \subset \mathcal{H}_K$, we can decompose f^* into

$$f^* = f_1 + f_2$$

with $f_1 \in \mathcal{F}_K$ and $f_2 \in \mathcal{F}_K^{\perp} \subset \mathcal{H}_K$. Using the reproducing property in $\mathcal{H}_K$ we have

$$\forall i \in [n], \qquad f_2(x_i) = \left\langle f_2, K_{x_i}(\,\cdot\,)\right\rangle_{H_K} = 0.$$

Therefore we have

$$\frac{1}{n} \sum_{i=1}^{n} \ell_{y_i}(f^*(x_i)) = \frac{1}{n} \sum_{i=1}^{n} \ell_{y_i}(f_1(x_i))$$

while the other term in the objective function is increasing in $\|\cdot\|_{H_K}$ with $\|f^*\|_{H_K} \geq \|f_1\|_{H_K}$ by Pythagoras' theorem. This shows that the objective function has a minimum in $\mathcal{F}_K$ and without loss of generality, f^* can be written as

$$f^*(\,\cdot\,) = \sum_{i=1}^{n} \alpha_i K(x_i, \cdot\,)$$

and the problem is therefore equivalent to solving:

$$\min_{\alpha \in \mathbb{R}^n} \left\{ \frac{1}{n} \sum_{i=1}^{n} \ell_{y_i}([K\alpha]_i) + \lambda^* \alpha^T K \alpha \right\},$$

where we have removed the term with the constant B which does not depend of the optimization variable. Taking $R : \mathbb{R}^n \to \mathbb{R}$, $x = (x_1, \ldots, x_n)^T \mapsto \frac{1}{n} \sum_{i=1}^{n} \ell_{y_i}(x_i)$ yields to the desired result.

5. we assume here that $\mathcal{H}_K \neq \{0\}$.

2. By definition of the *Fenchel-Legendre transform*, we have:

$$R^*(u) = \sup_{x\in\mathbb{R}^n} x^T u - R(x) = \sup_{x\in\mathbb{R}^n} \frac{1}{n}\sum_{i=1}^{n}\left\{ n x_i u_i - \ell_{y_i}(x_i)\right\}.$$

Since the terms appearing in the sum are independent, we can switch the supremum and the sum which yields to:

$$R^*(u) = \frac{1}{n}\sum_{i=1}^{n}\sup_{x_i\in\mathbb{R}}\left\{(nu_i)\cdot x_i - \ell_{y_i}(x_i)\right\} = \frac{1}{n}\sum_{i=1}^{n}\ell^*_{y_i}(nu_i).$$

3. *To be consistent with the notations of the subject, we will simply denote λ^* as λ.*
The Lagrangian of (21) is defined by

$$L(\alpha,u,\mu) = R(u) + \lambda\alpha^T K\alpha + \mu^T(u - K\alpha)$$

Let's compute its dual function $g(\mu) = \inf_{x,u\in\mathbb{R}^n} L(\alpha,u,\mu)$.

- L is quadratic in α:

$$\nabla_\alpha L = 2\lambda K\alpha - K\mu = 0 \quad \text{which implies} \quad K(2\lambda\alpha - \mu) = 0,$$

i.e. $\alpha = \dfrac{\mu}{2\lambda} + \epsilon$ with $\epsilon \in \operatorname{Ker} K$. A solution f to the initial problem represented by α or $\alpha+\epsilon$ is invariant if ϵ lives in $\operatorname{Ker} K$, so setting ϵ to zero yields to

$$\alpha = \frac{\mu}{2\lambda}. \tag{22}$$

A small computation shows that with such α

$$\lambda\alpha^T K\alpha - \mu^T K\alpha = -\frac{\mu^T K\mu}{4\lambda}.$$

- By considering the terms in the Lagrangian depending on u, we have

$$\inf_{u\in\mathbb{R}^n} R(u) + \mu^T u = \inf_{u\in\mathbb{R}^n} R(u) - (-\mu)^T u = -\sup_{u\in\mathbb{R}^n}(-\mu)^T u - R(u) = -R^*(-\mu).$$

By considering both equations, the dual problem can be written as

$$\sup_{\mu\in\mathbb{R}^n} -\frac{\mu^T K\mu}{4\lambda} - R^*(-\mu) \quad \text{subject to} \quad \mu \geq 0. \tag{23}$$

Once the dual problem (23) is solved in μ, we get a solution to the primal problem (21) with (22).

4. To explicit the dual problem, we need to explicit $R^*(\cdot)$ which requires to explicit $\ell^*_y(\cdot)$ for $y \in \{-1,1\}$.
   (a) We want to compute

$$\ell^*_y(u) = \sup_{x\in\mathbb{R}} \underbrace{ux - \log(1 + \mathbf{e}^{-y\cdot x})}_{:=\phi(x)}$$

Taking its first and second derivatives gives us

$$\forall x\in\mathbb{R},\quad \phi'(x) = u + \frac{y}{1+\mathbf{e}^{y\cdot x}} \quad \text{and} \quad \phi''(x) = -\frac{\mathbf{e}^{y\cdot x}}{(1+\mathbf{e}^{y\cdot x})^2} < 0$$

- We notice that the equation $\phi'(x) = 0$ as a solution for the concave function ϕ only if $u\cdot y < 0$ which yields to

$$-u\cdot y = \frac{1}{1+\mathbf{e}^{-y\cdot x}} \quad \text{which implies} \quad u\cdot y\in(-1,0) \quad \text{and} \quad x = y\log\left(\frac{1+u\cdot y}{-u\cdot y}\right).$$

Replacing x in $\phi(x)$ gives (after a small computation)

$$\ell^*_y(u) = (1+u\cdot y)\log(1+u\cdot y) - u\cdot y\log(-u\cdot y) \quad \text{for} \quad u\cdot y\in(-1,0).$$

- For $u \cdot y = 0$, *i.e.* $u = 0$

$$\phi(x) = -\log\left(1 + \mathbf{e}^{-y \cdot x}\right) \le 0 = \lim_{y \cdot x \to +\infty} \phi(x).$$

- For $u \cdot y > 0$, by using the fact that $1 = y^2$

$$\phi(x) = (u \cdot y) \cdot (y \cdot x) - \log\left(1 + \mathbf{e}^{-y \cdot x}\right) \to +\infty \quad \text{when} \quad y \cdot x \to +\infty$$

- For $u \cdot y < -1$ and using the same trick

$$\begin{aligned} \phi(x) &= (u \cdot y + 1) \cdot (y \cdot x) - \log\left(1 + \mathbf{e}^{-y \cdot x}\right) - y \cdot x \\ &= (u \cdot y + 1) \cdot (y \cdot x) + \log\left(\frac{\mathbf{e}^{-y \cdot x}}{1 + \mathbf{e}^{-y \cdot x}}\right) \to +\infty \quad \text{when} \quad y \cdot x \to -\infty \end{aligned}$$

- For $u \cdot y = -1$ and taking the same expression as above

$$\phi(x) = \log\left(\frac{\mathbf{e}^{-y \cdot x}}{1 + \mathbf{e}^{-y \cdot x}}\right) \le 0 = \lim_{y \cdot x \to -\infty} \phi(x)$$

We can finally summarize

$$\ell_y^*(u) = \begin{cases} (1 + u \cdot y)\log(1 + u \cdot y) - (u \cdot y)\log(-u \cdot y) \text{ if } u \cdot y \in [-1;0] \\ +\infty \text{ otherwise.} \end{cases}$$

with the convention $0 \ln(0) = 0$.
The dual can be written as

$$\sup_{\mu \in \mathbb{R}^n} \left\{ -\frac{\mu^T K \mu}{4\lambda} - \sum_{i=1}^n \ell_{y_i}^*(-n\mu_i) \right\} \quad \text{subject to} \quad \mu \ge 0.$$

which is therefore equivalent to

$$\sup_{\mu \in \mathbb{R}^n} \left\{ -\frac{\mu^T K \mu}{4\lambda} - \sum_{i=1}^n (1 - n\mu_i \cdot y_i)\log(1 - n\mu_i \cdot y_i) + (n\mu_i \cdot y_i)\log(n\mu_i \cdot y_i) \right\}$$
$$\text{subject to } \mu \ge 0 \quad \text{and} \quad 0 \le \operatorname{diag}(y)\mu \le \frac{1}{n} \cdot \mathbf{1}.$$

Two remarks:

- If $y_i = -1$, since we both have $n\mu_i y_i \ge 0$ and $\mu_i \ge 0$, we should have $\mu_i = 0$.
- By denoting $h : \mathbb{R}^n \to \mathbb{R}, x \mapsto -\sum_{i=1}^n x_i \log x_i$ the entropy of points x_i, the objective function above can be written as

$$-\frac{\mu^T K \mu}{4\lambda} + h(\mathbf{1} - n\operatorname{diag}(y)\mu) + h(n\operatorname{diag}(y)\mu).$$

(b) We want to compute

$$\ell_y^*(u) = \sup_{x \in \mathbb{R}} \{ux - \max(0, 1 - yx)^2\} = \sup_{x \in \mathbb{R}} \underbrace{\{(u \cdot y) \cdot x - \max(0, 1 - x)^2\}}_{:=\phi(x)}$$

where we used the fact that $y^2 = 1$ and $x \mapsto y \cdot x$ is a bijection of $\mathbb{R}$.

- If $u \cdot y = 0$, we have

$$\forall x \in \mathbb{R}, \qquad \phi(x) = -\max(0, 1 - x)^2 \le \phi(1) = 0.$$

- If $u \cdot y > 0$, we have $\phi(x) = (u \cdot y) \cdot x$ for x big enough, which clearly tends to $+\infty$.

- For $u \cdot y < 0$, ϕ is linear and strictly decreasing for $x \geq 1$, so

$$\sup_{x \in \mathbb{R}} \phi(x) \quad = \quad \sup_{x \leq 1} \phi(x) \quad = \quad \sup_{x \leq 1} \{ (u \cdot y) \cdot x - (1-x)^2 \}. \tag{24}$$

  We are then optimizing a (concave) quadratic function on a restricted domain. Setting its derivative to zero yields to

$$\phi'(x) = u \cdot y - 2(x-1) = 0 \quad \text{which is equivalent to} \quad x = 1 + \frac{u \cdot y}{2}. \tag{25}$$

  Since by hypothesis $u \cdot y < 0$, we verify that the optimal x above satisfy $x \leq 1$ as required. Replacing x by its expression above in $\phi(x)$ gives us

$$\phi(x) = u \cdot y \left(1 + \frac{u \cdot y}{4} \right) > u \cdot y = \phi(1). \tag{26}$$

We can finally summarize

$$\ell_y^*(u) = \begin{cases} u \cdot y \left(1 + \frac{u \cdot y}{4} \right) \text{ if } u \cdot y \leq 0 \\ +\infty \text{ otherwise.} \end{cases}$$

The dual can be written as

$$\sup_{\mu \in \mathbb{R}^n} \left\{ -\frac{\mu^T K \mu}{4\lambda} - \sum_{i=1}^{n} \ell_{y_i}^*(-n\mu_i) \right\} \quad \text{subject to} \quad \mu \geq 0.$$

which is therefore equivalent to

$$\sup_{\mu \in \mathbb{R}^n} \left\{ -\frac{\mu^T K \mu}{4\lambda} + \sum_{i=1}^{n} (n\mu_i \cdot y_i) \left(1 + \frac{(-n\mu_i \cdot y_i)}{4} \right) \right\}$$
$$\text{subject to } \mu \geq 0 \quad \text{and} \quad \operatorname{diag}(y)\mu \geq 0.$$

Remark:

— If $y_i = -1$, since we both have $y_i \cdot \mu_i \geq 0$ and $\mu_i \geq 0$, we get $\mu_i = 0$.