HW1

Gabriel ROMON

Exercise 1

1. Since $\cos(x-y) = \cos(x)\cos(y)+\sin(x)\sin(y) = \langle \begin{pmatrix}\cos(x)\\ \sin(x)\end{pmatrix}, \begin{pmatrix}\cos(y)\\ \sin(y)\end{pmatrix} \rangle_{\mathbb{R}^2}$, we have $K(x,y) = \langle \phi(x), \phi(y)\rangle_{\mathbb{R}^2}$ hence K is positive definite.

2. On $\mathcal{X}$, $|x^Ty| \leq \|x\|_2\|y\|_2 < 1$ hence $\dfrac{1}{1-x^Ty} = \sum_{k=0}^{\infty}(x^Ty)^k$. $(x,y)\mapsto x^Ty$ being the linear kernel, each $(x,y)\mapsto (x^Ty)^k$ is positive definite (as a product of p.d kernels) hence the partial sums $(x,y)\mapsto \sum_{k=0}^n (x^Ty)^k$ are positive definite (as sums of p.d kernels). K is therefore a pointwise limit of positive definite kernels, and is thus positive definite.

3. Let $\mathcal{F}$ be the space of measurable functions from $(\Omega,\mathcal{A})$ to $([-1,1],\mathcal{B}([-1,1]))$. On this space we consider the bilinear form $\langle f,g\rangle = \int fg dP$. It is non-negative in the sense that $\langle f,f\rangle = \int f^2 dP \geq 0$. Note that

$$K(A,B) = P(A\cap B) - P(A)P(B) = \int (1_A - P(A))(1_B - P(B))dP = \langle \phi(A), \phi(B)\rangle$$

Non-negativity of $\langle\cdot,\cdot\rangle$ shows that K is positive definite.

4. Let us prove first that $\mathbb{R}_+\times\mathbb{R}_+\to\mathbb{R}, (x,y)\mapsto \min(x,y)$ is a positive definite kernel. Let $N\in\mathbb{N}$, $x_1,\dots,x_N\in\mathbb{R}_+$, $a_1,\dots,a_N\in\mathbb{R}$ and note that

$$\sum_{i=1}^N\sum_{j=1}^N a_ia_j\min(x_i,x_j) = \sum_{i=1}^N\sum_{j=1}^N a_ia_j\int_0^\infty 1_{t\leq x_i}1_{t\leq x_j}dt = \int_0^\infty \left(\sum_{i=1}^N a_i 1_{t\leq x_i}\right)^2 dt \geq 0$$

Next, note that

$$\begin{aligned}\min(f(x)g(y), f(y)g(x)) &= 1_{g(x)>0}1_{g(y)>0}\min(\frac{f(x)}{g(x)}g(x)g(y), \frac{f(y)}{g(y)}g(x)g(y))\\ &= 1_{g(x)>0}1_{g(y)>0}g(x)g(y)\min(\frac{f(x)}{g(x)}, \frac{f(y)}{g(y)})\\ &= g(x)g(y)\min(1_{g(x)>0}\frac{f(x)}{g(x)}, 1_{g(y)>0}\frac{f(y)}{g(y)})\end{aligned}$$

$(x,y)\mapsto g(x)g(y)$ is a positive definite kernel. The function $x\mapsto 1_{g(x)>0}\frac{f(x)}{g(x)}$ has values in $\mathbb{R}_+$, and by positive definiteness of the min proved before,

$$(x,y)\mapsto \min(1_{g(x)>0}\frac{f(x)}{g(x)}, 1_{g(y)>0}\frac{f(y)}{g(y)})$$

is positive definite, hence $(x,y)\mapsto\min(f(x)g(y), f(y)g(x))$ is positive definite as a product.

5. Let us prove the claim for $\mathcal{X} = \mathcal{P}(E)\setminus\{\emptyset\}$ first. Let μ denote the counting measure on E and $\mathcal{F}$ be the space of measurable functions from $(E,\mathcal{P}(E))$ to $([0,1],\mathcal{B}([0,1]))$. On this space we consider the bilinear form $\langle f,g\rangle = \int fg d\mu$. It is non-negative in the sense that $\langle f,f\rangle = \int f^2 d\mu \geq 0$. Note that

$$|A\cap B| = \mu(A\cap B) = \int 1_A 1_B d\mu = \langle 1_A, 1_B\rangle$$

hence $(A,B) \mapsto |A \cap B|$ is a positive definite kernel. Next, note that

$$\frac{1}{|A \cup B|} = \frac{1}{|E| - |A^c \cap B^c|} = \frac{1}{|E|} \frac{1}{1 - \frac{|A^c \cap B^c|}{|E|}}$$

Since A and B are non-empty, $\frac{|A^c \cap B^c|}{|E|} < 1$ thus

$$\frac{1}{|A \cup B|} = \frac{1}{|E|} \sum_{k=0}^{\infty} \left(\frac{|A^c \cap B^c|}{|E|} \right)^k$$

Since $|A^c \cap B^c| = \langle 1_{A^c}, 1_{B^c} \rangle$, $(A,B) \mapsto |A^c \cap B^c|$ is a positive definite kernel, and by closedness arguments, $\frac{1}{|E|} \sum_{k=0}^{\infty} \left(\frac{|A^c \cap B^c|}{|E|} \right)^k$ also is. Finally, $(A,B) \mapsto \frac{|A \cap B|}{|A \cup B|}$ is the product of two positive definite kernels, and is thus positive definite.

In the general case, if A or B is empty, $K(A,B) = 0$. Returning to the basic definition of a positive definite kernel, removing the empty A_i and using the previous claim finishes the proof.

Exercise 2

1. Let $N \in \mathbb{N}$, $x_1, \dots, x_N \in \mathcal{X}$, $a_1, \dots, a_N \in \mathbb{R}$ and note that

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

$\alpha K_1 + \beta K_2$ is thus positive definite.

• Suppose first that $\alpha = \beta = 1$. Let H_1 and H_2 be the RKHS corresponding respectively to K_1 and K_2. Let $\mathcal{F}$ be the Hilbert direct sum of H_1 and H_2: $\mathcal{F} = H_1 \times H_2$ equipped with the inner product

$$\langle (f_1, f_2), (g_1, g_2) \rangle_{\mathcal{F}} := \langle f_1, g_1 \rangle_{H_1} + \langle f_2, g_2 \rangle_{H_2}$$

It is standard that $(\mathcal{F}, \langle \cdot, \cdot \rangle_{\mathcal{F}})$ is a Hilbert space.
Let $u : \left\{ \begin{array}{ccc} \mathcal{F} & \longrightarrow & H_1 + H_2 \\ (f_1, f_2) & \longmapsto & f_1 + f_2 \end{array} \right.$. u is linear, surjective and $\ker u$ is closed (easy to prove).
$\ker u^{\perp}$ is thus in direct sum with $\ker u$: $\mathcal{F} = \ker u \oplus \ker u^{\perp}$, and let $v := u_{|\ker u^{\perp}}$. A standard result of linear algebra shows that v is bijective. We can then define the following inner product on $H_1 + H_2$:

$$\langle f, g \rangle_{H_1 + H_2} := \langle v^{-1}(f), v^{-1}(g) \rangle_{\mathcal{F}}$$

Since v^{-1} is a linear isomorphism, $(H_1 + H_2, \langle \cdot, \cdot \rangle_{H_1 + H_2})$ is a Hilbert space. Let us show that it is the RKHS of $K_1 + K_2$. First, since $K_{1_x} \in H_1$ and $K_{2_x} \in H_2$, we have $(K_1 + K_2)_x = K_{1_x} + K_{2_x} \in H_1 + H_2$. Next, if $f \in H_1 + H_2$ we write $f = g + h$ with $g \in H_1$ and $h \in H_2$. Note that

$$\begin{aligned}
\langle f, (K_1 + K_2)_x \rangle_{H_1 + H_2} &= \langle g + h, K_{1_x} + K_{2_x} \rangle_{H_1 + H_2} \\
&= \langle v^{-1}(g + h), v^{-1}(K_{1_x} + K_{2_x}) \rangle_{\mathcal{F}} \\
&= \langle (g, h), (K_{1_x}, K_{2_x}) \rangle_{\mathcal{F}} \\
&= \langle g, K_{1_x} \rangle_{H_1} + \langle h, K_{2_x} \rangle_{H_2} \\
&= g(x) + h(x) \\
&= f(x)
\end{aligned}$$

• Let α, β be arbitrary positive reals. Let H_1 and H_2 be the RKHS corresponding respectively to K_1 and K_2. It is easy to prove that the RKHS of αK_1 is $(H_1, \frac{1}{\alpha}\langle\cdot,\cdot\rangle_{H_1})$ and that of βK_2 is $(H_2, \frac{1}{\beta}\langle\cdot,\cdot\rangle_{H_2})$. Applying the previous result to the sum, we let $\mathcal{F} = H_1 \times H_2$ equipped with the inner product

$$\langle (f_1, f_2), (g_1, g_2)\rangle_{\mathcal{F}} := \frac{1}{\alpha}\langle f_1, g_1\rangle_{H_1} + \frac{1}{\beta}\langle f_2, g_2\rangle_{H_2}$$

We define u, v and the inner product as before:

$$\langle f, g\rangle_{H_1+H_2} := \langle v^{-1}(f), v^{-1}(g)\rangle_{\mathcal{F}}$$

$(H_1 + H_2, \langle\cdot,\cdot\rangle_{H_1+H_2})$ is the RKHS of $\alpha K_1 + \beta K_2$.

2. Let $N \in \mathbb{N}$, $x_1, \dots, x_N \in \mathcal{X}$, $a_1, \dots, a_N \in \mathbb{R}$ and note that

$$\sum_{i=1}^{N}\sum_{j=1}^{N} a_i a_j \langle \Psi(x_i), \Psi(x_j)\rangle_{\mathcal{F}} = \left\|\sum_{i=1}^{N} a_i \Psi(x_i)\right\|_{\mathcal{F}}^2 \geq 0$$

hence K is a positive definite kernel.

Let $E = \{\sum_{i=1}^{N} a_i\Psi(x_i) | x_i \in \mathcal{X}, a_i \in \mathbb{R}, N \geq 1\}$. E is a linear subspace of $\mathcal{F}$.
For $z \in \overline{E}$, let $f_z : y \mapsto \langle z, \Psi(y)\rangle_{\mathcal{F}}$ and $H = \{f_z | z \in \overline{E}\}$. Note that $\left\{\begin{array}{ccc} \overline{E} & \longrightarrow & H \\ z & \longmapsto & f_z \end{array}\right.$ is a linear isomorphism. It is clearly linear and surjective.
Let us prove injectivity. Consider some $z \in \overline{E}$ such that $f_z = 0$. Let $\varepsilon > 0$. By definition of $\overline{E}$, there exists $r \in \mathcal{F}$ with $\|r\|_{\mathcal{F}} \leq \varepsilon$, $N \geq 1$, $x_1, \dots, x_N \in \mathcal{X}$ and $a_1, \dots, a_N \in \mathbb{R}$ such that $z = \sum_{i=1}^{N} a_i\Psi(x_i) + r$.
Since $f_z = 0$, for all $y \in \mathcal{X}$,

$$\langle \sum_{i=1}^{N} a_i\Psi(x_i) + r, \Psi(y)\rangle_{\mathcal{F}} = 0$$

Plugging $y = x_i$ and summing yields $\langle \sum_{i=1}^{N} a_i\Psi(x_i) + r, \sum_{i=1}^{N} a_i\Psi(x_i)\rangle_{\mathcal{F}} = 0$, hence

$$\left\|\sum_{i=1}^{N} a_i\Psi(x_i)\right\|^2 = -\langle r, \sum_{i=1}^{N} a_i\Psi(x_i)\rangle_{\mathcal{F}} \leq \|r\| \left\|\sum_{i=1}^{N} a_i\Psi(x_i)\right\|$$

thus $\left\|\sum_{i=1}^{N} a_i\Psi(x_i)\right\| \leq \|r\| \leq \varepsilon$, hence $\|z\| \leq 2\varepsilon$. This holds for any $\varepsilon > 0$, so $z = 0$.

Since $\overline{E}$ is a closed subspace of $\mathcal{F}$, it is a Hilbert space when equipped with the induced inner product. Consequently, H equipped with the inner product $\langle f_z, f_{z'}\rangle_H := \langle z, z'\rangle_{\overline{E}}$ is a Hilbert space. Let us prove that it is the RKHS of K. First note that $K_x = f_{\Psi(x)} \in H$. For the reproducing property, given $z \in \overline{E}$,

$$\langle f_z, K_x\rangle_H = \langle f_z, f_{\Psi(x)}\rangle_H = \langle z, \Psi(x)\rangle_{\overline{E}} = f_z(x)$$

Exercise 3

1. • Let us show first that $(\mathcal{H}, \langle\cdot,\cdot\rangle_{\mathcal{H}})$ is a Hilbert space. $\mathcal{H}$ is clearly a vector space and $\langle\cdot,\cdot\rangle_{\mathcal{H}}$ is a symmetric bilinear form.
Let us prove that it is positive definite: if $\langle f, f\rangle_{\mathcal{H}} = 0$ for some $f \in \mathcal{H}$, then $\int_0^1 f'^2(u)du = 0$, hence $f' = 0$ a.e. in $[0, 1]$. Since f is absolutely continuous,

$$\forall x \in [0, 1], f(x) = f(0) + \int_0^x f'(t)dt = f(0) = 0$$

Hence $f = 0$ and $\langle\cdot,\cdot\rangle_{\mathcal{H}}$ is an inner product.
It remains to prove that $\mathcal{H}$ is complete for the norm induced by its inner product. Let $(f_n)_n \in \mathcal{H}^{\mathbb{N}}$ be

a Cauchy sequence. Note that $\|f_n\|_{\mathcal{H}}^2 = \int_0^1 f_n'(t)^2 dt = \|f_n'\|_{L^2([0,1])}$, hence $(f_n')_n$ is Cauchy in $L^2([0,1])$, hence converges to some $g \in L^2([0,1])$. Let $x \in [0,1]$ and note that

$$|f_n(x) - f_m(x)| = \left|\int_0^x f_n'(t) - f_m'(t)dt\right| \leq \int_0^x |f_n'(t) - f_m'(t)|dt \leq \|f_n' - f_m'\|_{L^2([0,1])}\sqrt{x}$$

$(f_n(x))_n$ is thus Cauchy in $\mathbb{R}$ and we may define $f(x) := \lim_n f_n(x)$. It remains to prove that $f \in \mathcal{H}$ and $f_n \xrightarrow[\|\cdot\|_{\mathcal{H}}]{} f$. Note that $f(x) = \lim_n f_n(x) = \lim_n \int_0^x f_n'(t)dt$ and

$$\left|\int_0^x f_n'(t)dt - \int_0^x g(t)dt\right| \leq \int_0^x |f_n'(t) - g(t)|dt \leq \|f_n' - g\|_{L^2([0,1])}\sqrt{x} \xrightarrow[n\to\infty]{} 0$$

Hence $\lim_n \int_0^x f_n'(t)dt = \int_0^x g(t)dt$, thus $f(x) = \int_0^x g(t)dt$. This proves that f is absolutely continuous with $f' = g$ a.e, and $f(0) = 0$. Hence $f \in \mathcal{H}$.
Lastly, $\|f_n - f\|_{\mathcal{H}}^2 = \int_0^1 (f_n'(t) - f'(t))^2 = \int_0^1 (f_n'(t) - g(t))^2 = \|f_n' - g\|_{L^2([0,1])} \xrightarrow[n\to\infty]{} 0$.

• Let us prove that the reproducing kernel of $\mathcal{H}$ is $K : (x,y) \mapsto \min(x,y)$.
Let $x \in [0,1]$. Note that $K_x(0) = 0$ and K_x is Lipschitz (as a continuous piecewise affine function), hence absolutely continuous. It's easy to check that $K_x' \in L^2([0,1])$, hence $K_x \in \mathcal{H}$.
Besides, $\langle f, K_x\rangle_{\mathcal{H}} = \int f'(t)K_x'(t)dt = \int_0^x f'(t)dt = f(x)$, thus $K : (x,y) \mapsto \min(x,y)$ is the reproducing kernel of $\mathcal{H}$.

2. • Let us prove that $\mathcal{H}$ is a closed subspace of $\mathcal{G}$, the RKHS of question 1. Since $\mathcal{G}$ is closed, it suffices to prove that if $(f_n)_n \in \mathcal{H}^{\mathbb{N}}$ converges to some $f \in \mathcal{G}$ for the norm $\|\cdot\|_{\mathcal{G}}$, then $f(1) = 0$. Note that

$$|f_n(1) - f(1)| = \left|\int_0^1 f_n'(t) - f'(t)dt\right| \leq \int_0^1 |f_n'(t) - f'(t)|dt \leq \|f_n' - f'\|_{L^2([0,1])} = \|f_n - f\|_{\mathcal{G}}$$

Hence $f(1) = \lim_n f_n(1) = 0$. As a closed subspace of $\mathcal{G}$, $\mathcal{H}$ is a Hilbert space for the induced inner product.

• Let us prove that the reproducing kernel of $\mathcal{H}$ is $K : (x,y) \mapsto \min(x,y) - xy$. Let $x \in [0,1]$. Note that $K_x(0) = K_x(1) = 0$ and K_x is Lipschitz (as a continuous piecewise affine function), hence absolutely continuous. It's easy to check that $K_x' \in L^2([0,1])$, hence $K_x \in \mathcal{H}$.
Besides,

$$\langle f, K_x\rangle_{\mathcal{H}} = \int f'(y)K_x'(y)dy = \int_0^x (1-x)f'(y)dy + \int_x^1 -xf'(y)dy = (1-x)(f(x)-f(0)) + x(f(1)-f(x)) = f(x)$$

thus $K : (x,y) \mapsto \min(x,y) - xy$ is the reproducing kernel of $\mathcal{H}$.

3. I assume that $\mathcal{H}$ is the RHKS of **the second question**.

• Let us show first that $(\mathcal{H}, \langle\cdot,\cdot\rangle_{\mathcal{H}})$ is a Hilbert space. $\mathcal{H}$ is clearly a vector space and $\langle\cdot,\cdot\rangle_{\mathcal{H}}$ is a symmetric bilinear form.
Let us prove that it is positive definite: if $\langle f, f\rangle_{\mathcal{H}} = 0$ for some $f \in \mathcal{H}$, then $\int_0^1 f^2(u) + f'^2(u)du = 0$, hence $f' = 0$ a.e. in $[0,1]$. Since f is absolutely continuous,

$$\forall x \in [0,1], f(x) = f(0) + \int_0^x f'(t)dt = f(0) = 0$$

Hence $f = 0$ and $\langle\cdot,\cdot\rangle_{\mathcal{H}}$ is an inner product.
It remains to prove that $\mathcal{H}$ is complete for the norm induced by its inner product. Let $(f_n)_n \in \mathcal{H}^{\mathbb{N}}$ be a Cauchy sequence. Since $\|f_n\|_{\mathcal{H}}^2 = \|f_n\|_{L^2([0,1])}^2 + \|f_n'\|_{L^2([0,1])}^2$, $(f_n)_n$ and $(f_n')_n$ are Cauchy in $L^2([0,1])$. Similarly to what was done in question 1, $(f_n')_n$ converges in L^2 to some g and $(f_n)_n$ converges pointwise to some f with $f \in \mathcal{H}$ and $f' = g$ a.e. By uniqueness of the limit, $(f_n)_n$ must also converge in $L^2([0,1])$ to f, hence $\|f_n - f\|_{\mathcal{H}}^2 \xrightarrow[\|\cdot\|_{\mathcal{H}}]{} f$.

- Let us prove that the reproducing kernel of $\mathcal{H}$ is

$$K : (x, y) \mapsto \frac{1}{\sinh(1)} \sinh(\min(1-x, 1-y)) \sinh(\max(x, y))$$

Let $x \in [0,1]$. Note that $K_x(0) = K_x(1) = 0$ and K_x is Lipschitz (as a piecewise Lipschitz function), hence absolutely continuous. Since K_x' is piecewise continuous on a compact set, we have $K_x' \in L^2([0,1])$, hence $K_x \in \mathcal{H}$.
Tedious computations show that the reproducing property holds.

Exercise 4

a. Let $\mathcal{L}(f, \lambda) = \frac{1}{n}\sum_{i=1}^{n} \ell_{y_i}(f(x_i)) + \lambda(\|f\|_{\mathcal{H}_K} - B)$ be the Lagrangian of the optimization problem at hand.
Since ℓ_y is convex for any y, it is easy to prove that $f \mapsto \frac{1}{n}\sum_{i=1}^{n} \ell_{y_i}(f(x_i))$ is convex. The constraint $f \mapsto \|f\|_{\mathcal{H}_K} - B$ is also clearly convex, and Slater's condition is trivially true on the closed ball. Strong duality thus holds, so the minimum we are looking for is equal to $\min_{f \in \mathcal{H}_K} L(f, \lambda^*)$ for some $\lambda^* \geq 0$. This rewrites as

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

Removing the constant term and applying the representer's theorem yields

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

If we let $R(u) := \frac{1}{n}\sum_{i=1}^{n} \ell_{y_i}(u_i)$, the problem turns into $\min_{\alpha \in \mathbb{R}^n} R(K\alpha) + \lambda^* \alpha^T K \alpha$.

b. Note that

$$\begin{aligned}
R^*(\eta) &= \sup_{u \in \mathbb{R}^n} \left[\eta^T u - \frac{1}{n}\sum_{i=1}^{n} \ell_{y_i}(u_i) \right] \\
&= \frac{1}{n} \sup_{u \in \mathbb{R}^n} \left[\sum_{i=1}^{n} n\eta_i u_i - \ell_{y_i}(u_i) \right] \\
&= \frac{1}{n} \sum_{i=1}^{n} \sup_{u \in \mathbb{R}^n} \left[n\eta_i u_i - \ell_{y_i}(u_i) \right] \\
&= \frac{1}{n} \sum_{i=1}^{n} \ell_{y_i}^*(n\eta_i)
\end{aligned}$$

Swapping the finite sum and the supremum can be justified by the standard equality $\sup(A + B) = \sup A + \sup B$.

c. Let $\mathcal{L}(\alpha, u, \eta) = R(u) + \lambda\alpha^T K \alpha + \eta^T(K\alpha - u)$ be the Lagrangian of the problem. Note that

$$\begin{aligned}
\inf_{\alpha, u} \mathcal{L}(\alpha, u, \eta) &= \inf_{\alpha, u \in \mathbb{R}^n} \Big((R(u) - \eta^T u) + (\lambda\alpha^T K\alpha + \eta^T K\alpha) \Big) \\
&= \inf_{\alpha \in \mathbb{R}^n} \lambda\alpha^T K\alpha + \eta^T K\alpha - \sup_{u \in \mathbb{R}^n} [\eta^T u - R(u)] \\
&= \inf_{\alpha \in \mathbb{R}^n} \Big(\lambda\alpha^T K\alpha + \eta^T K\alpha \Big) - R^*(\eta)
\end{aligned}$$

$\alpha \mapsto \lambda\alpha^T K\alpha + \eta^T K\alpha$ is convex (as the sum of two convex functions) and differentiable, so it reaches a global minimum at any of its critical points. Setting the gradient to 0 yields $2\lambda K\alpha + K\eta = 0$, hence

$\lambda\alpha^T K\alpha = -\frac{1}{2}\eta^T K\alpha$, and $-\eta^T K\alpha = \frac{1}{2\lambda}\eta^T K\eta$.
As a result $\inf_{\alpha,u}\mathcal{L}(\alpha,u,\eta) = -\frac{1}{4\lambda}\eta^T K\eta - R^*(\eta)$, so the dual writes as

$$\min_{\eta\in\mathbb{R}^n} \frac{1}{4\lambda}\eta^T K\eta + R^*(\eta)$$

From the previous first order conditions we got $K(2\lambda\alpha+\eta)=0$. Once a solution η^* of the dual is computed, we can choose any $\alpha^* \in \frac{1}{2\lambda}\left(\ker(K) - \eta^*\right)$

d. For notational ease let $H_{y,p}(u) = pu - \ell_y(u)$, so that $\ell_y^*(p) = \sup_u H_{y,p}(u)$.

- For the logistic case $H_{y,p}(u) = pu - \log\left(1+e^{-yu}\right)$. Since $y^2=1$,

$$H_{y,p}(u) = py(yu) - \log\left(1+e^{-yu}\right) = (py+1)(yu) - \log\left(1+e^{yu}\right)$$

Thus $H_{y,p}(u) \to \infty$ when $py > 0$ and $yu \to \infty$ or $py < -1$ and $yu \to -\infty$. We can therefore constrain our attention to the case $-1 \leq py \leq 0$, where the concave function $H_{y,p}$ has an upper bound. Differentiating yields $H'_{y,p}(u^*) = 0 \iff p + y\frac{e^{-yu^*}}{e^{-yu^*}+1} \iff u^* = y\log\left(-1-\frac{1}{py}\right)$ and $H_{y,p}(u^*) = (py+1)\log(py+1) - py\log(-py)$. Hence

$$\ell_y^*(p) = \begin{cases} (py+1)\log(py+1) - py\log(-py) \text{ if } -1 \leq py \leq 0 \\ +\infty \text{ otherwise} \end{cases}$$

The dual writes as:

$$\min_{\eta\in\mathbb{R}^n} \frac{1}{4\lambda}\eta^T K\eta + \frac{1}{n}\sum_{i=1}^n \left[(n\eta_i y_i + 1)\log(n\eta_i y_i + 1) - n\eta_i y\log(-n\eta_i y_i)\right]$$

under the constraint $\forall i \in \{1,\ldots,n\},\ 0 \leq -\eta_i y_i \leq \frac{1}{n}$.

- For the squared hinge case, we deal with $H_{y,p}(u) = pu - \max(0, 1-yu)^2$. Since $H_{y,p}(u) = py(yu) - \max(0,1-yu)^2$ it boils down to maximizing $T_v(x) = vx - \max(0,1-x)^2$ where $v = py$ and $x = yu$. Since $T_v(x) \to \infty$ when $v > 0$, we can focus on the case $v \geq 0$. In this case, $T_v(x) = vx$ when $x > 1$ and the supremum is v. When $x \leq 1$, $T_v(x) = vx - (1-x)^2$. This quadratic function reaches its maximum when $x = 1+\frac{v}{2}$ and the value is $v+\frac{v^2}{4}$. Thus $\sup_x T_v(x) = +\infty$ if $v > 0$ and $v+\frac{v^2}{4}$ if $v \leq 0$. Hence:

$$\ell_y^*(p) = \begin{cases} py + \frac{p^2}{4} \text{ if } py \leq 0 \\ +\infty \text{ otherwise} \end{cases}$$

The dual writes as:

$$\min_{\eta\in\mathbb{R}^n} \frac{1}{4\lambda}\eta^T K\eta + \frac{1}{n}\sum_{i=1}^n \left[n\eta_i y_i + \frac{n^2\eta_i^2}{4}\right] \tag{1}$$

under the constraint $\forall i \in \{1,\ldots,n\},\ \eta_i y_i \leq 0$.