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

$$f^*(u) = \sup_{x \in \mathbb{R}^N} <x, u> - f(x).$$

Exercice 1.
Compute the Fenchel-Legendre transforms of the following functions defined for $u \in \mathbb{R}$ and indexed by a parameter $y \in \{-1, +1\}$

- Hinge loss:
  $$\ell_y(u) = \max(0, 1 - yu).$$

- Squared hinge loss:
  $$\ell_y(u) = \max(0, 1 - yu)^2.$$

- Logistic loss:
  $$\ell_y(u) = \log \left(1 + e^{-yu}\right).$$

- Exponential loss:
  $$\ell_y(u) = e^{-yu}.$$

Exercice 2.
Let $(x_1, y_1), \ldots, (x_n, y_n)$ 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 $i = 1, \ldots, n$. $\mathcal{H}_K$
denotes the RKHS of the kernel \( K \). We want to learn a function \( f : \mathcal{X} \mapsto \mathbb{R} \) by solving the following optimization problem:

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

where \( \ell_y \) is one of the loss functions defined in Exercise 1 and \( \lambda > 0 \) is a regularization parameter.

a. Show that the solution to problem (1) can be found by solving the following problem:

\[
\min_{\alpha \in \mathbb{R}^n} R(K \alpha) + \lambda \alpha^\top K \alpha,
\]

where \( K \) is the \( n \times n \) Gram matrix and \( R : \mathbb{R}^n \mapsto \mathbb{R} \) should be explicit.

b. Compute the Fenchel-Legendre transform \( R^* \) of \( R \) in terms of Fenchel-Legendre transform \( \ell_y^* \) of \( \ell_y \).

c. Adding the slack variable \( u = K \alpha \), the problem (1) can be rewritten as a constrained optimization problem:

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

Compute the dual problem of (3) in terms of \( R^* \), and explain how a solution to (3) can be found from a solution to the dual problem. c. Explicit the dual problem for the different loss functions defined in Exercise 1. For the hinge loss, how does it related to the formulation we saw during the course?