Problem 0.

(i) Give the definition of a soft classifier, and its associated hard classifier.

(ii) Give the definition of a convex surrogate loss function, and give three examples. Plot these functions on one graph.

(iii) Give a probabilistic meaning of the logistic loss, and then explain why the exponential loss has no probabilistic meaning.

Problem 1.
Consider the problem of binary classification, with response variable $Y \in \{-1, 1\}$.

(i) Show that function f^* minimising the risk $\mathbf{E}\{\varphi(-Yf(X))|X = x\}$ for the exponential loss $\varphi(z) = e^z$ is given by

$$f^*(x) = \frac{1}{2} \log \left(\frac{\mathbf{P}(Y = 1|X = x)}{\mathbf{P}(Y = -1|X = x)} \right) .$$

(ii) Show that the population minimiser derive in question *(i)* coincides with the population minimizer associated with the logistic loss $\varphi(z) = \log(1 + e^z)$, up to a constant.

Problem 2.
Let φ be a convex surrogate of the indicator function. For $\alpha \in \mathbb{R}$ and $\eta \in [0, 1]$, put

$$H_\eta(\alpha) := \eta\varphi(-\alpha) + (1 - \eta)\varphi(\alpha) ,$$

and

$$\tau(\eta) := \inf_{\alpha \in \mathbb{R}} H_\eta(\alpha) .$$

Zhang's result states that provided there exists $c > 0$ and $\gamma \in [0, 1]$ such that for all $\eta \in [0, 1]$,

$$|\eta - 1/2| \leq c\,(1 - \tau(\eta))^\gamma \ ,$$

then it is possible to use the excess risk of a soft classifier to bound the excess risk of the associated hard classifier (see page 7 of the lecture notes).

Show that this condition holds for the hinge loss $\varphi(z) = \max(z + 1, 0)$, and give the optimal values of c and γ in this case.