Machine Learning	Prof. Matthias Hein
Exercise Sheet 1 - 29.4.2021 (12 Points)	due: 6.5.2021

Exercise 1 - Bayes Optimal Function - warm-up

- **(3 Points)** Let $\mathcal{Y} = \mathbb{R}_+ = \{x \in \mathbb{R} | x \geq 0\}$ (regression with output on the positive part of $\mathbb{R}$) and suppose that $\mathbb{E}[Y|X = x] > 0$. Show that the Bayes optimal function, $f^*(x) = \arg\min_{c \in \mathbb{R}} \mathbb{E}[L(Y, c)|X = x]$ for the loss function $L(y, f(x)) = \log(f(x)) + \frac{y}{f(x)}$, is given by $f^*(x) = \mathbb{E}[Y|X = x]$. Discuss the properties of this loss function compared to least square loss. For what kind of noise model do you think is this loss function useful (note that the target space is the set of non-negative reals) ?

Exercise 2 - Bayes Optimal Function

Consider the following convex margin-based loss function for the binary classification problem

$$L_k(yf(x)) = \begin{cases} \max\{0, 1 - yf(x)\} & yf(x) \geq 0 \\ 1 - kyf(x) & \text{otherwise} \end{cases}$$

where k is a constant such that $k > 1$.

a. **(4 points)** Derive the Bayes optimal function for the loss L_k.

b. **(2 points)** Show that this loss function is not classification calibrated.

Hints: Since the loss function is not differentiable, first determine the ranges of $f(x)$ over which you can work without the max-operation and the case distinction of L_k. Minimize the loss over these ranges separately and then determine the global minimizer from these solutions.

Exercise 3 - Bayes error

We have a binary classification problem, $\mathcal{Y} = \{-1, 1\}$, with the following distribution on $\mathcal{X} = [0, 1]$,

$$\mathrm{P}(Y = 1|X = x) = \begin{cases} 0.1, & \text{if } 0 \leq x \leq \frac{1}{8}, \\ 0.9, & \text{if } \frac{1}{8} < x < \frac{7}{8}, \\ 0.1, & \text{if } \frac{7}{8} \leq x \leq 1. \end{cases}$$

and the marginal density is uniform, $p(x) = 1$, $\forall x \in [0, 1]$. We are using the 0-1-loss.

a. **(2 Points)** What is the Bayes optimal error of this problem ?

b. **(1 Points)** Determine the parameter(s) (w^*, b^*) and the error of the classifier(s) $f_{(w^*,b^*)}$,

$$f_{(w,b)} = \text{sign}\big(wx + b\big), \quad w, b \in \mathbb{R},$$

with smallest error.

Hint: You don't need to give a derivation in b)- just write down the optimal parameters and the corresponding error. If there is more than one optimal set of parameters, then provide all possible optimal parameters.