The Mathematics of Machine Learning
Homework Set 5

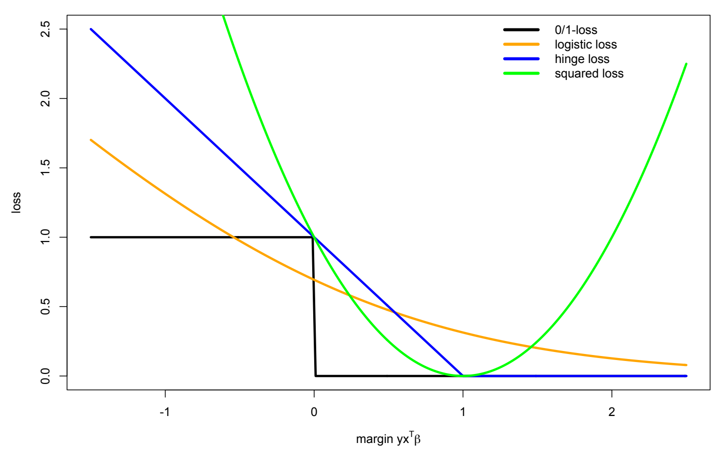
Due 23 March 2023 before 13:00
via Canvas

You are allowed to work on this homework in pairs. One person per pair submits the answers via Canvas. Make sure to put both names on the submission.

1. Consider the extended version of the 0/1-loss with outcomes $y \in \{-1, +1\}$ and predictions $\hat{y} \in \mathbb{R}$: $L(y, \hat{y})$ is 0 if $\text{sign}(\hat{y}) = y$ and 1 otherwise. Suppose that predictions are made by a linear model: $\hat{y} = x^\top \beta$. Show that the 0/1 loss is not convex as a function of the model parameters β.
Hint 1: Recall that a function f is convex if, for any two points z_0 and z_1 inside its domain,

$$f(z_\lambda) \leq (1 - \lambda) f(z_0) + \lambda f(z_1) \qquad \textit{for all } \lambda \in [0, 1],$$

where $z_\lambda := (1 - \lambda) z_0 + \lambda z_1$.
Hint 2: Note that it is sufficient to construct a counter example. When doing so, keep this picture in mind:



2. Consider the squared error loss $L(y, \hat{y}) = (y - \hat{y})^2$ for binary classification with $y \in \{-1, +1\}$, and suppose that predictions are made by a linear model: $\hat{y} = x^\top \beta$. Show that the loss can be expressed as a function of the margin $yx^\top \beta$, as shown in the picture above.

3. Suppose the training data for binary classification (with $y \in \{-1, +1\}$) are linearly separable with positive margin, meaning that there exists some β^* such that $y_i = \operatorname{sign}(x_i^\top \beta^*)$ and $x_i^\top \beta^* \neq 0$ for all $i = 1, \ldots, N$. Show that the minimum in the ERM objective for the logistic loss

$$\min_\beta \frac{1}{N} \sum_{i=1}^{N} \log(1 + e^{-y_i x_i^\top \beta})$$

is not achieved by any β with finite length $\|\beta\| < \infty$.
NB: In the lecture there was a related argument about β that classified all data points correctly, but now your argument also needs to cover β that do not classify all data points correctly.