Exercises 3

Discussion of solutions: Monday, Oct 30 in class

In this assignment, we are working with the hypothesis class of (homogeneous) linear classifiers in a \(d\)-dimensional euclidian space:

\[ \mathcal{H}_{\text{lin}} = \{ h_w : w \in \mathbb{R}^d \}, \]

where

\[ h_w(x) = \text{sign}(\langle x, w \rangle) = \text{sign}\left( \sum_{i=1}^{d} x_i w_i \right). \]

We often just write \( w \) instead of \( h_w \).

1. Convexity of losses

(a) Show that the empirical 0–1-loss on a sample can have local minima. That is, give an example of a sample \( S = ((x_1, y_1), \ldots, (x_n, y_n)) \) where the function

\[ f(w) = L_0^1_S(w) \]

has a local minimum.

(b) Can this also happen for the true loss, that is the function

\[ f(w) = L_0^1_P(w) \]

for some distribution \( P \) on \( \mathbb{R}^d \)?

2. Gradient descent

Let \( S = ((x_1, y_1), \ldots, (x_n, y_n)) \) be a sample that is linearly separable. Let \( w^* \) be a vector of minimal norm that separates the data with margin 1. Let \( R = \max_i \|x_i\| \).

Define a function \( f \) as follows:

\[ f(w) = \max_{i \in [n]} (1 - y_i \langle w, x_i \rangle) \]

(a) Explain why \( w^* \) exists.

(b) Is \( f \) convex?

(c) Show that \( \min_{w: \|w\| \leq \|w^*\|} f(w) = 0 \). Show that any \( w \) for which \( f(w) < 1 \) has \( L_0^1_S(w) = 0 \).

(d) Show how to calculate a subgradient of \( f \).

(e) Describe and analyze the subgradient descent algorithm for this function.
3. SVM

(a) Give an example of a data sample $S$ and a parameter $\lambda$, where the output of hard-SVM and soft-SVM are identical.

(b) Give an example of a separable data sample $S$ and a parameter $\lambda$, where the output of hard-SVM and soft-SVM are not identical.

(c) Proof or refute: There is a parameter $\lambda$ such that for all separable samples $S$, the output of soft-SVM and hard-SVM are identical.