Exercises 3

Disussion of solutions: Monday, Oct 30 in class

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

$$\mathcal{H}_{\text{lin}} = \{h_{\boldsymbol{w}} \ : \ w \in \mathbb{R}^d\},$$

where

$$h_{\boldsymbol{w}}(\boldsymbol{x}) = \text{sign}\left(\langle \boldsymbol{x}, \boldsymbol{w} \rangle\right) = \text{sign}\left(\sum_{i=1}^{d} x_i w_i\right).$$

We often just write $\boldsymbol{w}$ instead of $h_{\boldsymbol{w}}$.

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 $S = ((\boldsymbol{x}_1, y_1), \ldots, (\boldsymbol{x}_n, y_n))$ where the function

   $$f(\boldsymbol{w}) = L_S^{0-1}(\boldsymbol{w})$$

   has a local minimum.

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

   $$f(\boldsymbol{w}) = L_P^{0-1}(\boldsymbol{w})$$

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

2. **Gradient descent**
   Let $S = ((\boldsymbol{x}_1, y_1), \ldots, (\boldsymbol{x}_n, y_n))$ be a sample that is linearly separable. Let $\boldsymbol{w}^*$ be a vector of minimal norm that separates the data with margin 1. Let $R = \max_i \|\boldsymbol{x}_i\|$. Define a function f as follows:

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

   (a) Explain why $\boldsymbol{w}^*$ exists.

   (b) Is f convex?

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

   (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 λ, where the output of hard-SVM and soft-SVM are identical.

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

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