Machine Learning I: Foundations
Exercise Sheet 2

Prof. Marius Kloft   TA:Billy Joe Franks

11.05.2021
Deadline: 18.05.2020

1) (MANDATORY) 10 Points
Interestingly the linear hard-margin SVM, given by

$$\max_{\gamma,b\in\mathbb{R},\mathbf{w}\in\mathbb{R}^d} \gamma \tag{1}$$
$$\text{s.t. } y_i(\mathbf{w}^T\mathbf{x}_i + b) \geq \|\mathbf{w}\|\gamma,\ \forall i \in \{1,\ldots,n\},$$

requires only two (non-equal) training points (with opposite labels) to find a separating hyperplane. Let $X := \{\mathbf{x}_1, \ldots, \mathbf{x}_n\}$ and $Y := \{y_1, \ldots, y_n\}$, with $x_i \in \mathbb{R}^d$ and $y_i \in \{-1, 1\}$, be a dataset. Let γ^*, $\mathbf{w}^*$, and b^* be the optimal solution to the above optimization problem (1) on X, Y. You may assume $w_1 \neq 0$.

a) Find a minimal dataset (X', Y') with $|X'| = |Y'| = 2$ (consisting of only two data points) with the same hard-margin SVM solution (Eq. (1)) as for the dataset (X, Y) , that is, γ^*, $\mathbf{w}^*$, and b^*.

b) Prove that, for your choice of X' and Y' in a), γ, $\mathbf{w}^*$, and b^* are optimal solutions of (1).

c) How is this choice of X' and Y' related to the nearest centroid classifier (NCC)? (Answer this question with at most 5 sentences.)

2) Consider the soft-margin SVM as in the lecture. Now assume we do not optimize over b and it is fixed to $b = 0$. Construct a dataset for which any classifier learned (with $b = 0$) performs poorly. Does any classifier with b fixed to a different constant like $b = 1$ still perform poorly?

3) Construct a worst-case dataset for the nearest centroid classifier (NCC). This dataset should be easily (not necessarily linearly) separable and the NCC should behave as poorly as possible on this training dataset. **Hint:** To this end you will have to figure out for yourself how poorly the NCC can perform. Is it possible for the NCC to have 0% accuracy?

4) Solve programming task 2.