Exercise of Supervised Learning: SVM Part 2

Yawei Li

yawei.li@stat.uni-muenchen.de

December 15, 2023

Exercise 1: Kernelized Multiclass SVM

For a data set $\mathcal{D} = \{(\mathbf{x}^{(1)}, y^{(1)}), \dots, (\mathbf{x}^{(n)}, y^{(n)})\}$ with $y^{(i)} \in \mathcal{Y} = \{+1, -1\}$, assume that we are provided with a suitable feature map $\phi : \mathcal{X} \to \Phi$, where $\Phi \subset \mathbb{R}^d$. In the featureized SVM learning problem we are facing the following optimization problem:

$$
\begin{aligned}
&\min_{\boldsymbol{\theta}, \theta_0, \zeta^{(i)}} \frac{1}{2}\boldsymbol{\theta}^T\boldsymbol{\theta} + C\sum_{i=1}^{n}\zeta^{(i)} \\
&\text{s.t. } y^{(i)}\left(\left\langle \boldsymbol{\theta}, \phi^{(i)} + \theta_0 \right\rangle\right) \geq 1 - \zeta^{(i)} \qquad \forall i \in \{1, \dots, n\}, \\
&\text{and } \zeta^{(i)} \geq 0 \qquad i \in \{1, \dots, n\},
\end{aligned}
$$

where $C \geq 0$ is some constant.

(a) Argue that this is equivalent to the following ERM problem:

$$
\mathcal{R}_{\text{emp}}(\boldsymbol{\theta}) = \frac{1}{2}||\boldsymbol{\theta}||^2 + C\sum_{i=1}^{n}\max(1 - y^{(i)}(\boldsymbol{\theta}^T\phi^{(i)} + \theta_0)), 0).
$$

i.e., the regularized ERM problem for the hinge loss for the hypothesis space

$$
\mathcal{H} = \{f : \Phi \to \mathbb{R} \mid f(\boldsymbol{z}) = \boldsymbol{\theta}^T\boldsymbol{z} + \theta_0, \boldsymbol{\theta} \in \mathbb{R}^d, \theta_0 \in \mathbb{R}\}
$$

Solution to Exercise 1 (a)