University of Freiburg
Dept. of Computer Science
Prof. Dr. F. Kuhn
P. Schneider

Advanced Algorithms
Problem Set 6

Issued: Friday, May 31, 2019

Exercise 1: Learning a Linear Classifier

Assume that we are given m feature vectors $\boldsymbol{a_1}, \dots, \boldsymbol{a_m} \in \mathbb{R}^n$ and that each vector $\boldsymbol{a_i}$ has a label $\ell_i \in \{-1, +1\}$. Our goal will be to find non-negative weights $\boldsymbol{x} \in \mathbb{R}^n$, $\boldsymbol{x} \geq \boldsymbol{0}$, such that the weighted combination of the features matches the label, i.e., such that such that $\operatorname{sgn}(\boldsymbol{x}^\top \boldsymbol{a_i}) = \ell_i$ for all $i \in \{1, \dots, m\}$. Alternatively, we can define vectors $\boldsymbol{b_i} := \ell_i \boldsymbol{a_i}$ and we then require that $\boldsymbol{x}^\top \boldsymbol{b_i} \geq 0$ for all $i \in \{1, \dots, m\}$.

Concretely, we want to solve the following approximate version of the problem. Assume that there exists a non-negative vector $\boldsymbol{x^*}$ such that $\boldsymbol{b_i}^\top \boldsymbol{x^*} \geq 0$ for all i. W.l.o.g., we can assume that $\boldsymbol{x}^*$ is normalized such that $\boldsymbol{1}^\top \boldsymbol{x^*} = 1$, i.e., the entries of $\boldsymbol{x}^*$ sum up to 1. For a given parameter $\delta > 0$, our goal will be to find a vector $\boldsymbol{x}$, which is also normalized such that $\boldsymbol{1}^\top \boldsymbol{x} = 1$ such that $\boldsymbol{b_i}^\top \boldsymbol{x} \geq -\delta$ for all $i \in \{1, \dots, m\}$. In order to achieve this, we use the MWU algorithm as follows.

Assume that we have $\|\boldsymbol{b_i}\|_\infty \leq \rho$ for all $i \in \{1, \dots, m\}$ (i.e., all the absolute entries of the vectors $\boldsymbol{b_i}$ are upper bounded by ρ). We run the algorithm with n experts, one corresponding to each dimension. We interpret the vector $\boldsymbol{x}$ as a probability distribution on the n experts (dimensions) and initialize $\boldsymbol{x_1} := \frac{1}{n} \cdot \boldsymbol{1}$ to be the uniform distribution. In each step $t \geq 1$ of the MWU algorithm, we find a feature vector $\boldsymbol{b_i}$ for which $\boldsymbol{b_i}^\top \boldsymbol{x_t} < -\delta$ (if no such $\boldsymbol{b_i}$ exists, we are done and output the vector $\boldsymbol{x_t}$). We define the loss of expert $j \in \{1, \dots, n\}$ as $-b_{i,j}/\rho$ (where $b_{i,j}$ is the j^{th} entry of vector $\boldsymbol{b_i}$).

Show that after at most $O\big(\frac{\rho^2}{\delta^2} \log n\big)$ steps of the MWU algorithm, we have found a vector $\boldsymbol{x}$ for which $\boldsymbol{x}^\top \boldsymbol{b_i} \geq -\delta$ for all $i \in \{1, \dots, m\}$.