Bogdan Alexe
Machine Learning
Faculty of Mathematics and Computer Science
University of Bucharest

Assignment 2

Due: Wednesday, 5 June 2019

Problem 1 (1 point)

Let $\mathcal{X} = \mathbf{R}$ and consider $\mathcal{H}$ the class of 3-piece classifiers (signed intervals):

$\mathcal{H} = \{h_{a,b,s}: \mathbf{R} \rightarrow \{-1,1\}, a \leq b, s \in \{-1,+1\}\}$, where

$$h_{a,b,s}(x) = \begin{cases} s & \text{if } x \in [a,b] \\ -s & \text{if } x \notin [a,b] \end{cases}$$

Give an efficient ERM algorithm for class $\mathcal{H}$ and compute its complexity for each of the following cases:

a. realizable case.
b. agnostic case.

Justify the correctness of your algorithm.

Note: Solutions which do not give the most efficient algorithm in terms of complexity will be penalized.

Problem 2 (1 point) - problem 8.2 in the book "Understanding Machine Learning: From Theory to Algorithms".

Let $\mathcal{H}_1$, $\mathcal{H}_2$, … be a sequence of hypothesis classes for binary classification. Assume that there is a learning algorithm that implements the ERM rule in the realizable case such that the output hypothesis of the algorithm for each class $\mathcal{H}_n$ only depends on $O(n)$ examples out of the training set. Furthermore, assume that such a hypothesis can be calculated given these $O(n)$ examples in time $O(n^2)$, and that the empirical risk of each such hypothesis can be evaluated in time $O(mn)$. For example, if $\mathcal{H}_n$ is the class of axis aligned rectangles in $\mathbf{R}^n$, we saw that it is possible to find an ERM hypothesis in the realizable case that is defined by at most *2n* examples. Prove that in such cases, it is possible to find an ERM hypothesis for $\mathcal{H}_n$ in the unrealizable case in time $O(m \times n \times m^{O(n)})$.

Problem 3 (1.5 points)

Consider the boosting algorithm described (page 4) in the article "Rapid object detection using a boosted cascade of simple features", P. Viola and M. Jones, CVPR 2001. Consider that the number of positives is equal with the number of negative examples ($l = m$).

a. Prove that the distribution w_{t+1} obtained at round t + 1 based on the algorithm described in the article is the same with the distribution $\boldsymbol{D}^{(t+1)}$ obtained based on the procedure described in lecture 11 (slides 11-13).
b. Prove that the two final classifiers (the one described in the article and the one described in the lecture) are equivalent.
c. Assume that at each iteration t of AdaBoost, the weak learner returns a hypothesis h_t for which the error ε_t satisfies $\varepsilon_t \le 1/2 - \gamma$, $\gamma > 0$. What is the probability that the classifier h_t (selected as the best weak learner at iteration t) will be selected again at iteration t+1? Justify your answer.

Bonus Problem (1 point)

Consider H^d_{2DNF} the class of 2-term disjunctive normal form formulae consisting of hypothesis of the form h: $\{0,1\}^d \to \{0,1\}$,

$$h(\mathbf{x}) = A_1(\mathbf{x}) \vee A_2(\mathbf{x}),$$

where $A_i(\mathbf{x})$ is a Boolean conjunction of literals (in H^d_{conj}).

It is known that the class H^d_{2DNF} is not efficient properly learnable but can be learned improperly considering the class H^d_{2CNF}.

a. Give a γ-weak-learner algorithm for learning the class H^d_{2DNF} which is not a stronger PAC learning algorithm for H^d_{2DNF} (like the one considering H^d_{2CNF}). Prove that this algorithm is a γ-weak-learner algorithm for H^d_{2DNF}.

b. Compute the bound obtained with $L(\mathcal{B},T)$. Compare this bound with the bound based on VC-dimension of the class H^d_{2CNF}.

Hint: Find an algorithm that returns h(x)=0 or the disjunction of 2 literals.