CS60073: Advanced Machine Learning

End-semester Examination Spring 2019

Answer all FOUR questions. Answer all parts of a question together.
Time: 3 hrs. Total marks: 100.

1.(a) Write the steps of a generic Structural Risk Minimization algorithm. [5]

(b). Consider the input space to be the unit interval on the real line, $\mathcal{X} = [0,1] \subset \mathbb{R}$. A binning is a partition of the unit interval into m non-overlapping equi-sized sub-intervals (bins) $Q_j, j \in \{1, 2, \ldots, m\}$. Now, consider the class of classification rules that take either the value 0 or 1 in each of the sub-intervals Q_j, i.e.,

$$\mathcal{H}_m = \left\{ h : \mathcal{X} \to \{0,1\} : h(x) = \sum_{j=1}^{m} c_j \mathbf{1}(x \in Q_j), c_j = \{0,1\} \right\},$$

where, $\mathbf{1}()$ is the indicator function. Note that this class has 2^m elements. The histogram classifier is the element of this class obtained by doing a majority vote inside each sub-interval. Namely,

$$\hat{c}_j = \begin{cases} 1 & \text{if} \quad \frac{\sum_{i,x_i \in Q_j} y_i}{\sum_{i,x_i \in Q_j} 1} \geq 1/2, \\ 0 & \text{otherwise.} \end{cases}$$

Show that the histogram classifier is the Empirical Risk Minimizer for $\mathcal{H}_m$. [10]

(c). A band classifier in 2-dimension uses two "parallel" hyperplanes in 2-dimension, and labels the region between the hyperplanes as $y = 1$ and the remaining region as $y = 0$. Derive the VC dimension of the class of all possible band classifiers in 2-dimension. [10]

2. Consider the input space $X_n = \{0,1\}^n$ of n-bit vectors. Consider the following hypothesis class $\mathcal{H}_{mono} = \{0, 1, x_1, \bar{x_1}, x_2, \bar{x_2}, \ldots x_n, \bar{x_n}\}$. The hypothesis class contains $2n + 2$ functions. The functions "0" and "1" are constant and predict 0 and 1 on all instances in X_n. The function "x_i" evaluates to 1 on any $a \in \{0,1\}^n$ satisfying $a_i = 1$, and evaluate to 0 otherwise. Likewise the function "$\bar{x_i}$" evaluates to 1 on any $a \in \{0,1\}^n$ satisfying $a_i = 0$, and evaluates to 0 otherwise. Also, consider another function class $\mathcal{H}_{conjunction}$ of all possible conjunctions that can be defined over X_n. A function in $\mathcal{H}_{conjunction}$ is a conjunction of a set of literals, where each literal corresponds to either a variable x_i or its negation $\bar{x_i}$. Note that the number of literals in a conjunction function may vary.

(a). Show that the class $\mathcal{H}_{conjunction}$ is $\frac{1}{10n}$-weak learnable using $\mathcal{H}_{mono}$. [15]

(b). Let the function class $\mathcal{H}_{conjunction_k}$ denote the class of conjunctions on at most k literals. Show that the class $\mathcal{H}_{conjunction_k}$ is $\frac{1}{10k}$-weak learnable using $\mathcal{H}_{mono}$. [10]

3.(a). State the Halving algorithm for online learning. Let $\mathcal{H}$ be a finite hypothesis class. Show that the Halving algorithm enjoys a mistake bound $M_{Halving} \leq \log_2(|\mathcal{H}|)$. [10]

(b). Let $d \geq 2$, $\mathcal{X} = \{1, 2, \ldots, d\}$ and let $\mathcal{H} = \{h_j : j \in [d]\}$, $h_j(x) = \mathbf{1}_{[x=j]}$. Here $\mathbf{1}_{[x=j]}$ is an indicator function. Calculate the mistake $M_{Halving}(\mathcal{H})$. [10]

(c). Let $Ldim(\mathcal{H})$ denote the Littlestone dimension and $VCdim(\mathcal{H})$ denote the VC dimension of any function class $\mathcal{H}$. Show that $VCdim(\mathcal{H}) \leq Ldim(\mathcal{H})$. [5]

4.(a). Consider a set of non-negative weights $w_i \geq 0$ and a set of convex functions f_i, for $i = 1, \ldots, r$. Show that $g(x) = \sum_{i=1}^{r} w_i f_i(x)$ is also a convex function. [10]

(b). Consider the problem of learning halfspaces with hinge loss. We limit our domain to the Euclidean ball with radius R. The label set is $\mathcal{Y} = \{-1, +1\}$, and the hinge loss function l is defined by $l(\mathbf{w}, (\mathbf{x}, y)) = \max[0, 1 - y(\mathbf{w} \cdot \mathbf{x})]$. Show that the loss function is convex. Show that it is also R-Lipschitz. [15]

–BEST WISHES–