Homework 1
posted February 1, 2020
due March 2, 2020
Reminders:
• Solutions should include the statements of the problems.
• The preferred format is LaTeX.
• Solution must be your own; homework must be handed in class and on time.

1. Let $X = \mathbb{R}^2$ and $Y = \{0, 1\}$. The set of hypothesis $\mathcal{H}$ is the class of concentric circles in $\mathbb{R}^2$: namely, the hypothesis $h_r$ is the circle defined by $x^2 + y^2 \leq r^2$. A labeling function $f : X \rightarrow Y$ defined a point $P$ as a positive example if $f(P) = 1$ and a negative example if $f(P) = 0$. The realizability assumption means that a circle of radius $r^*$ exists that contains all positive example.

(a) Suppose that an ERM algorithm returns for a training sequence $S = \{(P_i, y_i) \mid 1 \leq i \leq m\}$ a circle $h$ of radius $\bar{r}$. Prove that the error of this prediction rule is bounded above by the probability that no point in $S$ belongs to the set $E = \{x \in \mathbb{R}^2 \mid \bar{r} \leq \|x\| \leq r^*\}$.

(b) Prove the inequality $(1 - \epsilon)^m \leq e^{-cm}$.

(c) Prove that $\mathcal{H}$ is PAC-learnable and its sample complexity is bounded by

$$m_{\mathcal{H}}(\epsilon, \delta) \leq \left\lceil \frac{\log \frac{1}{\delta}}{\epsilon} \right\rceil.$$ 

2. Consider the hypothesis class $\mathcal{H}$ of all Boolean conjunctions of $d$ variables. Define $X = \{0, 1\}^d$ and $Y = \{0, 1\}$. A literal over the variables $x_1, \ldots, x_d$ is a Boolean function such that $f(x) = x_i$ or $f(x) = \overline{x_i}$ for some $i$, $1 \leq i \leq d$, where $x = (x_1, \ldots, x_d)$. A conjunction is any product of literals (e.g., $h(x) = x_1\overline{x_2}$, where $x = (x_1, x_2)$.

Consider the hypothesis class of all conjunctions of literals over $d$ variables. The empty conjunction is interpreted as the all-positive hypothesis ($h(x) = 1$ for all $x$). Any conjunction which contains a variable
and its negation (like \( x_i \overline{x_j} \), etc.) is interpreted as the all-negative hypothesis.

We assume realizability, which in this context, means that there exists a Boolean conjunction that generates the labels. Thus, each example \((x, y) \in \mathcal{X} \times \mathcal{Y}\) consists of an assignment to the \(d\) Boolean variables and its truth value. For example, for \(d = 3\) and the true hypothesis \(f(x) = x_1 \overline{x_2}\), the training set \(S\) may contain

\[ ((1,1,1), 0), ((1,0,1), 1), ((0,1,0), 0), ((1,0,0), 1). \]

(a) Prove that \(|\mathcal{H}| = 3^d + 1\);

(b) Prove that the hypothesis class of all conjunctions over \(d\) variable is PAC learnable and bound its sample complexity \(m_{\mathcal{H}}(\epsilon, \delta)\).

(c) Design an algorithm that implements the ERM rule and whose time is polynomial in \(dm\).