Homework 1
posted September 25, 2017
due October 9, 2017

1. Ax axis alligned rectangle classifier in the plane is a classifier that assigns 1 to a point if and only it it is insider a certain rectangle. Formally, given real numbers a_1, b_1, a_2, b_2 such that $a_1 \leqslant b_1$ and $a_2 \leqslant b_2$ define $h_{(a_1,b_1,a_2,b_2}$ by

$$h(x_1, x_2) = \begin{cases} 1 & \text{if } a_1 \leqslant x_1 \leqslant b_1 \text{ and } a_2 \leqslant x_2 \leqslant b_2, \\ 0 & \text{otherwise.} \end{cases}$$

   The class $\mathcal{H}_{rec}$ of axis-alligned rectangles is

$$\mathcal{H}_{rec} = \{h_{(a_1,b_1,a_2,b_2)} \mid a_1 \leqslant b_1, a_2 \leqslant b_2\}.$$

   Some points are labeled as positive, others as negative. We hypothesise that the realizability assumption holds.

   (a) We hypothesised that the realizability assumption holds. What does this mean?

   (b) Let $\mathcal{A}$ be the algorithm that returns the smallest rectangle enclosing all positive examples in the training set. Show that $\mathcal{A}$ is an ERM.

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

   Consider the hypothesis class of all conjunctions of literals over d variables. The empty conjunction is interpreted as the all-positive hypothesis ($h(\mathbf{x}) = 1$ for all $\mathbf{x}$). Any conjunction which contains a variable and its negation (like $x_i\overline{x_i}x_j$, 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 $(\mathbf{x}, y) \in \mathcal{X} \times \mathcal{Y}$ consists of an assignment to the d Boolean variables and its truth value. For example, for $d = 3$ and the true hypothesis $f(\mathbf{x}) = x_1\overline{x_2}$, 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.