Model Solution for Homework 1

Statistical Learning Theory

1. Suggest ERM an algorithm for the functions class of conjunctions defined below:

Consider data of the form $\{x_1, \ldots, x_n\} \in \{0,1\}^n$. A *conjunction* is a function on some subset of the variables $\{x_1, x_2 \ldots, x_n, \bar{x}_1, \bar{x}_2 \ldots, \bar{x}_n\}$, where $\bar{x}_i = 1 - x_i$. It maps x to 1 if every variable in the subset is 1, and 0 otherwise. For example:

$$\begin{aligned} h(x) &= x_1 \wedge x_2 \wedge \bar{x}_5 \\ h(x) &= \bar{x}_3 \wedge \bar{x}_4 \\ h(x) &= x_1 \wedge x_2 \wedge \cdots \wedge x_r \end{aligned}$$

Solution:

ERM Algorithm -

Algorithm 1 LEARNCONJ

Maintain two sets of indices $S = \{x_1, x_2 \ldots, x_n\}$,$\bar{S} = \{\bar{x}_1, \bar{x}_2, \bar{x}_n\}$
for each input $x^{(i)}$ **do**
  if $l(x^{(i)}) = 1$ **then**
    For each $j \in [n]$, remove $\bar{x}_j$ from $\bar{S}$ if $x_j^{(i)} = 1$, and remove x_j from S if $x^{(i)} = 0$
  else
    Do nothing
  end if
end for
Output the conjunction of the remaining literals in S and $\bar{S}$.

At any point, the hypothesis $h(x) = \left(\bigwedge_{i \in S} x_j\right) \wedge \left(\bigwedge_{i \in \bar{S}} \bar{x}_j\right)$ is consistent with all examples seen so far.

2. Derive the sample complexity of the suggested ERM algorithm..

Solution:

Theorem 1. *With $m \geq \frac{1}{\epsilon}(n \log 3 + \log \frac{1}{\delta})$ examples, with probability $1 - \delta$,* LEARNCONJ *will be correct on a new example from $\mathcal{D}$ with probability $1 - \epsilon$.*

Proof. Let $h \in H$ be a hypothesis with $Pr_{x \sim \mathcal{D}}(h(x) \neq l(x)) > \epsilon$. The probability that h is not eliminated after m examples is at most $(1 - \epsilon)^m$.

The probability that any such 'bad h' survives after m examples is at most $3^n(1 - \epsilon)^m$.

Setting $3^n(1 - \epsilon)^m < \delta$ and taking the log of both sides, we see that $m \geq \frac{1}{\epsilon}(n \log 3 + \log \frac{1}{\delta})$ suffices.

□