Foundations of Machine Learning — Homework Assignment 1

Anirudhan J Rajagopalan
N18824115
ajr619

October 11, 2015

A. PAC Learning

1

Algorithm $\mathcal{A}$: Given a sample $\mathcal{S}$, the algorithm returns the tightest interval I' = I_s containing all points labeled with 1 in sample S.

Error intervals: We define error regions as the intervals that lie between [a,b] but outside I'.

Proof of PAC learnability :

Let I be the target concept — [a,b].
Let $\epsilon > 0$ be the error.
Let m be the number of samples in S.
Let I_s be the tightest interval formed by the the points labelled 1 from sample S.
Let $\Pr[I_s]$ denote the probability mass of the interval defined by I_s.

$$\Pr[I_s] > \epsilon$$

Lets assume that the interval I_s (denoted by [a',b']) has error ϵ. The error can be found in intervals [a,a') and (b',b] denoted by $I_1 and I_2$ respectively. If we assume that the error is equally distributed across the two regions, we can denote the error to be $\epsilon/2$ for each of I_1 and I_2. Each point in the error region has a probability of (1-$\frac{\epsilon}{2}$). So Probability of the error in sample being greater than ϵ can be written as

$$\begin{aligned}
\Pr_{S\sim\mathcal{D}^m}[R(I_S) > \epsilon] \leq &\sum_{i=1}^{2} \Pr_{S\sim\mathcal{D}^m}[I \bigcap I_i = \emptyset] \\
\leq &2(1-\epsilon/2)^m \\
\leq &2\mathrm{e}^{-m\epsilon/2}
\end{aligned}$$

Equating the RHS to δ gives us the sample complexity.

$$\begin{aligned}
2\mathrm{e}^{-m\epsilon/2} =&\delta \\
\frac{2}{\delta} =&\mathrm{e}^{m\epsilon/2} \\
\log\frac{2}{\delta} =&\frac{m\epsilon}{2} \\
m =&\frac{2}{\epsilon}\log\frac{2}{\delta}
\end{aligned}$$

2

Algorithm $\mathcal{A}$: Given a sample $\mathcal{S}$ for target concept C containing p closed intervals,

1. If there are p separate sequence of positively labeled points in the training data, then return the union of the p tightest intervals containing the positive points.
2. Otherwise, return (p - i) tightest intervsals, each containing a sequence of positie labels separfated by i negative labels. i can take the values 0 to (p -1).

Error intervals: Let $[a_1, b_1] \cup [a_2, b_2] \cup ...[a_p, b_p]$ be the target concept C. Let $\epsilon > 0$. We can assume that $\Pr[a_i, b_i] > \epsilon/(k+1)$. The actual error on the training set is gap between the target concept $[a_i, b_i]$ and the learned concept $[a_i', b_i']$. Each of the error regions r_i can occur with a probability $\epsilon/2(k+1)$. (Factor of 2 for $[a_i, a_i'] and [b_i', bi]$.

Proof of PAC learnability :

If error($\mathrm{h_S}$) $> \epsilon$, then either the union of the intervals misses one of the regions r_i or $\Pr[b_i, a_{i+1}] > \epsilon/(k+1)$. Thus by union bound, we have

$$\begin{aligned}
\Pr[\text{error}(\mathrm{h_S})] \leq& \Pr[\exists i \in [1,p] : h_S \text{ misses } r_i] + e^{-m\epsilon/(k+1)} \\
\leq& 2k(1 - \frac{\epsilon}{2(k+1)})^m + e^{-m\epsilon/(k+1)} \\
\leq& 2ke^{-m\epsilon/2(k+1)} + e^{-m\epsilon/(k+1)} \\
\leq& (2k+1)e^{-m\epsilon/2(k+1)}
\end{aligned}$$

Setting $\delta = 0$ and solving the RHS will give us a bound for m.

$$(2k+1)e^{-m\epsilon/2(k+1)} = \delta \tag{1}$$

$$\frac{2k+1}{\delta} = e^{m\epsilon/2(k+1)} \tag{2}$$

$$\log \frac{2k+1}{\delta} = \frac{m\epsilon}{2(k+1)} \tag{3}$$

$$m = \frac{2(k+1)}{\epsilon} \log \frac{2k+1}{\delta} \tag{4}$$

Therefore, for probability of error to be less than ϵ m should be greater than the value obtained in (4).

When p is 2 :

Substituting k = 2 in the equation (4) above gives us $m \geq \frac{6}{\epsilon} \log \frac{5}{\delta}$