PAC bounds

Soham Dan : sdan2

ECE 543 : University of Illinois Urbana-Champaign

1. Introduction

This report explores the work by Auer and Ortner (A new PAC bound for intersection-closed concept classes) to address the question of how many labeled examples are needed for a successful model in the PAC learning framework (success being measured by the $\epsilon - \delta$ metric of PAC)
The following results are known :

- Ehrenfeucht et al. (1989) : Every PAC algorithm needs $\Omega(\frac{1}{\epsilon}(d+\log(\frac{1}{\delta})))$
- Blumer et al. (1989) : Every consistent algorithm needs no more than $\emptyset(\frac{1}{\epsilon}(d\log(\frac{1}{\epsilon})) + \log(\frac{1}{\delta}))$ labeled examples.
- Auer and Ortner (2007) : This upper bound is tight for some consistent algorithms

In this paper : new PAC bounds for arbitrary intersection closed concept classes of $\emptyset(\frac{1}{\epsilon}(d\log d + \log\frac{1}{\delta}))$ with the closure algorithm. For the proofs, please refer to the original paper.

2. Definitions

2.1. Concept Class

A concept class over a (possibly infinite) set X is a subset $C \subseteq 2^X$. For $Y \subseteq X$ we set $C \cap Y := \{C \cap Y | C \in \mathbf{C}\}$. The VC-dimension of a concept class $C \subseteq 2^X$ is the cardinality. of a largest $Y \subseteq X$ for which $C \cap Y = 2^Y$.

2.2. Intersection Closed Concept Class

A concept class $C \subseteq 2^X$ is intersection-closed if for all $C_1, C_2 \in C$ the intersection $C_1 \cap C_2$ is in $\mathbf{C}$ as well.

2.3. Closure

For any set $Y \subseteq X$ and any concept class $C \subseteq 2^X$ we define the closure of Y (with respect to C) as the intersection of all concepts in C that contain Y , i.e. $clos_C(Y) := \cap_{Y \subseteq C \in \mathbf{C}} C$. there is no concept containing Y , then the closure is by definition of the nullary intersection the set X itself, so that $Y \subseteq clos(Y)$ holds in general. The following proposition provides an alternative definition of intersection-closed concept classes for finite X.

Proposition : A concept class $C \subseteq 2^X$ over finite X is intersection-closed if and only if for $Y \subseteq C \in \mathbf{C}$ one always has $clos(Y) \in C$.

Proposition : If **C** is intersection closed and $VC(\mathbf{C}) = d$, then $VC(\mathbf{C'}) = d$. Here **C'** is the concept class of all intersections of concepts in **C**.

A *spanning set* of Y (with respect to an intersection-closed concept class $C \subseteq 2^X$) is any set $S \subseteq Y$ such that $clos_C(S) = clos_C(Y)$. A spanning set S of Y is called *minimal* if no subset of S is a spanning set of Y . Finally, let $span_C(Y)$ denote the set of all minimal spanning sets of Y . Again we will often drop the index, if no ambiguity can arise. Note that if Y is finite, then $span(Y) = \phi$.

Theorem : Let $C \subseteq 2^X$ be an intersection closed class of VC dimension d. Let $Y \subseteq X$ be finite and contained in some concept of C. Then all minimal spans of Y have size atmost d.

Corollary : Let $C \subseteq 2^X$ be an intersection closed class of VC dimension d. Then all minimal spans of any finite $Y \subseteq X$ have size at most $d + 1$.

Sauer Shelah Lemma(1972):

$$|C| \leq \binom{|X|}{\leq d} = \sum_{i=0}^{d} \binom{X}{i}$$

3. Central Theorem of the Paper

Theorem : Let $\mathbf{C} \subseteq 2^X$ be a well-behaved. intersection closed concept class of VC dimension $d \geq 10$. Then **C** is PAC learnable from $m =$

$\max\{\frac{16}{\epsilon} d \log d, \frac{6}{\epsilon} \log \frac{7}{\delta}\}$ examples.
But, this does not hold for any consistent learning algorithm but the *closure algorithm.*

The **Closure Algorithm** essentially returns a hypothesis which is the closure of the set of positive examples and thus negative examples do not play a role in determining the output

Examples : For the class of convex sets (which is intersection closed), the algorithm simply returns the convex hull of the positive examples.

Proposition : The hypothesis generated by the closure algorithm classifies all negative examples correctly.

Proposition : Let **C** be some intersection closed concept class. If **C** satisfies k of given examples $x_1, x_2, ..., x_{2m}$ incorrectly then they are in $\cup_{i=1}^{k} S_i$

Lemma: Let $\mathbf{C} \subseteq 2^X$ be a well, behaved, intersection closed concept class of VC dimension d, P a probability distribution on X and $C \in \mathbf{C}$. Then for all $\epsilon \geq 0$ and for all $m \geq \frac{2}{\epsilon}$, given m independent random examples labeled by C and drawn according to P, the probability that the hypothesis h generated by the closure algorithm has error $\mathbf{er}_{C,P}(h) > \epsilon$ is at most $2\sum_{k=p}^{m} 2^{-k}\binom{kd}{\leq d}$ where $p = \epsilon m/2$.

4. Bound for Classes with additional properties

Theorem : If $d \geq 10$ and $m \geq \max\{\frac{16}{\epsilon} d \log d, \frac{6}{\epsilon} \log \frac{7}{\delta}\}$, then $2\sum_{k=p}^{m} 2^{-k}\binom{kd}{\leq d} < \delta$ where $p = \epsilon m/2$. An optimal PAC bound for intersection closed classes with homogeneous spans: For these classes, one can actually obtain $\emptyset(\frac{1}{\epsilon}(d + \log \frac{1}{\delta}))$ which matches the lower bound by Ehrenfeucht et al.,1989.

Definition : An intersection-closed concept class $C \subseteq 2^X$ of VC-dimension d is said to have homogeneous spans S if one can assign to each finite $Y \subseteq C \in \mathbf{C}$ a (not necessarily minimal) spanning set S(Y) of size at most d, such that for all $Y \subseteq X$ and all $x \in S(Y)$: $S(Y) \backslash x \subseteq S(Y \backslash x)$.

Proposition: Let $C \subseteq 2^X$ be an intersection-closed concept class $C \subseteq 2^X$ with homogeneous spans. Then for all finite $Y \subseteq X$ and all $Z \subseteq S(Y)$: $S(Y) \backslash Z \subseteq S(Y \backslash Z)$.

Theorem: Let C be a well-behaved, intersection-closed concept class of VC-dimension d. If C has homogeneous spans, then it is PAC learnable from $m = \emptyset(\frac{1}{\epsilon}(d + \log \frac{1}{\delta}))$

Theorem:Let X be an arbitrary set and $C_{X,d}$ the class of all subsets of X of size at most d. Furthermore, let A be an algorithm that chooses as its hypothesis not the smallest concept consistent with the given examples (as the closure algorithm does), but an arbitrarily selected largest consistent concept. Then A needs $\Omega(\frac{e}{\epsilon}(d + \log \frac{1}{\delta}))$examples to learn $C_{X,d}$.

5. Conclusion

The major direction to be explored in the future is to get rid of the homogeneous span restriction and finding similar results for general intersection closed concept classes. As the authors acknowledge, this is far from trivial and combinatorial properties of such classes need to be further explored.
The main takeaway from this paper is twofold: A new PAC bound for arbitrary intersection closed concept classes and a proof of the optimal bound for intersection closed classes with homogeneous spans.

6. References

Auer,P. & Ortner,R. A new PAC bound for intersection-closed concept classes.Mach Learn (2007) 66:151163
Simon, H. An Almost Optimal PAC Algorithm. JMLR: Workshop and Conference Proceedings vol 40:112, 2015
Auer, P. (1997). Learning nested differences in the presence of malicious noise. Theor. Comput. Sci., 185(1), 159175.
Auer, P., & Cesa-Bianchi, N. (1998). On-line learning with malicious noise and the closure algorithm. Ann. Math. Artif. Intell., 23 (12), 8399.
Auer, P., Long, P.M., &Srinivasan, A. (1998). Approximating hyper-rectangles: Learning and pseudorandom sets. J. Comput. Syst. Sci., 57(3), 376388.
Blumer, A., Ehrenfeucht, A., Haussler, D., &Warmuth, M. (1989). Learnability and the Vapnik-Chervonenkis dimension. J. ACM, 36(4), 929965.

Ehrenfeucht, A., Haussler, D., Kearns, M. J., & Valiant, L. G. (1989). A general lower bound on the number of examples needed for learning. Inf. Comput., 82(3), 247261.
Floyd, A.,&Warmuth, M. (1995). Sample compression, learnability, and theVapnik-Chervonenkis dimension. Machine Learning, 21(3), 269304.
Haussler, D., Littlestone, N., & Warmuth, M. (1994). Predicting 0,1-functions on randomly drawn points. Inf. Comput., 115(2), 248292.
Helmbold, D., Sloan, R., & Warmuth, M. (1990). Learning nested differences of intersection-closed concept classes. Machine Learning 5, 165196.
Sauer, N. (1972). On the density of families of sets. J. Combin. Theory Ser. A, 13, 145147.