k-fold unions of low-dimensional concept classes

David Eisenstat

September 2009

Abstract

We show that 2 is the minimum VC dimension of a concept class whose k-fold union has VC dimension $\Omega(k \log k)$.

Keywords: combinatorial problems, VC dimension, concept class, union of concepts, intersection of concepts

1 Introduction

The Vapnik-Chervonenkis (VC) dimension of a class of concepts measures the difficulty of learning a concept in that class [1, 6]. Many interesting classes consist of concepts that are unions or intersections of simpler concepts, and it is natural to ask whether the VC dimension of the primitive class can be used to bound the VC dimension of the combinations. Via a counting argument, Haussler and Welzl [4, Lemma 4.5] showed that the VC dimension of unions of k concepts from a fixed class is $O(k \log k)$, a result that holds for intersections as well by duality. The best known lower bound was $\Omega(k)$ until Eisenstat and Angluin [3, Theorem 1] used the probabilistic method to construct a class of VC dimension 5 whose k-fold union has VC dimension $\Theta(k \log k)$.

In this work, we reduce the VC dimension of the constructed class from 5 to 2 and show that no further reduction is possible. The main new difficulty is that to achieve VC dimension 4 or less, we now must let each new concept depend on the partially constructed class. We also simplify the previous reasoning by exposing the potential function as a parameter and by drawing the points in a concept with replacement instead of without.

2 Definitions

Let X be a set of *points*. A typical choice for X is $[n] = \{1, \ldots, n\}$, though X may be infinite. A *concept* on X is a subset of X. A *(concept) class* on X is a nonempty set of concepts on X. A class $\mathcal{C}$ on X *shatters* a set $S \subseteq X$ if for all sets $T \subseteq S$, there exists a concept $C \in \mathcal{C}$ such that $C \cap S = T$. The *VC dimension* of $\mathcal{C}$ is $\mathrm{VCdim}(\mathcal{C}) = \sup\{|S| : S \subseteq X,\ \mathcal{C} \text{ shatters } S\}$. The *$k$-fold union* of $\mathcal{C}$ is $\mathcal{C}^{k\cup} = \{C_1 \cup \cdots \cup C_k : C_1, \ldots, C_k \in \mathcal{C}\}$, another class on X.

We let $\log x$ stand for the natural logarithm of x.

3 Construction

In this section, we prove the following theorem.

Theorem 1. *There exists a class $\mathcal{C}$ such that $\mathrm{VCdim}(\mathcal{C}) \leq 2$ and $\mathrm{VCdim}(\mathcal{C}^{k\cup}) = \Omega(k \log k)$.*

Let $g(n) = O(n/\log n)$ be an increasing function to be chosen later. For all sufficiently large n, we construct a class $\mathcal{C}_n$ on $X = [n]$ such that $\mathrm{VCdim}(\mathcal{C}_n) \leq 2$ and $\mathrm{VCdim}(\mathcal{C}_n^{g(n)\cup}) = n$. Noting that $n = \Omega(g(n) \log g(n))$, we finish by assembling the classes $\mathcal{C}_n$ into one class.

3.1 Framework

We make two definitions to guide our choice of concepts for $\mathcal{C}_n$. One way to ensure that $\mathrm{VCdim}(\mathcal{C}_n) \leq 2$ is for each pair of concepts in $\mathcal{C}_n$ to have at most one point in common.

Definition 2. A class $\mathcal{C}$ is *d-almost disjoint* if for distinct $C, C' \in \mathcal{C}$, $|C \cap C'| < d$.

Lemma 3. *If $\mathcal{C}$ is d-almost disjoint, then* $\mathrm{VCdim}(\mathcal{C}) \leq d$.

Proof. Let S be a set of size $d+1$ and let $x \in S$. There do not exist concepts $C, C' \in \mathcal{C}$ such that $C \cap S = S$ and $C' \cap S = S \setminus \{x\}$, as then $|C \cap C'| \geq d$ in contradiction of the hypothesis. We conclude that $\mathcal{C}$ does not shatter any set of size $d+1$. □

Next we give a condition that implies that $\mathrm{VCdim}(\mathcal{C}_n^{g(n)\cup}) = n$.

Definition 4. Let $f : \mathbf{Z}^+ \to \mathbf{R}^+$ be a nondecreasing function from the positive integers to the positive real numbers. A concept C on X *f-covers* a finite nonempty set $S \subseteq X$ if $C \subseteq S$ and $|C| \geq f(|S|)$. A class $\mathcal{C}$ on X *f-covers* a finite nonempty set $S \subseteq X$ if some $C \in \mathcal{C}$ f-covers S, and $\mathcal{C}$ is *f-covering* if it f-covers all such sets.

Lemma 5. *If $\mathcal{C}$ is a class on $[n]$ that is f-covering and that contains the empty concept, then $\mathcal{C}^{k\cup}$ shatters $[n]$, where $k = \lfloor \sum_{t=1}^{n} 1/f(t) \rfloor$.*

Proof. Let $F(s) = \sum_{t=1}^{s} 1/f(t)$. It suffices to show that for all sets $S \subseteq [n]$, S is the union of at most $F(|S|)$ concepts in $\mathcal{C}$, as the empty concept can be used to bring the total up to k. We proceed by complete induction on $|S|$.

If $|S| = 0$, then the conclusion is immediate. Otherwise, since $\mathcal{C}$ is f-covering, there exists a concept $C \in \mathcal{C}$ such that $C \subseteq S$ and $|C| \geq f(|S|)$. Writing $S = (S \setminus C) \cup C$, we find by hypothesis that S is the union of at most $F(|S| - |C|) + 1$ concepts in $\mathcal{C}$. For all $t \leq |S|$, we have $1/f(t) \geq 1/f(|S|) \geq 1/|C|$ as f is positive and nondecreasing, and $\sum_{t=|S|-|C|+1}^{|S|} 1/f(t) \geq 1$. It follows that $F(|S| - |C|) + 1 \leq F(|S|)$. □

The function f should be large enough that $k = O(n/\log n)$ but small enough that it is possible to choose concepts with small intersections. Following some trial and error, we define

$$f(t) = \begin{cases} 1 & \text{if } t \leq n^{5/6}, \\ \log_{n/t} n^{1/6} & \text{if } n^{5/6} < t \leq n/2, \\ \log_2 n^{1/6} & \text{if } n/2 < t \end{cases} \qquad g(n) = \left\lfloor \sum_{t=1}^{n} 1/f(t) \right\rfloor,$$

and observe that f is continuous, positive and nondecreasing. When $n^{5/6} < |S| \leq n/2$, the probability that $\lceil f(|S|) \rceil$ random points all belong to S is about $1/n^{1/6}$, usually less but never less than $1/n^{1/3}$ as we show later. We make adequate progress without large concepts when $|S| > n/2$, and when $|S| \leq n^{5/6}$, we resort to singletons.

Lemma 6. $g(n) = O(n/\log n)$.

Proof. Since $1/f$ is nonincreasing, $g(n)$ is bounded above by the following integral:

$$\begin{aligned} \int_0^n 1/f(t)\,dt &= n^{5/6} + \frac{n/2}{\log_2 n^{1/6}} + \int_{n^{5/6}}^{n/2} \frac{\log(n/t)}{\log n^{1/6}}\,dt \\ &\leq O(n/\log n) + \frac{1}{\log n^{1/6}} \int_0^n \log(n/t)\,dt \\ &= O(n/\log n) + \frac{n}{\log n^{1/6}} \int_0^1 \log(1/u)\,du \\ &= O(n/\log n). \end{aligned}$$

□

3.2 Component classes

The class $\mathcal{C}_n$ is constructed by a natural greedy algorithm. Start with $\mathcal{C}_n = \{S : S \subseteq [n],\ |S| \leq 1\}$, a class that is 2-almost disjoint because it consists of the empty concept and all singletons. Since $f(t) = 1$ for all $t \leq n^{5/6}$, each nonempty set with at most $n^{5/6}$ elements is already f-covered. As long as some set is not f-covered, repeatedly add to $\mathcal{C}_n$ a concept C that f-covers as many new sets as possible, subject to the constraint that $\mathcal{C}_n \cup \{C\}$ be 2-almost disjoint. We show that this algorithm succeeds in creating a class $\mathcal{C}_n$ that is 2-almost disjoint and f-covering, implying by Lemma 3 that $\mathrm{VCdim}(\mathcal{C}_n) \leq 2$ and by Lemma 5 that $\mathrm{VCdim}(\mathcal{C}_n^{g(n)\cup}) = n$.

Let S be a set with $|S| > n^{5/6}$. Let ℓ be a random integer between 2 and $\lceil \log_2 n^{1/6} \rceil$ inclusive, let $x_1, \ldots, x_\ell$ sample $[n]$ uniformly *with replacement*, and let $C = \{x_1, \ldots, x_\ell\}$. With probability $\Theta(1/\log n)$, $\ell = \lceil f(|S|) \rceil$. In this case, assuming that $|S| < n$,

$$\begin{aligned}
\Pr[C \subseteq S] &= (|S|/n)^\ell = (|S|/n)^{\lceil f(|S|) \rceil} \geq (|S|/n)^{f(|S|)+1} \\
f(|S|) &\leq \log_{n/|S|} n^{1/6} \\
(|S|/n)^{f(|S|)} &\geq (|S|/n)^{\log_{n/|S|} n^{1/6}} = 1/n^{1/6} \\
(|S|/n)^1 &> 1/n^{1/6} \\
\Pr[C \subseteq S] &> 1/n^{1/3}.
\end{aligned}$$

If $|S| = n$, then $C \subseteq S$ with certainty.

The concept C may exhibit one of two defects. The first is when $|C| < \ell$, which implies that C is too small to f-cover S. The second is when there exists a previous concept $C' \in \mathcal{C}_n$ such that $|C \cap C'| \geq 2$, implying that $\mathcal{C}_n \cup \{C\}$ is not 2-almost disjoint. Both defects are the result of a bad joint outcome for a pair of random variables (x_i, x_j) with $i < j$, where the set of all bad outcomes is $B = \{(x, x) : x \in [n]\} \cup \{(x, y) : \{x, y\} \subseteq C' \in \mathcal{C}_n\}$. No concept in $\mathcal{C}_n$ is larger than $O(\log n)$ points, so B has at most $O(n + |\mathcal{C}_n| \log^2 n)$ elements. Assuming that $|\mathcal{C}_n| \leq n^{3/2}$, the probability that C is defective is $O((\log^2 n)|B|/n^2) = O((\log^4 n)/n^{1/2}) = o(1/n^{1/3})$, as there are $O(\log^2 n)$ pairs of variables and n^2 possible outcomes.

For each set S, then, with probability at least

$$q = \Theta(1/\log n) \cdot (n^{1/3} - o(n^{1/3})) = \Theta(1/(n^{1/3} \log n)),$$

a random concept f-covers S and can be added to $\mathcal{C}_n$. Accordingly, we can add a concept C

that reduces the number of sets not yet f-covered by a factor of $1/(1-q)$. After m greedily chosen concepts, at most $(2^n-1)(1-q)^m$ sets have not been f-covered, so the algorithm terminates when $m \leq \lceil \log_{1-q} 2^{-n} \rceil = \Theta(n/q) = o(n^{3/2})$ as required.

3.3 Assembly

We embed all of the classes $\mathcal{C}_n$ in a single class $\mathcal{C}$ via the *disjoint union* construction of Johnson [5, p. 55]. Let

$$X = \bigcup_n ([n] \times \{n\}) \qquad\qquad \mathcal{C} = \bigcup_n \{C \times \{n\} : C \in \mathcal{C}_n\},$$

so that $\mathcal{C}$ is a class on X. The counterpart of each concept $C \in \mathcal{C}_n$ is a concept on $[n] \times \{n\}$, so $\mathcal{C}$, like $\mathcal{C}_n$, is 2-almost disjoint. It follows by Lemma 3 that $\mathrm{VCdim}(\mathcal{C}) \leq 2$, and for all sufficiently large n, $\mathrm{VCdim}(\mathcal{C}^{g(n)\cup}) \geq \mathrm{VCdim}(\mathcal{C}_n^{g(n)\cup}) = n$, which is $\Omega(g(n) \log g(n))$.

4 VC dimension one

Theorem 7 ([3])**.** *If $\mathcal{C}$ is a class of VC dimension 1, then for all $k \geq 1$, $\mathrm{VCdim}(\mathcal{C}^{k\cup}) \leq k$.*

Proof. Let S be a set of $k+1$ points. If there exists a concept $C \in \mathcal{C}^{k\cup}$ such that $C \cap S = S$, then by the pigeonhole principle, there exists a concept $C' \in \mathcal{C}$ such that $|C' \cap S| \geq 2$. Letting $\{x, y\} \subseteq C'$, we observe that since $\mathrm{VCdim}(\mathcal{C}) = 1$, there exists a set $T \in \{\emptyset, \{x\}, \{y\}\}$ such that for all concepts $C'' \in \mathcal{C}$, $C'' \cap S \neq T$. As the class $\{\emptyset, \{x\}, \{y\}, \{x, y\}\} \setminus \{T\}$ is closed under union, no concept $C'' \in \mathcal{C}^{k\cup}$ satisfies $C'' \cap S = T$, and we conclude that $\mathcal{C}^{k\cup}$ does not shatter S. □

A similar result of Wenocur and Dudley [7, Theorem 2.1] states that when $\mathcal{C}_1, \ldots, \mathcal{C}_k$ are classes whose concepts are linearly ordered by inclusion, the class $\{C_1 \cup \cdots \cup C_k : C_i \in \mathcal{C}_i\}$ has VC dimension at most k. While VC dimension 1 is a more inclusive condition than linearly ordered [7, Corollary 2.2], Theorem 7 requires identical $\mathcal{C}_i$, so the results are incomparable. The simultaneous generalization of both results is false, as evidenced by the following counterexample where $\mathcal{C}_1$ and $\mathcal{C}_2$ have VC dimension 1 but $\{C_1 \cup C_2 : C_1 \in \mathcal{C}_1,\ C_2 \in \mathcal{C}_2\}$ has VC dimension 3:

$$\mathcal{C}_1 = \{\emptyset, \{1\}, \{2\}, \{3\}\} \qquad\qquad \mathcal{C}_2 = \{\emptyset, \{1\}, \{2\}, \{2, 3\}\}.$$

5 Discussion

Classes of VC dimension 3 or 4 with properties similar to $\mathcal{C}$ can be constructed by taking the disjoint union of $\mathcal{C}$ with a class of the appropriate VC dimension.

One of the motivations for this work is determining the VC dimension of intersections of k halfspaces in $\mathbf{R}^d$. The case $d < 4$ is treated by Dobkin and Gunopulos [2, Lemmas 4 and 6], who attribute to Emo Welzl the proof that the VC dimension is exactly $2k + 1$ when $d = 2$ and at most $4k$ when $d = 3$. Their technique requires that polytopes have large independent sets of vertices, but when $d \geq 4$, there exist polytopes whose 1-skeletons are complete graphs on n vertices (neighborly polytopes). The best known bound in this case is still $O(dk \log k)$. The author attempted to obtain a matching lower bound by embedding $\mathcal{C}$ as a collection of lines in $\mathbf{R}^d$, but it appears that the algebraic structure of the latter imposes significant restrictions.

Besides the results described in Section 4, there is much yet unknown about when the VC dimension of Boolean combinations of concepts can grow only as the sum of the VC dimensions of the primitive classes. Johnson [5, Ch. 2] investigates some of these questions from a model-theoretic perspective.

6 Acknowledgments

Part of this work was done while the author was a graduate student at Brown University. The author wishes to thank the anonymous referee for his or her valuable suggestions.

References

[1] Anselm Blumer, Andrzej Ehrenfeucht, David Haussler, and Manfred K. Warmuth. Learnability and the Vapnik-Chervonenkis dimension. *Journal of the ACM*, 36(4):929–965, 1989.

[2] David P. Dobkin and Dimitrios Gunopulos. Concept learning with geometric hypotheses. In *COLT '95: Proceedings of the Eighth Annual Conference on Computational Learning Theory*, pages 329–336, New York, NY, USA, 1995. ACM.

[3] David Eisenstat and Dana Angluin. The VC dimension of k-fold union. *Information Processing Letters*, 101(5):181–184, 2007.

[4] David Haussler and Emo Welzl. Epsilon-nets and simplex range queries. *Discrete & Computational Geometry*, 2:127–151, 1987.

[5] Hunter R. Johnson. *Definable families of finite Vapnik Chervonenkis dimension.* PhD thesis, University of Maryland, 2008. `http://hdl.handle.net/1903/8174`.

[6] Vladimir Vapnik and Alexey Ya. Chervonenkis. On the uniform convergence of relative frequencies of events to their probabilities. *Theory of Probability and its Applications*, 16(2):264–280, 1971.

[7] Roberta S. Wenocur and Richard M. Dudley. Some special Vapnik-Chervonenkis classes. *Discrete Mathematics*, 33(3):313–318, 1981.