Probably Approximately Correct Learning - I

Prof. Dan A. Simovici

UMB

Collections of Subsets of Finite Sets

Let S be a finite set and let $\mathcal{C}$ be a VC-class with $VCD(\mathcal{C}) = d$. We know that $2^d \leqslant |\mathcal{C}|$.
Claim:

$$|\mathcal{C}| \leqslant (|S| + 1)^d.$$

Definition

Let $\mathcal{C}$ be a collection of subsets of a finite set S, let $s \in S$, and let $\mathcal{C}_{(s)}$ be the collection $\mathcal{C}_{S-\{s\}}$.
$\mathcal{C}$ has a pair (A, B) at s if $A, B \in \mathcal{C}$, $B \subseteq A$, and $A - B = \{s\}$.

If (A, B) is a pair of $\mathcal{C}$ at s, then $B = A - \{s\}$.
Define the subcollections $\mathcal{P}'(\mathcal{C}, s)$ and $\mathcal{P}''(\mathcal{C}, s)$ of $\mathcal{C}$ as

$$\begin{aligned}
\mathcal{P}'(\mathcal{C}, s) &= \{A \in \mathcal{C} \mid (A, B) \text{ is a pair at } s \text{ for some } B \in \mathcal{C}\} \\
&= \{A \in \mathcal{C} \mid s \in A, A - \{s\} \in \mathcal{C}\}, \\
\mathcal{P}''(\mathcal{C}, s) &= \{B \in \mathcal{C} \mid (A, B) \text{ is a pair at } s \text{ for some } A \in \mathcal{C}\} \\
&= \{B \in \mathcal{C} \mid s \notin B, B \cup \{s\} \in \mathcal{C}\}.
\end{aligned}$$

Lemma

Let $\mathcal{C}$ be a collection of subsets of a finite set S, let $s \in S$, and let $\mathcal{C}_{(s)}$ be the collection $\mathcal{C}_{S-\{s\}}$.
The following statements hold:

- *if (A_1, B_1), (A_1, B_2), (A_2, B_1), (A_2, B_2) are pairs at s of $\mathcal{C}$, then $A_1 = A_2$ and $B_1 = B_2$;*
- *we have $|\mathcal{P}'(\mathcal{C}, s)| = |\mathcal{P}''(\mathcal{C}, s)|$;*
- $|\mathcal{C}| - |\mathcal{C}_{(s)}| = |\mathcal{P}''(\mathcal{C}, s)|$;
- $\mathcal{C}_{(s)} = \{C \in \mathcal{C} \mid s \notin C\} \cup \{C - \{s\} \mid C \in \mathcal{C} \text{ and } s \in C\}$;
- $|\mathcal{C}_{(s)}| = |\mathcal{C}| - |\mathcal{P}''(\mathcal{C}, s)|$.

Lemma

Let $\mathcal{C}$ be a collection of sets of a non-empty finite set S and let s_0 be an element of S. If $\mathcal{P}''(\mathcal{C}, s_0)$ shatters a subset T of $S - \{s_0\}$, then $\mathcal{C}$ shatters $T \cup \{s_0\}$.

Proof: Since $\mathcal{P}''(\mathcal{C}, s_0)$ shatters T, for every subset U of T there is $B \in \mathcal{P}''(\mathcal{C}, s_0)$ such that $U = T \cap B$. Let W be a subset of $T \cup \{s_0\}$. If $s_0 \notin W$, then $W \subseteq T$ and by the previous asumption, there exists $B \in \mathcal{P}''(\mathcal{C}, s_0)$ such that $W = (T \cup \{s_0\}) \cap B$. If $s_0 \in W$, then there exists $B_1 \in \mathcal{P}''(\mathcal{C}, s_0)$ such that for $W_1 = W - \{s_0\}$ we have $W_1 = T \cap B_1$. By the definition of $\mathcal{P}''(\mathcal{C}, s_0)$, $B_1 \cup \{s_0\} \in \mathcal{C}$ and

$$(T \cup \{s_0\}) \cap (B_1 \cup \{s_0\}) = (T \cap B) \cup \{s_0\} = W_1 \cup \{s_0\} = W.$$

Thus, $\mathcal{C}$ shatters $T \cup \{s_0\}$.

Theorem

Let $\mathcal{C}$ be a collection of sets of a non-empty finite set S with $VCD(\mathcal{C}) = d$. We have

$$2^d \leqslant |\mathcal{C}| \leqslant (|S| + 1)^d.$$

Proof: For the second inequality the argument is by induction on $|S|$. The basis case, $|S| = 1$ is immediate.

Suppose that the inequality holds for collections of subsets with no more than n elements and let S be a set containing $n+1$ elements. Let s_0 be an arbitrary but fixed element of S.

By a previous Lemma we have $|\mathcal{C}| = |\mathcal{C}_{(s_0)}| + |\mathcal{P}''(\mathcal{C}, s_0)|$.

The collection $\mathcal{C}_{(s_0)}$ consists of subsets of $S - \{s_0\}$. Since $VCD(\mathcal{C}) = d$, it is clear that $VCD(\mathcal{C}_{(s_0)}) \leqslant d$ and, by inductive hypothesis $|\mathcal{C}_{(s_0)}| \leq (|S - \{s_0\}| + 1)^d$.

Proof (cont'd)

We claim that $VCD(\mathcal{P}''(\mathcal{C}, s_0)) \leqslant d - 1$. Suppose that

$$\mathcal{P}''(\mathcal{C}, s_0) = \{B \in \mathcal{C} \mid s_0 \notin B, B \cup \{s_0\} \in \mathcal{C}\}$$

shatters a set T, where $T \subseteq S$ and $|T| \geqslant d$. Then, by a previous Lemma, $\mathcal{C}$ shatters $T \cup \{s_0\}$; since $|T \cup \{s_0\}| \geq d + 1$, this would lead to a contradiction. Therefore, we have $VCD(\mathcal{P}''(\mathcal{C}, s_0)) \leqslant d - 1$ and, by the inductive hypothesis, $|\mathcal{P}''(\mathcal{C}, s_0))| \leqslant (|S - \{s_0\}| + 1)^{d-1}$. These inequalities imply

$$\begin{aligned}
|\mathcal{C}| &= |\mathcal{C}_{(s_0)}| + |\mathcal{P}''(\mathcal{C}, s_0)| \\
&\leqslant (|S - \{s_0\}| + 1)^d + (|S - \{s_0\}| + 1)^{d-1} \\
&= (|S - \{s_0\}| + 1)^{d-1}(|S - \{s_0\}| + 2) \\
&\leqslant (|S| + 1)^d.
\end{aligned}$$

Asymptotic VCD Dimension

Let $\mathcal{C}$ be a class of concepts on $\{0,1\}^*$ and let $C \in \mathcal{C}$. The projection $C^{(n)}$ of C is

$$C^{(n)} = C \cap \{0,1\}^n.$$

The projection of the class $\mathcal{C}$ is

$$\mathcal{C}^{(n)} = \{C \cap \{0,1\}^n | C \in \mathcal{C}\}.$$

The asymptotic VCD of $\mathcal{C}$ is the function *vcd* that maps $(\mathcal{C}, n)$ into $VCD(\mathcal{C}^{(n)})$.

Lemma

Let $d = vcd(\mathcal{C}, n)$. We have

$$2^d \leqslant |\mathcal{C}^{(n)}| \leqslant 2^{(n+1)d}.$$

Proof: The concept class $\mathcal{C}^{(n)}$ is a collection of subsets of $\{0,1\}^n$. Therefore,

$$2^d \leqslant |\mathcal{C}^{(n)}| \leqslant (2^n + 1)^d.$$

Taking into account that $2^n + 1 \leq 2^{(n+1)}$ it follows that

$$2^d \leqslant |\mathcal{C}^{(n)}| \leqslant 2^{(n+1)d}.$$

Learning Concepts in VC Classes

Algorithm 2.1: A Generic Learning Algorithm for VC-Classes

Data: n, ϵ, δ and the example space
Result: A hypothesis H

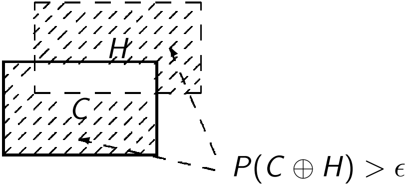
let $m = \frac{1}{\epsilon}\left((n+1)vcd(\mathcal{C})\ln 2 + \ln\frac{1}{\delta}\right)$;
generate m examples to form a set S;
pick a hypothesis H consistent with the examples in S;
return H;

Let C be the target concept.

- We require that with probability greater than $1 - \delta$ the hypothesis H generated by the algorithm be such that

$$P(C \oplus H) \leqslant \epsilon.$$

- Equivalently: With probability less than δ the hypothesis H is such that $P(C \oplus H) > \epsilon$.

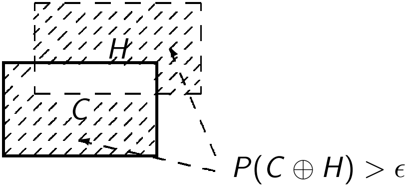


Examples space is $\{0,1\}^n$.

- Let $H \subseteq \{0,1\}^n$ be such that $P(H \oplus C) > \epsilon$;
- the probability that the next example will be consistent with H is at most $1 - \epsilon$; hence, the probability that m examples are consistent with H is at most $(1-\epsilon)^m$;
- the probability that m examples are consistent with some choice of H is at most $2^n(1-\epsilon)^m$.

Thus, to ensure that with probability less than δ the hypothesis H is such that $P(C \oplus H) > \epsilon$ it suffices to have

$$2^n(1-\epsilon)^m \leqslant \delta.$$



To satisfy

$$2^n(1-\epsilon)^m \leqslant \delta,$$

since $|\mathcal{C}| \leq 2^{(n+1)VCD(\mathcal{C})}$ it suffices to demand that

$$2^{(n+1)VCD(\mathcal{C})}(1-\epsilon)^m \leq \delta.$$

Applying logarithms:

$$(n+1)VCD(\mathcal{C})\ln 2 + m\ln(1-\epsilon) \leqslant \ln\delta$$

Since $\ln(1+\alpha) \leq \alpha$, it suffices if

$$(n+1)VCD(\mathcal{C})\ln 2 - \frac{m}{\epsilon} \leqslant \ln\delta,$$

so

$$m \geqslant \frac{1}{\epsilon}\left((n+1)VCD(\mathcal{C})\ln 2 + \ln\frac{1}{\delta}\right)$$

- if m examples are drawn, with probability at least $1-\delta$, any hypothesis H consistent with the examples will be such that $P(H \oplus C) \leqslant \epsilon$.
- a hypothesis H is consistent with all examples drawn if and only if C is consistent with all of them;
- with probability at least $1-\delta$ any H that is consistent with all m examples will be such that $P(H \oplus C) \leqslant \epsilon$.