MACHINE LEARNING - CS671 - Part 2a
The Vapnik-Chervonenkis Dimension

Prof. Dan A. Simovici

UMB

- The *Vapnik-Chervonenkis* dimension of a collection of sets was introduced in [3] and independently in [2].
- Its main interest for ML is related to one of the basic models of machine learning, the probably approximately correct PAC learning paradigm as was shown in [1].

The Trace of a Collection of Sets on a Set

Let $\mathcal{C} \subseteq \mathcal{P}(U)$.
The *trace* of $\mathcal{C}$ on K is the collection of sets

$$\mathcal{C}_K = \{K \cap C \mid C \in \mathcal{C}\}.$$

If $\mathcal{C}_K$ equals $\mathcal{P}(K)$, then we say that K is *shattered by $\mathcal{C}$*. This means that there are concepts in $\mathcal{C}$ that split K is all $2^{|K|}$ possible ways. concepts.

Main Definition

The *Vapnik-Chervonenkis dimension* of the collection $\mathcal{C}$ (called the VC-dimension for brevity) is the largest size of a set K that is shattered by $\mathcal{C}$ and is denoted by $VCD(\mathcal{C})$.

Example

The VC-dimension of the collection of intervals in $\mathbb{R}$ is 2.

Remarks

- If $VCD(\mathcal{C}) = d$, then there exists a set K of size d such that for each subset L of K there exists a subset $C \in \mathcal{C}$ such that $L = K \cap C$.
- Since there exist 2^d subsets of K, there are at least 2^d sets in $\mathcal{C}$, so $2^d \leqslant |\mathcal{C}|$. Thus,

$$VCD(\mathcal{C}) \leqslant \log_2 |\mathcal{C}|.$$

- If $\mathcal{C}$ is finite, then $VCD(\mathcal{C})$ is finite. The converse is false: there exist infinite collections $\mathcal{C}$ that have a finite VC-dimension.

The tabular form of $\mathcal{C}_K$

Let $U = \{u_1, \ldots, u_n\}$, and let $\theta = (T_{\mathcal{C}}, u_1 u_2 \cdots u_n, \mathbf{r})$ be a table, where $\mathbf{r} = (t_1, \ldots, t_p)$. The domain of each of the attributes u_i is the set $\{0, 1\}$. Each tuple t_k corresponds to a set C_k of $\mathcal{C}$ and is defined by

$$t_k[u_i] = \begin{cases} 1 & \text{if } u_i \in C_k, \\ 0 & \text{otherwise,} \end{cases}$$

for $1 \leqslant i \leqslant n$. Then, $\mathcal{C}$ shatters K if the content of the projection $\mathbf{r}[K]$ consists of $2^{|K|}$ distinct rows.

Example

Let $U = \{u_1, u_2, u_3, u_4\}$ and let $\mathcal{C}$ be the collection of subsets of U given by $\mathcal{C} = \{\{u_2, u_3\}, \{u_1, u_3, u_4\}, \{u_2, u_4\}, \{u_1, u_2\}, \{u_2, u_3, u_4\}\}$.

$T_{\mathcal{C}}$

u_1	u_2	u_3	u_4
0	1	1	0
1	0	1	1
0	1	0	1
1	1	0	0
0	1	1	1

$K = \{u_1, u_3\}$ is shattered by $\mathcal{C}$ because

$$\mathbf{r}[K] = ((0,1),(1,1),(0,0),(1,0),(0,1))$$

contains the all four necessary tuples $(0,1)$, $(1,1)$, $(0,0)$, and $(1,0)$. On the other hand, it is clear that no subset K of U that contains at least three elements can be shattered by $\mathcal{C}$ because this would require $\mathbf{r}[K]$ to contain at least eight tuples. Thus, $VCD(\mathcal{C}) = 2$.

Remarks

- Every collection of sets shatters the empty set.
- If $\mathcal{C}$ shatters a set of size n, then it shatters a set of size p, where $p \leqslant n$.

VC Classes

For $\mathcal{C}$ and for $m \in \mathbb{N}$, let $\Pi_{\mathcal{C}}[m]$ be the largest number of distinct subsets of a set having m elements that can be obtained as intersections of the set with members of $\mathcal{C}$, that is,

$$\Pi_{\mathcal{C}}[m] = \max\{|\mathcal{C}_K| \mid |K| = m\}.$$

We have $\Pi_{\mathcal{C}}[m] \leqslant 2^m$; however, if $\mathcal{C}$ shatters a set of size m, then $\Pi_{\mathcal{C}}[m] = 2^m$.

Definition

A *Vapnik-Chervonenkis class* (or a *VC class*) is a collection $\mathcal{C}$ of sets such that $VCD(\mathcal{C})$ is finite.

Example

Example

Let $\mathcal{S}$ be the collection of sets $\{(-\infty, t) \mid t \in \mathbb{R}\}$.

- Any singleton is shattered by $\mathcal{S}$. Indeed, if $S = \{x\}$ is a singleton, then $\mathcal{P}(\{x\}) = \{\emptyset, \{x\}\}$. Thus, if $t \geqslant x$, we have $(-\infty, t) \cap S = \{x\}$; also, if $t < x$, we have $(-\infty, t) \cap S = \emptyset$, so $\mathcal{S}_S = \mathcal{P}(S)$.
- There is no set S with $|S| = 2$ that can be shattered by $\mathcal{S}$. Indeed, suppose that $S = \{x, y\}$, where $x < y$. Then, any member of $\mathcal{S}$ that contains y includes the entire set S, so $\mathcal{S}_S = \{\emptyset, \{x\}, \{x, y\}\} \neq \mathcal{P}(S)$.

This shows that $\mathcal{S}$ is a VC class and $VCD(\mathcal{S}) = 1$.

Example

Consider the collection $\mathcal{I} = \{[a,b] \mid a, b \in \mathbb{R}, a \leqslant b\}$ of closed intervals. We claim that $VCD(\mathcal{I}) = 2$.

- There exists a set $S = \{x, y\}$ such that $\mathcal{I}_S = \mathcal{P}(S)$: consider the intersections

$$\begin{aligned}
&[u, v] \cap S = \emptyset, \text{ where } v < x,\\
&[x - \epsilon, \tfrac{x+y}{2}] \cap S = \{x\},\\
&[\tfrac{x+y}{2}, y] \cap S = \{y\},\\
&[x - \epsilon, y + \epsilon] \cap S = \{x, y\},
\end{aligned}$$

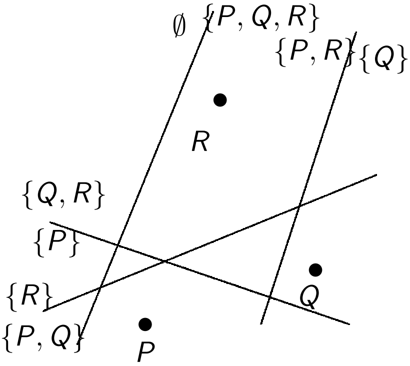
  which show that $\mathcal{I}_S = \mathcal{P}(S)$.
- No three-element set can be shattered by $\mathcal{I}$: Let $T = \{x, y, z\}$ be a set that contains three elements. Note that any interval that contains x and z also contains y, so it is impossible to obtain the set $\{x, z\}$ as an intersection between an interval in $\mathcal{I}$ and the set T.

Example: Three-point sets shattered by half-planes

Let $\mathcal{H}$ be the collection of closed half-planes in $\mathbb{R}^2$, that is, the collection of sets of the form

$$\{x = (x_1, x_2) \in \mathbb{R}^2 \mid ax_1 + bx_2 - c \geqslant 0, a \neq 0 \text{ or } b \neq 0\}.$$

We claim that $VCD(\mathcal{H}) = 3$.
Let $P, Q, R \in \mathbb{R}^2$ be non-colinear. The family of lines shatters the set $\{P, Q, R\}$, so $VCD(\mathcal{H})$ is at least 3.

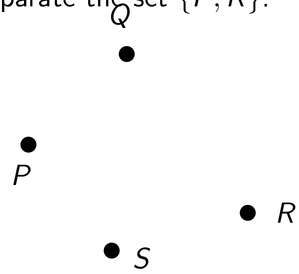


No set that contains at least four points can be shattered by $\mathcal{H}$.

Let $\{P, Q, R, S\}$ be a set such that no three points of this set are collinear. If S is located inside the triangle P, Q, R, then every half-plane that contains P, Q, R will contain S, so it is impossible to separate the subset $\{P, Q, R\}$.

Thus, we may assume that no point is inside the triangle formed by the remaining three points.

Any half-plane that contains two diagonally opposite points, for example, P and R, will contain either Q or S, which shows that it is impossible to separate the set $\{P, R\}$.

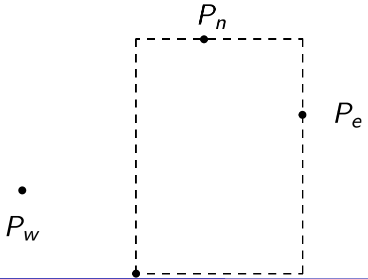


No set that contains four points may be shattered by $\mathcal{H}$, so $VCD(\mathcal{H}) = 3$.

Example

Let $\mathbb{R}^2$ be equipped with a system of coordinates and let $\mathcal{R}$ be the set of rectangles whose sides are parallel with the axes x and y. Each such rectangle has the form $[x_0, x_1] \times [y_0, y_1]$.

There is a set S with $|S| = 4$ that is shattered by $\mathcal{R}$. Indeed, let S be a set of four points in $\mathbb{R}^2$ that contains a unique "northernmost point" P_n, a unique "southernmost point" P_s, a unique "easternmost point" P_e, and a unique "westernmost point" P_w. If $L \subseteq S$ and $L \neq \emptyset$, let R_L be the smallest rectangle that contains L. For example, we show the rectangle R_L for the set $\{P_n, P_s, P_e\}$.



Example (cont'd)

This collection cannot shatter a set of points that contains at least five points.

Indeed, let S be a set of points such that $|S| \geqslant 5$ and, as before, let P_n be the northernmost point, etc. If the set contains more than one "northernmost" point, then we select exactly one to be P_n. Then, the rectangle that contains the set $K = \{P_n, P_e, P_s, P_w\}$ contains the entire set S, which shows the impossibility of separating the set K.

Major Result

If a collection of sets $\mathcal{C}$ is not a VC class (that is, if the Vapnik-Chervonenkis dimension of $\mathcal{C}$ is infinite), then

$$\Pi_{\mathcal{C}}[m] = 2^m$$

for all $m \in \mathbb{N}$.
However, we shall prove that if $VCD(\mathcal{C}) = d$, then $\Pi_{\mathcal{C}}[m]$ is bounded asymptotically by a polynomial of degree d.

The Function ϕ

For $n, k \in \mathbb{N}$ and $0 \leqslant k \leqslant n$ define the number $\binom{n}{\leqslant k}$ as

$$\binom{n}{\leqslant k} = \sum_{i=0}^{k} \binom{n}{i}$$

Clearly, $\binom{n}{\leqslant 0} = 1$ and $\binom{n}{\leqslant n} = 2^n$.

Theorem

Let $\phi : \mathbb{N}^2 \longrightarrow \mathbb{N}$ be the function defined by

$$\phi(d, m) = \begin{cases} 1 & \text{if } m = 0 \text{ or } d = 0 \\ \phi(d, m-1) + \phi(d-1, m-1) & \text{otherwise.} \end{cases}$$

We have

$$\phi(d, m) = \binom{m}{\leqslant d}$$

for $d, m \in \mathbb{N}$.

Proof

The argument is by strong induction on $s = i + m$.
The base case, $s = 0$, implies $m = d = 0$.
Suppose that the equality holds for $\phi(d', m')$, where $d' + m' < d + m$. We have:

$$
\begin{aligned}
\phi(d, m) &= \phi(d, m-1) + \phi(d-1, m-1) \\
&\quad \text{(by definition)} \\
&= \sum_{i=0}^{d} \binom{m-1}{i} + \sum_{i=0}^{d-1} \binom{m-1}{i} \\
&\quad \text{(by inductive hypothesis)} \\
&= \sum_{i=0}^{d} \binom{m-1}{i} + \sum_{i=0}^{d} \binom{m-1}{i-1} \\
&\quad \text{(since } \binom{m-1}{-1} = 0\text{)} \\
&= \sum_{i=0}^{d} \left(\binom{m-1}{i} + \binom{m-1}{i-1} \right) = \sum_{i=0}^{d} \binom{m}{i} = \binom{m}{\leqslant d}.
\end{aligned}
$$

Sauer-Shelah Theorem

Theorem

If $\mathcal{C}$ is a collection of subsets of S that is a VC-class such that $VCD(\mathcal{C}) = d$, then $\Pi_{\mathcal{C}}[m] \leqslant \phi(d, m)$ for $m \in \mathbb{N}$, where ϕ is the function defined above.

Proof

The argument is by strong induction on $s = d + m$, the sum of the VCD of $\mathcal{C}$ and the size of the set.

- **For the base case**, $s = 0$ we have $d = m = 0$ and this means that the collection $\mathcal{C}$ shatters only the empty set. Thus, $\Pi_{\mathcal{C}}[0] = |\mathcal{C}_{\emptyset}| = 1$, and this implies $\Pi_{\mathcal{C}}[0] = 1 = \phi(0, 0)$.
- **The inductive case:** Suppose that the statement holds for pairs (d', m') such that $d' + m' < s$ and let $\mathcal{C}$ be a collection of subsets of S such that $VCD(\mathcal{C}) = d$.

Proof (cont'd)

Let K be a set of cardinality m and let k_0 be a fixed (but, otherwise, arbitrary) element of K.
Consider the trace $\mathcal{C}_{K-\{k_0\}}$. Since $|K - \{k_0\}| = m - 1$, we have, by the inductive hypothesis, $|\mathcal{C}_{K-\{k_0\}}| \leqslant \phi(d, m-1)$.
Let $\mathcal{C}'$ be the collection of sets given by

$$\mathcal{C}' = \{G \in \mathcal{C}_K \mid k_0 \notin G, G \cup \{k_0\} \in \mathcal{C}_K\}.$$

- $\mathcal{C}' = \mathcal{C}'_{K-\{k_0\}}$ because $\mathcal{C}'$ consists only of subsets of $K - \{k_0\}$.
- The VCD of $\mathcal{C}'$ is less than d. Indeed, let K' be a subset of $K - \{k_0\}$ that is shattered by $\mathcal{C}'$. Then, $K' \cup \{k_0\}$ is shattered by $\mathcal{C}$, hence $|K'| < d$.

By the inductive hypothesis, $|\mathcal{C}'| = |\mathcal{C}_{K-\{k_0\}}| \leq \phi(d-1, m-1)$.

The collection of sets $\mathcal{C}_K$ is a collection of subsets of K that can be regarded as the union of two disjoint collections:

- those subsets in $\mathcal{C}_K$ that do not contain the element k_0, that is $\mathcal{C}_{K-\{k_0\}}$;
- those subsets of K that contain k_0.

If L is a second type of subset, then $L - \{k_0\} \in \mathcal{C}'$. Thus,

$$|\mathcal{C}_K| = |\mathcal{C}_{K-\{k_0\}}| + |\mathcal{C}'_{K-\{k_0\}}|,$$

so $|\mathcal{C}_K| \leqslant \phi(d, m-1) + \phi(d-1, m-1)$, which is the desired conclusion.

Lemma 1

Lemma

For $d \in \mathbb{N}$ and $d \geqslant 2$ we have

$$2^{d-1} \leqslant \frac{d^d}{d!}.$$

The argument is by induction on d. The basis step, $d = 2$ is immediate. Suppose the inequality holds for d. We have

$$\begin{aligned}
\frac{(d+1)^{d+1}}{(d+1)!} &= \frac{(d+1)^d}{d!} = \frac{d^d}{d!} \cdot \frac{(d+1)^d}{d^d} \\
&= \frac{d^d}{d!} \cdot \left(1 + \frac{1}{d}\right)^d \geqslant 2^d \cdot \left(1 + \frac{1}{d}\right)^d \geqslant 2^d
\end{aligned}$$

(by inductive hypothesis)

because

$$\left(1 + \frac{1}{d}\right)^d \geqslant 1 + d\frac{1}{d} = 2.$$

Lemma 2

Lemma

We have $\phi(d, m) \leqslant 2\frac{m^d}{d!}$ *for every* $m \geqslant d$ *and* $d \geqslant 1$*.*

The argument is by induction on d and n. If $d = 1$, then $\phi(1, m) = m + 1 \leqslant 2m$ for $m \geqslant 1$, so the inequality holds for every $m \geqslant 1$, when $d = 1$.

If $m = d \geqslant 2$, then $\phi(d, m) = \phi(d, d) = 2^d$ and the desired inequality follows immediately from the previous Lemma.

Proof (cont'd)

Suppose that the inequality holds for $m > d \geqslant 1$. We have

$$\begin{aligned}
\phi(d, m+1) &= \phi(d, m) + \phi(d-1, m) \\
&\quad \text{(by the definition of } \phi\text{)} \\
&\leqslant 2\frac{m^d}{d!} + 2\frac{m^{d-1}}{(d-1)!} \\
&\quad \text{(by inductive hypothesis)} \\
&= 2\frac{m^{d-1}}{(d-1)!}\left(1 + \frac{m}{d}\right).
\end{aligned}$$

It is easy to see that the inequality

$$2\frac{m^{d-1}}{(d-1)!}\left(1 + \frac{m}{d}\right) \leqslant 2\frac{(m+1)^d}{d!}$$

is equivalent to

$$\frac{d}{m} + 1 \leqslant \left(1 + \frac{1}{m}\right)^d$$

and, so it is valid. This yields the inequality.

Theorem

The function ϕ satisfies the inequality:

$$\phi(d, m) < \left(\frac{em}{d}\right)^d$$

for every $m \geqslant d$ and $d \geqslant 1$.

Proof

By Lemma 2, $\phi(d, m) \leqslant 2\frac{m^d}{d!}$. Therefore, we need to show only that

$$2\left(\frac{d}{e}\right)^d < d!.$$

The argument is by induction on $d \geqslant 1$.
The basis case, $d = 1$ is immediate.

Suppose that $2\left(\frac{d}{e}\right)^d < d!$. We have

$$\begin{aligned} 2\left(\frac{d+1}{e}\right)^{d+1} &= 2\left(\frac{d}{e}\right)^d\left(\frac{d+1}{d}\right)^d\frac{d+1}{e} \\ &= \left(1+\frac{1}{d}\right)^d\frac{1}{e}\cdot 2\left(\frac{d}{e}\right)^d(d+1) < 2\left(\frac{d}{e}\right)^d(d+1), \end{aligned}$$

because

$$\left(1+\frac{1}{d}\right)^d < e.$$

The last inequality holds because the sequence $\left(\left(1+\frac{1}{d}\right)^d\right)_{d\in\mathbb{N}}$ is an increasing sequence whose limit is e. Since $2\left(\frac{d+1}{e}\right)^{d+1} < 2\left(\frac{d}{e}\right)^d(d+1)$, by inductive hypothesis we obtain:

$$2\left(\frac{d+1}{e}\right)^{d+1} < (d+1)!.$$

Corollary

If m is sufficiently large we have $\phi(d, m) = O(m^d)$.

A. Blumer, A. Ehrenfeucht, D. Haussler, and M. Warmuth.
Learnability and the Vapnik-Chervonenkis dimension.
Journal of the ACM, 36:929–965, 1989.

N. Sauer.
On the density of families of sets.
Journal of Combinatorial Theory (A), 13:145–147, 1972.

V. N. Vapnik and A. Y. Chervonenkis.
On the uniform convergence of relative frequencies of events to their probabilities.
Theory of Probability and Applications, 16:264–280, 1971.