24. VC-Dim

Wednesday, November 3, 2021 12:28 PM

Statistical classification

Define: A set system $(X, \mathcal{H})$ consists of a set X and a class $\mathcal{H}$ of subsets of X.

Ex. X = set of all possible emails

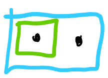
$\mathcal{H} = \{h_0, h_1, \dots\}$ where $h_0 \subseteq X$ is the set of all spam emails
$h_1 \subseteq X$ is the set of marketing emails
$\vdots$

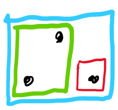
Define: A set system $(X, \mathcal{H})$ **shatters** a set $A \subseteq X$ if each subset of A can be expressed as $A \cap h$ for some $h \in \mathcal{H}$.

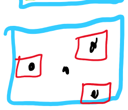
Define: The Vapnik-Chervonenkis (VC) dimension of $\mathcal{H}$ is the size of the largest set shattered by $\mathcal{H}$.

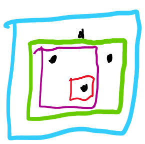
Ex. Let $X = \mathbb{R}^2$ and $\mathcal{H} = \{\text{axis-parallel rectangles}\}$

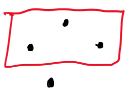

Trivial to shatter


Almost trivial to shatter


Again, can shatter 3 pts


Cannot shatter this set, because cannot get just 3 outer pts


But there does exist a subset we can shatter, so $VC(\mathcal{H}) \geq 4$


Can we shatter all 5-pt sets?
Consider a rectangle containing all pts, and shrink to get all sets of size 4.

Can remove only pts that are uniquely extreme to get 4-pt subsets.

$\Rightarrow VC(\{\text{axis-parallel rectangles}\}) = 4$

Ex. $X = \mathbb{R}$, $\mathcal{H} = \{[a,b] \mid a, b \in \mathbb{R}\}$. $VC(\mathcal{H}) = 2$

Ex. $X = \mathbb{R}$, $\mathcal{H} = \{[a,b], [c,d] \mid a,b,c,d \in \mathbb{R}\}$ $VC(\mathcal{H}) = 4$

Ex. $X = \mathbb{R}$, $\mathcal{H} = \{A \subseteq \mathbb{R} \mid |A| < \infty\}$ $VC(\mathcal{H}) = \infty$

Ex. $X = \mathbb{R}$, $\mathcal{H} = \{ A \subseteq \mathbb{R} \mid |A| < \infty \}$ $\qquad VC(\mathcal{H}) = \infty$

Ex. $X = \mathbb{R}^2$, $\mathcal{H} = \{ \text{convex polygons} \}$ $\qquad VC(\mathcal{H}) = \infty$ $\qquad$ (via circle)

Ex. $X = \mathbb{R}^d$, $\mathcal{H} = \{ \vec{x} \mid \vec{w}^T \vec{x} \geq t \}$ $\qquad VC(\mathcal{H}) = d+1$

affine half-spaces

Lemma: $VC(\mathcal{H}) \geq d+1$ where $\mathcal{H} = \{ \vec{x} \mid \vec{w}^T \vec{x} \geq t \}$.

proof. Let $S = \{0, e_1, e_2, \dots, e_d\}$ where $e_i = (0, \dots, 0, \underset{\uparrow \text{ pos } i}{1}, 0, \dots, 0)^T$

Let $A \subseteq S$ be a subset. WLOG assume $0 \in A$.

Let $\vec{w} = (1, \dots, 1)^T - \sum_{a \in A} a$

Then $A \subseteq \{ \vec{x} \mid \vec{w}^T \vec{x} \leq 0 \}$ and $S - A \subseteq \{ \vec{x} \mid \vec{w}^T \vec{x} > 0 \}$

So $\mathcal{H}$ shatters a $d+1$ pt set $\Rightarrow VC(\mathcal{H}) \geq d+1$. $\boxed{}$

Thm 5.9 (Radon) Any set $S \subseteq \mathbb{R}^d$ with $|S| \geq d+2$ can be partitioned into two disjoint subsets A and B such that $\text{convex}(A) \cap \text{convex}(B) \neq \emptyset$.

proof. WLOG, assume $|S| = d+2$ with $S = \{ \vec{a}_1, \dots, \vec{a}_{d+2} \}$.

Let $A = [\vec{a}_1 \ \ \vec{a}_2 \ \cdots \ \vec{a}_{d+2}] \in \mathbb{R}^{d \times (d+2)}$

Let $B = \begin{bmatrix} A \\ \vec{1}^T \end{bmatrix}$. $\text{rank}(B) \leq d+1$, so cols are lin dep. Let $B = [\vec{b}_1 \ \cdots \ \vec{b}_{d+2}] \in \mathbb{R}^{(d+1) \times (d+2)}$

Let $\vec{x} = (x_1, x_2, \dots, x_{d+2})^T$ s.t. $B\vec{x} = 0$.

WLOG, say $x_1, \dots, x_s \geq 0$ and $x_{s+1}, \dots, x_{d+2} < 0$.

Normalize $\vec{x}$ s.t. $\sum_{i=1}^{s} |x_i| = 1$.

Then $\sum_{i=1}^{s} |x_i| \vec{b}_i = \sum_{i=s+1}^{d+2} |x_i| \vec{b}_i$

$\Rightarrow \sum_{i=1}^{s} |x_i| \vec{a}_i = \sum_{i=s+1}^{d+2} |x_i| \vec{a}_i$ and $\sum_{i=1}^{s} |x_i| = \sum_{i=s+1}^{d+2} |x_i| = 1$.

Thus, both sides are convex combinations of disjoint cols of A.

$\Rightarrow$ The convex hulls of the two sets of corresponding pts intersect. $\boxed{}$

But, then it is impossible to have a linear separator of these two sets, so half-planes cannot shatter a $(d+2)$-pt set.

$$\Rightarrow VC(\mathcal{H}) < d+2$$

$$\Rightarrow VC(\mathcal{H}) = d+1.$$

Ex. $X = \mathbb{R}^d$, $\mathcal{H} = \{\vec{x} \mid |\vec{x} - \vec{x}_0| \le r\}$ spheres.

$VC(\mathcal{H}) < d+2$, because if we can put spheres around two disjoint sets then those sets are also divided by a hyperplane, and $VC(\{\text{half-spaces}\}) = d+1$.

$VC(\mathcal{H}) \ge d+1$ by the same construction as half-spaces.
let $S = \{\vec{0}, \vec{e}_1, \ldots, \vec{e}_d\}$.

Given a subset $A \subseteq S$, choose the ball center $\vec{a}_0 = \sum_{\vec{a} \in A} \vec{a}$

Then $|\vec{a}_0 - \vec{a}| = \sqrt{|A| - 1}$ $\forall \vec{a} \in A$ and $\vec{a} \ne 0$.

$|\vec{a}_0 - \vec{a}| = \sqrt{|A| + 1}$ $\forall \vec{a} \notin A$ and $\vec{a} \ne 0$.

$|\vec{a}_0| = \sqrt{|A|}$.

So we can choose a radius s.t. this ball contains exactly A.

Define: For any set system $(X, \mathcal{H})$, the shatter function

$$\pi_{\mathcal{H}}(n) = \max_{|A| = n} \left| \{A \cap h\}_{h \in \mathcal{H}} \right|,$$

the maximum number of subsets of any set A of size n that can be expressed as $A \cap h$ for $h \in \mathcal{H}$.

Note: $\pi_{\mathcal{H}}(n) = 2^n$ for $n \le VC(\mathcal{H})$ (because by def., can completely shatter at least one set of size $\le$ VC-dim)

Notation: $\binom{n}{\le d} = \binom{n}{0} + \binom{n}{1} + \cdots + \binom{n}{d} \le n^d + 1$ ($\le n^d$ if $d > 1$)
(because consider choosing n items d times with duplication)

Lemma 5.10 (Sauer) For any set system $(X, \mathcal{H})$ of VC-dim $\le d$,

$$\pi_{\mathcal{H}}(n) \le \binom{n}{\le d} \quad \forall n.$$

proof. Note that for $n \geq 1$ and $d \geq 1$,

$$\binom{n}{\leq d} = \underbrace{\binom{n-1}{\leq d-1}}_{\substack{\text{if we choose 1st element,} \\ \text{then we can choose at most} \\ d-1 \text{ of the remaining } n-1}} + \underbrace{\binom{n-1}{\leq d}}_{\substack{\text{if we don't choose the 1st element,} \\ \text{then we can choose at most } d \text{ of} \\ \text{the remaining } n-1.}}$$

We will prove by induction on n and d.

Base case: $d = 0$ (i.e. VC-dim $= 0$, so can only shatter $\emptyset$)

Suppose $h_1 \neq h_2 \in \mathcal{H}$. Then $\exists\, a \in X$ s.t. $a \in h_1$, $a \notin h_2$.

But then $\mathcal{H}$ shatters $\{a\} \Rightarrow d \geq 1$.

$\Rightarrow \mathcal{H} = \{h\}$.

Any set $A \subseteq X$ has $|2^A| = 2^{|A|}$ subsets, so only set of size 0 can be shattered if $|\mathcal{H}| = 1$. i.e. only $\emptyset$ can be shattered

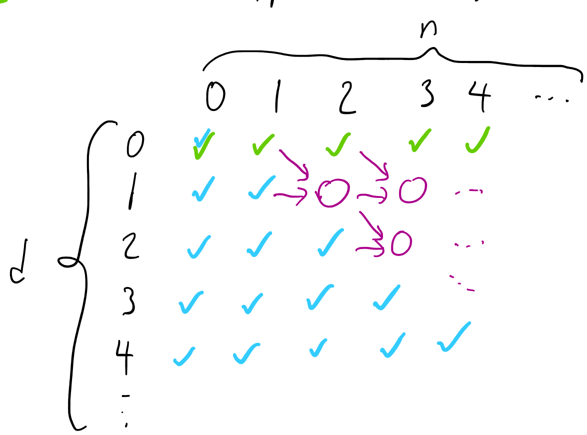
$$\Rightarrow \pi_{\mathcal{H}}(n) = 1 = \binom{n}{\leq 0}.$$

Base case: $n \leq d$

By def, $\exists$ a set A of size n that can be shattered, so $\pi_{\mathcal{H}}(n) = 2^n$.

$$\binom{n}{\leq d} = \underbrace{\binom{n}{0} + \binom{n}{1} + \dots + \binom{n}{n}}_{2^n} + \underbrace{\binom{n}{n+1} + \dots + \binom{n}{d}}_{0} = 2^n$$

General n, d: Induction hypothesis: Assume true for $(n-1, d-1)$ and $(n-1, d)$.



Select $A \subseteq X$ with $|A| = n$ s.t. $\pi_{\mathcal{H}}(n)$ subsets of A can be expressed as $A \cap h$, $h \in \mathcal{H}$ $\Big\}$ a "maximally shatterable" A

WLOG, assume $X = A$ and replace $h \in \mathcal{H}$ by $h \cap A$, removing duplicate sets.

Then $|\mathcal{H}| = \pi_{\mathcal{H}}(n)$ and each $h \in \mathcal{H}$ is a subset of A (i.e. $\mathcal{S} = (A, \{h \cap A\}_{h \in \mathcal{H}})$)

Need only control $|\mathcal{H}| \leq \binom{n}{\leq d}$ — our modified set system

Create a new set system $\mathcal{S}_1$ by removing u from A & each set in $\mathcal{H}$.

i.e. $\mathcal{S}_1 = (A - \{u\}, \mathcal{H}_1)$, where $\mathcal{H}_1 = \{h - \{u\} \mid h \in \mathcal{H}\}$

For $h \subseteq A - \{u\}$, if exactly one of h and $h \cup \{u\}$ is in $\mathcal{H}$, then h contributes one set to both $\mathcal{H}$ & $\mathcal{H}_1$.

If both h & $h \cup \{u\}$ are in $\mathcal{H}$, then h counts for two sets in $\mathcal{H}$ but only one set in $\mathcal{H}_1$.

Thus, $|\mathcal{H}| - |\mathcal{H}_1| = \left| \{(h_1, h_2) \mid h_1 - h_2 = \{u\}\}_{h_1, h_2 \in \mathcal{H}} \right|$ i.e. pairs of sets that differ only by u.

Let $\mathcal{S}_2 = (A - \{u\}, \mathcal{H}_2)$, where $\mathcal{H}_2 = \{h \mid u \notin h,\ h \in \mathcal{H},\ h \cup \{u\} \in \mathcal{H}\}$

Then $|\mathcal{H}| = |\mathcal{H}_1| + |\mathcal{H}_2|$

$\Rightarrow \pi_{\mathcal{H}}(n) = \pi_{\mathcal{H}_1}(n-1) + \pi_{\mathcal{H}_2}(n-1)$

each set in $\mathcal{H}_1$ by construction corresponds to a unique $(A-u) \cap h$, $h \in \mathcal{H}_1$.

Note: $VC(\mathcal{H}_1) \leq d$ because otherwise, $\mathcal{H}_1$ would shatter a set of cardinality $d+1$, and $\mathcal{H}$ would also shatter that set.

Note: $VC(\mathcal{H}_2) \leq d-1$ because if $\mathcal{H}_2$ shattered $B \subseteq A - \{u\}$ with $|B| \geq d$, then $B \cup \{u\}$ would be shattered by $\mathcal{H}$.

By induction hypothesis, $\pi_{\mathcal{H}_1}(n-1) \leq \binom{n-1}{\leq d}$

$$\pi_{\mathcal{H}_2}(n-1) \leq \binom{n-1}{\leq d-1}$$

Thus, $\pi_{\mathcal{H}}(n) \leq \binom{n-1}{\leq d} + \binom{n-1}{\leq d-1} = \binom{n}{\leq d}$. $\boxed{}$