VC Dimension

$\mathcal{X}$: sample space

Suppose $\mathcal{H}$ is a set of hypotheses $\mathcal{X} \to \{0,1\}$

Let S be a finite subset of $\mathcal{X}$ (the sample)

define $\Pi_{\mathcal{H}}(S) = \{ h|_S \mid h \in \mathcal{H} \}$

Restrict h to S & see which different ones are present.

$\Pi_{\mathcal{H}}(S)$: the set of <u>dichotomies</u> that are realized by $\mathcal{H}$.

$$\Pi_{\mathcal{H}}(m) = \max_{S : |S| = m} |\Pi_{\mathcal{H}}(S)|$$

Note $\Pi_{\mathcal{H}}(m) \le 2^m$

<u>EXAMPLES</u> (1) $\mathcal{X} = \mathbb{R}$ $\mathcal{H} = \{$ threshold functions $\}$.

$\Pi_{\mathcal{H}}(m) = m+1$

(2) $\mathcal{X} = \mathbb{R}$ $\mathcal{H} = \{ (l, u) \mid l < u \in \mathbb{R} \}$

Then $m + (m-1) + \cdots + 1 = O(m^2) = \Pi_{\mathcal{H}}(m)$.

Note $m^2 \ll 2^m$

<u>VC Dimension</u> We say a set S is <u>shattered</u> by $\mathcal{H}$ if every function from $S \to \{0,1\}$ can be realized as a restriction of $\mathcal{H}$ to S.

<u>def</u> The VC dimension of $\mathcal{H}$ is the cardinality of the largest set that can be shattered by $\mathcal{H}$.

<u>Rem</u> Note the implicit existential quantifier. It is possible that a set of size k can be shattered whereas a set of size $k-1$ cannot be shattered.

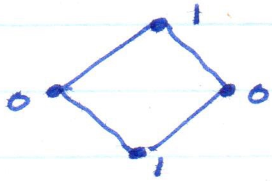
To show that VC dim($\mathcal{H}$) $= d$ we must find __a__ set of size d that can be shattered. It does not matter if there is another set that cannot be shattered even one that is smaller than d. We must also __prove__ that no set of size $(d+1)$ can be shattered.

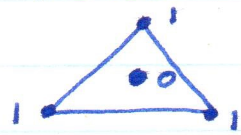
__EXAMPLES__ (1) Intervals in $\mathbb{R}$ can shatter any set of size 2 but no set of size 3 as we cannot have 

VC dim $= 2$

(2) $\mathcal{H}$: lines in $\mathbb{R}^2$: 3 non collinear points can be shattered but ~~a~~ 4 points can never be shattered

VC dim $= 3$





(3) Linear threshold functions in $\mathbb{R}^n$

$$f : \mathbb{R}^n \longrightarrow \{+1, -1\}$$

$\vec{x} \in \mathbb{R}^n$
$\vec{a} \in \mathbb{R}^n$, $b \in \mathbb{R}$

$$f_{\vec{a},b}(\vec{x}) = \begin{cases} +1 & \vec{a}\cdot\vec{x} \geq b \\ - & \vec{a}\cdot\vec{x} < b \end{cases}$$

Take $\vec{x}_0 = (0,0,\cdots,0)$ origin of $\mathbb{R}^n$

$i = 1,\dots n$ $\quad \vec{x}_i := (0,0,\cdots,1,\cdots 0)$ 1 in position i

Let U be __any__ subset of these $(n+1)$ points.

Let $\vec{w} = (\chi_U(\vec{x}_1), \chi_U(\vec{x}_2), \dots \chi_U(\vec{x}_n))$

where $\chi_U(\vec{x}) = \begin{cases} +1 \text{ if } \vec{x} \in U \\ -1 \text{ if } \vec{x} \notin U \end{cases}$

Now define $f_{\vec{w}, \chi_U(x_0)/2}$ as the function

Write $y_i = \chi_U(x_i)$ so $\vec{\omega} = (y_1, y_2, \cdots y_n)$

$$\text{sign}\left(\vec{\omega}\cdot\vec{x}_i + {}^{y_0}\!/_2\right) = \text{sign}\left(y_i + {}^{y_0}\!/_2\right) = y_i$$

$$\text{So } f_{\vec{\omega}, y_0/2}(\vec{x}_i) = \begin{cases} +1 & \text{if } x_i \in U \\ -1 & \text{if } x_i \notin U \end{cases}$$

Hence it realizes the labelling U.

Now we need to prove that <u>no</u> set of $(n+1)$ points can be shattered. If we have two sets of points with intersecting convex hulls no hyperplane can separate them.

<u>Radon's Thm</u> Any subset of $n+2$ points in R^n can be partitioned into two subsets S_1, S_2 such that ~~S_1~~ the convex hulls of S_1 & S_2 intersect.

<u>Proof</u> Let $S = \{\vec{x}_1, \ldots \vec{x}_{n+2}\}$.

$$\text{Consider } \underbrace{\sum_{i=1}^{n+2} \lambda_i \vec{x}_i = 0}_{n \text{ equs}} \quad \& \quad \underbrace{\sum_{i=1}^{n+2} \lambda_i = 0}_{1 \text{ equ}}$$

This is a system of $n+1$ equations in $(n+2)$ variables. So we can always find a <u>non trivial</u> solution for the λ_i. Define $I_1 = \{i \mid \lambda_i > 0\}$

$I_2 = \{i \mid \lambda_i < 0\}$ Both sets have to be non-empty since $\sum \lambda_i = 0$

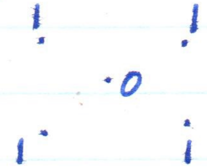
$$\text{Define } \lambda = \sum_{i \in I_1} \lambda_i = -\sum_{j \in I_2} \lambda_j > 0$$

$$\sum_{i \in I_1} \frac{\lambda_i}{\lambda} \vec{x}_i = \sum_{i \in I_2} -\frac{\lambda_i}{\lambda} \vec{x}_i$$

so this point is in both convex hulls.

VC dim $= n+1$

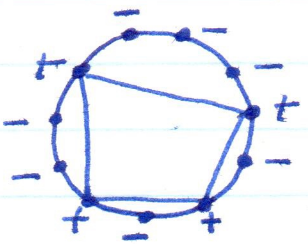
Axis aligned rectangles VC dim $= 4$



A set of 5 points can never be shattered.

Convex polygons ~~of~~ of d sides

VC dim $\geq (2d+1)$



Choose $(2d+1)$ points on a circle. If there are more negative points use the chords between the positive points. If there are more positive points than negative points use the tangents to the negative points. Hard to prove the upper bound. But it is clear that if we do not restrict d VC dim $= \infty$

$|\mathcal{H}| < \infty$ VC dim $(\mathcal{H}) \leq \log_2 |\mathcal{H}|$

If VC dim $= d$ then some set of d points can be shattered so $|\mathcal{H}| \geq 2^d$ so

$d \leq \log_2 |\mathcal{H}|$.

<u>Thm</u> (Sauer's lemma) If $VC(\mathcal{H}) = d$

$$\Pi_{\mathcal{H}}(m) \leq \sum_{i=0}^{d} \binom{m}{i}$$

<u>Proof</u> By induction on m. Base case is $m = 2$ & $d = 1$ so not all dichotomies of S can be realized so $\Pi_{\mathcal{H}}(m) \leq 3 = \binom{2}{0} + \binom{2}{1}$

Induction case $S \subseteq X$ $|S| = m-1$ & $x \notin S$ so $|S \cup \{x\}| = m$. Let us look at all dichotomies realized by $\mathcal{H}$ on $S \cup \{x\}$.

$\mathcal{H} = \mathcal{H}_1 \cup \mathcal{H}_2$ where

$\mathcal{H}_1 = \{ h \in \mathcal{H} \mid \forall h' \in \mathcal{H} \;\; h|_S = h'|_S \Rightarrow h(x) = h'(x) \}$

Once h has assigned values to S it has no freedom to choose the value of x.

$\mathcal{H}_2 = \{ h \in \mathcal{H} \mid \exists h' \in h \;\; h|_S = h'|_S \wedge h(x) \neq h'(x) \}$

$\mathcal{H}_1, \mathcal{H}_2$ partition $\mathcal{H}$

$$|\Pi_{\mathcal{H}}(S \cup \{x\})| = |\Pi_{\mathcal{H}_1}(S \cup \{x\})| + |\Pi_{\mathcal{H}_2}(S \cup \{x\})|$$
$$= |\Pi_{\mathcal{H}_1}(S)| + 2|\Pi_{\mathcal{H}_2}(S)|$$
$$= |\Pi_{\mathcal{H}}(S)| + |\Pi_{\mathcal{H}_2}(S)|.$$

Now note if $S' \subseteq S$ is shattered by $\mathcal{H}_2$ then so is $S' \cup \{x\}$ by $\mathcal{H}$. $VC(\mathcal{H}_2) \leq d-1$ otherwise $VC(\mathcal{H}) > d$ [If $|S'| \geq d$ $|S' \cup \{x\}| \geq d+1$]

Then by IH

$$\cancel{\Pi_{\mathcal{H}}(m) + \Pi_{\mathcal{H}_2}}$$

$$|\Pi_{\mathcal{H}}(S)| + |\Pi_{\mathcal{H}_2}(S)| \leq \sum_{i=0}^{d} \binom{m-1}{i} + \sum_{i=0}^{d-1} \binom{m-1}{i}$$

$j = i+1$

$$= 1 + \sum_{i=1}^{d} \binom{m-1}{i} + \sum_{j=1}^{d} \binom{m-1}{j-1}$$

$$= 1 + \sum_{i=1}^{d} \left[\binom{m-1}{i} + \binom{m-1}{i-1} \right] = \sum_{i=0}^{d} \binom{m}{i}$$

<u>Cor</u> $\quad \forall m \geq d$

$$\Pi_{\mathcal{H}}(m) \leq \left(\frac{em}{d}\right)^d = O(m^d)$$ where $VC(\mathcal{H}) = d$.

<u>Proof</u>

$$\Pi_{\mathcal{H}}(m) \leq \sum_{i=0}^{d} \binom{m}{i} \quad [\text{Sauer}]$$

$$\leq \sum_{i=0}^{d} \binom{m}{i} \left(\frac{m}{d}\right)^{d-i} \quad [m \geq d]$$

$$\leq \sum_{i=0}^{m} \binom{m}{i} \left(\frac{m}{d}\right)^{d-i}$$

$$= \left(\frac{m}{d}\right)^d \left[\sum_{i=0}^{m} \binom{m}{i} \left(\frac{d}{m}\right)^i \right]$$

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

$$\leq \left(\frac{m}{d}\right)^d \left(e^{d/m}\right)^m \quad [1 + x \leq e^x]$$

$$= \left(\frac{m}{d}\right)^d e^d = \left(\frac{em}{d}\right)^d \quad \blacksquare$$

So when you have finite VC dimension the growth rate is only polynomial. If VC dimension is infinite then $\Pi_{\mathcal{H}}(m)$ grows as 2^m.