3.11 VC dimension

[Definition 11] VC dimension
The VC dimension of a family of functions H with boolean outputs is the maximum number of points that can be shattered by H: $\quad VC(H) = \sup\{n : s(H, n) = 2^n\}$

Note: To show a class H has VC dimension d,

(i) upper bound: show $d+1$ points can't be shattered;

(ii) lower bound: show d points can be shattered.

[Theorem 10] finite-dimentional function class
Let $F \subseteq \{f : \mathcal{X} \to \mathbb{R}\}$. Let $H = \{x \mapsto \mathbb{1}\{f(x) \geq 0\} : f \in F\}$.
Then we have $\quad VC(H) \leq \dim(F)$

Pf: for any $n > \dim(F)$, $x_1, \cdots, x_n$ are given.

Consider $M(f) := [f(x_1), \cdots, f(x_n)] \in \mathbb{R}^n$

$M := \{M(f) : f \in F\}$ is linear space. $\dim(M) \leq \dim(F)$.

Since $n > \dim(F) \geq \dim(M)$, $\exists\, 0 \neq c \subset \mathbb{R}^n$ s.t. $M(f) \cdot c = 0$ for all $f \in F$. Without loss of generality, $\{c_i > 0\} \neq \phi$. Then

$$\sum_{c_i > 0} c_i f(x_i) + \sum_{c_i \leq 0} c_i f(x_i) = 0 \qquad \text{for all } f \in F.$$

Suppose H shatters $\{x_1, \cdots, x_n\}$, we could find a $h \in H$ s.t.

$h(x_i) = 1$ whenever $c_i \geq 0$ and $h(x_i) = 0$ whenever $c_i < 0$

we have $\sum_{c_i \geq 0} c_i h(x_i) + \sum_{c_i < 0} c_i h(x_i) > 0$, but $h \in F$, which is a contradiction.

Therefore, H can't shatter $\{x_1, \cdots, x_n\}$ for any choise of

$VC(H) \leq \dim(F)$ $\square$

$x_1, \cdots, x_n$, so $VC(H) = \dim(F)$ $\square$

Application: Half-spaces passing through the origin.

Let $H = \{x \mapsto \mathbb{1}\{w \cdot x \geq 0\} : w \in \mathbb{R}^d\}$ be the set of half-spaces in d-dim.

By Th.10, $VC(H) \leq d$. The lower bound can be obtained by construction: creating d points:

$x_1 = [1, 0, 0, \cdots 0]$, $x_2 = [0, 1, 0, \cdots, 0]$, $\cdots$, $x_d = [0, 0, 0, \cdots, 1]$

Given any $I \subseteq \{1, \cdots, d\}$, we can construct w s.t.

$w_i = 1$ if $i \in I$, $w_i = -1$ if $i \notin I$.

Then for any $I_1 \neq I_2$, $I_1, I_2 \subseteq \{1, \cdots, d\}$, $w^{(1)}, w^{(2)}$

$$w^{(i)} \cdot x_j = w_j^{(i)} = \begin{cases} 1, & j \in I_i \\ -1, & j \in I_i \end{cases}$$

we have $(w^{(1)} x_j)_j \neq (w^{(2)} x_j)_j$, that is, $s(H, d) = 2^d$.

(Since $|2^I| = 2^d$).

[Lemma 6] Sauer's lemma.

For a class H be a class with VC dimension d.

Then $s(H, n) \leq \sum_{i=1}^{d} \binom{n}{i} \leq \begin{cases} 2^n & \text{if } n \leq d \\ \left(\frac{en}{d}\right)^d & \text{if } n > d \end{cases}$

Pf: Consider such a table

T

x_1	x_2	x_3	x_4
0	1	0	1
0	0	0	1
1	1	1	0
1	0	1	0

$\Longrightarrow$ T'

x_1	x_2	x_3	x_4
0	1	0	1
0	0	0	1
0	1	0	0
0	0	0	0

① A canonical form:

(i) Pick a column j.

(ii) For each row r with $r_j = 1$, set $r_j = 0$ if the resulting r doesn't exist in the table.

(iii) Repeat until no more changes are possible.

Note: the number of rows is the same and all the rows are distinct, so $s(\mathcal{H}, n) = s(\mathcal{H}', n)$

② Now we show that $VC(H') \le VC(H)$.

The transformations proceed one column at a time:

$$T \to T_1 \to \cdots T_k \xrightarrow{\text{trans col } j} T_{k+1} \to \cdots \to T'.$$

Claim: After transforming any column j, if some $S \subseteq \{1, \cdots, n\}$ of points is shattered (all $2^{|S|}$ labelings exist on those columns) after transformation, then S was also shattered before transformation.

- Trival when $j \notin S$.
- If $j \in S$:

For any row i with 1 in j, there is a row i' with 0 in column j s.t. $T_{k+1}(i, j') = T_{k+1}(i', j')$ for $j' \neq j$, but $T_{k+1}(i, j) = 1$ and $T_{k+1}(i', j) = 0$

Note that $T_k(i, j) = 1$, $T_k(i', j) = 0$ (other-wise $T_k(i, \cdot) = T_k(i', \cdot)$)

Then all $2^{|S|}$ labelings on S existed before transformation.

③ Each row of T' must contain at most d ones.

Suppose if T' has a row with k ones in $S \subseteq \{1, \cdots, n\}$

Then for each $j \in S$, $\exists$ another row with $k-1$ ones in $S/\{j\}$

Reasoning recursively, all 2^k subset must exist.

$\Rightarrow$ $k \le d$. Based on simple counting, we have

$$S(\mathcal{H}, n) \le \sum_{i=0}^{d} \binom{n}{i}$$ at most d ones in a row of $l = n$.

which completes the first ineq. Observe that for $n \geqslant d$

$$
\begin{aligned}
\sum_{i=1}^{d} \binom{n}{i} &\leqslant \left(\frac{n}{d}\right)^d \sum_{i=0}^{d} \binom{n}{i} \left(\frac{d}{n}\right)^i \\
&\leq \left(\frac{n}{d}\right)^d \sum_{i=0}^{n} \binom{n}{i} \left(\frac{d}{n}\right)^i \\
&= \left(\frac{n}{d}\right)^d \left(1 + \frac{d}{n}\right)^n \\
&\leq \left(\frac{n}{d}\right)^d e^d
\end{aligned}
$$

$\square$.