Learning Infinite Classes II

In this lecture we introduce the notion of VC dim (Vapnik - Chervonenkis dimension) and discuss the fundamental theorem of learning theory.

An important result will be proved on the growth rate of hyp classes when the VC dim is finite (Sauer's lemma).

#.

Recap from last time.

$C = \{c_1, c_2 \ldots c_m\} \subset \mathcal{X}$

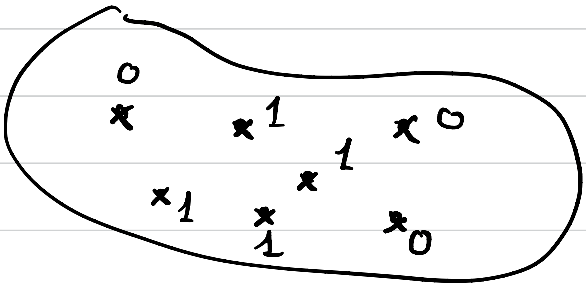
$\mathcal{H}$ set of hypothesis $h : \mathcal{X} \to \mathcal{Y} = \{0,1\}$

$\mathcal{H}_C$ = restriction of functions in $\mathcal{H}$ to $C \to \mathcal{Y}$

It is a good idea to view $\mathcal{H}_C$ as the set of binary assignments

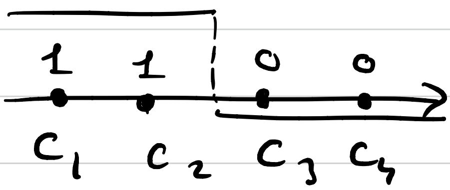
$$\{h(c_1)\ h(c_2)\ \dots\ h(c_m)\} \subset \{0,1\}^m$$

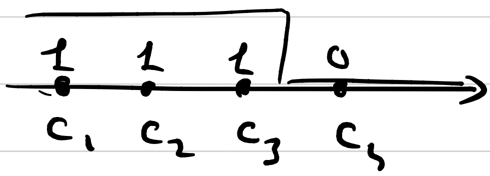
or labelings of elements of C.



$C \subset \mathcal{X}$.

Example: Threshold fct $\{h_a = \mathbb{1}_{x \leq a}\} = \mathcal{H}$





etc... $\mathcal{H}_C$ here has 5 functions (5 possible labellings with threshold fcts among all possible which are 16)

Definition of Shattering.

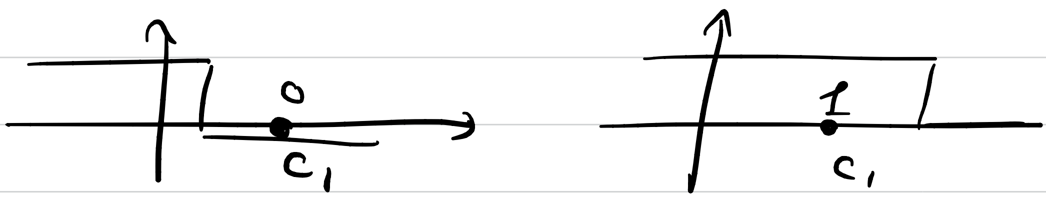
$\mathcal{H}$ is said to shatter a set $C \subset \mathcal{X}$ if $\mathcal{H}_C$ contains all possible labelings

i.e. all possible fcts $C \to \{0,1\}$ i.e. if $|\mathcal{H}_C| = 2^{|C|}$.

Examples.

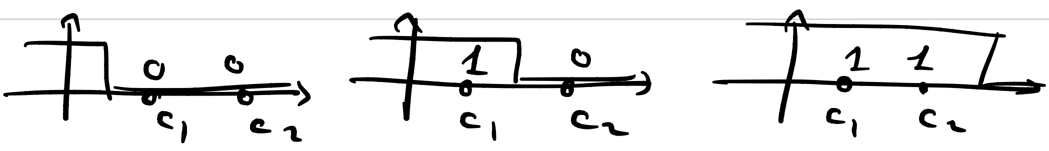
- In the previous example above obviously $\mathcal{H}$ = Threshold fcts does not shatter $C = \{c_1 c_2 c_3 c_4\}$. since we have $|\mathcal{H}_C| = 5$ and $2^{|C|} = 16$.

- $C = \{c_1\}$ Threshold fcts shatter C, Indeed



so we get all possible labelings.

- $C = \{c_1, c_2\}$ Threshold fcts do not shatter C.



but $(0,1)$ is missing.

Definition: **VC dimension of an hypothesis class.**

$VC\dim(\mathcal{H}) = d$ is the largest integer d such that we can find some set $C \subset \mathcal{X}$ with $|C| = d$ which is shattered by $\mathcal{H}$.

Remark: $VC\dim(\mathcal{H}) = d$ is the integer such that

- $\exists\, C$, $|C| = d$, shattered by $\mathcal{H}$ so, s.t $\mathcal{H}_C$ has all possible labelings.

AND:

- $\forall\, C$, $|C| \geq d+1$, are not shattered by $\mathcal{H}$ so, s.t $\mathcal{H}_C$ does not contain all possible labelings.

$$d = \max\{|C| : \text{such that } |\mathcal{H}_C| = 2^{|C|}\}.$$

Examples of computation of VC dim.

a) Threshold fcts $\mathcal{X} = \mathbb{R}$. $\mathcal{H} = \{ \mathbb{1}_{x \leq a} = h_a \}$.

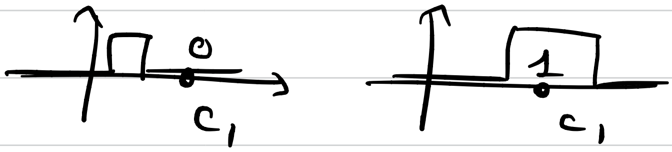
$C = \{c_1\}$ is shattered $\Rightarrow \mathrm{VCdim}(\mathcal{H}) \geq 1$

$C = \{c_1, c_2\}$ is not shattered $\Rightarrow \mathrm{VCdim}(\mathcal{H}) < 2$

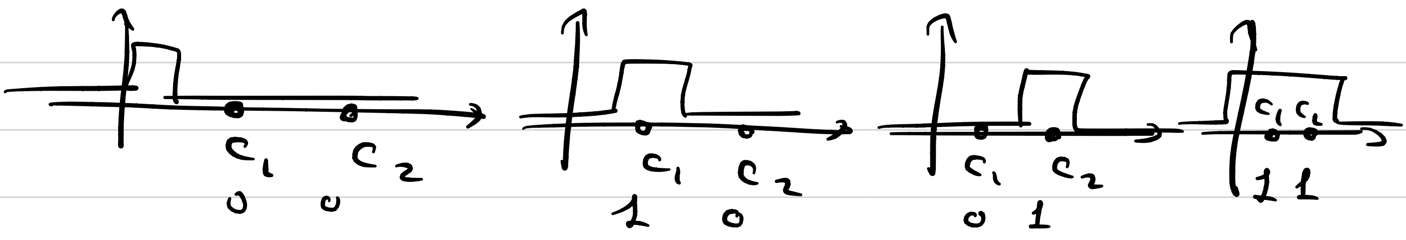
Thus $\mathrm{VCdim}(\mathcal{H}) = 1$.

b) Interval fcts. $\mathcal{H} = \{ h_{a,b} = \mathbb{1}_{a \leq x \leq b} \}$

$C = \{c_1\}$ is shattered



$C = \{c_1, c_2\}$ is also shattered



$\Rightarrow \mathrm{vcdim}\,\mathcal{H} \geq 2$.

$C = \{c_1, c_2, c_3\}$ is not shattered. Indeed the labeling 1 0 1 is not obtained by rect fcts.

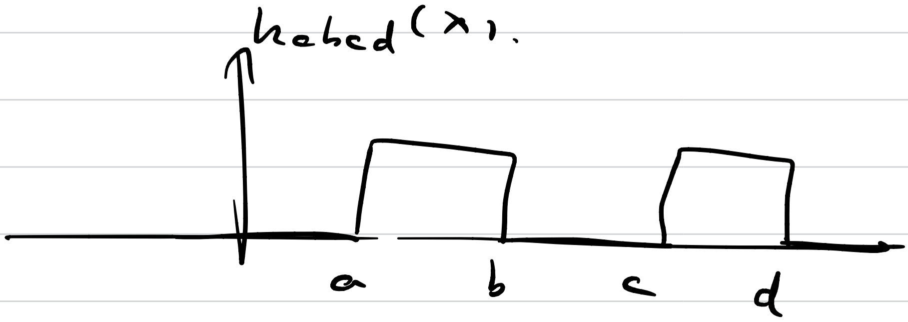
$\Rightarrow \mathrm{vcdim}\,\mathcal{H} < 3$

$\boxed{\text{Thus } \mathrm{VCdim}(\mathcal{H}) = 2}$

c) Exercise:

Let $\mathcal{H}_{abcd} = \{ h_{abcd} = \mathbb{1}_{a \le x \le b} \cdot \mathbb{1}_{c \le x \le d} \}$

for $a < b < c < d$.



Show that $\mathrm{VCdim}(\mathcal{H}) = 4$.

d) Exercises.

Let $\mathcal{H} = \{ h_\theta : h_\theta(x) = \frac{1}{2} \lceil \sin \theta x \rceil, \theta \in \mathbb{R} \}$

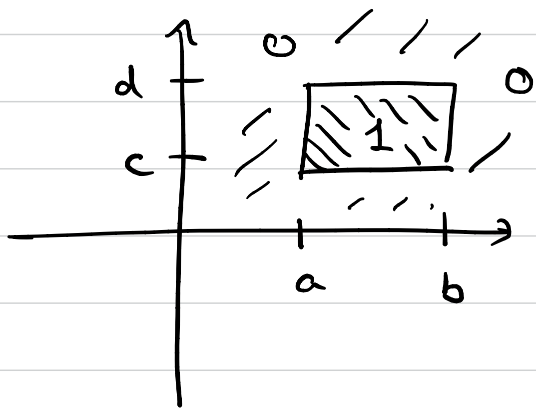
upper integer part.

Note this class is parametrized by one parameter $\theta \in \mathbb{R}$.

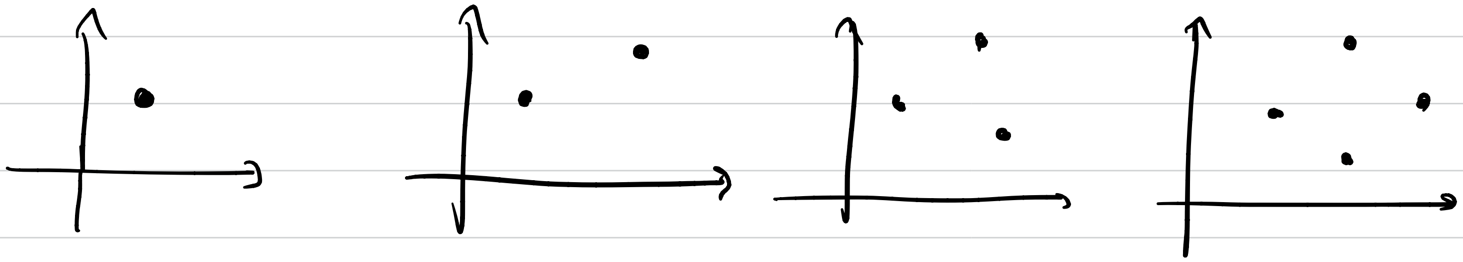
Show that $\mathrm{VCdim}(\mathcal{H}) = +\infty$. In other words given any m, one can construct C with $|C| = m$ s.t. $\mathcal{H}_C$ contains any labeling.

e) Axis aligned rectangles.

$\mathcal{X} = \mathbb{R}^2 \qquad h_{[a,b]\times[c,d]}(x) = \begin{cases} 1 & x \in [a,b]\times[c,d] \\ 0 & \end{cases}$



fct equal to 1 inside rect.



These sets C are all shattered $\Rightarrow$ VCdim ≥ 4.

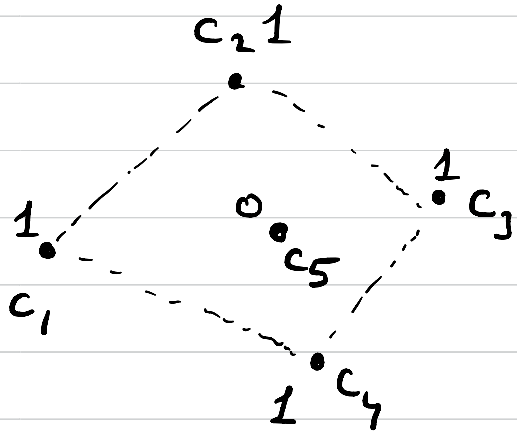
but All sets with $|C| = 5$ cannot be shattered

$c_1 = (c_1^x, c_1^y) \qquad c_1^x$ min

$c_2 = (c_2^x, c_2^y) \qquad c_2^y$ max

$c_3 = (c_3^x, c_3^y) \qquad c_3^x$ max

$c_4 = (c_4^x, c_4^y) \qquad c_4^y$ min



VCdim < 5.

$c_5 \to$ Necessarily inside convex hull

A Rectangle enclosing $c_1 c_2 c_3 c_4$ will also enclose c_5 Hence.

The case of finite classes.

Consider a finite class $|\mathcal{H}| < +\infty$ and look at sets $|C|$ such that $\underline{\underline{2^{|C|} > |\mathcal{H}|}}$.

Obviously these sets cannot be shattered since there are more potential labelings than hyp fcts.

Thus sets C with $|C| > \log_2 |\mathcal{H}|$ cannot be shattered. $\boxed{\text{This means } \quad \mathrm{VC\,dim}(\mathcal{H}) \leq \log_2 |\mathcal{H}|}$

No Free Lunch Thm revisited.

Let $VCdim(\mathcal{H}) = +\infty$. This means for any integer $2m$, any set $|C| = 2m$ is shattered and $|\mathcal{H}_C| = 2^{2m}$ i.e C contains all possible labelling fcts. We can redo the proof of the No Free Lunch Thm which was based on distr constructed out of all these labeling fcts. Conclusion was:

for any Alg $A(S)$ receiving $|S| \leq m$ $\exists \mathcal{D}_i$ over $\mathcal{X} \times \{0,1\}$ and $f_i \in \mathcal{H}_C$ s.t

(a) $L_{\mathcal{D}_i}(f_i) = 0$

(b) $\mathbb{P}\left(L_{\mathcal{D}_i}(A(S)) \geq \frac{1}{8} \right) \geq \frac{1}{7}$

and hence $\mathcal{H}$ is not PAC learnable.

> **Thm:** $VCdim(\mathcal{H}) = +\infty \Rightarrow \mathcal{H}$ is not PAC learnable
>
> $\mathcal{H}$ is PAC learnable $\Rightarrow VCdim(\mathcal{H}) < +\infty$.

Fundamental Theorem of PAC learning.

Let $\mathcal{H} \ni h : \mathcal{X} \to \{0,1\}$.

Let loss fct ℓ s.t $0 \leq \ell(h, z) \leq 1$.

The following are all equivalent:

1. $\mathcal{H}$ has the uniform conv property
2. $\mathcal{H}$ is agnostic PAC learnable with the ERM rule
3. $\mathcal{H}$ is agnostic PAC learnable.
4. $\mathcal{H}$ is PAC learnable.
5. $\mathcal{H}$ is PAC learnable with the ERM rule
6. $\mathcal{H}$ is finite VC dimension.

Proof: $1 \Rightarrow 2$ (previous classes)

$2 \Rightarrow 3$, $3 \Rightarrow 4$, $2 \Rightarrow 5$ (trivial!)

$4 \Rightarrow 6$ (No free lunch thm just discussed).

$\boxed{6 \Rightarrow 1 \text{ Remains to be shown.}}$

Now we discuss the proof of (6) => (1).

Main idea first:

Recap from last time $\underbrace{\tau_{\mathcal{H}}(m) = \max_{|C|=m} |\mathcal{H}_C|}_{\text{"growth rate"}}$

and we proved

Thm: $\forall \mathcal{D}$ we have for $\forall \delta > 0$

$$\mathbb{P}_{S \sim \mathcal{D}^m} \left\{ \sup_{h \in \mathcal{H}} | \underbrace{L_{\mathcal{D}}(h)}_{\text{true Risk}} - \underbrace{L_S(h)}_{\text{empirical risk}} | \leq \frac{4 + \sqrt{\log \tau_{\mathcal{H}}(m)}}{\delta \sqrt{2m}} \right\} \geq 1 - \delta$$

This thm implies unif conv property holds if we have $\frac{4 + \sqrt{\log \tau_{\mathcal{H}}(m)}}{\delta\sqrt{2m}} < \epsilon$ and thus $\mathcal{H}$ is agnostic PAC learnable.

This can be achieved for $m \geq m_{\mathcal{H}}^{UC}(\epsilon, \delta) = C \, \frac{d + \log 1/\delta}{\epsilon^2}$

as long as $\tau_{\mathcal{H}}(m) = \mathrm{poly}(m)$.

So what remains to be understood is the behavior of $\tau_{\mathcal{H}}(m)$ and notably when is it $\mathrm{poly}(m)$?

- Note that if $m \leq d = \mathrm{vcdim}(\mathcal{H})$ then $\exists C$ with $|C| = m$ shattered by $\mathcal{H}$ (by definition of VC)

$\Rightarrow |\mathcal{H}_C| = 2^m \Rightarrow \boxed{\tau_{\mathcal{H}}(m) = 2^m \text{ for } m \leq d.}$

- The real question is what happens for $m \geq d+1$?

Lemma: Sauer - Shelah - Perles

Let $\mathrm{vcdim}(\mathcal{H}) = d < +\infty$. Then for $m \geq d+1$ we have

$$\tau_{\mathcal{H}}(m) \leq \sum_{i=0}^{d} \binom{m}{i} \leq \left(\frac{em}{d}\right)^d.$$

Proof of Lemma.

- The inequ $\displaystyle\sum_{i=0}^{d} \binom{m}{i} \leq \left(\frac{em}{d}\right)^d$ for $m \geq d+1$

is "calculus" and is left as exercise or

(see Appendix of UML).

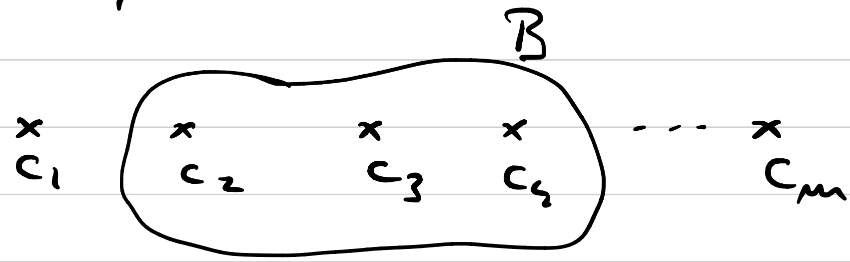
- We show here $\tau_{\mathcal{H}}(m) \leq \displaystyle\sum_{i=0}^{d} \binom{m}{i}$ for $m \geq d+1$.

First remark: it suffices to prove

$$|\mathcal{H}_C| \leq \left| \{ B \subset C \text{ such that } \mathcal{H} \text{ shatters } B \} \right|$$

Number of subsets of C that are shattered by $\mathcal{H}$.

Indeed:



B - subsets have size $|B| \leq d$ since they are shattered
of B-subsets of size i is $\binom{m}{i}$.

(Note for $i=0$ we take $B = \emptyset$ which is shattered by convention).

We will prove the inequ by induction.

$\underline{\underline{m=1}}$: $C = \{c\}$ possibilities for $\mathcal{H}_C$ are

$\mathcal{H}_C = \{h(c)=0\}$ $\mathcal{H}_C = \{h(c)=1\}$

and $\mathcal{H}_C = \{h_1(c) = 0, h_2(c) = 1\}$.

- If $\mathcal{H}_C = \{h(c)=0\}$ or $\mathcal{H}_C = \{h(c)=1\}$

$|\mathcal{H}_C| = 1$ so $\tau_{\mathcal{H}}(m=1) = 1$.

subsets $B \subset C$ that are shattered is only $\emptyset$.
(since $C = \{c\}$ is not shattered here).

$|\{B \subset C,\ B \text{ is shattered by } \mathcal{H}\}| = 1$.

$1 \leq 1$ ✓

- If $\mathcal{H}_C = \{h_1(c)=0, h_2(c)=1\}$ then

$|\mathcal{H}_C| = 2$ so $\tau_{\mathcal{H}}(m=2) = 2$.

subsets $B \subset C$ that are shattered are $\emptyset$, $C = \{c\}$

$|\{B \subset C,\ B \text{ is shattered by } \mathcal{H}\}| = 2$

$2 \leq 2$ ✓

Thus base case $\underline{m=1}$ works.

<u>induction hypothesis</u>: we assume the inequality holds for all sets C' of size $1 \leq k \leq m-1$

We must prove it then holds for C of size m.

$C' = \{c_2 \dots c_m\}$ $\quad \overset{c_2}{\times} \quad \overset{c_3}{\times} \quad \dots \quad \overset{c_m}{\times}$

add a point c_1

$C = \{c_1 ; c_2 \dots c_m\}$ $\quad \overset{c_1}{\times} \quad \overset{c_2}{\times} \quad \times \quad \dots \quad \overset{c_m}{\times}$

<u>Possibilities for $\mathcal{H}_C$:</u>

$$\left.\begin{array}{ccc} c_1 & c_2 & c_m \\ \times & \times \; \dots & \times \\ 0 & y_2 & y_m \\ \text{OR} & & \\ 1 & & \end{array}\right\}$$

<u>labelings</u> of C' extend to only one possibility for c_1.

This set of <u>labelings</u> is called Y_0.

$$\left.\begin{array}{cccc} c_1 & c_2 & & c_m \\ \times & \times & \dots & \times \\ 0 & y_2 & \dots & y_m \\ 1 & & & \end{array}\right\}$$

<u>labelings</u> of C' extend to both possibilities for c_1

This set of labelings is called Y_1.

We have obviously

$$|\mathcal{H}_C| = |Y_0| + |Y_1|.$$

Inequality on $|Y_0|$:

$$|Y_0| = |\mathcal{H}_{C'}|$$

$$\leq \left| \{ B' \subset C' \text{ s.t } B' \text{ is shattered by } \mathcal{H} \} \right|$$

by induction hypothesis.

$$= \left| \{ B \subset C \text{ s.t } c_1 \notin B \text{ and } B \text{ is shattered by } \mathcal{H} \} \right|$$

trivial

in summary

$$\boxed{|Y_0| \leq \left| \{ B \subset C \text{ s.t } c_1 \notin B \text{ and } B \text{ is shattered by } \mathcal{H} \} \right|}$$

Inequality on Y_1 :

Let $\tilde{\mathcal{H}} \subset \mathcal{H}$ where $\mathcal{H}'$ contains pairs of functions $\bar{h}, \bar{\bar{h}}$ which are equal on C' and opposite on c_1.

$$\tilde{\mathcal{H}} = \Big\{ \bar{h} \in \mathcal{H} \text{ such that } \exists\, \bar{\bar{h}} \in \mathcal{H} \text{ with } \big(\bar{h}(c_1)\ \bar{h}(c_2) \dots \bar{h}(c_m)\big) = \big(1-\bar{\bar{h}}(c_1),\ \bar{\bar{h}}(c_2) \dots \bar{\bar{h}}(c_m)\big) \Big\}$$

agree

$\bar{h}(c_1) \neq \bar{\bar{h}}(c_1)$.

$$|Y_1| = |\tilde{\mathcal{H}}_{C'}|$$

$$\underset{\text{induction}}{\leq} \left| \{ B' \subset C' \text{ such that } B' \text{ is shattered by } \tilde{\mathcal{H}} \} \right|$$

$$= \left| \{ B' \subset C' \text{ s.t } B' \cup \{c_1\} \text{ is shattered by } \tilde{\mathcal{H}} \} \right|$$

since $\mathcal{H}'$ contains pairs of hyp which differ on c_1 .

$$\underset{\text{trivial}}{=} \left| \left\{ B \subset C \text{ such that } c_1 \in B \text{ and } B \text{ is shattered by } \tilde{\mathcal{H}} \right\} \right|$$

$$\le \left| \left\{ B \subset C \text{ s.t } c_1 \in B \text{ and } B \text{ is shattered by } \mathcal{H} \right\} \right|$$

because $\tilde{\mathcal{H}} \subset \mathcal{H}$ so <u>more sets</u> are shattered by $\mathcal{H}$ since we have more fcts in $\mathcal{H}$.

in summary we have obtained

$$|Y_0| \le \left| \left\{ B \subset C \text{ s.t } c_1 \notin B \text{ \& } B \text{ shattered by } \mathcal{H} \right\} \right|$$

$$|Y_1| \le \left| \left\{ B \subset C \text{ s.t } c_1 \in B \text{ \& } B \text{ shattered by } \mathcal{H} \right\} \right|$$

$$\Rightarrow \boxed{\begin{aligned} |\mathcal{H}_C| &= |Y_0| + |Y_1| \\ &\le \left| \left\{ B \subset C \text{ s.t } B \text{ shattered by } \mathcal{H} \right\} \right| \end{aligned}}$$

CQFD.