Machine Learning from Data

Lecture 6: Spring 2021

Today’s Lecture

- Bounding the Growth Function $m_{\mathcal{H}}(N)$
- Models are either Good or Bad
- The VC Bound

Putting Everything Together

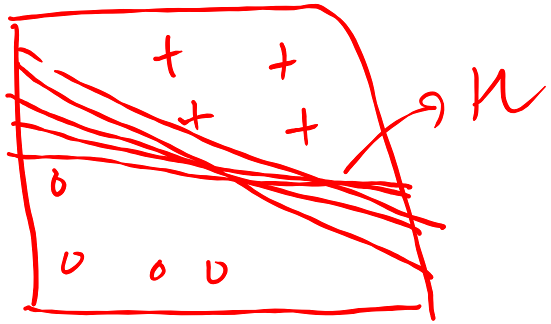
- The growth function:

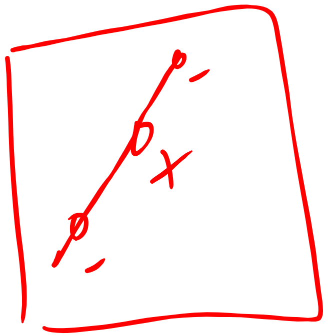
The *growth function* $m_{\mathcal{H}}(N)$ considers the worst possible $\mathbf{x}_1, \ldots, \mathbf{x}_N$.

$$m_{\mathcal{H}}(N) = \max_{\mathbf{x}_1,\ldots,\mathbf{x}_N} |\mathcal{H}(\mathbf{x}_1, \ldots, \mathbf{x}_N)|.$$

$\mathcal{H}$

growth function





1) Can we bound $m_H(N)$ by a polynomial in N?

2) Can we replace $|H|$ by $m_H(N)$ in the generalization bound?

Perceptron: $N=4$ (16) $\longrightarrow$ 14

1-D positive ray: $N=2$ (4) $\longrightarrow$ 3

2-D rectangle: $N=5$ $\boxed{m_H(5) < 2^5}$

$\rightarrow$ $m_H(N)$ drops below 2^N

$\rightarrow$ A break-point is any k for which $m_H(k) < 2^k$

Quiz Q1

I give you a set of k^* points $\mathbf{x}_1, \ldots, \mathbf{x}_{k^*}$ on which $\mathcal{H}$ implements $< 2^{k^*}$ dichotomys.

(a) k^* is a break point.

(b) k^* is not a break point.

(c) all break points are $> k^*$.

(d) all break points are $\leq k^*$.

(e) we don't know anything about break points.

Quiz Q2

For every set of k^* points $\mathbf{x}_1, \ldots, \mathbf{x}_{k^*}$, $\mathcal{H}$ implements $< 2^{k^*}$ dichotomys.

(a) k^* is a break point. → definition

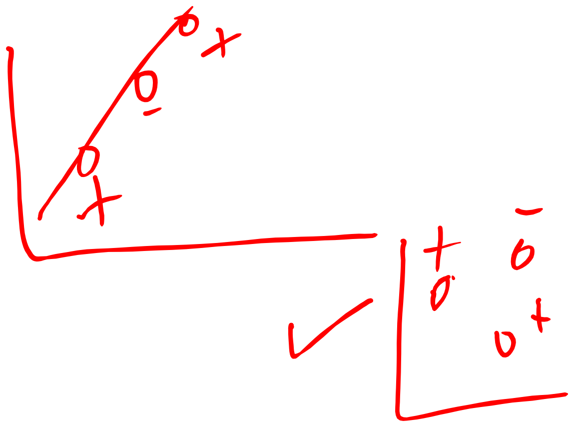
(b) k^* is not a break point.

(c) all $k \geq k^*$ are break points.

(d) all $k < k^*$ are break points.

(e) we don't know anything about break points.

Quiz Q3



To show that k is *not* a break point for $\mathcal{H}$:

a, e

$\mathcal{H} \xrightarrow{\text{shatters}} x_1 \dots x_k$

(a) Show a set of k points $\mathbf{x}_1, \dots \mathbf{x}_k$ which $\mathcal{H}$ can shatter.

(b) Show $\mathcal{H}$ can shatter any set of k points.

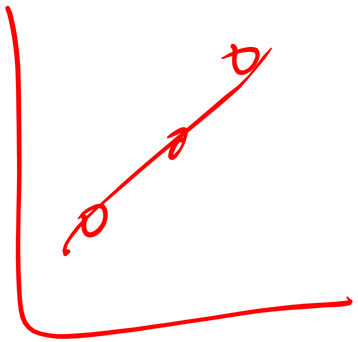
(c) Show a set of k points $\mathbf{x}_1, \dots \mathbf{x}_k$ which $\mathcal{H}$ cannot shatter.

(d) Show $\mathcal{H}$ cannot shatter any set of k points.

(e) Show $m_{\mathcal{H}}(k) = 2^k$. (similar to (a))

Quiz Q4

To show that k *is* a break point for $\mathcal{H}$:



(a) Show a set of k points $\mathbf{x}_1, \ldots \mathbf{x}_k$ which $\mathcal{H}$ can shatter.

(b) Show $\mathcal{H}$ can shatter any set of k points.

(c) Show a set of k points $\mathbf{x}_1, \ldots \mathbf{x}_k$ which $\mathcal{H}$ cannot shatter.

(d) Show $\mathcal{H}$ cannot shatter any set of k points.

(e) Show $m_{\mathcal{H}}(k) > 2^k$.

Back to the puzzle

How many dichotomies can you list on 4 points so that no 2 is shattered.

x_2 & x_3 are shattered!

4 points

$\mathbf{x}_1$	$\mathbf{x}_2$	$\mathbf{x}_3$	$\mathbf{x}_4$
○	○	○	○
○	○	○	●
○	○	●	○
○	●	○	○
●	○	○	○
o	●	●	o

$\boxed{\text{5 dichotomies}}$

N, k

Bound

The combinatorial relationship

3 points

max. no. dichotomies

$\mathbf{x}_1$	$\mathbf{x}_2$	$\mathbf{x}_3$
○	○	○
○	○	●
○	●	○
●	○	○

4

$B(3,2) = 4$

4 points

$B(4,2) = 5$

$\mathbf{x}_1$	$\mathbf{x}_2$	$\mathbf{x}_3$	$\mathbf{x}_4$
○	○	○	○
○	○	○	●
○	○	●	○
○	●	○	○
●	○	○	○

5

$B(N,k)$: The maximum no. of dichotomies on data points such that no subset of size k is shattered.

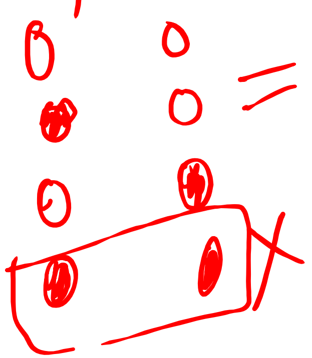
$B(N, k)$

N \ k	1	2	3	4	5	6
1	1	2	2	2	2	2
2	1	3	4	4	4	4
3	1		7	8	8	8
4	1			15	16	16
5	1				31	32
6	1					63

$B(N, N) = 2^N - 1$

$B(2,2)$

$x_1 \quad x_2$



$= 2^2 - 1$

How many dichotomies can you list on 4 points so that no subset of 3 is shattered.

$B(N, k)$

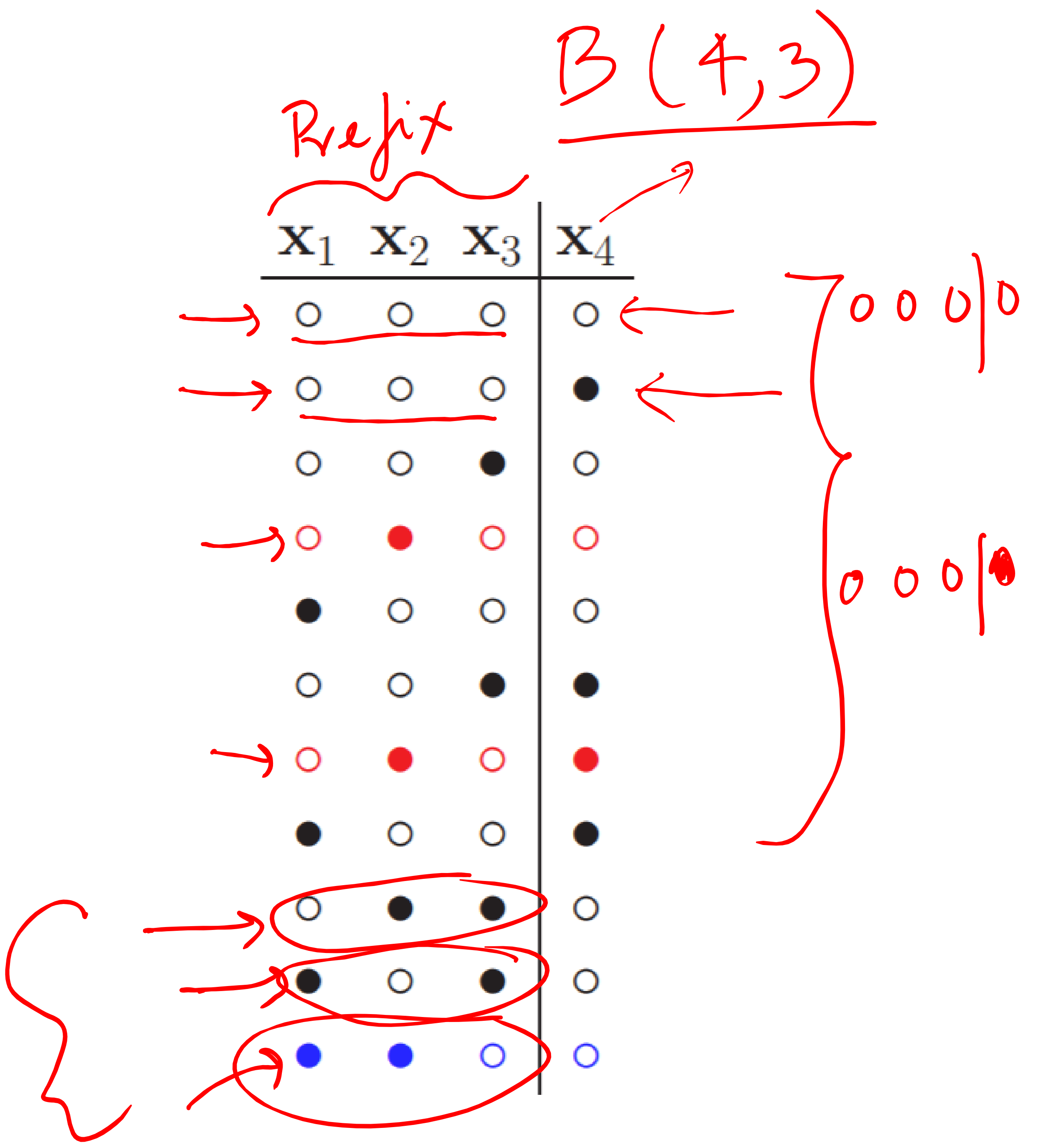
$B(4,3) = 11$

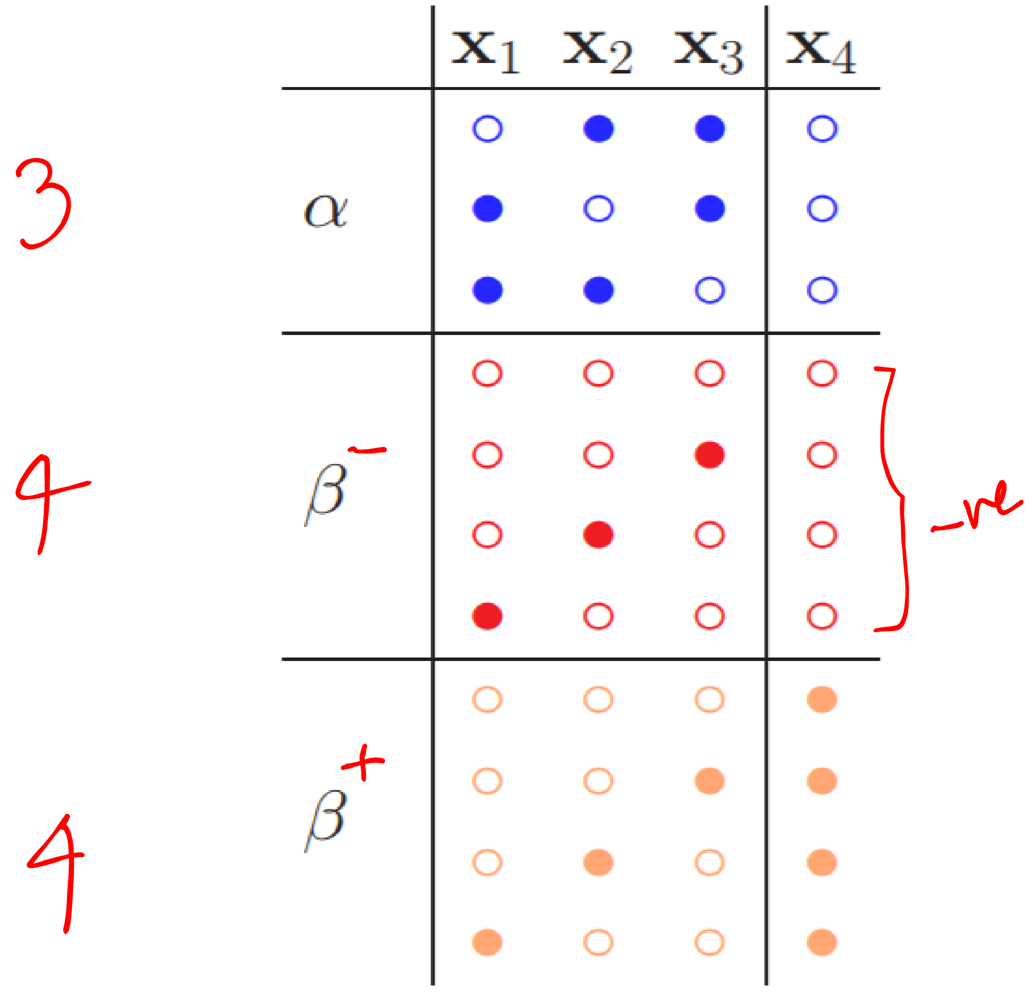
$\mathbf{x}_1$	$\mathbf{x}_2$	$\mathbf{x}_3$	$\mathbf{x}_4$
○	○	○	○
○	○	○	●
○	○	●	○
○	●	○	○
●	○	○	○
○	○	●	●
○	●	○	●
●	○	○	●
○	●	●	○
●	○	●	○
●	●	○	○

11

Based on prefix's we will segregate the dichotomies.

α





α: prefix appears once

β: prefix appears twice

$$B(4,3) = \alpha + 2\beta = 11$$

$$B(4,3) = \alpha + 2\beta$$

	$\mathbf{x}_1$	$\mathbf{x}_2$	$\mathbf{x}_3$	$\mathbf{x}_4$
α	○	●	●	○
	●	○	●	○
	●	●	○	○
$\bar{\beta}$	○	○	○	○
	○	○	●	○
	○	●	○	○
	●	○	○	○
β	○	○	○	●
	○	○	●	●
	○	●	○	●
	●	○	○	●

N

$\boxed{B(4,3)}$

$\boxed{\alpha + \bar{\beta} \le B(\overset{N}{3},\overset{k}{3})}$

Suppose a pair is shattered.

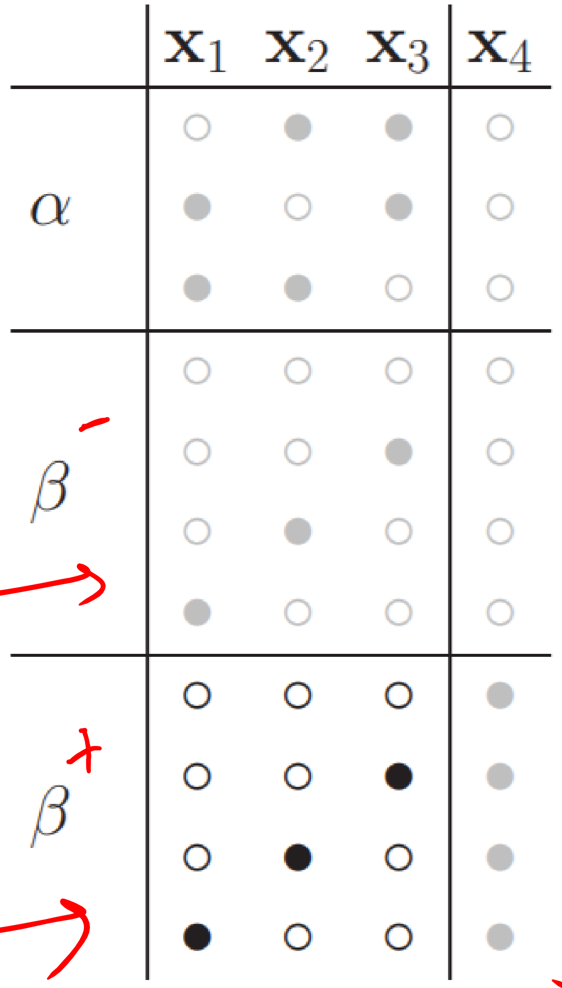
x_1 x_2 x_4

β

β^+

Impossible!

$B(4,3)$



$\beta^+ \le B(3,2)$

$$\therefore\ B(\underset{N}{4}, \underset{k}{3}) = \alpha + 2\beta$$

$$= \underbrace{\alpha + \beta} + \beta$$

$$\leq B(3,3) + B(3,2)$$

$$\boxed{B(\underline{N}, \underline{k}) \leq \underline{B(N-1, k) + B(N-1, k-1)}}$$

Fill the table values

	k						
N	1	2	3	4	5	6	$\cdots$
1	1						
2	1	3					
3	1		7				
4	1			15			
5	1				31		
6	1					63	
$\vdots$	$\vdots$						$\ddots$

$$B(N, 1) = 1$$

$$B(N, N) = 2^N - 1$$

$$\boxed{B(N,k) \leq B(N-1,k) + B(N-1,k-1)}$$

N \ k	1	2	3	4	5	6	$\cdots$
1	1						
2	1	3					
3	1	**4**	7				
4	1	5	$\boxed{11}$	15			
5	1				31		
6	1					63	
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\ddots$

$B(3,2) \leq B(2,2) + B(2,1)$

$B(4,3)$

$$\boxed{B(N,k)} \leq B(N-1,k) + B(N-1,k-1)$$

		k						
		1	2	3	4	5	6	...
	1	1						
	2	1	3					
N	3	1	**4**	7				
	4	1	**5**	$\boxed{\mathbf{11}}$	15			
	5	1	**6**	**16**	**26**	31		
	6	1	**7**	**22**	**42**	**57**	63	
	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋱

$B(4,3)$

Analytic Bound

- Theorem: $\boxed{B(N,k) \le \sum_{i=0}^{k} \binom{N}{i}}$?

$$B(N,k) \le B(N-1,k) + B(N-1,k-1)$$

$$\boxed{\binom{N}{k} = \binom{N-1}{k} + \binom{N-1}{k-1}} \checkmark$$

Proof: 1 ✓ o $\underbrace{\text{o o o . — — —}}$ N items. $\binom{N-1}{k-1}$

2 × o $\underbrace{\text{o o . .}}$ N items $\binom{N-1}{k}$

base case

Proof: (Induction on N.)

N+1's & all k's

1. Verify for $N = 1$: $B(1,1) \le \binom{1}{0} = 1$ ✓

Assume

2. Suppose $B(N,k) \le \sum_{i=0}^{k-1} \binom{N}{i}$.

Lemma. $\boxed{\binom{N}{k} + \binom{N}{k-1} = \binom{N+1}{k}}$.

Top

$$\boxed{B(N+1,k) \le B(N,k) + B(N,k-1)}$$

we proved this already

$$
\begin{aligned}
&\le \sum_{i=0}^{k-1} \binom{N}{i} + \sum_{i=0}^{k-2} \binom{N}{i} \\
&= \sum_{i=0}^{k-1} \binom{N}{i} + \sum_{i=1}^{k-1} \binom{N}{i-1} \\
&= 1 + \sum_{i=1}^{k-1} \left(\binom{N}{i} + \binom{N}{i-1} \right) \\
&= 1 + \sum_{i=1}^{k-1} \binom{N+1}{i} \\
&= \sum_{i=0}^{k-1} \binom{N+1}{i}
\end{aligned}
$$

Re-arrangement

B(N+1, k)

Theorem

i) $\sum_{i=0}^{k-1} \binom{N}{i} \leq N^{k-1} + 1$ ✓

ii) $\sum_{i=0}^{k-1} \binom{N}{i} \leq \left(\frac{eN}{k-1}\right)^{k-1}$ ✓

$\therefore \quad B(N,k) \leq N^{k-1} + 1$

$$\boxed{B(N,k) \leq \text{Polynomial in } N}$$

N	1	2	3	4	5	6
2-D perceptron	2	4	8	14		
1-D Ray $(N+1)$	2	3	8	16	$< 2^5$	
2-D Rectangle	2	4				
2^N	2	4	8	16	32	

$m_H(N) =$ max. no. of dichotomies that can be implemented on N points by a H.

$k \longrightarrow$ break points

Subset of size k

$k \rightarrow$ break point

Any subset of k points is not shattered by these $m_H(N)$ dichotomies

$$\underbrace{B(N,k)} \longrightarrow \text{max.}$$

$$\boxed{m_H(N) \le B(N,k)}$$

Theorem : Let $\mathcal{H}$ be any hypothesis set:

i) $\mathcal{H}$ is extremely complex (Never breaks)
i.e. $m_{\mathcal{H}}(N) = 2^N \;\forall N$ (Not very useful)

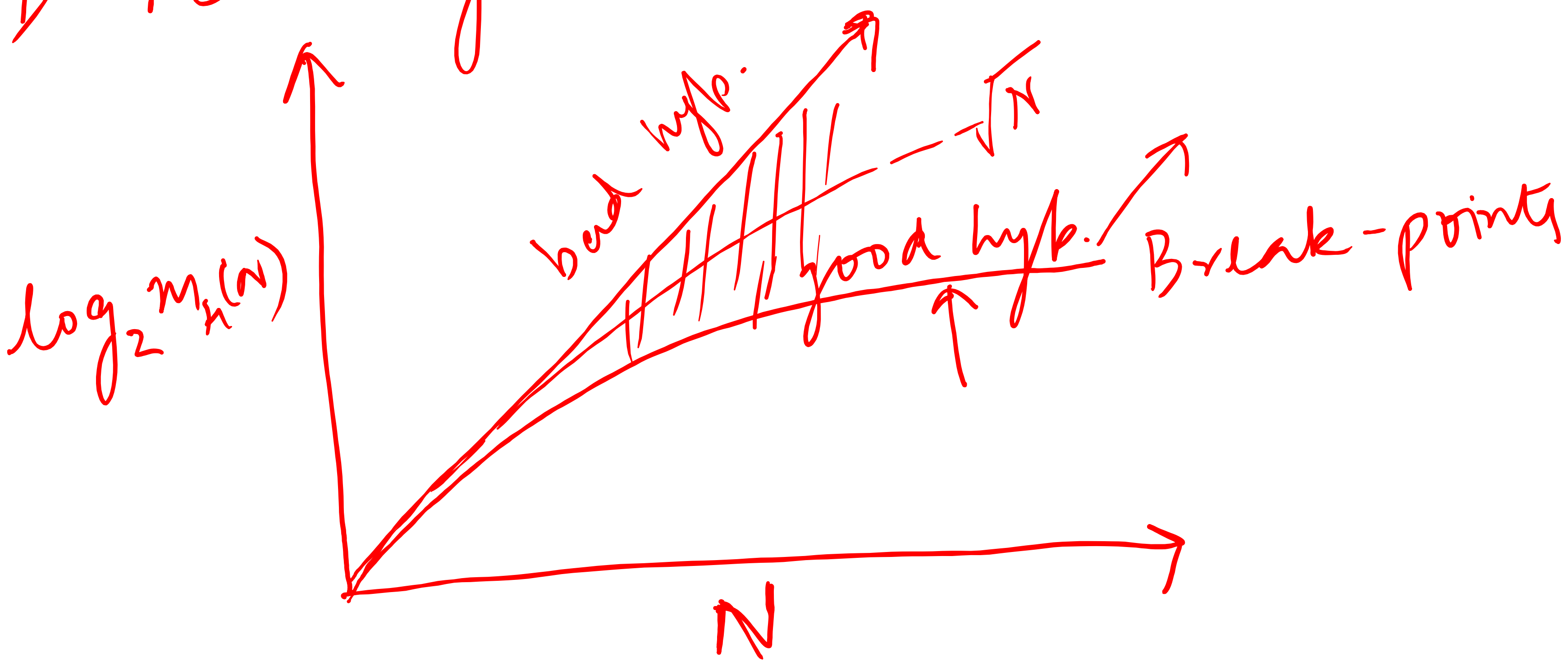
ii) $\mathcal{H}$ does get broken. In this case it has a breakpoint (k). Then:

$$\underline{m_{\mathcal{H}}(N)} \leq \underline{\underline{B(N,k)}} \leq N^{k-1} + 1$$

2-D perceptron $\longrightarrow N^{k-1}+1 \xrightarrow{\text{Bound}} \le N^{3}+1$

1-D ray $\longrightarrow N^{k-1}+1 \xrightarrow{\text{Bound}} \le N^{1}+1$

2-D rectangle $\longrightarrow \le N^{5-1}+1 = N^{4}+1$



Error bar

$$E_{out} \le E_{in} + O\left(\frac{\ln |H|}{N}\right)$$

$\ln m_H(N)$

$\downarrow$

2^N

$\ne 0$

ii) $\frac{\ln (N)}{N} \longrightarrow 0$

$E_{out} \simeq E_{in}$

✓ Can we get a polynomial bound on $m_{\mathcal{H}}(N)$ even for infinite $\mathcal{H}$?

Yes

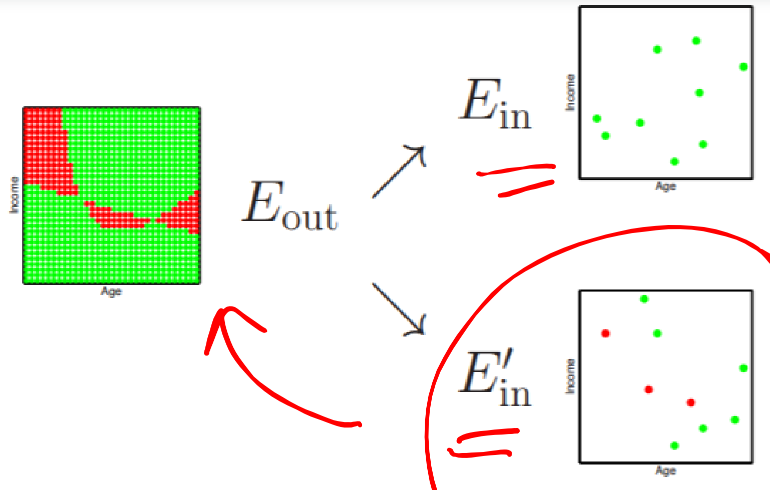
Can we replace $|\mathcal{H}|$ with $m_{\mathcal{H}}(N)$ in the generalization bound?

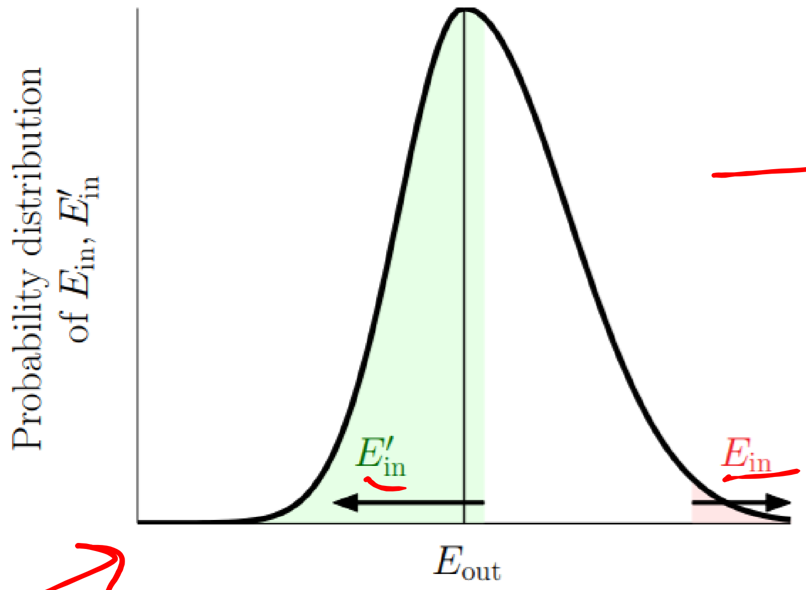
Yes

$E_{out} \le E_{in} +$ [illegible]

Test

The *ghost data set*: a 'fictitious' data set $\mathcal{D}'$:





E'_{in} is like a test error on N new points.

E_{in} deviates from E_{out} implies E_{in} deviates from E'_{in}.

E_{in} and E'_{in} have the same distribution.

$\mathbb{P}[(E'_{\text{in}}(g), E_{\text{in}}(g))$ "deviate"$] \geq \frac{1}{2}\,\mathbb{P}\,[(E_{\text{out}}(g), E_{\text{in}}(g))$ "deviate"$]$

E_{in} E'_{in}

$\mathbf{x}_1$	$\mathbf{x}_2$	$\mathbf{x}_3$	...	$\mathbf{x}_N$	$\mathbf{x}_{N+1}$	$\mathbf{x}_{N+2}$	$\mathbf{x}_{N+3}$	...	$\mathbf{x}_{2N}$
○	○	●	...	○	●	●	○	...	○

Number of dichotomys is at most $m_{\mathcal{H}}(2N)$. (does not affect the polynomial nature of the bound)

Up to technical details, analyze a "hypothesis set" of size at most $m_{\mathcal{H}}(2N)$.

The Vapnik-Chervonenkis Bound (VC Bound)

$$\mathbb{P}\left[|E_{\text{in}}(g) - E_{\text{out}}(g)| > \epsilon\right] \leq 4m_{\mathcal{H}}(2N)e^{-\epsilon^2 N/8}, \qquad \text{for any } \epsilon > 0.$$

$$\mathbb{P}\left[|E_{\text{in}}(g) - E_{\text{out}}(g)| \leq \epsilon\right] \geq 1 - 4m_{\mathcal{H}}(2N)e^{-\epsilon^2 N/8}, \qquad \text{for any } \epsilon > 0.$$

$$E_{\text{out}}(g) \leq E_{\text{in}}(g) + \sqrt{\frac{8}{N}\log\frac{4m_{\mathcal{H}}(2N)}{\delta}}, \qquad \text{w.p. at least } 1 - \delta.$$

Eout $\approx$ Ein

Thanks!