

# Machine Learning Foundations

## (機器學習基石)



### Lecture 6: Theory of Generalization

Hsuan-Tien Lin (林軒田)

htlin@csie.ntu.edu.tw

Department of Computer Science  
& Information Engineering

National Taiwan University  
(國立台灣大學資訊工程系)



# Roadmap

- 1 When Can Machines Learn?
- 2 **Why** Can Machines Learn?

## Lecture 5: Training versus Testing

**effective** price of choice in training: **(wishfully)**  
**growth function**  $m_{\mathcal{H}}(N)$  with **a break point**

## Lecture 6: Theory of Generalization

- Restriction of Break Point
- Bounding Function: Basic Cases
- Bounding Function: Inductive Cases
- A Pictorial Proof

- 3 How Can Machines Learn?
- 4 How Can Machines Learn Better?

# The Four Break Points

growth function  $m_{\mathcal{H}}(N)$ : max number of dichotomies

- positive rays:

$$m_{\mathcal{H}}(N) = N + 1$$

○ ×  $m_{\mathcal{H}}(2) = 3 < 2^2$ : break point at 2

- positive intervals:

$$m_{\mathcal{H}}(N) = \frac{1}{2}N^2 + \frac{1}{2}N + 1$$

○ × ○  $m_{\mathcal{H}}(3) = 7 < 2^3$ : break point at 3

- convex sets:

$$m_{\mathcal{H}}(N) = 2^N$$



$m_{\mathcal{H}}(N) = 2^N$  always: no break point

- 2D perceptrons:

$m_{\mathcal{H}}(N) < 2^N$  in some cases



$m_{\mathcal{H}}(4) = 14 < 2^4$ : break point at 4

break point  $k \implies$  break point  $k + 1, \dots$

**what else?**

## Restriction of Break Point (1/2)

what 'must be true' when **minimum break point**  $k = 2$

- $N = 1$ : every  $m_{\mathcal{H}}(N) = 2$  by definition
- $N = 2$ : every  $m_{\mathcal{H}}(N) < 4$  by definition  
(so **maximum possible** = 3)

maximum possible  $m_{\mathcal{H}}(N)$  when  $N = 3$  and  $k = 2$ ?

1 dichotomy , shatter any two points? **no**

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

## Restriction of Break Point (1/2)

what 'must be true' when **minimum break point**  $k = 2$

- $N = 1$ : every  $m_{\mathcal{H}}(N) = 2$  by definition
- $N = 2$ : every  $m_{\mathcal{H}}(N) < 4$  by definition  
(so **maximum possible** = 3)

maximum possible  $m_{\mathcal{H}}(N)$  when  $N = 3$  and  $k = 2$ ?

2 dichotomies, shatter any two points? **no**

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

## Restriction of Break Point (1/2)

what 'must be true' when **minimum break point**  $k = 2$

- $N = 1$ : every  $m_{\mathcal{H}}(N) = 2$  by definition
- $N = 2$ : every  $m_{\mathcal{H}}(N) < 4$  by definition  
(so **maximum possible** = 3)

maximum possible  $m_{\mathcal{H}}(N)$  when  $N = 3$  and  $k = 2$ ?

3 dichotomies, shatter any two points? **no**

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

# Restriction of Break Point (1/2)

what 'must be true' when **minimum break point**  $k = 2$

- $N = 1$ : every  $m_{\mathcal{H}}(N) = 2$  by definition
- $N = 2$ : every  $m_{\mathcal{H}}(N) < 4$  by definition  
(so **maximum possible** = 3)

maximum possible  $m_{\mathcal{H}}(N)$  when  $N = 3$  and  $k = 2$ ?

4 dichotomies, shatter any two points? **yes**

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

## Restriction of Break Point (1/2)

what 'must be true' when **minimum break point**  $k = 2$

- $N = 1$ : every  $m_{\mathcal{H}}(N) = 2$  by definition
- $N = 2$ : every  $m_{\mathcal{H}}(N) < 4$  by definition  
(so **maximum possible** = 3)

maximum possible  $m_{\mathcal{H}}(N)$  when  $N = 3$  and  $k = 2$ ?

4 dichotomies, shatter any two points? **no**

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



# Restriction of Break Point (1/2)

what 'must be true' when **minimum break point**  $k = 2$

- $N = 1$ : every  $m_{\mathcal{H}}(N) = 2$  by definition
- $N = 2$ : every  $m_{\mathcal{H}}(N) < 4$  by definition  
(so **maximum possible** = 3)

maximum possible  $m_{\mathcal{H}}(N)$  when  $N = 3$  and  $k = 2$ ?

5 dichotomies, shatter any two points? **yes**

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

# Restriction of Break Point (1/2)

what 'must be true' when **minimum break point**  $k = 2$

- $N = 1$ : every  $m_{\mathcal{H}}(N) = 2$  by definition
- $N = 2$ : every  $m_{\mathcal{H}}(N) < 4$  by definition  
(so **maximum possible** = 3)

maximum possible  $m_{\mathcal{H}}(N)$  when  $N = 3$  and  $k = 2$ ?

5 dichotomies, shatter any two points? **yes**

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

# Restriction of Break Point (1/2)

what 'must be true' when **minimum break point**  $k = 2$

- $N = 1$ : every  $m_{\mathcal{H}}(N) = 2$  by definition
- $N = 2$ : every  $m_{\mathcal{H}}(N) < 4$  by definition  
(so **maximum possible** = 3)

maximum possible  $m_{\mathcal{H}}(N)$  when  $N = 3$  and  $k = 2$ ?

5 dichotomies, shatter any two points? **yes**

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

# Restriction of Break Point (1/2)

what 'must be true' when **minimum break point**  $k = 2$

- $N = 1$ : every  $m_{\mathcal{H}}(N) = 2$  by definition
- $N = 2$ : every  $m_{\mathcal{H}}(N) < 4$  by definition  
(so **maximum possible** = 3)

maximum possible  $m_{\mathcal{H}}(N)$  when  $N = 3$  and  $k = 2$ ?

maximum possible so far: **4 dichotomies**

$\mathbf{x}_1$	$\mathbf{x}_2$	$\mathbf{x}_3$
○	○	○
○	○	×
○	×	○
×	○	○
⋮-(	⋮-(	⋮-(

## Restriction of Break Point (2/2)

what 'must be true' when **minimum break point**  $k = 2$

- $N = 1$ : every  $m_{\mathcal{H}}(N) = 2$  by definition
- $N = 2$ : every  $m_{\mathcal{H}}(N) < 4$  by definition  
(so **maximum possible** = 3)
- $N = 3$ : **maximum possible** =  $4 \ll 2^3$

—break point  $k$  **restricts maximum possible**  $m_{\mathcal{H}}(N)$  **a lot** for  $N > k$

idea:

$$\begin{aligned} m_{\mathcal{H}}(N) &\leq \text{maximum possible } m_{\mathcal{H}}(N) \text{ given } k \\ &\leq \text{poly}(N) \end{aligned}$$

## Fun Time

When minimum break point  $k = 1$ , what is the maximum possible  $m_{\mathcal{H}}(N)$  when  $N = 3$ ?

① 1

② 2

③ 4

④ 8

Reference Answer: ①

Because  $k = 1$ , the hypothesis set cannot even shatter one point. Thus, every 'column' of the table cannot contain both  $\circ$  and  $\times$ . Then, after including the first dichotomy, it is not possible to include any other different dichotomy. Thus, the maximum possible  $m_{\mathcal{H}}(N)$  is 1.

$\mathbf{x}_1$	$\mathbf{x}_2$	$\mathbf{x}_3$
$\circ$	$\times$	$\circ$
$\circ$	$\times$	$\times$

# Bounding Function

**bounding function**  $B(N, k)$ :

maximum possible  $m_{\mathcal{H}}(N)$  when break point =  $k$

- combinatorial quantity:  
maximum number of length- $N$  vectors with ( $\circ$ ,  $\times$ )  
while ‘**no shatter**’ **any length- $k$**  subvectors
- irrelevant of the details of  $\mathcal{H}$   
e.g.  $B(N, 3)$  bounds both
  - positive intervals ( $k = 3$ )
  - 1D perceptrons ( $k = 3$ )

new goal:  $B(N, k) \leq \text{poly}(N)$ ?

## Table of Bounding Function (1/4)

$B(N, k)$		$k$						
		1	2	3	4	5	6	...
$N$	1							
	2		3					
	3		4					
	4							
	5							
	6							
	$\vdots$							

## Known

- $B(2, 2) = 3$  (maximum  $< 4$ )
- $B(3, 2) = 4$  ('pictorial' proof previously)



# Table of Bounding Function (2/4)

$B(N, k)$		$k$						
		1	2	3	4	5	6	...
$N$	1	1						
	2	1	3					
	3	1	4					
	4	1						
	5	1						
	6	1						
	$\vdots$	$\vdots$						

## Known

- $B(N, 1) = 1$  (see previous quiz)

## Table of Bounding Function (3/4)

		$k$						
$B(N, k)$		1	2	3	4	5	6	...
$N$	1	1	2	2	2	2	2	...
	2	1	3	4	4	4	4	...
	3	1	4		8	8	8	...
	4	1				16	16	...
	5	1					32	...
	6	1						...
	$\vdots$	$\vdots$						

## Known

- $B(N, k) = 2^N$  for  $N < k$   
—including all dichotomies not violating ‘breaking condition’

## Table of Bounding Function (4/4)

		$k$						
$B(N, k)$		1	2	3	4	5	6	...
$N$	1	1	2	2	2	2	2	...
	2	1	3	4	4	4	4	...
	3	1	4	7	8	8	8	...
	4	1			15	16	16	...
	5	1				31	32	...
	6	1					63	...
	$\vdots$	$\vdots$						$\ddots$

## Known

- $B(N, k) = 2^N - 1$  for  $N = k$   
 — **removing a single dichotomy** satisfies 'breaking condition'

**more than halfway done! :-)**

# Fun Time

For the 2D perceptrons, which of the following claim is true?

- ① minimum break point  $k = 2$
- ②  $m_{\mathcal{H}}(4) = 15$
- ③  $m_{\mathcal{H}}(N) < B(N, k)$  when  $N = k =$  minimum break point
- ④  $m_{\mathcal{H}}(N) > B(N, k)$  when  $N = k =$  minimum break point

Reference Answer: ③

As discussed previously, minimum break point for 2D perceptrons is 4, with  $m_{\mathcal{H}}(4) = 14$ . Also, note that  $B(4, 4) = 15$ . So bounding function  $B(N, k)$  can be 'loose' in bounding  $m_{\mathcal{H}}(N)$ .

# Estimating $B(4, 3)$

		$k$						
$B(N, k)$		1	2	3	4	5	6	...
$N$	1	1	2	2	2	2	2	...
	2	1	3	4	4	4	4	...
	3	1	4	7	8	8	8	...
	4	1		?	15	16	16	...
	5	1				31	32	...
	6	1					63	...
	$\vdots$	$\vdots$						$\ddots$

## Motivation

- $B(4, 3)$  shall be related to  $B(3, ?)$   
—‘adding’ one point from  $B(3, ?)$

next: reduce  $B(4, 3)$  to  $B(3, ?)$

# 'Achieving' Dichotomies of $B(4, 3)$

after checking all  $2^{2^4}$  sets of dichotomies, **the winner is ...**

	$x_1$	$x_2$	$x_3$	$x_4$
01	○	○	○	○
02	×	○	○	○
03	○	×	○	○
04	○	○	×	○
05	○	○	○	×
06	×	×	○	×
07	×	○	×	○
08	×	○	○	×
09	○	×	×	○
10	○	×	○	×
11	○	○	×	×

		$k$					
$B(N, k)$		1	2	3	4	5	6
$N$	1	1	2	2	2	2	2
	2	1	3	4	4	4	4
	3	1	4	7	8	8	8
	4	1		11	15	16	16
	5	1				31	32
	6	1					63

how to reduce  $B(4, 3)$  to  $B(3, ?)$  cases?

# Reorganized Dichotomies of $B(4, 3)$

after checking all  $2^{2^4}$  sets of dichotomies, **the winner is ...**

	$x_1$	$x_2$	$x_3$	$x_4$
01	○	○	○	○
02	×	○	○	○
03	○	×	○	○
04	○	○	×	○
05	○	○	○	×
06	×	×	○	×
07	×	○	×	○
08	×	○	○	×
09	○	×	×	○
10	○	×	○	×
11	○	○	×	×



	$x_1$	$x_2$	$x_3$	$x_4$
01	○	○	○	○
05	○	○	○	×
02	×	○	○	○
08	×	○	○	×
03	○	×	○	○
10	○	×	○	×
04	○	○	×	○
11	○	○	×	×
06	×	×	○	×
07	×	○	×	○
09	○	×	×	○

**orange:** pair; **purple:** single

# Estimating Part of $B(4, 3)$ (1/2)

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

	$\mathbf{x}_1$	$\mathbf{x}_2$	$\mathbf{x}_3$
$\alpha$	○	○	○
	×	○	○
	○	×	○
	○	○	×
$\beta$	×	×	○
	×	○	×
	○	×	×

- $\alpha + \beta$ : dichotomies on  $(\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3)$
- $B(4, 3)$  'no shatter' any 3 inputs  
 $\implies \alpha + \beta$  'no shatter' any 3

	$\mathbf{x}_1$	$\mathbf{x}_2$	$\mathbf{x}_3$	$\mathbf{x}_4$
$2\alpha$	○	○	○	○
	○	○	○	×
	×	○	○	○
	×	○	○	×
	○	×	○	○
	○	×	○	×
	○	○	×	○
	○	○	×	×
$\beta$	×	×	○	×
	×	○	×	○
	○	×	×	○

$$\alpha + \beta \leq B(3, 3)$$



# Estimating Part of $B(4, 3)$ (2/2)

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

	$\mathbf{x}_1$	$\mathbf{x}_2$	$\mathbf{x}_3$
$\alpha$	○	○	○
	×	○	○
	○	×	○
	○	○	×

- $\alpha$ : dichotomies on  $(\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3)$  with  $\mathbf{x}_4$  **paired**
- $B(4, 3)$  'no shatter' any 3 inputs  $\implies \alpha$  'no shatter' any 2

	$\mathbf{x}_1$	$\mathbf{x}_2$	$\mathbf{x}_3$	$\mathbf{x}_4$
$2\alpha$	○	○	○	○
	○	○	○	×
	×	○	○	○
	×	○	○	×
	○	×	○	○
	○	×	○	×
	○	○	×	○
	○	○	×	×
$\beta$	×	×	○	×
	×	○	×	○
	○	×	×	○

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

# Putting It All Together

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

$$\alpha + \beta \leq B(3, 3)$$

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

$$\Rightarrow B(4, 3) \leq B(3, 3) + B(3, 2)$$

		$k$					
$B(N, k)$		1	2	3	4	5	6
$N$	1	1	2	2	2	2	2
	2	1	3	4	4	4	4
	3	1	4	7	8	8	8
	4	1	$\leq 5$	11	15	16	16
	5	1	$\leq 6$	$\leq 16$	$\leq 26$	31	32
	6	1	$\leq 7$	$\leq 22$	$\leq 42$	$\leq 57$	63

now have **upper bound** of bounding function

# Putting It All Together

$$B(N, k) = 2\alpha + \beta$$

$$\alpha + \beta \leq B(N - 1, k)$$

$$\alpha \leq B(N - 1, k - 1)$$

$$\Rightarrow B(N, k) \leq B(N - 1, k) + B(N - 1, k - 1)$$

		$k$					
$B(N, k)$		1	2	3	4	5	6
$N$	1	1	2	2	2	2	2
	2	1	3	4	4	4	4
	3	1	4	7	8	8	8
	4	1	$\leq 5$	11	15	16	16
	5	1	$\leq 6$	$\leq 16$	$\leq 26$	31	32
	6	1	$\leq 7$	$\leq 22$	$\leq 42$	$\leq 57$	63

now have **upper bound** of bounding function

# Bounding Function: The Theorem

$$B(N, k) \leq \underbrace{\sum_{i=0}^{k-1} \binom{N}{i}}_{\text{highest term } N^{k-1}}$$

- simple induction using **boundary and inductive formula**
- for fixed  $k$ ,  $B(N, k)$  upper bounded by  $\text{poly}(N)$   
 $\implies m_{\mathcal{H}}(N)$  is  $\text{poly}(N)$  if break point exists

‘ $\leq$ ’ can be ‘ $=$ ’ actually,  
**go play and prove it if math lover! :-)**

# The Three Break Points

$$B(N, k) \leq \underbrace{\sum_{i=0}^{k-1} \binom{N}{i}}_{\text{highest term } N^{k-1}}$$

- positive rays:  $m_{\mathcal{H}}(N) = N + 1 \leq N + 1$   
 $\circ \times \quad m_{\mathcal{H}}(2) = 3 < 2^2$ : break point at 2
- positive intervals:  $m_{\mathcal{H}}(N) = \frac{1}{2}N^2 + \frac{1}{2}N + 1 \leq \frac{1}{2}N^2 + \frac{1}{2}N + 1$   
 $\circ \times \circ \quad m_{\mathcal{H}}(3) = 7 < 2^3$ : break point at 3
- 2D perceptrons:  $m_{\mathcal{H}}(N) = ? \leq \frac{1}{6}N^3 + \frac{5}{6}N + 1$   
 $\times \begin{array}{c} \circ \\ \circ \end{array} \times \quad m_{\mathcal{H}}(4) = 14 < 2^4$ : break point at 4

can bound  $m_{\mathcal{H}}(N)$  by only **one break point**

# Fun Time

For 1D perceptrons (positive and negative rays), we know that  $m_{\mathcal{H}}(N) = 2N$ . Let  $k$  be the minimum break point. Which of the following is not true?

- ①  $k = 3$
- ② for some integers  $N > 0$ ,  $m_{\mathcal{H}}(N) = \sum_{i=0}^{k-1} \binom{N}{i}$
- ③ for all integers  $N > 0$ ,  $m_{\mathcal{H}}(N) = \sum_{i=0}^{k-1} \binom{N}{i}$
- ④ for all integers  $N > 2$ ,  $m_{\mathcal{H}}(N) < \sum_{i=0}^{k-1} \binom{N}{i}$

Reference Answer: ③

The proof is generally trivial by listing the definitions. For ②,  $N = 1$  or  $2$  gives the equality. One thing to notice is ④: the upper bound can be 'loose'.

BAD Bound for General  $\mathcal{H}$ 

want:

$$\mathbb{P}\left[\exists h \in \mathcal{H} \text{ s.t. } |E_{\text{in}}(h) - E_{\text{out}}(h)| > \epsilon\right] \leq 2 m_{\mathcal{H}}(N) \cdot \exp\left(-2 \epsilon^2 N\right)$$

actually, when  $N$  large enough,

$$\mathbb{P}\left[\exists h \in \mathcal{H} \text{ s.t. } |E_{\text{in}}(h) - E_{\text{out}}(h)| > \epsilon\right] \leq 2 \cdot 2 m_{\mathcal{H}}(2N) \cdot \exp\left(-2 \cdot \frac{1}{16} \epsilon^2 N\right)$$

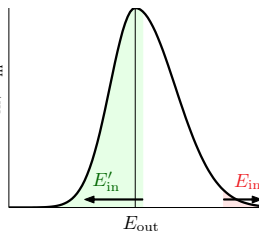
next: **sketch** of proof

# Step 1: Replace $E_{\text{out}}$ by $E'_{\text{in}}$

$$\begin{aligned} & \frac{1}{2} \mathbb{P} \left[ \exists h \in \mathcal{H} \text{ s.t. } |E_{\text{in}}(h) - E_{\text{out}}(h)| > \epsilon \right] \\ \leq & \mathbb{P} \left[ \exists h \in \mathcal{H} \text{ s.t. } |E_{\text{in}}(h) - E'_{\text{in}}(h)| > \frac{\epsilon}{2} \right] \end{aligned}$$

- $E_{\text{in}}(h)$  finitely many,  $E_{\text{out}}(h)$  infinitely many  
—replace the evil  $E_{\text{out}}$  first
- how? sample verification set  $\mathcal{D}'$  of size  $N$  to calculate  $E'_{\text{in}}$
- BAD  $h$  of  $E_{\text{in}} - E_{\text{out}}$   
probably  $\implies$  BAD  $h$  of  $E_{\text{in}} - E'_{\text{in}}$

Probability distribution  
of  $E_{\text{in}}, E'_{\text{in}}$



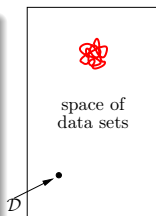
evil  $E_{\text{out}}$  removed by  
verification with 'ghost data'



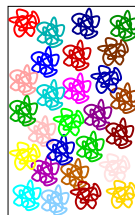
## Step 2: Decompose $\mathcal{H}$ by Kind

$$\begin{aligned} \text{BAD} &\leq 2\mathbb{P}\left[\exists h \in \mathcal{H} \text{ s.t. } |E_{\text{in}}(h) - E'_{\text{in}}(h)| > \frac{\epsilon}{2}\right] \\ &\leq 2m_{\mathcal{H}}(2N)\mathbb{P}\left[\text{fixed } h \text{ s.t. } |E_{\text{in}}(h) - E'_{\text{in}}(h)| > \frac{\epsilon}{2}\right] \end{aligned}$$

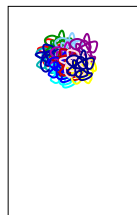
- $E_{\text{in}}$  with  $\mathcal{D}$ ,  $E'_{\text{in}}$  with  $\mathcal{D}'$   
—now  $m_{\mathcal{H}}$  comes to play
- how? infinite  $\mathcal{H}$  becomes  
 $|\mathcal{H}(\mathbf{x}_1, \dots, \mathbf{x}_N, \mathbf{x}'_1, \dots, \mathbf{x}'_N)|$   
kinds
- union bound on  $m_{\mathcal{H}}(2N)$  kinds



(a) Hoeffding Inequality



(b) Union Bound



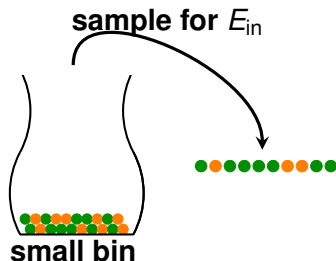
(c) Now

use  $m_{\mathcal{H}}(2N)$  to calculate BAD-overlap properly

# Step 3: Use Hoeffding without Replacement

$$\begin{aligned}
 \text{BAD} &\leq 2m_{\mathcal{H}}(2N)\mathbb{P}\left[\text{fixed } h \text{ s.t. } |E_{\text{in}}(h) - E'_{\text{in}}(h)| > \frac{\epsilon}{2}\right] \\
 &\leq 2m_{\mathcal{H}}(2N) \cdot 2 \exp\left(-2\left(\frac{\epsilon}{4}\right)^2 N\right)
 \end{aligned}$$

- consider bin of  $2N$  examples, choose  $N$  for  $E_{\text{in}}$ , leave others for  $E'_{\text{in}}$   
 $|E_{\text{in}} - E'_{\text{in}}| > \frac{\epsilon}{2} \Leftrightarrow \left|E_{\text{in}} - \frac{E_{\text{in}} + E'_{\text{in}}}{2}\right| > \frac{\epsilon}{4}$
- so? just 'smaller bin', 'smaller  $\epsilon$ ', and Hoeffding without replacement



use Hoeffding after zooming to fixed  $h$

## That's All!

## Vapnik-Chervonenkis (VC) bound:

$$\begin{aligned} & \mathbb{P}\left[\exists h \in \mathcal{H} \text{ s.t. } |E_{\text{in}}(h) - E_{\text{out}}(h)| > \epsilon\right] \\ & \leq 4m_{\mathcal{H}}(2N) \exp\left(-\frac{1}{8}\epsilon^2 N\right) \end{aligned}$$

- replace  $E_{\text{out}}$  by  $E'_{\text{in}}$
- decompose  $\mathcal{H}$  by kind
- use Hoeffding without replacement

## 2D perceptrons:

- break point? 4
- $m_{\mathcal{H}}(N)$ ?  $O(N^3)$

learning with 2D perceptrons feasible! :-)

## Fun Time

For positive rays,  $m_{\mathcal{H}}(N) = N + 1$ . Plug it into the VC bound for  $\epsilon = 0.1$  and  $N = 10000$ . What is VC bound of BAD events?

$$\mathbb{P}\left[\exists h \in \mathcal{H} \text{ s.t. } |E_{\text{in}}(h) - E_{\text{out}}(h)| > \epsilon\right] \leq 4m_{\mathcal{H}}(2N) \exp\left(-\frac{1}{8}\epsilon^2 N\right)$$

①  $2.77 \times 10^{-87}$

②  $5.54 \times 10^{-83}$

③  $2.98 \times 10^{-1}$

④  $2.29 \times 10^2$

Reference Answer: ③

Simple calculation. Note that the BAD probability bound is not very small even with 10000 examples.

# Summary

① When Can Machines Learn?

② **Why** Can Machines Learn?

Lecture 5: Training versus Testing

Lecture 6: Theory of Generalization

- Restriction of Break Point  
**break point 'breaks' consequent points**
- Bounding Function: Basic Cases  
 **$B(N, k)$  bounds  $m_{\mathcal{H}}(N)$  with break point  $k$**
- Bounding Function: Inductive Cases  
 **$B(N, k)$  is  $\text{poly}(N)$**
- A Pictorial Proof  
 **$m_{\mathcal{H}}(N)$  can replace  $M$  with a few changes**

• **next: how to 'use' the break point?**

③ How Can Machines Learn?

④ How Can Machines Learn Better?